

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Дніпропетровський національний університет імені Олеся Гончара

**АКТУАЛЬНІ ПРОБЛЕМИ
АВТОМАТИЗАЦІЇ
ТА ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ**

ТОМ 17

Збірник наукових праць

Дніпропетровськ
Видавництво
Дніпропетровського національного університету
2013

УДК 519.689:519.254

ББК 32.973.202

A43

Описані задачі автоматизації обробки даних, інформаційна технологія сучасних процедур більш адекватного відображення експериментальних даних у системах моніторингу. Досліджені обчислювані процедури сплайн-перетворень, які застосовані у системах ГІС та прийняття рішень, прикладні задачі інформаційних технологій.

Для науковців, аспірантів та фахівців у галузі інформатики.

A43 Актуальні проблеми автоматизації та інформаційних технологій : зб. наук. пр. / наук. ред. О.Г. Байбуз – Д. : Вид-во Дніпропетр. нац. ун-ту, 2013. – Т.17. – 124 с.
ISBN 978-966-551-362-9

Описаны задачи автоматизированной обработки данных, информационная технология современных вычислительных процедур более адекватного отображения экспериментальных данных в системах мониторинга. Исследованы вычислительные процедуры сплайн-преобразований, примененных в системах ГИС и принятия решений, прикладные задачи информационных технологий.

Для научных сотрудников, аспирантов и специалистов в области информатики.

УДК 519.689:519.254

ББК 32.973.202

Рекомендовано вченою радою

Дніпропетровського національного університету імені Олеся Гончара

Науковий редактор

доктор технічних наук, професор **О.Г. Байбуз**

Редакційна колегія:

д-р техн. наук, проф. **О.М. Карпов** (заст. наук. ред.); д-р фіз.-мат. наук, проф. **О.М. Ахметшин**; д-р техн. наук, проф. **В.П. Малайчук**; д-р техн. наук, проф. **О.М. Петренко**; д-р техн. наук, проф. **С.В. Голуб**; д-р техн. наук, проф. **В.О. Доровський**; д-р техн. наук, проф. **Бейсенби Мамірбек Аукербайули**; канд. техн. наук, доц. **О.М. Мацуга** (відп. секр.)

Рецензенти:

д-р техн. наук, проф. **Б.І. Мороз**

д-р техн. наук, проф. **М.О. Шутко**

© Дніпропетровський національний
університет імені Олеся Гончара, 2013
© Видавництво ДНУ, оформлення, 2013
© Автори статей, 2013

УДК 519.682:336.763

В.Б. Говоруха, О.В. Дубель

Академія митної служби України

МЕХАНІЗАЦІЯ І АВТОМАТИЗАЦІЯ ДІЛОВОДНИХ ПРОЦЕСІВ

Стаття присвячена питанням автоматизації документообігу. Розвиток глобальних комунікацій у діловому і повсякденному житті зумовив появу нової сфери взаємовідносин, предметом яких є електронний обмін даними. Програма електронного документообігу сьогодні активно впроваджується в державних установах та органах державної влади, що істотно розширює можливості розвитку електронного документообігу. Тому впровадження повністю електронного документообігу є завданням досить актуальним.

Ключові слова: документ, електронний документообіг, електронне діловодство, інформація.

Статья посвящена вопросам автоматизации документооборота. Развитие глобальных коммуникаций в деловой и повседневной жизни привело к появлению новой области взаимоотношений, предметом которых является электронный обмен данными. Программа электронного документооборота сегодня активно внедряется в государственных учреждениях и органах государственной власти, что существенно расширяет возможности развития электронного документооборота. Поэтому внедрение полностью электронного документооборота является задачей достаточно актуальной.

Ключевые слова: документ, электронный документооборот, электронное делопроизводство, информация.

The article deals with automation of the document exchange. The development of global communications in business and everyday life was at the bottom of appearing a new sphere of relationship. It is electronic exchange of datum. The program of electronic exchange of documents is actively applied in state institutions, government bodies that favours the development of such electronic changing papers. That is why completely electronic exchange of documents is enough timely and urgent task.

Keywords: document, electronic document circulation, electronic office work, information.

Електронний документообіг – високотехнологічний і прогресивний підхід до суттєвого підвищення ефективності роботи органів державної влади і місцевого самоврядування. Гарантією успішної роботи органів влади завжди є ефективна діяльність державних службовців. Але для якісного обслуговування потреб громадян учорашні методи обробки інформації вже не є найкращими. Сьогодні необхідно мати доступ до інформаційних ресурсів і скоротити часові витрати на розв’язання задач, не пов’язаних з обслуговуванням громадян.

Відсутність необхідності вручну розмножувати документи, відслідковувати переміщення паперових документів усередині організації, контролювати порядок передачі конфіденційної інформації істотним образом знижує трудовитрати діловодів. Наскрізний автоматичний контроль виконання на всіх етапах роботи з документами кардинально підвищує якість роботи виконавців, робить терміни підготовки документів більш прогнозованими й керованими [1].

Спільне використання систем електронного діловодства і сховищ інформації дозволяє систематизувати й поєднувати інформацію, що полегшує її аналіз і складання звітів. Для пошуку прихованих закономірностей у великих масивах даних можна приймати більш ефективні рішення та дії, що базується на відповідних технологіях дістання інформації з даних (data mining techniques). Усе це можливо тільки в системі управління, побудованій на основі цілком електронного документообігу. Інформаційні безпаперові технології полегшують процес управління знаннями. Вони створюють основу рішень, що забезпечують автоматизований і централізований обмін знаннями і дістають лише необхідну інформацію з усіх доступних джерел.

Системи електронного діловодства і документообігу можуть сприяти створенню нової організаційної культури в органах влади, зробити роботу державних службовців більш легкою, цікавою та значимою. Інформаційні технології дозволяють державним службовцям працювати не тільки над виконанням внутрішньовідомчих задач, але й спільними зусиллями вирішувати більш широкий спектр державних проблем. Інформаційні технології можуть виступати як і каталізатор, завдяки якому органи влади перейдуть на новий рівень взаємовідносин із населенням, коли державні службовці будуть прямо відповідати на запити громадян і ставитися до них як до клієнтів, а не як до надокучливих відвідувачів [5].

Саме тому ефективність управління підприємствами й організаціями не в останню чергу залежить від коректного рішення задач оперативного і якісного формування електронних документів, контролю їхнього виконання, а також продуманої організації їхнього збереження, пошуку й використання. Потреба в ефективному керуванні електронними документами і призвела до створення систем електронного документообігу (СЕД).

Основним завданням є перш за все дослідити систему електронного документообігу та програмного забезпечення, за допомогою якого він здійснюється, визначити доцільність впровадження СЕД для підвищення ефективності документообігу в установі, створити ефективне середовище керування та функціонування організації, класифікувавши та визначивши основні властивості СЕД.

Організація роботи з документами залежить від обсягу документообігу, організаційної структури та форми роботи з документами, прийнятої технології обробки документів. Робота з документами складається з ряду поступових операцій: прийняття документів, їх розподілення і направлення на виконання, виконання та відправка [3].

Незалежно від обсягу діяльності організації операції із прийняття та відправки документів зазвичай здійснюються централізовано. У структурі служби діловодства великих організацій для цього передбачають створення спеціалізованої дільниці – експедиції як одного зі структурних підрозділів служби. У канцеляріях і загальних відділах організацій створюється дільниця із прийняття та обробки документів із тими ж функціями. В організаціях, які не створюють свою службу діловодства, прийняття та обробка документів, операції з їх відправки виконуються централізовано секретарем.

При надходженні електронних документів обробляється тільки супроводжувана документація, машинні носії в упаковці передаються за призначенням.

Експедиція веде облік усіх вхідних документів: на документі проставляється штамп про їх отримання, де вказується найменування організації, дата надходження, порядковий обліковий номер документа. Зазвичай на штампі передбачено місце для наступного відбитку вхідного зареєстрованого індексу.

При первинній експедиційній обробці документ сортується по структурних підрозділах. Головний критерій цього сортування – адресування документів. Як правило документи, адресовані

керівництву, передаються до канцелярії для підготовки до розгляду, туди ж надходять документи, адресовані організації (без вказування прізвища посадової особи). Документи, адресовані структурним підрозділам, сортуються за їх назвами.

Передача документів між структурними підрозділами (із експедиції до канцелярії, із канцелярії до структурних підрозділів і т.д.) здійснюється через секретаря цих підрозділів або тих осіб, що відповідають у них за роботу з документами [6].

Автоматизація діловодних процесів є обов'язковою умовою раціональної організації діловодства в кожній установі, засобом підвищення продуктивності та здешевлення управлінської праці.

Автоматизація діловодних процесів має здійснюватися на основі упорядкованої системи документування управлінської діяльності, уніфікації та скорочення кількості форм використовуваних документів.

Механізація і автоматизація впроваджуються на всіх етапах діловодного процесу: підготовка документів, їх копіювання, оперативне зберігання й транспортування, контроль за виконанням тощо.

Засоби механізації і автоматизації діловодних процесів мають бути сумісними і передбачати можливість їх об'єднання в єдину систему.

Склад програмного забезпечення і засобів, що використовуються, залежить від конкретних умов роботи апарату управління з документами, зокрема від організаційної структури апарату управління, його розміщення, умов праці співробітників, кількості та змісту документів, потреб в оперативній і ретроспективній інформації, ступеня централізації робіт із документами.

З метою більш раціонального використання технічних засобів, ПК та запобігання їх простою засоби, які технічно неможливо або економічно недоцільно встановлювати на кожному робочому місці, слід використовувати централізовано.

Комплекс технічних засобів має забезпечувати збирання й передачу інформації, її запис на машинні носії, введення інформації у ПК, виведення результатів її обробки у формі машинограм або відеограм, сумісність з іншими інформаційними системами, а також можливість об'єднання їх в єдину систему.

Під час впровадження нових технологій роботи з документами необхідно враховувати:

- доцільність впровадження технічних засобів;
- можливість придбання технічних засобів у певні терміни;

- наявність придатних приміщень;
- можливість підготовки чи залучення спеціалістів до обслуговування техніки та її ремонту.

Керівник установи несе відповідальність за ефективність використання механізованої і автоматизованої технології роботи з документами, здійснює перевірку знання посадовими особами технічних інструкцій для роботи з відповідною технікою [1].

Підсумовуючи зазначене, необхідно зауважити, що впровадження СЕД варто здійснювати поступово, починаючи з найбільш важливої ланки документообігу: опису, автоматизація якого дозволить швидко одержати позитивний ефект. У процесі впровадження СЕД необхідно забезпечити можливість роботи як за новою, так і за старою технологією, щоб не заважати повсякденній діяльності підприємства. Дуже важливу роль відіграє реальна підтримка керівництвом підприємства проекту впровадження (так званий фактор першої особи). При відсутності такої підтримки у кращому випадку система буде впроваджена лише в окремих підрозділах підприємства (навіть від цього можна чекати будь-якого помітного повернення інвестицій). Проблему «опору змінам» можна вирішити шляхом поступового й планомірного впровадження елементів електронного документообігу, починаючи з найпростішого (наприклад, навчання співробітників роботі з електронною поштою й internet-мережею, організацією електронного архіву та ін.) і проведення необхідної роз'яснювальної роботи [2]. У ході впровадження СЕД обов'язково варто організувати тренінги для співробітників підприємства, а також консультації для його помічників з організації переходу на електронну форму діловодства.

Термін «документообіг» сьогодні вживається доволі широко й зустрічається в описі практично кожної інформаційної системи. Автоматизація документообігу реалізується багатьма системами, і кожна з них претендує на «комплексний підхід до автоматизації документообігу». При цьому розмаїтість систем є надзвичайно широкою: від систем складського обліку до ПЗ контролю за технологічними процесами.

Терміни «документообіг», «діловодство» та ін. трактуються винятково з використанням загальнодержавних нормативних актів у застосуванні до роботи органів влади. Відсутність необхідності вручну розмножувати документи, відслідковувати переміщення паперових документів усередині організації, контролювати порядок передачі

конфіденційної інформації істотним образом знижує трудовитрати діловодів. Наскрізний автоматичний контроль виконання на всіх етапах роботи з документами кардинально підвищує якість роботи виконавців, робить терміни підготовки документів більш прогнозованими й керованими [4]. Спільне використання системи електронного діловодства і сховищ інформації дозволяє систематизувати й поєднувати інформацію, що полегшує її аналіз і складання звітів. Для пошуку прихованих закономірностей у великих масивах даних можна приймати більш ефективні рішення та дії, що базується на відповідних технологіях діставання інформації з даних. Усе це можливе тільки в системі управління, побудованій на основі цілком електронного документообігу.

Бібліографічні посилання

1. **Баласян В.Э.** От традиционного делопроизводства к электронному документообороту / В.Э. Баласян // Справочник секретаря и офис-менеджера. – 2003. – № 4. – С. 14–18.
2. **Беспянська Г.В.** Організація роботи з документами : навч. посіб. для дистанційного навчання / Г.В. Беспянська. – К. : Ун-т «Україна», 2006. – С. 40–59.
3. **ДСТУ 4163-2003.** Державна уніфікована система документації. Уніфікована система організаційно-розпорядчої документації. Вимоги до оформлювання документів // Держспоживстандарт України, 2003. – 22 с.
4. **Комова М.В.** Діловодство : навч. посіб. / М.В. Комова. – Львів: Тріада плюс, 2006. – 217 с.
5. **Філіпова Л.Я.** Системи управління електронним документообігом: загальні поняття термінології, організація, технології (зарубіжний досвід) / Л.Я. Філіпова // Вісн. Кн. палати. – 2001. – № 4. – С. 15–18.
6. **Шеремета А.В.** Первинні документи та документообіг на підприємстві / А.В. Шеремета // Секретар-референт. – 2004. – № 10. – С. 20–23.

Надійшла до редколегії 18.06.2013 р.

УДК 534.4:612.78

А.А. Карпов, О.М. Карпов, Д.С. Журба

Дніпропетровський національний університет імені Олеся Гончара

**АЛГОРИТМИ МОДЕЛЮВАННЯ БІОМЕХАНІКИ
АРТИКУЛЯЦІЇ ГУБ ЛЮДИНИ НА ОСНОВІ
В-СПЛАЙН-ПОВЕРХОНЬ**

Наведено опис практичної реалізації алгоритму моделювання біомеханіки артикуляції губ людини.

Ключові слова: *алгоритм, артикуляція, поверхневі сплайни.*

Даётся описание практической реализации моделирования биомеханики артикуляции губ человека.

Ключевые слова: *алгоритм, артикуляция, сплайны поверхности.*

It is considered the description of the practical realization of the algorithm of modeling of human articulation of lips in the article.

Keywords: *algorithm, articulation, surface spline.*

Вступ. Математичне моделювання тіла, складеного з простих геометричних форм (сфер, циліндрів або конусів) не є складним. Але дуже часто цього недостатньо; кузова автомобілів, поверхні літаків, фюзеляжі і багато чого іншого не так просто описати. Процедура, зазвичай використовувана в цих випадках, полягає в такому:

- поверхня покривається двома уявними групами ліній; перша йде в поздовжньому напрямку, друга – трансверсально до першої. Ця сітка ліній визначає безліч осередків, кожна з яких (у разі гладкої поверхні) буде обмежена чотирма гладкими кривими;
- координати вузлів цієї уявної сітки вимірюються на моделі чи на наборі креслень поперечних перетинів поверхні;
- за допомогою інтерполяції (усереднення) математично описуються ці дві групи ліній, що утворюють сітку.

Можна будувати досить гладкі криві і поверхні з використанням поліномів. Припустимо, що ми хочемо побудувати поверхню у вигляді графіка функції $z = z(x, y)$. Лінія $y = \text{const}$ на цій поверхні буде представлена лінією $z = z(x)$, вона буде проходити через послідовність

точок $(x_0, z_0), (x_1, z_1), \dots, (x_n, z_n)$ з $0 < \dots < x_i < \dots < x_n$. Наша мета – провести через ці точки складову криву $f(x)$, що має такі властивості:

- на кожному інтервалі $x_{i-1} \leq x \leq x_i, i = 1, 2, \dots, n$, функція $f(x)$ є кубічним поліномом;
- її перші й другі похідні неперервні у вузлах.

Отримана гладка крива називається кубічним сплайном. Термін «сплайн» виник за аналогією з назвою креслярського інструмента – тонкої металевої лінійки, яка може згинатися так, щоб проходити через задані точки. Фізично така крива мінімізує енергію внутрішніх напружень. Математично – має мінімальну середньоквадратичну кривизну, тобто вона найбільш гладка. Сплайни мають багато застосувань у конструюванні криволінійних форм. Однак вони мають і деякі обмеження:

- локальна зміна тягне за собою обчислення заново всього сплайна;
- можуть виникати проблеми при апроксимації прямої, що має розриви других похідних (наприклад, сполучення прямої лінії і дуги кола);
- із точки зору естетики не завжди прийнятні, оскільки кривизна поверхні, сконструйованої за допомогою сплайнів, змінюється іноді нерівномірно, що призводить до спотворень (наприклад, химерні спотворення предметів, відбитих від кузова автомобіля).

Ці обмеження можна усунути за допомогою В-сплайна [1].

Загальна частина. В-сплайн-поверхні. Природним розширенням поняття поверхні Безье є декартовий добуток В-сплайн-поверхонь, що визначається виразом

$$Q(u, v) = \sum_{i=1}^{n+1} \sum_{j=1}^{m+1} B_{i,j} * N_{i,k}(u) * M_{j,l}(v),$$

де $N_{i,k}(u)$ і $M_{j,l}(v)$ – базисні функції В-сплайна у біпараметричних напрямках u та v відповідно.

$$\begin{cases} N_{i,k}(u) = 1, x_i \leq u < x_{i+1} \\ N_{i,k}(u) = 0, \text{інакше} \end{cases}$$

$$N_{i,k}(u) = \frac{(u - x_i)N_{i,k-1}(u)}{x_{i+k-1} - x_i} - \frac{(x_{i+k} - u)N_{i+1,k-1}(u)}{x_{i+k} - x_{i+1}},$$

Аналогічно

$$\begin{cases} M_{j,i}(v) = 1, y_j \leq u < y_{j+1} \\ M_{j,i}(v) = 0, \text{інакше} \end{cases}$$

$$M_{j,i}(v) = \frac{(v - y_j)M_{j,i-1}(v)}{y_{j+1-i} - y_j} - \frac{(y_{j+1} - v)M_{j+2,i-1}(v)}{y_{j+1} - y_{j+1}}$$

де x_i та y_j є елементами вузлових векторів, $B_{i,j}$ є вершинами визначаючої полігональної сітки [1].

Для моделювання поверхні голови було обрано однорідну В-сплайн-поверхню, замкнену вздовж одного параметра та незамкнену вздовж іншого.

Для моделювання артикуляції та міміки обличчя було реалізовано функцію залежності переміщення точки моделі, побудованої внаслідок реалізації сплайн-поверхні, від топологічної відстані цієї точки до обраних початкових точок, положення яких відповідає положенню мімічних м'язів на обличчі людини:

$$\begin{cases} \Delta\alpha = \alpha_0 * F(i, i_{center}, \sigma) * F(j, j_{center}, \sigma) \\ \Delta\beta = \beta_0 * F(i, i_{center}, \sigma) * F(j, j_{center}, \sigma) \\ \Delta r = r_0 * F(i, i_{center}, \sigma) * F(j, j_{center}, \sigma), \end{cases}$$

де

$$F(i, i_{center}, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(i-i_{center})^2}{2\sigma^2}}$$

Для вказівки положення точок використана сферична система координат (α, β, r) , де

$$\begin{cases} x = r * \cos(\beta) * \sin(\alpha) \\ y = r * \sin(\beta) \\ z = r * \cos(\beta) * \cos(\alpha). \end{cases}$$

Для моделювання руху м'язів обличчя людини було записано декілька положень обличчя, кожне з яких відповідає вимові певного звуку російської мови. Рух здійснюється за рахунок наближення поточного стану обличчя до стану, що відповідає вимові наступного звуку. Кожна точка моделі, побудованої за допомогою В-сплайн-поверхні, наближається до кінцевої за час, що відповідає часу вимови даного звуку (прийнято вважати за 0,2 секунди на кожний звук).

Реалізація. У процесі виконання роботи була розроблена та реалізована програма, що імітує вимову звуків російської мови артикуляційним апаратом людини [2]. Для створення тривимірної

моделі голови людини була побудована В-сплайн-поверхня із заданим числом вузлів в обох напрямках, замкнена вздовж параметричної координати u та незамкнена вздовж координати v .

Для моделювання руху артикуляційного апарату були використані тривимірні колоколовидні функції, що визначають зміщення точки поверхні залежно від відстані до заданих точок, а також поступова зміна положення точок поверхні між певними вузловими положеннями, що відповідають вимові певних звуків.

Вхідними даними для виконання роботи програми є вхідний текст, вимову якого потрібно змодельовати.

Вихідними даними є параметри моделі артикуляційного апарату людини у кожний момент часу, а також візуалізація цієї моделі.

Середовище розробки програми – Microsoft Visual Studio 2010, Графічна бібліотека – OpenGL, Платформа – Tao Framework. Програма моделює функціонування артикуляції та міміки обличчя людини (рис. 1).

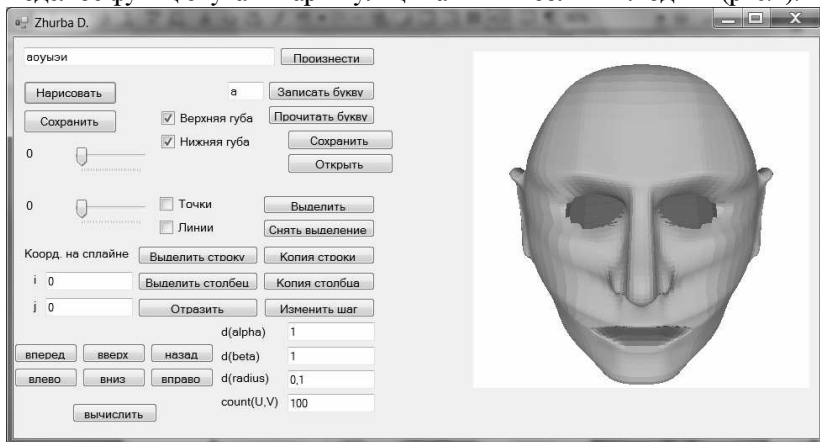


Рис.1. Моделювання

Для запуску програми слід спочатку завантажити певну модель (кнопка «Нарисувати»), а потім ввести символний рядок, вимові якого відповідає артикуляція, що моделюватиметься у програмі, та натиснути кнопку «Произнести». Результат моделювання буде відображений графічно (рис.2).

Програма дозволяє моделювати В-сплайн-поверхню, за якою виконується побудова обличчя. Для цього слід увімкнути режим сітки та скористатися відповідними інструментами (рис.3).

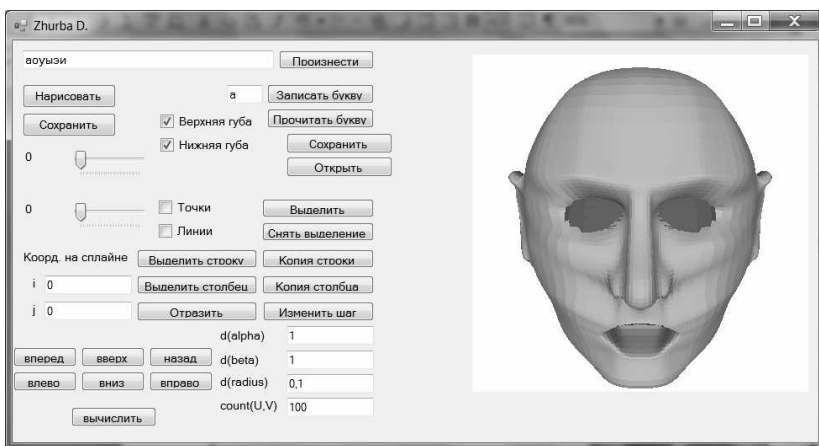


Рис.2. Моделювання процесу вимови звуку

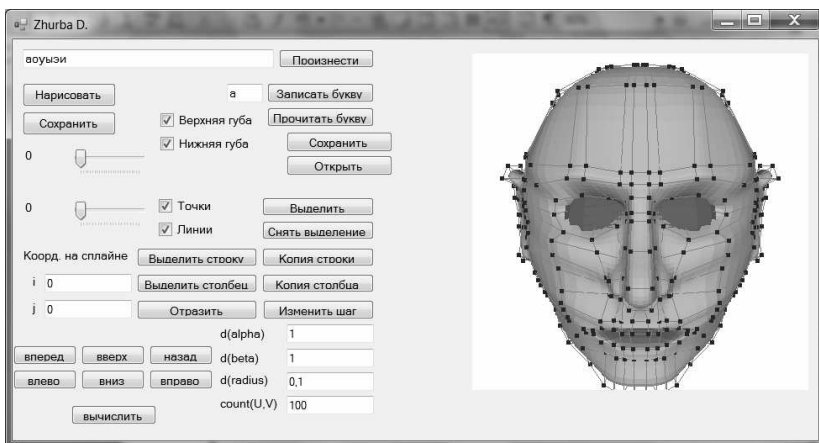


Рис.3. Моделювання обличчя. Тестування

Введемо наступний рядок (рис.4), запустимо програму.

Програма розраховує параметри, наближені до реальних при вимові даних звуків. Введемо іншу послідовність звуків, для яких форма та положення ротових м'язів значно відрізняються. При цьому програма відобразить інший результат, що відповідає заданій послідовності звуків.

За допомогою відповідних інструментів можна записати певний стан міміки обличчя та присвоїти це значення певній літері.

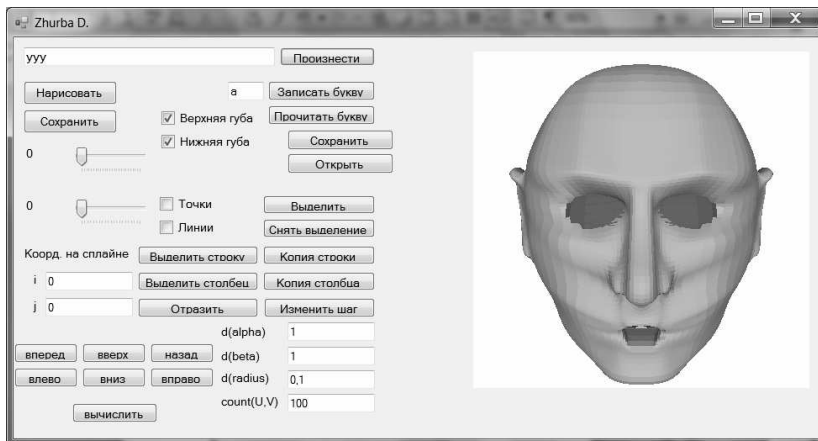


Рис.4. Тестування програми

Висновки. У результаті виконання роботи були досягнуті такі результати:

- досліджені способи математичного моделювання функціонування артикуляції та міміки обличчя;
- досліджені сучасні методи побудови тривимірної моделі та візуалізації голови людини;
- розроблені методи наближеного моделювання функціонування артикуляції та міміки обличчя;
- реалізована програма демонстрації роботи артикуляційного апарату.

Бібліографічні посилання

1. **Роджерс Д.** Математические основы машинной графики / Д.Роджерс. – М.: Мир, 2001.
2. **Журба Д.С.** Алгоритми моделювання біомеханіки артикуляції обличчя / Д.С. Журба, О.М. Карпов // Проблеми математичного моделювання: тези доп. Міждерж. наук.-метод. конф. 13–15 червня 2013р. – Дніпродзержинськ: ДДТУ, 2012. – С.114–115.

Надійшла до редколегії 23.06.2013 р.

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПЕРЕВІРКИ ГІПОТЕЗИ ПРО ПАРАМЕТРИ НОРМАЛЬНОГО РОЗПОДІЛУ НА ОСНОВІ ПОСЛІДОВНОГО АНАЛІЗУ

Розглядається схема побудови оперативної характеристики та функції середнього обсягу вибірки для послідовного критерію Вальда у випадку нормального розподілу. Наведено обчислювальні процедури методу послідовного аналізу для випадку, коли одночасно два параметри нормального розподілу є невідомими.

Ключові слова: *послідовний аналіз, нормальний розподіл.*

Рассматривается схема построения оперативной характеристики и функции среднего объема выборки для последовательного критерия Вальда в случае нормального распределения. Приведены вычислительные процедуры метода последовательного анализа для случая, когда одновременно два параметра нормального распределения неизвестны.

Ключевые слова: *последовательный анализ, нормальное распределение.*

In this paper we considered schemes to computation operating characteristic function and average sample number function for the Wald's sequential criteria in case of normal distribution. There are sequential computation procedures in case of both normal distraction parameters unknown.

Keywords: *sequential analysis, normal distraction.*

Постановка проблеми в загальному вигляді. У даній роботі розглядається застосування критерію послідовного аналізу до випадку гіпотези про одночасно задані обидва параметри нормального розподілу. Для оцінки якості послідовного критерію потрібно побудувати оперативну характеристику та функцію середнього обсягу вибірки.

Аналіз останніх досягнень. Основний внесок у теорію послідовного аналізу зробив А. Вальд, який запропонував критерій відношення ймовірностей. У наступних роботах А.М. Бендерського,

І.С. Гродзенської, А.Е. Башарінова цей метод набув подальшого розвитку. У роботі Р.-Г.Егера та Е.Б.Цоя [2] був розглянутий загальний випадок перевірки гіпотез для випадку багатьох параметрів.

Мета роботи. Необхідно розробити обчислювальну технологію перевірки гіпотези про одночасно задані два параметри нормального розподілу.

Основна частина. Нехай $x_1, x_2, \dots$ – послідовність однаково розподілених випадкових величин із щільністю

$$f(x, m, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Будемо вважати елементи вибірки незалежними. Розглянемо гіпотезу вигляду

$$H_0 : m = m_0; \sigma^2 = \sigma_0^2,$$

$$H_1 : m = m_1; \sigma^2 = \sigma_1^2,$$

де $m_0 \neq m_1$ та/або $\sigma_0^2 \neq \sigma_1^2$. У випадку, коли $\sigma_0^2 = \sigma_1^2$, σ_0^2 відоме, маємо звичайний випадок гіпотези про середнє з відомою (фіксованою) дисперсією, що був розглянутий у [1].

Аналогічно у випадку, коли $m_0 = m_1$, m_0 відоме, маємо випадок гіпотези про дисперсію у випадку з відомим (фіксованим) середнім.

Складаємо за звичайною схемою відношення правдоподібності, отримуємо

$$l(x_1, x_2, \dots, x_n) = \prod_{i=1}^n \frac{f(x_i, m_1, \sigma_1^2)}{f(x_i, m_0, \sigma_0^2)} = \prod_{i=1}^n e^{(a + bx_i + cx_i^2)},$$

де

$$a = \ln \frac{\sigma_0}{\sigma_1} + \frac{1}{2} \left(\frac{m_0^2}{m_1^2} - \frac{m_1^2}{m_0^2} \right), b = \frac{m_1}{\sigma_1^2} - \frac{m_0}{\sigma_0^2}, c = \frac{1}{2} \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right).$$

Послідовний критерій відношення ймовірностей здійснюється аналогічним чином до наведених у [1] гіпотез, тобто шляхом використання загальної наведеної Вальдом схеми перевірки. На кожному етапі експерименту визначається відношення $l(x_1, x_2, \dots, x_n)$.

Додаткові спостереження проводяться до тих пір, поки

$$B \leq l(x_1, x_2, \dots, x_n) \leq A,$$

де A, B – апроксимації Вальда:

$$A = \frac{1-\beta}{\alpha},$$

$$B = \frac{\beta}{1-\alpha}.$$

Логарифмуємо

$$\ln B \leq \ln l(x_1, x_2, \dots, x_n) \leq \ln A.$$

У випадку, коли

$$\ln l(x_1, x_2, \dots, x_n) \leq \ln A,$$

приймаємо головну гіпотезу; якщо

$$\ln l(x_1, x_2, \dots, x_n) \geq \ln B,$$

то головна гіпотеза відхиляється.

У випадку, якщо виконується нерівність

$$\begin{aligned} & \sigma_1^2 \sigma_0^2 \ln B - n \sigma_1^2 \sigma_0^2 \left(\ln \frac{\sigma_0}{\sigma_1} + \frac{1}{2} \left(\frac{m_0^2}{\sigma_0^2} - \frac{m_1^2}{\sigma_1^2} \right) \right) \leq \\ & \leq \frac{\sigma_1^2 - \sigma_0^2}{2} \sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i (\sigma_0^2 m_1 - \sigma_1^2 m_0) \leq \\ & \leq \sigma_1^2 \sigma_0^2 \ln A - n \sigma_1^2 \sigma_0^2 \left(\ln \frac{\sigma_0}{\sigma_1} + \frac{1}{2} \left(\frac{m_0^2}{\sigma_0^2} - \frac{m_1^2}{\sigma_1^2} \right) \right), \end{aligned}$$

продовжуємо спостереження.

Найбільш важливими проблемами щодо опису статистичних властивостей послідовного критерію є обчислення його характеристик: оперативної характеристики та функції середнього обсягу вибірки.

Оперативна характеристика $L(m, \sigma^2)$ визначає ймовірність того, що процес завершиться прийняттям гіпотези H_0 , коли істинне значення параметрів дорівнює m, σ^2 . Вальдом [1] отримані наближені формули для оперативної характеристики у випадку, коли було знехтувано перевищенням величиною $l(x_1, x_2, \dots, x_n)$ границь A та B у момент завершення процесу. Почнемо з відношень, що визначають оперативну характеристику, яка у послідовному випадку має, окрім самостійного теоретичного значення, ще й важливу допоміжну роль, оскільки за її допомогою будується функція середнього обсягу випробувань послідовної процедури. В [1] було запропоновано для

графічного представлення оперативної характеристики розв'язувати рівняння відносно $h(m, \sigma^2)$:

$$\int_{-\infty}^{\infty} \left[\frac{f(x, m_1, \sigma_1^2)}{f(x, m_0, \sigma_0^2)} \right]^h f(x, m, \sigma^2) dx = 1.$$

Але такий метод для будь-якої гіпотези відносно декількох параметрів є трудомістким, тож у даній роботі було використано поняття конгруентних пар параметрів, як запропоновано у [2].

Можна продемонструвати, що так званий метод конгруентних пар параметрів може бути застосований до даної гіпотези. Оперативна характеристика та функція середнього обсягу вибірки для послідовного критерію Вальда може бути обчислена наближено у сенсі так званої апроксимації Вальда у значеннях конгруентних пар параметрів.

Дві пари параметрів (m, σ^2) та (m', σ'^2) називаються конгруентними, якщо існує дійсне число $h \neq 0$, причому

$$l_{n, m, \sigma^2, m', \sigma'^2} = l_{n, m_0, \sigma_0^2, m_1, \sigma_1^2}^h, n = 1, 2, \dots,$$

та позначаються

$$(m, \sigma^2) \xleftrightarrow{h} (m', \sigma'^2).$$

Вальдом було доведено, що побудований критерій завершує дію з імовірністю 1.

Це співвідношення еквівалентне

$$l_{1, m, \sigma^2, m', \sigma'^2} = l_{1, m_0, \sigma_0^2, m_1, \sigma_1^2}^h,$$

покладемо тепер

$$Z_{1, m, \sigma^2, m', \sigma'^2} = \ln l_{1, m, \sigma^2, m', \sigma'^2}; Z_{1, m_0, \sigma_0^2, m_1, \sigma_1^2} = \ln l_{1, m_0, \sigma_0^2, m_1, \sigma_1^2}$$

$$Z_{1, m_0, \sigma_0^2, m_1, \sigma_1^2} = a + b'x_1 + c'x_1^2$$

та

$$Z_{1, m, \sigma^2, m', \sigma'^2} = a' + b'x_1 + c'x_1^2.$$

Таким чином, маємо

$$a' + b'x_1 + c'x_1^2 = ha + hb'x_1 + hc'x_1^2.$$

Порівнюючи коефіцієнти, за означенням конгруентних пар параметрів $(m, \sigma^2, m', \sigma'^2) \xleftrightarrow{h} (m_0, \sigma_0^2, m_1, \sigma_1^2)$, як у [2], маємо

$$a' = ha, b' = hb, c' = hc.$$

Отже, для заданих m та σ^2 ці три рівності утворюють нелінійну систему рівнянь із невідомими m', σ'^2 та h . Розглядаючи оперативну характеристику та функцію середнього обсягу вибірки, вважаємо, що достатньо виконати обчислення лише значень h .

Таким чином, шляхом перетворень отримуємо

$$\sigma'^2 = \frac{1}{\frac{1}{\sigma^2} - 2hc}; m' = \sigma^2(hb + \frac{m}{\sigma^2})$$

– вирази для h , що залежить від m та σ^2 , для цих параметрів задаємо бажаний діапазон та для кожної пари обчислюємо значення h методом золотого перерізу, діапазон для невідомого отримуємо шляхом підбору.

Таким чином, результуючий вираз для h має вигляд

$$h = \frac{\ln \frac{\sigma}{\sigma'} + \frac{1}{2}(\frac{m^2}{\sigma^2} - \frac{m'^2}{\sigma'^2})}{\sigma}.$$

Після обчислення значення h побудова оперативної характеристики не викликає труднощів. Ця побудова виконується за стандартною формулою Вальда

$$L(\theta) \approx \frac{A^h - 1}{A^h - B^h}.$$

Середній обсяг вибірки для винесення остаточного рішення, коли істинне значення параметрів m, σ^2 і $E\{n\} \neq 0$, має вигляд

$$E\{n\} \approx \frac{L(\theta) \ln B + [1 - L(\theta)] \ln A}{\int_{-\infty}^{\infty} \left[\frac{f(x, m_1, \sigma_1^2)}{f(x, m_0, \sigma_0^2)} \right] f(x, m, \sigma^2) dx}.$$

Обчислення інтеграла у знаменнику виконується якимось чисельним методом, наприклад методом Сімпсона.

Приклад. Розглянемо такі гіпотези про параметри нормального розподілу:

$$H_0 : m = 0; \sigma^2 = 1;$$

$$H_1 : m = 0,5; \sigma^2 = 1,6.$$

Для оцінки якості послідовного критерію Вальда треба побудувати оперативну характеристику та функцію середнього обсягу вибірки.

Ймовірності похибок 1-го та 2-го роду обрано такими: $\alpha = 0.01$; $\beta = 0.01$.

Обрані діапазони параметрів для побудови:

$$m = -3..1.4,$$

$$\sigma^2 = 0.5..5,$$

$$h = -2..2.$$

Результати побудови оперативної характеристики та функції середнього обсягу вибірки представлено на рис. 1 та 2.

Побудована оперативна характеристика задовольняє необхідним умовам та є достатньою, тобто для всіх параметрів, що знаходяться в області прийняття головної гіпотези, оперативна характеристика є більшою за $1-\alpha$, а для параметрів, що знаходяться в області відхилення головної гіпотези, оперативна характеристика менша за β . Функція середнього обсягу вибірки демонструє, яке у середньому число випробувань необхідне для критерію Вальда при заданих гіпотезах та значеннях параметрів m, σ^2 .

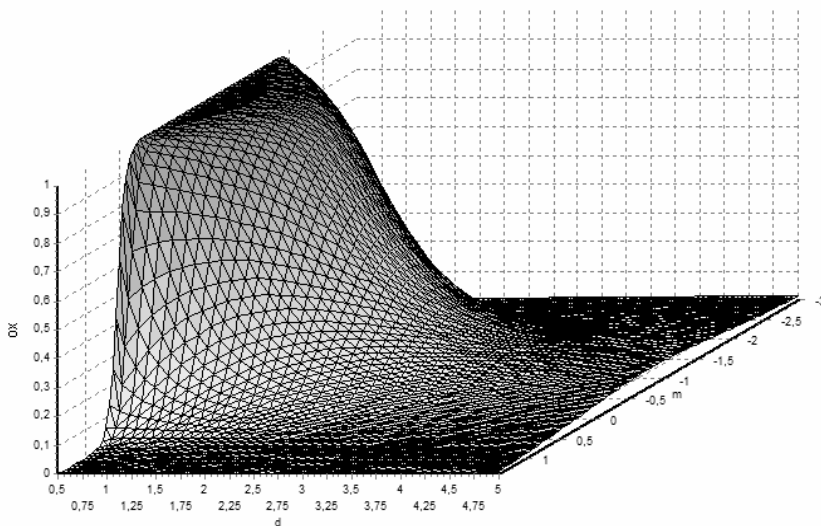


Рис.1. Оперативна характеристика

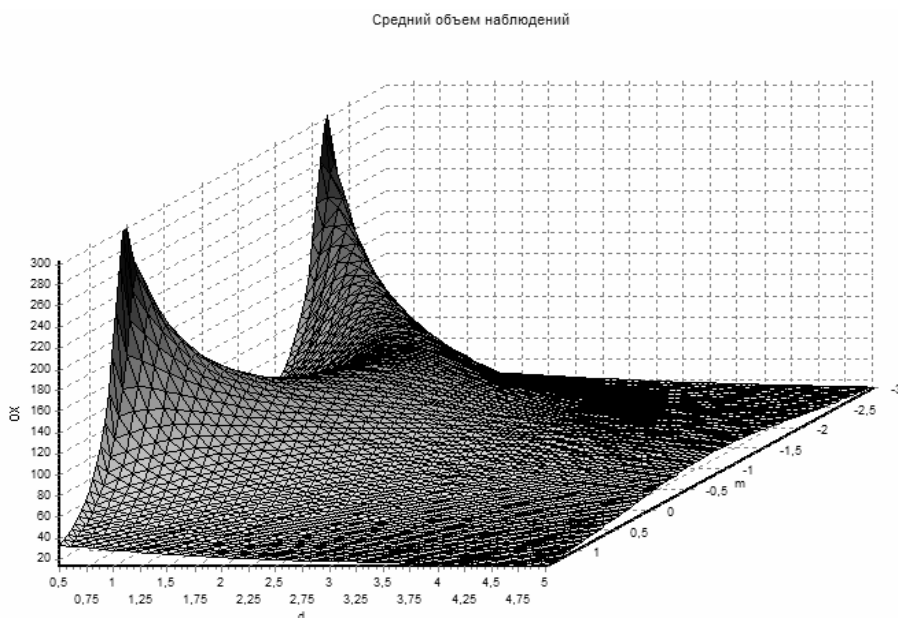


Рис.2. Функція середнього обсягу вибірки

Висновки та перспективи подальшого розвитку. Було розроблено обчислювальну схему перевірки гіпотези двох параметрів нормального розподілу одночасно на основі послідовного критерію відношення ймовірностей Вальда. Надалі вважаємо за доцільне розробити обчислювальні схеми розрахунку точних границь послідовного критерію для реалізації на ЕОМ.

Бібліографічні посилання

1. **Вальд А.** Последовательный анализ / А. Вальд. – М.: Физматгиз, 1960. – 307 с.
2. **Eger R.-H.** Robustness of Sequential Probability Ratio Tests in Case of Nuisance Parameters / R.-H.Eger, E.Б.Цой // The third international forum on strategic technologies: proc. of IFOST 2008, Novosibirsk; Tomsk (Russia), 23–29 june 2008. – P. 266–270.

Надійшла до редколегії 25.06.2013 р.

ЗАСТОСУВАННЯ АНСАМБЛІВ АЛГОРИТМІВ ДЛЯ ПІДВИЩЕННЯ СТІЙКОСТІ РЕЗУЛЬТАТІВ КЛАСТЕРИЗАЦІЇ

Здійснено огляд існуючих підходів застосування ансамблів алгоритмів у кластерному аналізі. Запропоновано інформаційну технологію підвищення стійкості результатів кластеризації даних медичного обстеження пацієнтів.

Ключові слова: кластерний аналіз, ансамблі алгоритмів, інформаційна технологія.

Осуществлен обзор существующих подходов применения ансамблей алгоритмов в кластерном анализе. Предложена информационная технология повышения устойчивости результатов кластеризации данных медицинского обследования пациентов.

Ключевые слова: кластерный анализ, ансамбли алгоритмов, информационная технология.

A review of existing cluster ensembles techniques has been conducted. The information technology of enhance the stability of the clustering results of medical examinations of patients has been offered.

Keywords: cluster analysis, cluster ensembles, information technology.

Вступ. На даний момент серед великої кількості підходів до розв'язання задачі кластеризації не існує універсальних методів. Кожен підхід має свої переваги та недоліки. Результат суттєво залежить від структури досліджуваних даних, вибору системи ознак, мір близькості, способів формалізації уявлень про схожість об'єктів та кластерів. Випадковий необгрунтований вибір методу може призвести до того, що отримане ним розбиття буде зовсім відмінним від природної, притаманної досліджуваним даним кластерної структури. Тому актуальними проблемами кластерного аналізу є оцінювання результатів для пошуку розбиття, що найкраще відповідає структурі даних, та розробка колективних методів кластерного аналізу (побудова ансамблю алгоритмів), що дозволить отримувати найбільш узгоджені та стійкі розв'язки.

Застосування ансамблів алгоритмів у кластерному аналізі є досить актуальним напрямом досліджень, оскільки на основі даного підходу може бути вирішено багато задач, таких як підвищення точності та стійкості результатів, зменшення простору ознак, кластеризація різнотипних даних, розпаралелювання обчислень та ін.

Аналіз публікацій. У загальному вигляді ідею застосування ансамблю алгоритмів для отримання колективного розв'язку задачі кластеризації можна подати у вигляді схеми (див. рисунок).



Рис. Загальна ідея ансамблевого підходу

Таким чином, можна виділити основні етапи ансамблевого підходу до кластеризації:

- 1) отримання індивідуальних розв'язків;
- 2) групування результатів (ансамбль кластеризацій);
- 3) визначення узагальненого (результуючого) розв'язку задачі.

Отримання індивідуальних розв'язків. На цьому етапі отримують набір $G = \{G_1, G_2, \dots, G_T\}$ різних розв'язків кластеризації для подальшого їх об'єднання, де $G_t, t = \overline{1, T}$ – варіант розбиття (угруповання) об'єктів на K_t кластерів. Залежно від задачі способи отримання індивідуальних розв'язків можуть бути такими: застосування до вихідних даних різних алгоритмів кластеризації; застосування одного алгоритму з різними значеннями параметрів (у тому числі кількістю кластерів) та/або початкових значень (центрів, степенів приналежності тощо); розбиття простору ознак на підгрупи та по чергове застосування алгоритму кластеризації до даних, що характеризуються кожною підгрупою ознак, та ін.

Групування результатів та визначення узагальненого розв'язку. Серед найбільш популярних (найчастіше вживаних у літературі та на практиці) можна виділити такі підходи:

1. *Прямий підхід*, або перевизначення кластерів. Даний підхід передбачає визначення відповідності між позначеннями кластерів різних індивідуальних розв'язків та застосування алгоритму голосування для отримання результуючого розбиття [1–4]. Найчастіше для перевизначення кластерів застосовують угорський алгоритм. Слід зазначити, що кількість кластерів у кожному вихідному розбитті та узагальненому розв'язку має бути однаковою. Ще одним недоліком даного підходу є значна складність.

2. *Графовий підхід*. Основною ідеєю даного підходу є перетворення результатів індивідуальних кластеризацій у граф (гіперграф) та застосування алгоритмів розділення графа (гіперграфа) для отримання результуючого розбиття. У роботі [5] запропоновано три графово-ансамблевих алгоритми: CSPA (The cluster-based similarity partitioning algorithm), HGPA (Hypergraph partitioning algorithm) та MCLA (Meta-clustering algorithm). Робота [6] присвячена методу HBGF (Hybrid Bipartite Graph Formulation), що полягає у побудові двостороннього графа, вершинами якого є як об'єкти вихідної вибірки, так і отримані кластери. Також графові підходи розглядаються у роботах [3; 7; 8].

3. *Матричний підхід*. Даний підхід є найбільш поширеним у застосуванні на практиці [3; 9; 10]. У загальному випадку складається із двох етапів: об'єднання результатів індивідуальних кластеризацій у вигляді матриці та визначення на її основі результуючого розбиття шляхом застосування будь-якого методу кластерного аналізу. Існує багато варіацій та різновидів даного підходу залежно від способу побудови узагальненої матриці та вибору визначального алгоритму кластеризації. У роботі [11] проведено експериментальне порівняння 24 методів даної групи на 24 різних наборах даних.

4. *Імовірнісний підхід*. Даний підхід ґрунтується на імовірнісних моделях, статистичних властивостях вихідних даних та результатів індивідуальних кластеризацій. У роботі [8] пропонується представляти результати набору окремих кластеризацій у вигляді нових атрибутів, що характеризують об'єкти вихідних даних. Задача знаходження узагальненого розв'язку розглядається як задача відновлення суміші поліноміальних розподілів та вирішується за допомогою ЕМ-алгоритму. Ще один метод імовірнісного ансамблевого підходу BCE (Bayesian Cluster Ensembles) запропоновано у роботі [12].

У роботі [13] подається огляд основних проблем, задач та методів ансамблевого підходу кластеризації.

Досить мало робіт присвячено ансамблям нечітких методів кластеризації. Проте в останні роки все більше уваги приділяється даному напрямку [7; 14; 15].

Постановка задачі. Запропонувати інформаційну технологію кластеризації із застосуванням ансамблевого підходу для підвищення стійкості результатів кластерного аналізу даних медичного обстеження пацієнтів, хворих на серцеву недостатність. Дані представлені у вигляді матриці $X = \{x_{ij}; i = 1, N, j = 1, p\}$, де N – кількість пацієнтів, p – кількість досліджуваних ознак, x_{ij} – значення j -ї ознаки, що спостерігається в i -го пацієнта, дійсне число.

Основний матеріал. Для вирішення поставленої задачі пропонується обчислювальна технологія, що складається з таких етапів:

1. Попередня обробка даних. Перш ніж застосовувати алгоритми кластерного аналізу, для підвищення їхньої точності необхідно провести попередню обробку даних, що полягає у відборі інформативних ознак та стандартизації даних.

2. Визначення набору індивідуальних роз'язків задачі кластеризації. Застосовуючи різні методи кластерного аналізу до вихідних даних, отримуємо набір угруповань (індивідуальних розв'язків) $G = \{G_1, G_2, \dots, G_T\}$, де $G_t = \{g_1^{(t)}, g_2^{(t)}, \dots, g_{K_t}^{(t)}\}$, $t = \overline{1, T}$,

$g_l^{(t)} = \{x_{lj}\}$, $i = \overline{1, K_t}$, $l = \overline{1, N_i^{(t)}}$, $x_{lj} = \{x_{lj}\}$, $j = \overline{1, p}$, K_t – кількість кластерів у t -му угрупованні, $N_i^{(t)}$ – кількість об'єктів у i -му кластері t -го

угруповання. $\sum_{i=1}^{K_t} N_i^{(t)} = N$, $\bigcup_{i=1}^{K_t} g_i^{(t)} = X$, $g_i^{(t)} \cap g_j^{(t)} = \emptyset$, $i, j = \overline{1, K_t}, i \neq j$.

Для того щоб набір індивідуальних роз'язків містив найбільш повну інформацію про угруповання, притаманне досліджуваним даним, пропонується застосовувати різноманітні алгоритми кластерного аналізу, а саме: алгоритми швидкої ієрархічної агломеративної кластеризації (одиничного зв'язку (ближнього сусіда), повного зв'язку (дальнього сусіда), середнього зв'язку, центрального та Уорда), графовий метод найкоротшого незамкненого шляху, енетичний алгоритм, методи розділової кластеризації К-середніх у варіантах Болла–Холла та Мак-Кіна, а також метод Forel [16; 17].

3. Оцінка якості отриманих результатів для виключення з подальшого аналізу неякісних розв'язків. Існують різноманітні функціонали та індекси якості, які дозволяють порівнювати отримані різними методами розбиття за певним обраним критерієм. Найуживанішими є сума внутрішньокластерних дисперсій за всіма ознаками, сума квадратів відстаней до центрів класів, сума внутрішньокластерних відстаней, відношення середньої внутрішньокластерної і середньої міжкластерної відстаней, індекси Каліфського – Гарабача, Данна, Беджека – Данна та ін. [17]. Для отримання багатокритеріальної оцінки пропонується агрегувати значення індексів якості за допомогою колективних методів прийняття рішень, наприклад Борда, Коупленда та ін.

4. Побудова ансамблю кластеризацій. Для отримання стійкого угруповання об'єднуємо результати індивідуальних кластеризацій у матрицю узгодженості таким чином:

а) створюємо матрицю $S = \{s_{ij}\}; i, j = \overline{1, N}$ та ініціалізуємо її нулями: $s_{ij} = 0; i, j = \overline{1, N}$;

б) розглядаємо по черзі угруповання з набору індивідуальних кластеризацій $G_t; t = \overline{1, T}$. Якщо i -й та j -й об'єкти у t -му угрупованні належать до одного кластера, то s_{ij} збільшуємо на одиницю: $s_{ij} = s_{ij} + 1$, інакше значення s_{ij} залишаємо без змін;

в) зводимо елементи матриці подібності до одиничної шкали: $s_{ij} = \frac{s_{ij}}{T}; i, j = \overline{1, N}$. Після такого перетворення s_{ij} набувають значень на відріжку від 0 до 1;

г) здійснюємо перетворення елементів матриці: $s_{ij} = 1 - s_{ij}; i, j = \overline{1, N}$;

д) визначення узагальнюючого розв'язку. Підсумкове розбиття $G' = \{g_1, g_2, \dots, g_{K'}\}, \bigcup_{i=1}^{K'} g_i = X, g_i \cap g_j = \emptyset,$

$i, j = \overline{1, K'}, i \neq j$ можна отримати, застосовуючи до матриці S алгоритми кластерного аналізу, за вихідну інформацію використовують матрицю відстаней між об'єктами (наприклад, ієрархічні або графові методи).

Для визначення нечіткого розбиття на кластери, тобто коли кожен об'єкт відноситься до кожного кластера з певним ступенем приналежності $\mu_{li} \in [0,1]$, $i = \overline{1, N}$, $l = \overline{1, K'}$, $\sum_{l=1}^{K'} \mu_{li} = 1$, слід до

матриці S застосовувати нечіткі методи кластерного аналізу, які за вихідну інформацію використовують матрицю відстаней між об'єктами (наприклад, Уіндхем, Дева–Сена).

Розглянемо результати застосування запропонованої технології до даних медичного обстеження хворих на серцеву недостатність. Обчислювальні схеми усіх складових частин технології реалізовано в авторському програмному забезпеченні «Medisa».

Досліджувані дані були отримані за допомогою Допплер-Ехокардіографії та зібрані в Українському державному науково-дослідному інституті медико-соціальних проблем інвалідності. У 394 пацієнтів замірялися такі показники: кінцево-діастолічний розмір лівого шлуночка, кінцево-сistolічний розмір лівого шлуночка, кінцево-діастолічний об'єм, кінцево-сistolічний об'єм, фракція викиду та систолічний тиск легеневої артерії. Метою аналізу було розподілення пацієнтів на п'ять груп, що відповідають стадіям серцевої недостатності та відсутності хвороби (див. табл.).

Таблиця

Результати кластеризації

Метод	Індекс Камфського – Гарабача	Індекс Данна	Індекс Беджека – Данна
Швидкий ієрархічний повного зв'язку (дальнього сусіда)	936,2	0,04	0,61
Швидкий ієрархічний Уорда	1129,1	0,05	0,57
К-середніх Болла – Холла	860,8	0,02	0,51
К-середніх Мак-Кіна	1081,7	0,01	0,64
Forel	865,8	0,06	0,57
Генетичний	773,54	0,01	0,04
Узагальнений розв'язок (швидкий ієрархічний метод середнього зв'язку)	1040,5	0,04	0,64

У таблиці представлено оцінки якості отриманих результатів після відсіювання неякісних угруповань. Чим більше значення індекса, тим якісніший результат. В останньому рядку наведено оцінки

узагальненого розв'язку, отриманого швидким ієрархічним методом середнього зв'язку. Як бачимо, даний результат є, можливо, не найкращий, проте один із найякісніших, а головн – стійкий. Таким чином, застосовуючи ансамблевий підхід, ми зменшуємо ризик отримання неякісного розбиття в умовах невизначеності.

Для більш детального аналізу кластерної структури досліджуваних даних був отриманий нечіткий узагальнений розв'язок методом Уіндхема, що демонструє степені приналежності об'єктів до кожного з кластерів.

Висновки. У даній роботі проведено огляд існуючих підходів застосування ансамблів алгоритмів у кластерному аналізі. Запропоновано інформаційну технологію підвищення стійкості результатів та отримання нечіткого узагальненого розв'язку кластеризації даних медичного обстеження пацієнтів, хворих на серцеву недостатність. Розроблено обчислювальні схеми та програмне забезпечення, проведено практичну апробацію на реальних даних. Дана технологія може бути застосована і в інших предметних галузях.

Бібліографічні посилання

1. **Dudoit S.** Bagging to improve the accuracy of a clustering procedure / S. Dudoit, J. Fridlyand // *Bioinformatics oxford university*. – 2003. – Vol. 19, No. 9. – P. 1090–1099.
2. **Fischer B.** Bagging for path-based clustering / B. Fischer, J. M. Buhmann // *IEEE Transaction on Pattern Analysis and Machine Intelligence*. –2003. – Vol. 25, No.11. – 7 p.
3. **Бирюков А. С.** Решение задач кластерного анализа коллективами алгоритмов / А.С. Бирюков, В.В. Рязанов, А.С. Шмаков // *Журнал вычислительной математики и математической физики*. – 2008. – Т. 48, № 1. – С. 176–192.
4. **Topchy A.** Adaptive clustering ensembles / A. Topchy, B. Minaei Bidgoli, A. K. Jain, W. Punch Proceeding International Conference on Pattern Recognition (ICPR), Cambridge, UK, 2004. – P. 272–275.
5. **Strehl A.** Cluster ensembles – a knowledge reuse framework for combining multiple partitions / A. Strehl, J. Ghosh // *Journal on Machine Learning Research*. – Feb. 2002. – P. 583–617.
6. **Fern X. Z.** Solving cluster ensemble problems by bipartite graph partitioning / X. Z. Fern, C. E. Brodley // *Proceedings of the 21 st International Conference on Machine Learning, Canada, 2004*.
7. **Ahmadzadeh M. A.** Graph Based Approach for Clustering Ensemble of Fuzzy Partitions / M. Ahmadzadeh, Z. Golestan, J. Vahidi, B. Shirazi // *Journal of mathematics and computer Science* –2013. – Vol. 6. – P. 154–165.

8. **Topchy A.** A mixture model for clustering ensembles / A. Topchy, A. Jain, W. Punch // SIAM International Conference on Data Mining, 2004. – P. 379–390.
9. **Fred A.** Combining Multiple Clusterings Using Evidence Accumulation / A. Fred, A. Jain // IEEE Transaction on Pattern Analysis and Machine Intelligence –2005. – Vol. 27, No. 6. – P. 835–850.
10. **Pestunov I. A.** Ensemble of Clustering Algorithms for Large Datasets / I. A. Pestunov, V. B. Berikov, E. A. Kulikova, S. A. Rylov // Optoelectronics, Instrumentation and Data Processing. – 2011, Vol. 47, No. 3. – P. 245–252.
11. **Kuncheva L. I.** Experimental comparison of cluster ensemble methods / L. I. Kuncheva, S. T. Hadjitodorov, L. P. Todorova, B. Sofia // Proc. 9th International Conference on Information Fusion, July 2006. – 7 p.
12. **Wang H.** Bayesian Clustering Ensemble / H. Wang, H. Shan, A. Banerjee // SIAM international conference on data mining (SDM). Sparks, Nevada, USA. 29April – 2 May, 2009. – P. 211–222.
13. **Reza G.** A Survey: Clustering Ensembles Techniques / G. Reza, S. Md. Nasir, I. Hamidah, M. Norwati // Proceedings of World Academy of Science, Engineering And Technology. – February 2009 – Vol. 26. – P. 636–645.
14. **Li Taoying** Fuzzy Clustering Ensemble with Selection of Number of Clusters / Taoying Li, Yan Chen // Journal Of Computers. – July 2010. – Vol. 5, №. 7. – P. 1112–1119.
15. **Avogadri R.** Fuzzy ensemble clustering based on random projections for DNA microarray data analysis / R. Avogadri, G. Valentini // Artificial Intelligence in Medicine, 2008. – P. 173–183.
16. **Мандель И. Д.** Кластерный анализ / И. Д. Мандель. – М., 1988. – 176 с.
17. **Айвазян С. А.** Классификация многомерных наблюдений / С. А. Айвазян, З. И. Бежаева, О. В. Староверов. – М., 1974. – 240 с.

Надійшла до редколегії 26.06.2013 р.

О.М. Мацуга*, І.В. Дроздова**

**Дніпропетровський національний університет імені Олеся Гончара*

***Державна установа «Український держаний науково-дослідний інститут медико-соціальних проблем інвалідності»*

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ КОРОТКОСТРОКОВОГО ПРОГНОЗУВАННЯ ПЕРВИННОЇ ІНВАЛІДНОСТІ В УКРАЇНІ

Розроблено інформаційну технологію прогнозування первинної інвалідності в Україні на базі комплексного застосування адаптивних методів короткострокового прогнозування.

Ключові слова: *інформаційна технологія, прогнозування, первинна інвалідність, адаптивна модель, часовий ряд, програмне забезпечення.*

Разработана информационная технология прогнозирования первичной инвалидности в Украине на основе комплексного применения адаптивных методов краткосрочного прогнозирования.

Ключевые слова: *информационная технология, прогнозирование, первичная инвалидность, адаптивная модель, временной ряд, программное обеспечение.*

The information technology of primary disability forecasting in Ukraine is developed. It is based on combined using of short-term forecasting adapt methods.

Keywords: *information technology, forecasting, primary disablement, adaptive model, time series, software.*

Постановка проблеми. Існуючий моніторинг інвалідності в Україні забезпечує отримання даних щодо первинної інвалідності в розрізі адміністративних територій, класів хвороб, груп інвалідності, верств населення тощо. Аналіз цих даних обмежений обчисленням середніх значень показників за певні роки, оцінюванням відсотка змін щодо минулого року та відносного ризику їх зростання [1]. Такий арсенал методів недостатній для поглиблення досліджень і не може забезпечити якісного аналізу динаміки інвалідності.

Актуальною є задача прогнозування інвалідності та впровадження інформаційних технологій для її розв'язання. Саме цьому присвячена дана робота.

Аналіз останніх досліджень і публікацій. Важливими етапами впровадження інформаційних технологій є розробка математичного та програмного забезпечення для розв'язання задачі, у даному випадку задачі аналізу та прогнозування первинної інвалідності.

Нині розроблено більш ніж 150 методів прогнозування. Серед них найбільш вживані методи екстраполяції. Екстраполяція базується на припущенні про збереження у майбутньому тенденцій минулого, а її основою є передбачення щодо стабільності факторів впливу. Минула тенденція представляється у вигляді часового ряду.

Під час прогнозування інвалідності знайшов застосування здебільшого класичний підхід до екстраполяції, основу якого становить виділення тренда часового ряду. Так, у роботах [2–5] автори використали апарат регресійного аналізу для відновлення тенденції зміни показника первинної інвалідності. У роботі [6] проводили згладжування ряду для визначення тенденції первинної інвалідності. У цілому застосування класичного підходу виправдано, якщо характер динаміки інвалідності незмінний і його можливо апроксимувати деякою функцією та мають місце середні або довгі часові ряди.

Серед методів екстраполяції також варто виділити методи ковзного середнього, медіанного, експоненціального згладжування, адаптивні, Бокса–Дженкінса, «Гусениці»–SSA, нейромережеві, нечіткі тощо [7–9]. З них адаптивні методи вирізняються відносною простотою застосування та можливістю давати надійні результати й бути застосованими до коротких часових рядів.

Актуальним є підхід, що базується на використанні декількох методів прогнозування у комплексі. Він дозволяє уникнути слабких місць кожного з методів окремо та підвищити ефективність прогнозу.

Серед програмних засобів для прогнозування варто відзначити універсальні статистичні пакети Statistica та SPSS, спеціалізовані програмні продукти для прогнозування EBPICTA, ОЛИМП: СтатЕксперт, ForecastPRO тощо. Усі вони забезпечують досить широкі можливості з прогнозування, проте здебільшого складні щодо застосування, не адаптовані під вирішення задач медичного спрямування та дорого коштують. Крім того, у більшості пакетів майже відсутня можливість комплексного застосування декількох методів з автоматичним підбиранням найкращого прогнозу.

Аналіз сучасного стану математичного та програмного забезпечення для прогнозування інвалідності свідчить, що розробку інформаційної технології перспективно проводити на основі

комплексу методів зі створенням власного програмного забезпечення. Слід зазначити, що результати даної роботи є продовженням авторських досліджень, опублікованих у [10]. Запропоновану раніше інформаційну технологію модифіковано шляхом додавання ще одного етапу та реалізовано в новому вдосконаленому програмному забезпеченні.

Постановка задачі. Державною установою «Український державний науково-дослідний інститут медико-соціальних проблем інвалідності МОЗ України» та відділом медико-соціальної експертизи та проблем інвалідності МОЗ України ведеться моніторинг інвалідності, у ході якого фіксуються значення первинної інвалідності на 27 адміністративних територіях та в цілому в Україні за 47 нозологічними класами хвороб серед дорослого та працездатного населення. Тим самим мають місце $28 \times 2 \times 47$ часових рядів вигляду

$$\{x'_t; t = \overline{1, n}\},$$

де x'_t – значення первинної інвалідності внаслідок хвороби x , зафіксоване у t -му році на певній адміністративній території або в цілому по Україні; n – кількість років, упродовж яких проводився моніторинг інвалідності.

Доступні дані моніторингу з 1992 по 2012 рік, тобто за 21 рік.

Ставиться задача спрогнозувати первинну інвалідність на наступний рік.

Для вирішення цієї задачі необхідно розробити інформаційну технологію короткострокового прогнозування первинної інвалідності та реалізувати її у вигляді програмного забезпечення.

Ураховуючи, що прогнозування має проводитися за короткими часовими рядами ($n = 21$), інформаційну технологію вирішено будувати на основі адаптивних методів короткострокового прогнозування [7]. З огляду на актуальність комплексного підходу до прогнозування проноз має базуватися на сумісному застосуванні декількох адаптивних моделей.

Основний матеріал. Запропонована інформаційна технологія прогнозування первинної інвалідності містить п'ять ключових етапів.

1. *Попереднє дослідження часового ряду.* Даний етап є основою для обґрунтування моделі прогнозування інвалідності. На даному етапі здійснюється перевірка часового ряду на випадковість [9] і, у випадку відхилення головної гіпотези відносно випадковості на користь альтернативної гіпотези щодо наявності тренда, проводиться відновлення тренда ряду [11]. У роботі для перевірки ряду на

випадковість використовуються критерії Спірмена, Манна, Фостера – Стюарта та критерій, заснований на знаках різниць, які є найбільш потужні за альтернативи щодо наявності тренда, а також критерій екстремальних точок, який забезпечує перевірку головної гіпотези за альтернативи щодо наявності циклічності в ряді [9].

II. *Згладжування часового ряду* з метою вилучення шумів. Введення даного етапу становить відмінність від інформаційної технології, яку автори описали у [10].

У роботі для згладжування використаний поліноміальний сплайн на основі *B*-сплайна, близький до інтерполяційного у середньому [12], переваги якого в низькій обчислювальній складності та високій якості апроксимації. Згладжені значення обчислюються на основі сплайна $S_{4,0}$ згідно із [12]:

$$x_t = \frac{1}{384} (x'_{t-2} + 76x'_{t-1} + 230x'_t + 76x'_{t+1} + x'_{t+2}), \quad t = \overline{3, n-2},$$

$$x_1 = x_2 = x'_1, \quad x_{n-1} = x_n = x'_n.$$

Слід зазначити, що раніше під час прогнозування показників гідрогеохімічного моніторингу було показано, що попереднє згладжування за допомогою даного сплайна дозволяє зменшити помилку прогнозування [13].

III. *Налаштування параметрів моделей прогнозування інвалідності*. Налаштування проводиться для кожної моделі окремо. Значення параметрів моделі змінюються з певним кроком (наприклад, кроком 0,01), і за кожних варіантів значень параметрів здійснюється прогнозування. Найкращі значення параметрів обираються за найменшою середньою, середньоквадратичною або максимальною абсолютною/відносною помилкою прогнозування або за результатами множинного аналізу [8; 11].

IV. *Прогнозування первинної інвалідності* за адаптивними моделями. Застосовуються дев'ять моделей [7], кожна з яких визначається типом тенденції у часовому ряді та наявністю циклічних коливань. Прогнозні значення інвалідності в усіх моделях розраховуються за загальною формулою

$$\hat{x}_\tau(t) = P_\tau,$$

де під $\hat{x}_\tau(t)$ мається на увазі прогноз на τ кроків наперед, зроблений у t -му році; P_τ розраховується для кожної моделі способом, що наведений у табл. 1.

Таблиця 1

Формули для обчислення прогнозів P_t [7]

Характер сезонності Тип тенденції	Сезонний ефект відсутній (1)	Адитивний сезонний ефект (2)	Мультиплікативний сезонний ефект (3)
Тенденція відсутня (А)	$\hat{a}_{1,t}$	$\hat{a}_{1,t} + g_{t+\tau-l}$	$\hat{a}_{1,t} f_{t+\tau-l}$
Лінійна тенденція (В)	$\hat{a}_{1,t} + \hat{a}_{2,t} \tau$	$\hat{a}_{1,t} + \hat{a}_{2,t} \tau + g_{t+\tau-l}$	$(\hat{a}_{1,t} + \hat{a}_{2,t} \tau) f_{t+\tau-l}$
Експоненційна тенденція (С)	$\hat{a}_{1,t} \hat{r}_t^\tau$	$\hat{a}_{1,t} \hat{r}_t^\tau + g_{t+\tau-l}$	$\hat{a}_{1,t} \hat{r}_t^\tau f_{t+\tau-l}$

Величини у табл. 1, за якими обчислюються прогнозні значення:

$\hat{a}_{1,t}$ – коефіцієнт, що характеризує поточний рівень ряду, обраховується за формулою

$$\hat{a}_{1,t} = \beta_1 d_1 + (1 - \beta_1) d_2, \quad 0 < \beta_1 < 1,$$

де d_1 , d_2 – величини, значення яких для кожної моделі прогнозування наведено у табл. 2;

$\hat{a}_{2,t}$ – коефіцієнт лінійного зростання, визначається за формулою

$$\hat{a}_{2,t} = \beta_2 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \beta_2) \hat{a}_{2,t-1}, \quad 0 < \beta_2 < 1;$$

$\hat{r}_t$ – коефіцієнт експоненціального зростання, який обраховується за формулою

$$\hat{r}_t = \beta_r \frac{\hat{a}_{1,t}}{\hat{a}_{1,t-1}} + (1 - \beta_r) \hat{r}_{t-1}, \quad 0 < \beta_r < 1;$$

$\hat{g}_t$ – адитивна сезонна компонента, що обчислюється за формулою

$$\hat{g}_t = \beta_g (x_t - \hat{a}_{1,t}) + (1 - \beta_g) \hat{g}_{t-1}, \quad 0 < \beta_g < 1,$$

де l – період сезонного явища;

$\hat{f}_t$ – мультиплікативний коефіцієнт сезонності, який розраховується за формулою

$$\hat{f}_t = \beta_f \frac{x_t}{\hat{a}_{1,t}} + (1 - \beta_f) \hat{f}_{t-1}, \quad 0 < \beta_f < 1.$$

Таблиця 2

Формули для обчислення величин d_1 та d_2 [7]

Характер сезонності Тип тенденції		Сезонний ефект відсутній (1)	Адитивний сезонний ефект (2)	Мультиплікативний сезонний ефект (3)
Тенденція відсутня (А)	d_1	x_t	$x_t - g_{t-l}$	$x_t / \hat{f}_{t-l}$
	d_2	$\hat{a}_{1,t-1}$	$\hat{a}_{1,t-1}$	$\hat{a}_{1,t-1}$
Лінійна тенденція (В)	d_1	x_t	$x_t - g_{t-l}$	$x_t / \hat{f}_{t-l}$
	d_2	$\hat{a}_{1,t-1} + \hat{a}_{2,t-1}$	$\hat{a}_{1,t-1} + \hat{a}_{2,t-1}$	$\hat{a}_{1,t-1} + \hat{a}_{2,t-1}$
Експоненційна тенденція (С)	d_1	x_t	$x_t - g_{t-l}$	$x_t / \hat{f}_{t-l}$
	d_2	$\hat{a}_{1,t-1} \hat{f}_{t-1}$	$\hat{a}_{1,t-1} \hat{f}_{t-1}$	$\hat{a}_{1,t-1} \hat{f}_{t-1}$

Для застосування наведених адаптивних моделей прогнозування потрібно знати початкові значення оцінок коефіцієнтів моделей.

Початкове значення оцінки $\hat{a}_{1,0}$ становить середнє арифметичне перших k членів часового ряду:

$$\hat{a}_{1,0} = \frac{1}{k} \sum_{t=1}^k x_t.$$

Для оцінки $\hat{a}_{2,0}$ використовується формула

$$\hat{a}_{2,0} = \frac{x_k - x_1}{k - 1}.$$

При цьому, як правило, $k = 2$ або $k = 3$.

Початкові оцінки адитивної сезонної компоненти визначаються як

$$\hat{g}_t = 0 \quad \text{або} \quad \hat{g}_t = x_t - \bar{x}, \quad t = \overline{1, l},$$

а оцінки мультиплікативного коефіцієнта сезонності обчислюються як

$$\hat{f}_t = 1 \quad \text{або} \quad \hat{f}_t = \frac{x_t}{\bar{x}}, \quad t = \overline{1, l}.$$

В. Прийняття рішення щодо найбільш адекватної моделі прогнозування. Прийняття рішення може бути здійснене одним із трьох способів.

1. За результатами попереднього дослідження часового ряду (перший етап) обирається адаптивна модель без тренда (А), якщо за критеріями Спірмена, Манна, Фостера–Стюарта та заснованого на знаках різниць була прийнята головна гіпотеза про випадковість ряду. В іншому випадку, якщо подальші дослідження показали, що найбільш адекватним є лінійний тренд, обирається відповідна модель (В), інакше перевага надається моделі з експоненційним трендом (С). У разі прийняття головної гіпотези щодо випадковості ряду за критерієм екстремальних точок обирається модель без сезонності (1), інакше – із сезонністю (2 або 3). Вибір між адитивним та мультиплікативним сезонним ефектами може бути зроблений за візуальним аналізом часового ряду.

2. Обирається модель із найменшою середньою, середньоквадратичною, максимальною абсолютною або відносною помилкою прогнозування.

Для фіксованого моменту часу t визначається помилка прогнозування:

- абсолютна

$$e_t = |x_t - \hat{x}_1(t-1)|,$$

де x_t – спостережуване значення ряду в момент t ; $\hat{x}_1(t-1)$ – прогнозоване значення (прогноз робиться в момент часу $(t-1)$ на один крок уперед);

- або відносна

$$e_t = \frac{|x_t - \hat{x}_1(t-1)|}{x_t} \cdot 100\%.$$

Помилка прогнозування визначається для моментів часу $t = \overline{k, n}$, де $k \geq [n/2]$. Тим самим формується масив абсолютних або відносних помилок $\{e_t; t = \overline{k, n}\}$, за яким обчислюється:

- середнє

$$o_1 = \frac{1}{n - k + 1} \sum_{t=k}^n e_t,$$

- середньоквадратичне відхилення

$$o_2 = \frac{1}{n-k} \sum_{t=k}^n (e_t - o_1)^2,$$

- максимальне значення

$$o_3 = \max_{t=k, n} e_t.$$

3. Обирається найбільш адекватна модель на основі множинного аналізу [8; 11], у якому під час вибору враховуються всі види помилок прогнозування. Експертами при цьому є помилки прогнозування. Кожна модель оцінюється шістьма експертами, і формується матриця оцінок $\{o_{h,j}; h = \overline{1,9}, j = \overline{1,6}\}$. Оцінки зводяться до єдиного масштабу, наприклад

$$o_{h,j} = \frac{o_{h,j}}{\sum_{h=1}^9 o_{h,j}}.$$

Тоді оцінки адекватності моделей μ_h , $h = \overline{1,9}$ та адекватності експертів λ_j , $j = \overline{1,6}$ визначаються ітераційно за формулами:

$$\mu_h^{(i)} = \sum_{j=1}^6 o_{h,j} \lambda_j^{(i-1)}, h = \overline{1,9},$$

$$\rho^{(i)} = \sum_{h=1}^9 \sum_{j=1}^6 o_{h,j} \mu_h^{(i)},$$

$$\lambda_j^{(i)} = \frac{1}{\rho^{(i)}} \sum_{h=1}^9 o_{h,j} \mu_h^{(i)}, j = \overline{1,6},$$

де $i = 1, 2, \dots$ – номер ітерації; $\lambda_j^{(0)} = \frac{1}{6}$, $j = \overline{1,6}$; $\sum_{j=1}^6 \lambda_j^{(i)} = 1$.

Ітераційний процес завершується, коли значення адекватностей перестають змінюватися. Найкращою вважається модель прогнозування, що має мінімальну оцінку адекватності μ_h , $h = \overline{1,9}$.

Інформаційну технологію реалізовано у вигляді програмного забезпечення «AFORD» («Analysis and FORecast of Disability»), яке відрізняється від створеного раніше «DisStat» більшою зручністю й наочністю для користувача та внутрішньою структурою. Програмне забезпечення створено у середовищі візуальної розробки програм Embarcadero RAD Studio Delphi 10.0.

Нижче як приклад наведено результати застосування запропонованої технології для прогнозування первинної інвалідності внаслідок виразки шлунка та дванадцятипалої кишки серед дорослого населення Дніпропетровської області.

У процесі перевірки часового ряду на випадковість (табл. 3) встановлено, що у Дніпропетровській області має місце тенденція до спадання первинної інвалідності внаслідок даної нозології серед дорослого населення та наявна циклічність.

Таблиця 3

Результати перевірки гіпотези про випадковість часового ряду

Критерій	Статистика	Квантиль ($\alpha = 0,1$)	Висновок
Спірмена	4,161	1,64	Існує тенденція до спадання
Манна	-4,952	1,64	Існує тенденція до спадання
Фостера – Стюарта	-1,739	1,64	Існує тенденція до спадання
Заснований на знаках різниць	-3,693	1,64	Існує тенденція до спадання
Екстремальних точок	-3,068	1,64	Ряд циклічний

За результатами відновлення тренда (табл. 4) найбільш значущою виявилася параболічна модель, для якої F -статистика дорівнює 118,70. Дана модель пояснює 86,20% дисперсії показника інвалідності. Відновлений тренд описується рівнянням (усі параметри значущі)

$$x(t) = 0.557 - 0.042t + 0.001t^2.$$

Незначно поступається їй за адекватністю експоненційна модель

$$x(t) = 0.704 \exp(-0.132t).$$

Вона пояснює 85,12% варіабельності показника, а F -статистика для неї становить 108,69.

Слід зазначити, що відновлений тренд також може бути використаний для прогнозування, хоча поряд із ним варто було б відновити й сезонну складову часового ряду. Ураховуючи, що ряд є короткий, у роботі прогноз будували за адаптивними методами.

Таблиця 4

Результати відновлення тренда часового ряду

Модель тренда	F-статистика	Коефіцієнт детермінації
$x(t) = a + bt$	94,48	83,26%
$x(t) = a + bt + ct^2$	118,70	86,20%
$x(t) = a + b \ln t$	68,45	78,27%
$x(t) = a \exp(bt)$	108,69	85,12%
$x(t) = at^b$	36,36	65,66%

Було побудовано дев'ять адаптивних моделей, параметри кожної налаштовано за допомогою множинного аналізу. Після цього обрано найкращу модель на основі множинного аналізу. У даному разі найкращою виявилася модель з експоненційною тенденцією та адитивним сезонним ефектом (див. рисунок). Результати вибору моделі на основі множинного аналізу узгоджуються з результатами попереднього дослідження ряду.

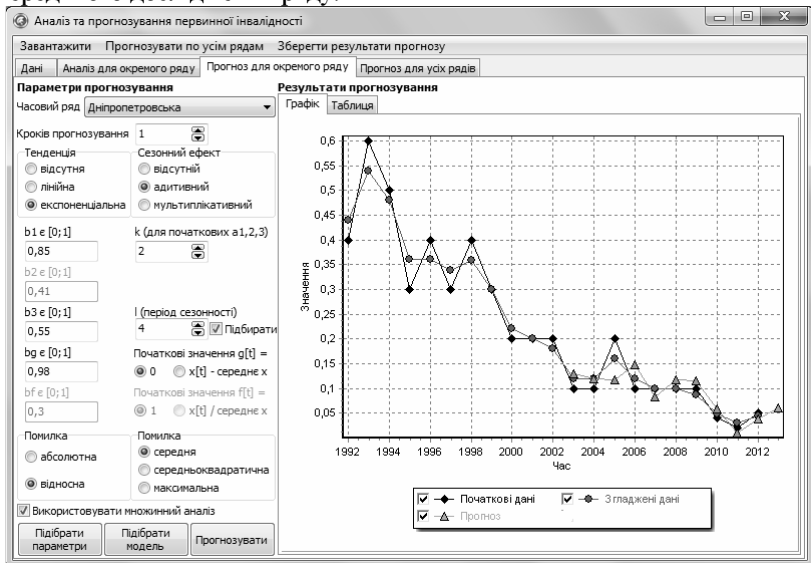


Рис. Результати прогнозування після підбору моделі

Згідно з обраною моделлю на 2013 рік прогнозується незначне зростання первинної інвалідності внаслідок виразки шлунка та дванадцятипалої кишки серед дорослого населення Дніпропетровської області (до 0,062 на 10 тис. населення порівняно з 0,05 у 2012 році).

Середня відносна помилка прогнозування становить 18,8%, середньоквадратичне відхилення відносної помилки – 10,5%. Без попереднього згладжування ряду сплайном $S_{4,0}$ середня відносна помилка прогнозування була майже на 10% вища та становила 28,3%, середньоквадратичне відхилення дорівнювало 15,1%.

Середня відносна помилка прогнозування для інших областей, класів хвороб та типу населення при використанні згладжування сплайном $S_{4,0}$ становила від 5% до 35%. Без попереднього згладжування рядів помилка була вищою на 1–10%.

Зазначимо, що розроблене програмне забезпечення дозволяє проводити прогнозування первинної інвалідності автоматично для всіх адміністративних територій України (для обраного класу хвороб та типу населення).

Висновки. Було створено інформаційну технологію прогнозування первинної інвалідності в Україні, яка базується на адаптивних методах, що дозволяє застосовувати її для коротких часових рядів.

Інформаційну технологію реалізовано у вигляді програмного забезпечення «AFORD», яке дозволяє аналізувати та прогнозувати первинну інвалідність за основними класами хвороб у розрізі адміністративних територій та верств населення. Середня відносна помилка прогнозування становить від 5% до 35% залежно від структури ряду.

Дана технологія дозволяє одержувати якісну статистичну інформацію про стан інвалідності населення України і є основою для утворення доказової бази медико-соціологічних досліджень та оцінки динаміки інвалідності.

Бібліографічні посилання

1. Основні показники інвалідності та діяльності медико-соціальних експертних комісій України за 2012 рік : аналіт. – інформ. довідник / [С.И.Черняк, А.В.Ипатов, Е.Н.Мороз та ін.]; за ред. Н.К. Хобзея. – Д. : Пороги, 2013. – 149 с.
2. **Гришина Л.П.** Анализ инвалидности в Российской Федерации за 1979–1999 годы и ее прогноз до 2015 года / Л.П. Гришина, Н.Д. Талалаева,

- Э.К. Амирова // Медико-социальная экспертиза и реабилитация. – 2001. – № 2. – С. 27–31.
3. **Землянова Е.В.** Прогноз показателей инвалидности детского населения России на период до 2020 года / Е.В. Землянова, Ж.И. Войцеховская // Социальные аспекты здоровья населения: электрон. журн. – 2009. – № 1 (9). – Режим доступа: http://vestnik.mednet.ru/content/view/112/30/lang_ru.
 4. **Мавликаева Ю.А.** Динамика общей инвалидности взрослого населения Пермского края и ее прогноз / Ю.А. Мавликаева, М.Я. Подлужная // Общественное здоровье и здравоохранение. – 2011. – № 2. – С. 5–9.
 5. **Магзумова Ш.Ш.** Прогнозирование инвалидности вследствие психозов в республике Узбекистан / Ш.Ш. Магзумова, А.А. Ганиханов // Рос. психиатр. журн. – 2009. – № 3. – С. 13–15.
 6. Динамика и прогнозирование распространенности первичной инвалидности детей вследствие врожденных пороков развития и хромосомных аномалий в Алтайском крае / [Б.А. Самарин, А.Г. Гончаренко, Е.И. Акимов, Т.А. Акопян, и др.] // Медико-социальная экспертиза и реабилитация. – 2009. – № 1. – С. 37–38.
 7. **Лукашин Ю.П.** Адаптивные методы краткосрочного прогнозирования временных рядов: учеб. пособие / Ю.П. Лукашин. – М.: Финансы и статистика, 2003. – 416 с.
 8. Принятие решений в системах мониторинга / Т.Г. Емельяненко, А.В. Зберовский, А.Ф. Приставка, Б.Е. Собко. – Д.: РИК НГУ, 2005. – 224 с.
 9. **Кендалл М.** Многомерный статистический анализ и временные ряды / М. Кендалл, А. Стьюарт. – М.: Наука, 1976. – 375 с.
 10. Інвалідність: первинна, скрита, прогнозована: монографія / упоряд. А.В. Іпатов, І.В. Дроздова, І.Я. Ханюкова, О.М. Мацуга, О.Л. Чуйко. – Д.: Пороги, 2012. – 386 с.
 11. **Бабак В.П.** Статистична обробка даних / В.П. Бабак, А.Я. Білецький, О.П. Приставка, П.О. Приставка. – К.: МІВВІЦ, 2001. – 388 с.
 12. **Приставка П.О.** Поліноміальні сплайни при обробці даних / П.О. Приставка. – Д.: РВВ ДНУ, 2004. – 236 с.
 13. **Єрещенко Н.М.** Інформаційно-аналітичне забезпечення аналізу та прогнозування показників гідрогеохімічного моніторингу при техногенному навантаженні: автореф. дис. на здобуття наук. ступеня канд. техн. наук: спец. 05.13.06 «Інформаційні технології» / Н.М. Єрещенко. – Д., 2011. – 20 с.

Надійшла до редколегії 26.06.2013 р.

ДОСЛІДЖЕННЯ *B*-СПЛАЙНА ШОСТОГО ПОРЯДКУ ТА ЇХ ЛІНІЙНОЇ КОМБІНАЦІЇ

Отримано властивості *B*-сплайна шостого порядку, визначеного на рівномірному розбитті. Досліджено властивості поліноміального сплайна на основі *B*-сплайнів шостого порядку, що є близьким до інтерполяційного у середньому.

Ключові слова: *B*-сплайн, поліноміальний сплайн, близький до інтерполяційного у середньому, норма сплайну.

Получены свойства *B*-сплайна шестого порядка, определённого на равномерном разбиении. Исследовано свойства полиномиального сплайна на основе *B*-сплайнов шестого порядка, близкого к интерполяционному в среднем.

Ключевые слова: *B*-сплайн, полиномиальный сплайн, близкий к интерполяционному в среднем, норма сплайна.

This article is the solution of practical research of the *B*-spline of six order. We obtain and research properties of polynomial spline based on the *B*-splines of six order that, related to the interpolator on average.

Keywords: *B*-spline, polynomial spline that, related to interpolator on average, norm of spline.

Постановка проблеми. Зазвичай при фіксації аналогового сигналу має місце таке. Нехай $\phi(t)$ – функція імпульсного відклику системи, що реєструє сигнал $p(t)$. У силу суто технічних властивостей систем реєстрації результатом згортки сигналу та функції відклику буде значення, усереднене на інтервалі дискретизації, зокрема

$$(p * \phi)(ih) = \frac{1}{h} \int_{ih-\frac{h}{2}}^{ih+\frac{h}{2}} p(\tau)\phi(\tau - ih)d\tau = \bar{p}_i.$$

Тоді цифровий сигнал може мати таке подання:

$$p_i = \bar{p}_i + \varepsilon_i, \quad i \in \mathbb{Z}, \quad (1)$$

де ε_i – випадкова вада. Стосовно вади ε_i , $i \in \mathbb{Z}$ можна припускати будь-який розподіл, проте за замовченням вважають, що має місце розподіл Гауса з нульовим математичним сподіванням та деякою дисперсією σ_ε^2 . Тому в задачах обробки цифрових сигналів, що задані співвідношенням (1), як модель $p(t)$ є потреба використовувати наближення із властивостями імпульсного нерекурсивного низькочастотного фільтра, наприклад лінійні комбінації B -сплайнів, близькі до інтерполяційних у середньому [1].

Відомо [2; 3], що будь-який B -сплайн порядку вище першого може бути використаний для визначення коротковіконного перетворення Фур'є; при цьому вже починаючи з п'ятого порядку B -сплайн у частотній області фактично мало чим відрізняється від гаусіана, хоча обрахунок B -сплайна п'ятого порядку [4] потребує менше обчислювальних витрат. Саме тому актуальним є дослідження лінійних комбінацій B -сплайнів вищого порядку, ніж п'ятий: від подібних сплайнів варто очікувати більш високі згладжувальні властивості при відносно низькій обчислювальній складності.

Аналіз досліджень та постановка задачі. Задачу відтворення гладких функцій на основі лінійних комбінацій B -сплайнів висвітлено в роботах І. Шоенберга, К. Де Бора, М.П. Корнійчука та ін. Увагу поліноміальним сплайнам, визначеним на локальних носіях, близьким до інтерполяційних у середньому, приділено А.О.Лигуном, О.А.Шумейко, В.В.Кармазіною [5 – 7] та у авторських дослідженнях [8].

Нехай із кроком $h > 0$ задане розбиття дійсної осі $\Delta_h : t_i = ih$, $i \in \mathbb{Z}$, у кожній точці якого отримано значення деякої неперервної функції $p(t) \in C^r$, $r \geq 2$, визначеної на $\mathbb{R}_1(-\infty; \infty)$. Вважають, що інформація про функцію $p(t)$, яка підлягає відтворенню, задана у

вузлах розбиття Δ_h у вигляді інтеграла $\bar{p}_i = \frac{1}{h} \int_{(i-0,5)h}^{(i+0,5)h} p(t) dt$; при

цьому істинне значення функції $p(t)$ у вузлах визначається аналогічно виразу (1).

Для апроксимації функції $p(t)$ за значеннями типу (1) у вузлах

розбиття Δ_h вводяться такі поліноміальні сплайни на основі B -сплайнів, що є близькими до інтерполяційних у середньому:

$$S_{r,0}(p, t) = \sum_{i \in \mathbb{Z}} p_i B_{r,h}(t - (i + 0,5)h), \quad r = 2, 4,$$

$$S_{r,0}(p, t) = \sum_{i \in \mathbb{Z}} p_i B_{r,h}(t - ih), \quad r = 3, 5,$$

де B -сплайн $B_{r,h}(t)$ порядку r ($r \geq 1$) визначається рекурентно таким чином: якщо

$$B_{0,h}(t) = \begin{cases} 0, & t \notin [-h/2; h/2], \\ 1, & t \in [-h/2; h/2], \end{cases}$$

то

$$B_{r,h}(t) = \frac{1}{h} \int_{t-h/2}^{t+h/2} B_{r-1,h}(\tau) d\tau. \quad (2)$$

Наприклад, B -сплайн п'ятого порядку є такий:

$$B_{5,h}(t) = \begin{cases} 0, & t \notin [-3h; 3h], \\ \frac{1}{120} \left(3 + \frac{t}{h} \right)^5, & t \in [-3h; -2h], \\ -\frac{1}{24} \left(\frac{t}{h} \right)^5 - \frac{3}{8} \left(\frac{t}{h} \right)^4 - \frac{5}{4} \left(\frac{t}{h} \right)^3 - \frac{7}{4} \left(\frac{t}{h} \right)^2 - \frac{5}{8} \left(\frac{t}{h} \right) + \frac{51}{120}, & t \in [-2h; -h], \\ \frac{1}{12} \left(\frac{t}{h} \right)^5 + \frac{1}{4} \left(\frac{t}{h} \right)^4 - \frac{1}{2} \left(\frac{t}{h} \right)^3 + \frac{11}{20}, & t \in [-h; 0], \\ -\frac{1}{12} \left(\frac{t}{h} \right)^5 + \frac{1}{4} \left(\frac{t}{h} \right)^4 - \frac{1}{2} \left(\frac{t}{h} \right)^3 + \frac{11}{20}, & t \in [0; h], \\ \frac{1}{24} \left(\frac{t}{h} \right)^5 - \frac{3}{8} \left(\frac{t}{h} \right)^4 + \frac{5}{4} \left(\frac{t}{h} \right)^3 - \frac{7}{4} \left(\frac{t}{h} \right)^2 + \frac{5}{8} \left(\frac{t}{h} \right) + \frac{51}{120}, & t \in [h; 2h], \\ \frac{1}{120} \left(3 - \frac{t}{h} \right)^5, & t \in [2h; 3h]. \end{cases} \quad (3)$$

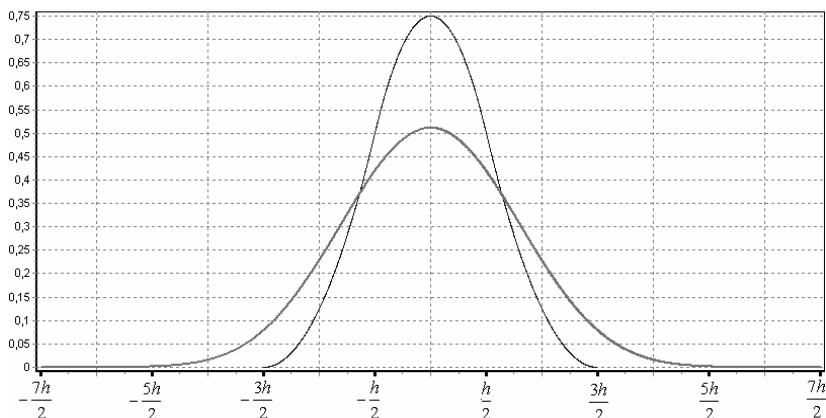
Про похибку відтворення функції $p(t)$ за використанням сплайнів $S_{r,0}(p,t)$, $r=2,3,4,5$ свідчать твердження [4–8]: при $h \rightarrow 0$ для довільної функції $p(t) \in C^2$ виконується

$$\|p(t) - S_{r,0}(p,t)\| \leq \frac{(r+2)h^2}{24} \|p''\| + \varepsilon \|p\| + o(h^2). \quad (4)$$

Поставимо за мету даної роботи отримати подання B -сплайна шостого порядку та дослідити відповідний поліноміальний сплайн, що є лінійною комбінацією зазначених B -сплайнів.

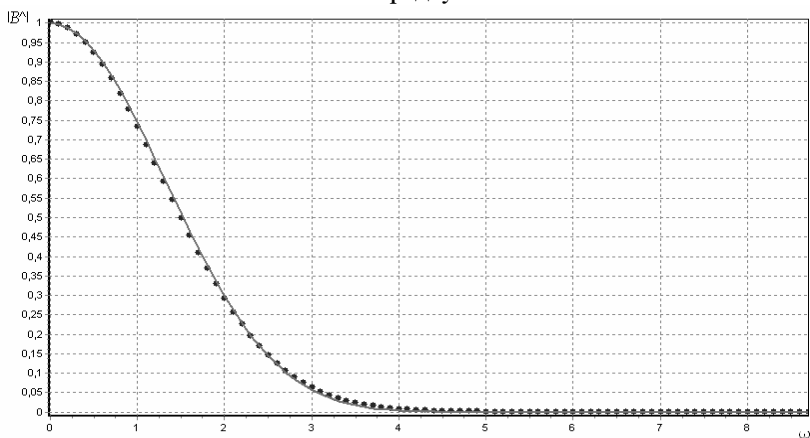
Виклад основного матеріалу. На підставі виразів (2), (3) неважко отримати такий вираз для B -сплайна шостого порядку (рис. 1):

$$B_{6,h}(t) = \begin{cases} 0, & t \notin \left[\frac{7h}{2}, \frac{7h}{2} \right], \\ \frac{1}{720} \left(\frac{t}{h} - \frac{7}{2} \right)^6, & t \in \left[\frac{7h}{2}, \frac{5h}{2} \right], \\ -\frac{1}{120} \left(\frac{t}{h} \right)^6 - \frac{7}{60} \left(\frac{t}{h} \right)^5 - \frac{21}{32} \left(\frac{t}{h} \right)^4 - \frac{133}{72} \left(\frac{t}{h} \right)^3 - \frac{329}{128} \left(\frac{t}{h} \right)^2 - \frac{1267}{960} \left(\frac{t}{h} \right) + \frac{1379}{7680}, & t \in \left[\frac{5h}{2}, \frac{3h}{2} \right], \\ \frac{1}{48} \left(\frac{t}{h} \right)^6 + \frac{7}{48} \left(\frac{t}{h} \right)^5 + \frac{21}{64} \left(\frac{t}{h} \right)^4 + \frac{35}{288} \left(\frac{t}{h} \right)^3 - \frac{91}{256} \left(\frac{t}{h} \right)^2 + \frac{7}{768} \left(\frac{t}{h} \right) + \frac{7861}{15360}, & t \in \left[\frac{3h}{2}, \frac{h}{2} \right], \\ -\frac{1}{36} \left(\frac{t}{h} \right)^6 + \frac{7}{48} \left(\frac{t}{h} \right)^4 - \frac{77}{192} \left(\frac{t}{h} \right)^2 + \frac{5887}{11520}, & t \in \left[\frac{h}{2}, \frac{h}{2} \right], \\ \frac{1}{48} \left(\frac{t}{h} \right)^6 - \frac{7}{48} \left(\frac{t}{h} \right)^5 + \frac{21}{64} \left(\frac{t}{h} \right)^4 - \frac{35}{288} \left(\frac{t}{h} \right)^3 - \frac{91}{256} \left(\frac{t}{h} \right)^2 - \frac{7}{768} \left(\frac{t}{h} \right) + \frac{7861}{15360}, & t \in \left[\frac{h}{2}, \frac{3h}{2} \right], \\ -\frac{1}{120} \left(\frac{t}{h} \right)^6 + \frac{7}{60} \left(\frac{t}{h} \right)^5 - \frac{21}{32} \left(\frac{t}{h} \right)^4 + \frac{133}{72} \left(\frac{t}{h} \right)^3 - \frac{329}{128} \left(\frac{t}{h} \right)^2 + \frac{1267}{960} \left(\frac{t}{h} \right) + \frac{1379}{7680}, & t \in \left[\frac{3h}{2}, \frac{5h}{2} \right], \\ \frac{1}{720} \left(\frac{t}{h} - \frac{7}{2} \right)^6, & t \in \left[\frac{5h}{2}, \frac{7h}{2} \right]. \end{cases} \quad (5)$$



**Рис.1. В-сплайни: порядку $r = 2$ (тонка лінія);
порядку $r = 6$ (товста лінія)**

На графіку (рис.2), порівняно з гаусіаном, продемонстровано властивості В-сплайна шостого порядку в частотній області.



**Рис.2. Частотна область: В-сплайн шостого порядку (суцільна лінія);
гаусіан (крапки) з $\sigma = 0,782$**

Із виразу (5) зрозуміло, що сплайн $B_{6,h}(t)$ є симетричною функцією відносно носія $d_6 = [-7h/2; 7h/2]$ із такими властивостями:

$$B_{6,h}(0) = \frac{5887}{11520}, \quad B_{6,h}\left(\pm \frac{h}{2}\right) = \frac{151}{360}, \quad B_{6,h}(\pm h) = \frac{10543}{46080},$$

$$B_{6,h}\left(\pm \frac{3h}{2}\right) = \frac{19}{240}, \quad B_{6,h}(\pm 2h) = \frac{361}{23040}, \quad B_{6,h}\left(\pm \frac{5h}{2}\right) = \frac{1}{720},$$

$$B_{6,h}(\pm 3h) = \frac{1}{46080}, \quad B_{6,h}\left(\pm \frac{kh}{2}\right) = 0, \quad k > 6,$$

$$\frac{1}{h} \int_{-h/2}^{h/2} B_{6,h}(t) dt = \frac{151}{315}, \quad \frac{1}{h} \int_{(\pm 1-1/2)h}^{(\pm 1+1/2)h} B_{6,h}(t) dt = \frac{397}{1680},$$

$$\frac{1}{h} \int_{(\pm 2-1/2)h}^{(\pm 2+1/2)h} B_{6,h}(t) dt = \frac{1}{42}, \quad \frac{1}{h} \int_{(\pm 3-1/2)h}^{(\pm 3+1/2)h} B_{6,h}(t) dt = \frac{1}{5040},$$

$$\frac{1}{h} \int_{(\pm k-1/2)h}^{(\pm k+1/2)h} B_{6,h}(t) dt = 0, \quad k > 3,$$

$$\frac{1}{h} \int_{(\pm 1/2-1/2)h}^{(\pm 1/2+1/2)h} B_{6,h}(t) dt = \frac{259723}{645120}, \quad \frac{1}{h} \int_{(\pm 3/2-1/2)h}^{(\pm 3/2+1/2)h} B_{6,h}(t) dt = \frac{20219}{215040},$$

$$\frac{1}{h} \int_{(\pm 5/2-1/2)h}^{(\pm 5/2+1/2)h} B_{6,h}(t) dt = \frac{2179}{645120}, \quad \frac{1}{h} \int_{(\pm 7/2-1/2)h}^{(\pm 7/2+1/2)h} B_{6,h}(t) dt = \frac{1}{645120},$$

$$\frac{1}{h} \int_{(\pm k/2-1/2)h}^{(\pm k/2+1/2)h} B_{6,h}(t) dt = 0, \quad k > 7.$$

За аналогією зі сплайнами $B_{r,h}(t)$, $r = 2, 3, 4, 5$ можна стверджувати, що якщо $S_r(\Delta_h)$ – множина всіх сплайнів мінімального дефекту за розбиттям Δ_h і $B_{6,h}(t) \in S_6(\Delta_h)$, то лінійна комбінація сплайнів $B_{6,h}(t)$ також буде належати множині сплайнів мінімального дефекту $S_6(\Delta_h)$.

Для апроксимації функції $p(t)$ за значеннями типу (1) у вузлах розбиття Δ_h введемо таку лінійну комбінацію B -сплайнів (5):

$$S_{6,0}(p, t) = \sum_{i \in Z} p_i B_{6,h}(t - (i + 0,5)h). \quad (6)$$

Подамо сплайн $S_{6,0}(p, t)$ у зручному для реалізації в обчислювальному середовищі вигляді. Якщо ввести заміну $x = 2(t - (i + 0,5)h)/h$, $|x| \leq 1$, то вираз (6) можна записати так:

$$\begin{aligned} S_{6,0}(p, t) = & \frac{1}{46080} \left((1-x)^6 p_{i-3} + \right. \\ & + (722 - 1416x + 1110x^2 - 400x^3 + 30x^4 + 24x^5 - 6x^6) p_{i-2} + \\ & + (10543 - 8670x + 1185x^2 + 860x^3 - 255x^4 - 30x^5 + 15x^6) p_{i-1} + \\ & + (23548 - 4620x^2 + 420x^4 - 20x^6) p_i + \\ & + (10543 + 8670x + 1185x^2 - 860x^3 - 255x^4 + 30x^5 + 15x^6) p_{i+1} + \\ & \left. + (722 + 1416x + 1110x^2 + 400x^3 + 30x^4 - 24x^5 - 6x^6) p_{i+2} + (1+x)^6 p_{i+3} \right). \quad (7) \end{aligned}$$

Нехай у вузлах розбиття Δ_h для значень деякої гладкої неперервної функції $p(t)$ виконується умова (1). Тоді для сплайна (6) має місце рівність

$$S_{6,0}(p, t) = S_{6,0}(\bar{p}, t) + S_{6,0}(\varepsilon, t).$$

Оцінка якості відтворення функції $p(t)$ зводиться до оцінки відхилення

$$|p(t) - S_{6,0}(p, t)| = |p(t) - S_{6,0}(\bar{p}, t) - S_{6,0}(\varepsilon, t)|$$

або для $|\varepsilon_i| < \varepsilon$, $i \in Z$, до оцінки нерівності

$$|p(t) - S_{6,0}(p, t)| \leq |p(t) - S_{6,0}(\bar{p}, t)| + \varepsilon \|S_{6,0}\|, \quad (8)$$

де $\|S_{6,0}\| = \sup_{|\varepsilon_i|} \max_t |S_{6,0}(\varepsilon, t)|$ – норма сплайн-оператора $S_{6,0}(p, t)$.

Подальша задача оцінки якості відтворення $p(t)$ складається із двох етапів: знаходження норми сплайн-оператора $S_{6,0}(p, t)$ і задача визначення похибки відтворення.

З урахуванням розкладу $S_{6,0}(\bar{p}, t)$ та $p(t) \in C^6$ у ряд Тейлора в околі точки $((i-0,5)h)$ нескладно стверджувати, що для $\forall p(t) \in C^2$ при $h \rightarrow 0$ рівномірно по t має місце асимптотична рівність

$$p(t) - S_{6,0}(\bar{p}, t) = -\frac{1}{3} p''(t) h^2 + o(h^2).$$

Отже, при $h \rightarrow 0$ для довільної функції $p(t) \in C^2$ виконується

$$\|p(t) - S_{5,0}(\bar{p}, t)\| = \frac{7h^2}{24} \|p''\| + o(h^2).$$

Проведемо оцінку норми сплайна (6).

Теорема 1. Для сплайна $S_{6,0}(p, t)$ є вірним

$$\|S_{6,0}\| = \|p\|.$$

Доведення. Із представлення сплайна $S_{6,0}(p, t)$ у вигляді (7) випливає, що

$$\|S_{6,0}\| \leq \frac{1}{46080} \|p\| \max_{|x| \leq 1} A(x),$$

де

$$\begin{aligned} A(x) = & \left| (1-x)^6 \right| + \left| 722 - 1416x + 1110x^2 - 400x^3 + 30x^4 + 24x^5 - 6x^6 \right| + \\ & + \left| 10543 - 8670x + 1185x^2 + 860x^3 - 255x^4 - 30x^5 + 15x^6 \right| + \\ & + \left| 23548 - 4620x^2 + 420x^4 - 20x^6 \right| + \\ & + \left| 10543 + 8670x + 1185x^2 - 860x^3 - 255x^4 + 30x^5 + 15x^6 \right| + \\ & + \left| 722 + 1416x + 1110x^2 + 400x^3 + 30x^4 - 24x^5 - 6x^6 \right| + \left| (1+x)^6 \right|. \end{aligned}$$

Ураховуючи, що функція $A(x)$ є парною, для знаходження її максимуму достатньо розглянути її для $x \in [0; 1]$. Для цих x , зважаючи, що вирази під знаком модуля більші за нуль, маємо

$$\max_{x \in [0; 1]} A(x) = 46080.$$

Таким чином,

$$\|S_{6,0}\| \leq \|p\|. \quad (9)$$

З іншого боку, для будь-якого частинного випадку

$$\|S_{6,0}(p, t)\| \geq \|S_{6,0}(p^*, t)\|.$$

Тоді для $x = 0$ маємо

$$S_{6,0}(p, t) = \frac{1}{46080} (p_{i-3}^* + 722p_{i-2}^* + 10543p_{i-1}^* + 23548p_i^* + 10543p_{i+1}^* + 722p_{i+2}^* + p_{i+3}^*).$$

Якщо

$$p_{i-3}^* = p_{i-2}^* = p_{i-1}^* = p_i^* = p_{i+1}^* = p_{i+2}^* = p_{i+3}^* = \|p\|,$$

то

$$\|S_{6,0}\| \geq \|S_{6,0}(p^*, t)\| \geq \|S_{6,0}(p^*, 0)\| = \|p\|,$$

а отже, з урахуванням (9), приходимо до висновку, що

$$\|S_{6,0}\| = \|p\|.$$

Теорему доведено.

Наслідок 2. Для $\forall p(t) \in C^2$ має місце

$$\|p(t, q) - S_{6,0}(p, t)\| \leq \frac{h^2}{3} \|p''\| + \varepsilon \|p\| + o(h^2).$$

Висновки. У роботі отримано подання та властивості B -сплайна шостого порядку, визначеного на рівномірній сітці вузлів. Отримано явний вигляд та проведено дослідження поліноміального сплайна, близького до інтерполяційних у середньому, що є лінійною комбінацією B -сплайнів шостого порядку.

Як слід було очікувати, у порівнянні зі сплайнами $S_{r,0}(p, t)$, $r = 2, 3, 4, 5$ сплайн $S_{6,0}(p, t)$ має більш виражені властивості згладжування; підтверджено справедливість нерівності (4) для випадку, коли $r = 6$.

Подальші дослідження можуть полягати в отриманні часткових випадків сплайна $S_{6,0}(p, t)$ задля реалізації їх у задачах обробки

цифрових сигналів та узагальнення теоретичних результатів досліджень на дво- та багатовимірний випадок.

Бібліографічні посилання

1. **Приставка П.О.** Лінійні комбінації B -сплайнів, близькі до інтерполяційних у середньому, в задачі моделювання аналогових сигналів / П.О. Приставка // Актуальні проблеми автоматизації та інформаційних технологій: зб. наук. пр. – Д.: Вид-во ДНУ. – 2011. – Т.15. – С.4–17.
2. **Чун Ч.** Введение в вэйлеты: пер. с. англ. / Ч. Чуи – М.: Мир, 2001. – 412 с.
3. **Unser M.** Splines: A Perfect Fit for Signal and Image Processing, IEEE Signal Processing Magazine / M/Unser5 // – 1999. – Vol. 16, No. 6. – P. 22–38.
4. **Приставка П.О.** Дослідження B -сплайну п'ятого порядку та їх лінійної комбінації / П.О.Приставка, О.Г.Чолипкіна // Математичне моделювання. 2007 – Дніпродзержинськ, ДДТУ. – №1(16). – С.14–17.
5. **Лигун А.А.** Асимптотические методы восстановления кривых. / А.А.Лигун, А.А.Пумейко. – К.: ІМ НАН України, 1996. – 358 с.
6. **Лигун А.А.** О восстановлении эмпирической функции плотности распределения с помощью гистосплайнов второго порядка/ А.А.Лигун, В.В.Кармазина. – Днепродзержинск, Днепродзерж. индустр. ин-т., 1989. – 30 с. – Деп. в УкрНИИИТИ 8.06.89г., N1559- Ук89.
7. **Лигун А.А.** Восстановление функций плотности распределения и их производных с помощью кубических гистосплайнов // А.А.Лигун, В.В.Кармазина / Днепродзерж. индустр. ин-т. – Днепродзержинск, 1989. – 38 с. – Деп. в УкрНИИИТИ 13.11.89г., N2569 – Ук89.
8. **Приставка П.О.** Поліноміальні сплайни при обробці даних / П.О.Приставка. – Д.: Вид-во ДНУ, 2004. – 236 с.

Надійшла до редколегії 01.07.2013 р.

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ РОЗПІЗНАВАННЯ ОБЛИЧ

Виконано аналіз ефективності застосування інформаційної технології розпізнавання облич людей із використанням одновимірних прихованих марківських моделей за допомогою створеного програмного продукту, розробленого на об'єктно-орієнтованій мові програмування Java у середовищі Eclipse.

Ключові слова: *прихована марківська модель, класифікація, алгоритм прямого-зворотного проходу, алгоритм Вітербі, алгоритм Баума–Велша.*

Выполнен анализ эффективности применения информационной технологии распознавания лиц людей с использованием одномерных марковских моделей с помощью созданного программного продукта, разработанного на объектно-ориентированном языке программирования Java в среде Eclipse.

Ключевые слова: *скрытая марковская модель, классификация, алгоритм прямого-обратного хода, алгоритм Витерби, алгоритм Баума–Велша.*

It is made the analysis of the of information technology effectiveness facial recognition of people using the one-dimensional Markov models. Analysis is based on using the created software product developed with object-oriented programming language Java within Eclipse.

Keywords: *hidden Markov model, classification, forward-backward algorithm, Viterbi algorithm, Baum–Welch algorithm.*

Вступ. Системи розпізнавання облич можуть бути використані для контролю доступу осіб до зон з обмеженим доступом, доступу до комп'ютерів, пошуку та розпізнавання осіб за спеціалізованими базами даних (наприклад, бази даних злочинців). Система розпізнавання облич має бути стійкою до умов, які змінюються (наприклад зміна освітлення, різні кути нахилу та повороту облич тощо).

Аналіз останніх досліджень і публікацій. Для розпізнавання облич найчастіше використовуються методи на основі головних компонент (власних облич) [1], лінійного дискримінантного аналізу [2], нейронних мереж [3; 4], прихованих марківських моделей [5; 6], метод еластичних графів [7].

Постановка задачі. Виконати аналіз ефективності застосування одновимірних прихованих марківських моделей для розпізнавання облич, провівши алгоритмізацію та створивши програмне забезпечення.

Основний матеріал. Прихована марківська модель складається з таких елементів: кінцевої множини прихованих станів системи $S = \{s_1, s_2, \dots, s_N\}$; кінцевої множини спостережуваних станів: $V = \{v_1, v_2, \dots, v_M\}$; матриці перехідних імовірностей $A = \{a_{ij}\}$, де $a_{ij} = P(q_t = s_j | q_{t-1} = s_i)$, $i, j = \overline{1, N}$, $t = \overline{1, T}$, q_t – стан моделі у момент часу t ; послідовності $Q = q_1, q_2, \dots, q_T$ випадкових величин зі значеннями із S ; вектора початкових імовірностей станів системи: $\pi = \{\pi_i\}$; матриці ймовірностей спостережуваних символів $B = \{b_i(k)\}$, $b_i(k) = P(o_t = v_k | q_t = s_i)$, $i = \overline{1, N}$, $k = \overline{1, M}$. o_t – спостережуваний символ у момент часу $t = \overline{1, T}$.

Кожній послідовності прихованих станів $s_{i_1}, s_{i_2}, \dots$ ставиться у відповідність послідовність випадкових величин $o_1, o_2, \dots$ зі значеннями з V . При цьому ймовірність спостереження деякого символу залежить тільки від того, в якому стані знаходиться система.

Навчання системи проводиться на зображеннях облич різних людей, кожній людині ставиться у відповідність одна прихована марківська модель, яка буде використана під час класифікації нових зображень. Етап навчання системи полягає в налаштуванні параметрів прихованої марківської моделі для максимізації ймовірності розпізнавання відповідної людини на основі максимізації функції $P(O^{(r)} | \lambda^{(r)})$ та $P(O^{(r)}, Q | \lambda^{(r)})$, де $\lambda^{(r)}$ – модель, яка відповідає r -й людині, $O^{(r)}$ – множина зображень однієї людини, Q – послідовність прихованих станів системи. Визначити послідовність прихованих станів дозволяє алгоритм динамічного програмування Вітербі. Для максимізації ймовірності спостереження послідовності $O^{(r)}$ за використанням відповідної прихованої марківської моделі $\lambda^{(r)}$ використовується алгоритм Баума – Велша. Після навчання системи $\lambda^{(r)}$ використовується для класифікації нових зображень. Навчання системи складається із двох етапів (рис. 1 та 2 відповідно [8]).

Початкові значення параметрів моделі можна вибирати довільно, ураховуючи умови нормування. Доведено, що алгоритм завжди збігається, при цьому майже завжди до точки локального максимуму.

Алгоритм динамічного програмування Вітербі використовується для визначення максимізуючої послідовності прихованих станів, при фіксованій послідовності спостережуваних символів.

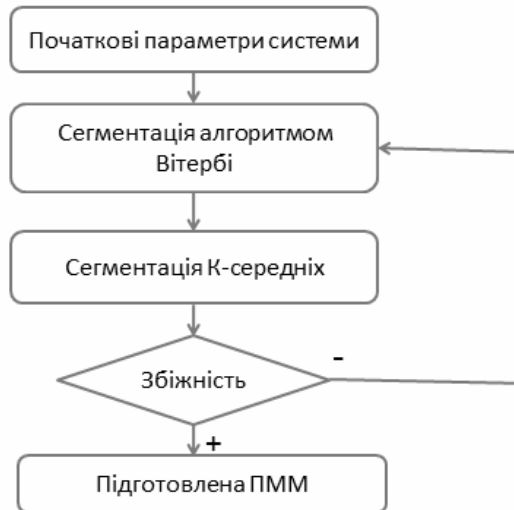


Рис.1. Початкова сегментація зображення

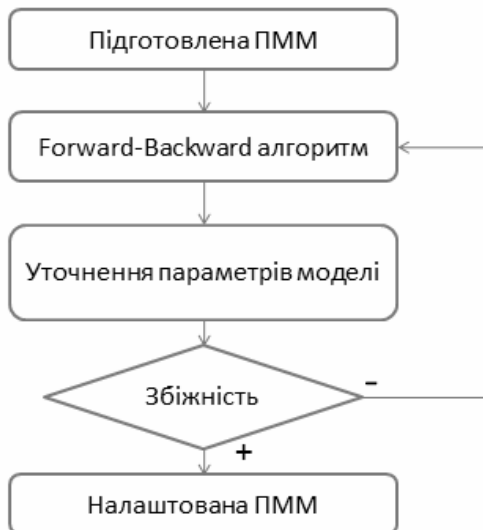


Рис.2. Уточнення параметрів моделі

Сегментація зображення за допомогою алгоритму К-середніх відбувається в межах кожного стану окремо, в якості початкового наближення алгоритму об'єкти рівномірно розподіляють за кластерами, для розрахунку відстані між об'єктами використовується метрика простору L_2 : $\|x\|^2 = \sum_{i=1}^{\dim(x)} x_i^2$.

Згідно з [9] для ефективного обчислення ймовірності появи послідовності символів без будь-яких умов використовується forward-backward-алгоритм [10], оскільки розрахункова формула

$$P(O|\lambda) = \sum_{q_1, q_2, \dots, q_T} \left(\pi_{q_1} \prod_{t=1}^T a_{q_t, q_{t+1}} \prod_{t=1}^T b_{q_t}(o_t) \right)$$

має високу обчислювальну складність, а саме $2TN^T$. Forward-backward-алгоритм дозволяє провести розрахунок $P(O|\lambda)$ за $3TN^2$ обчислень.

Уточнення параметрів моделі відбувається таким чином:

$$\begin{aligned} \pi_i^* &= \frac{1}{K} \sum_{k=1}^K \gamma_1^{(k)}(i), \\ a_{ij}^* &= \frac{\sum_{k=1}^K \sum_{t=1}^{T^k-1} \xi_t^{(k)}(i, j)}{\sum_{k=1}^K \sum_{t=1}^{T^k} \gamma_t^{(k)}(i)}, \\ c_{im}^* &= \frac{\sum_{k=1}^K \sum_{t=1}^{T^k} \gamma_t^{(k)}(i, m)}{\sum_{k=1}^K \sum_{t=1}^{T^k} \sum_{\tilde{m}}^{M_i} \gamma_t^{(k)}(i, \tilde{m})}, \\ \mu_{im}^* &= \frac{\sum_{k=1}^K \sum_{t=1}^{T^k} \gamma_t^{(k)}(i, m) o_t^{(k)}}{\sum_{k=1}^K \sum_{t=1}^{T^k} \gamma_t^{(k)}(i, m)}, \\ \Sigma_{im}^* &= \frac{\sum_{k=1}^K \sum_{t=1}^{T^k} \gamma_t^{(k)}(i, m) (o_t^{(k)} - \mu_{im}^*)(o_t^{(k)} - \mu_{im}^*)^T}{\sum_{k=1}^K \sum_{t=1}^{T^k} \gamma_t^{(k)}(i, m)}. \end{aligned}$$

$$\text{Де } \gamma_t(i) = P(q_t = s_i | O, \lambda) = \frac{a_t(i) \beta_t(i)}{\sum_{j=1}^N a_t(j) \beta_t(j)}, \quad i = \overline{1, N}, \quad t = \overline{1, T},$$

$$\xi_t(i, j) = P(q_t = s_i, q_{t+1} = s_j | O, \lambda) = \frac{a_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{h=1}^N \sum_{p=1}^N a_t(h) a_{hp} b_p(o_{t+1}) \beta_{t+1}(p)},$$

$$i, j = \overline{1, N}, \quad t = \overline{1, T-1},$$

$$\gamma_t(i, m) = \gamma_t(i) \left(\frac{c_{im} N(o_t, \mu_{im}, \Sigma_{im})}{\sum_{h=1}^{M_i} c_{ih} N(o_t, \mu_{ih}, \Sigma_{ih})} \right)$$

Для зменшення розмірності простору символів використовується двовимірне дискретне косинус-перетворення, яке переводить зображення до частотної області.

Створена інформаційна технологія реалізована у вигляді програмного продукту, розробленого на об'єктно-орієнтованій мові програмування Java у середовищі Eclipse.

Практична апробація виконана на базі зображень Olivetty Research Ltd (ORL), яка складається із зображень 40 людей, по 10 зображень на людину. База містить напівтонові нормалізовані зображення розміром 112x92 пікселів.

Тестування проводилися з використанням таких параметрів: розмір вікна сканування 10x10 пікселів; перекриття вікон сканування 7x7 пікселів; 16 коефіцієнтів дискретного косинус-перетворення.

Під час тестування програмного продукту отримані такі результати.

1. Досить висока якість розпізнавання (табл.1).

Таблиця 1

Залежність якості розпізнавання від розміру навчальної вибірки

Кількість зображень на персону – для навчання	Кількість зображень на персону – для екзамену	Час навчання, с	Час розпізнавання, с	Якість, %
1	9	861	4724	86
2	8	1494	4121	92
3	7	2459	3642	92
4	6	3815	3102	94
5	5	4542	2590	94

2. Використання діагональної дисперсійно-коваріаційної матриці дозволяє суттєво збільшити швидкість роботи системи, з незначним погіршенням якості розпізнавання (табл.2).

Таблиця 2

Залежність якості розпізнавання від розміру навчальної вибірки з використанням діагональної дисперсійно-коваріаційної матриці

Кількість зображень на персону – для навчання	Кількість зображень на персону – для екзамену	Час навчання, с	Час розпізнавання, с	Якість, %
1	9	133	520	86
2	8	225	480	90
3	7	327	401	93
4	6	398	346	92
5	5	525	290	91

3. Вдалий вибір параметрів впливає на швидкість та якість роботи системи (табл.3). У нашому випадку це: розмір вікна сканування 16x16 пікселів; перекриття вікон сканування 8x8

пікселів; використання 16 коефіцієнтів дискретного косинус-перетворення.

Таблиця 3

Залежність якості розпізнавання від розміру навчальної вибірки з використанням діагональної матриці

Кількість зображень на персону – для навчання	Час навчання, с	Час розпізнавання, с	Якість, %
1	32	77	80
2	40	70	92
3	53	60	94
4	61	51	95
5	78	43	95

4. Можливі два варіанти налаштування системи: для збільшення швидкодії системи з незначним погіршенням якості розпізнавання та для максимізації ймовірності розпізнавання зі значним програвом у швидкодії (табл.4).

Таблиця 4

Залежність якості розпізнавання від параметрів вікна сканування

Розмір вікна сканування	Перекриття	Кількість коефіцієнтів 2DDCT	Час навчання, с	Час розпізнавання, с	Якість, %
12×12	8×8	3×3	48	175	82
12×12	8×8	4×4	84	294	85
12×12	8×8	5×5	138	494	85
12×12	10×10	3×3	205	699	83
12×12	10×10	4×4	340	1189	86
12×12	10×10	5×5	340	1634	86
16×16	8×8	3×3	19	45	75
16×16	8×8	4×4	36	76	80
16×16	8×8	5×5	54	125	77
16×16	10×10	3×3	31	78	80
16×16	10×10	4×4	49	135	81
16×16	10×10	5×5	81	219	82

Висновки. Розроблена інформаційна технологія ідентифікації облич людини з використанням одновимірних марківських моделей дозволяє, із гарно налаштованими параметрами, досить швидко та

якісно розв'язувати задачу класифікації облич та може бути пристосована, залежно від необхідних умов, для збільшення швидкодії або якості розпізнавання.

Бібліографічні посилання

1. **Muller N.** Singular Value Decomposition, Eigenfaces, and 3D Reconstructions / N. Muller, L. Magaia, B.M. Herbst // SIAM Review. – 2004. – Vol. 46, No 3. – P. 518 – 545.
2. **Kim T.-K.** Component-based LDA Face Descriptor for Image Retrieval / Tae-Kyun Kim, Hyunwoo Kim, Wonjun Hwang // British Machine Vision Conference(BMVC), Cardiff, UK, 2002.
3. **Yahagi T.** Face Recognition using neural networks with multiple combinations of categories / T.Yahagi, H.Takano // International Journal of Electronics Information and Communication Engineering. – 1994. – Vol. J77-D-II, No.11. – P. 2151–2159.
4. **Latha P.** Face Recognition using Neural Networks / P.Latha, Dr.L.Ganesan, S.Annadurai // Signal Processing: An International Journal. – 2009. – Vol. 3. – Issue 5. – P. 153–160.
5. **Bicego M.** Using hidden Markov models and wavelets for face recognition / M. Bicego, U. Castellani, V. Murino // Image Analysis and Processing. – 2003. – P. 52–56.
6. **Nefian A.V.** Hidden Markov Models For Face Recognition / A.V. Nefian, M.H. Hayes // ICA SSP98. – 1998. – Vol. 5. – P. 2721–2724.
7. **Wiskott L.** Face Recognition by Elastic Bunch Graph Matching . Wiskott L., Fellous J.M., Kruger N. // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1997. – 19(7). – P. 775–779.
8. **Kinscher J.** The Munster Tagging Project Mathematical Background / J.Kinscher, H.Trebbe // Arbeitsbereich Linguistik University of Miinster, 1995.
9. **Rabiner L. R.** A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition / Proceedings of the IEEE. – 1989. – Vol. 77, No. 2. – P. 257–285.
10. **Dawei Shen** Some Mathematics for HMM / October 13th, 2008.

Надійшла до редколегії 10.07.2013 р.

ОБЗОР МЕТОДОВ И ПРОГРАММНЫХ ПРОДУКТОВ ПРОГНОЗИРОВАНИЯ ЗЕМЛЕТРЯСЕНИЙ

Зроблено огляд методів прогнозування за різними провісниками землетрусів і програмних продуктів роботи із сейсмічними даними.

Ключові слова: *сейсмічність, провісники, передбачення землетрусів.*

Выполнен обзор методов прогнозирования по различным предвестникам землетрясений и программных продуктов работы с сейсмическими данными.

Ключевые слова: *сейсмичность, предвестники, предсказание землетрясений.*

It was made the review of forecasting methods for various precursors of earthquakes and seismic software.

Keywords: *seismicity, precursors, earthquake prediction.*

Введение. Несмотря на многолетний опыт анализа сейсмических данных учеными геофизики, математики, геоинформатики, астрологии, сейсмологии в области прогноза землетрясений, на сегодняшний день нет ни одного метода, позволяющего на короткий промежуток времени достаточно точно предсказывать разрушительные стихийные бедствия. Поэтому актуальной является задача анализа существующих подходов к прогнозированию сейсмической активности, выделения общих характеристик, достоинств и недостатков используемых программных продуктов и методов, реализованных в них.

Анализ исследований в данной области. Прогноз землетрясений – это вероятностная характеристика места, времени и силы сейсмического события. Прогнозы делятся на следующие виды (стадии): долгосрочный (годы–десятилетия); среднесрочный (месяцы–годы); краткосрочный (дни–недели); оперативный или непосредственный (минуты–часы). На сегодняшний день определенный успех достигнут в области долгосрочных прогнозов. Долгосрочные прогнозы опираются на историю сейсмоактивности на определенной территории, делаются

с помощью статистических и физических методов, карт сейсмического районирования. Но на вопрос, когда может произойти беда и с какой вероятностью, долгосрочные прогнозы не дают точного ответа. Результаты долгосрочных прогнозов могут быть использованы при экономическом планировании городских территорий и промышленных комплексов, возведении гидротехнических, газонефтяных, химических сооружений на сейсмоопасных территориях. Задачу среднесрочного прогноза в определенной мере можно считать также решенной современной наукой. Если на исследуемой территории установлено достаточное количество геофизических устройств, то теоретически землетрясение определенной силы можно спрогнозировать за год-два до сейсмического события. Этого времени достаточно для проверки зданий и обучения населения. Краткосрочное прогнозирование на сегодняшний день считается задачей еще нерешенной. Передачу данных необходимо осуществлять в реальном режиме времени. При удачном построении краткосрочного прогноза можно своевременно остановить производства и эвакуировать население. Возможность непосредственного прогноза вообще отсутствует. Существует много методов прогноза землетрясений, основанных на наблюдении аномалий геофизических полей. Так, большое количество слабых толчков (сейсмическая активность), крупномасштабные деформации поверхности Земли (движение земной коры), поднятие и опускание участков земной коры, изменение наклонов поверхности, изменение деформации горных пород, изменение уровня грунтовых вод, изменение скоростей сейсмических волн, изменение вариаций геомагнитного поля, изменение электрического сопротивления Земли, изменение количества радона в грунтовых водах, необычное поведение животных, птиц, рыб могут свидетельствовать о возможности наступления сильного землетрясения.

Формулировка целей статьи. Целью данной статьи является анализ существующих методов и программных продуктов предсказания землетрясений.

Основной материал исследования. Прогноз землетрясений тесно связан с изучением аномальных явлений в геофизических полях, которые называются предвестниками, их насчитывается несколько сотен. Предвестники возникают в зоне готовящегося сейсмического события накануне его возникновения. Классификация предвестников приведена в таблице.

Таблица

Классификация предвестников землетрясений

Предвестники землетрясений	
Геологические	<i>Сейсмодислокации:</i> напряженно-деформационное состояние среды, формирование магистральных разрывов, сдвиговых подвижек, раскрытие и визуализация трещин, развитие оползней, склоновых, лавинных, селезневых и других процессов, изменения разрывно-трещинной сети, структурные геологические и геоморфологические изменения, повышение температуры поверхности Земли
Геофизические	<i>Сейсмологические:</i> сейсмические бреши, сейсмические затишья, миграция сейсмичности, рой землетрясений, плотность сейсмологических разрывов, локализация сейсмического процесса, форшоки; <i>деформационные:</i> деформации поверхности Земли, земной коры, плотности пород; <i>электромагнитные:</i> световые явления в атмосфере, возмущение электрического и магнитного потенциалов; <i>геомагнитные:</i> электрокинетические явления (аномалии электромагнитного поля), пьезомагнитные процессы (напряжение земной коры); <i>гравитационные;</i> <i>ионосферные:</i> аномальные увеличения интенсивности – очень низкочастотные и крайне низкочастотные сверхдлинные волны, аномальные изменения спорадических слоев, возникновение диффузных областей, увеличение интенсивностей линий кислорода; <i>гидроакустические;</i> <i>геотермические:</i> изменение температуры подземных вод и почвы
Геохимические	Изменение концентрации газов, изменение химического состава природных вод (увеличение содержания радона, гелия, углекислого газа, водорода, РН в термоминеральной воде), повышение концентрации приземного озона, радона и гелия
Гидродинамические	Пластовые подавления подземных вод, предельный объем воды, который можно получать из скважины за час, спонтанного газа, степень освобождения, водопроводимость водоносных горизонтов и трещин зон, повышение радона в атмосфере
Геодезические	Наклоны земной поверхности, движения земной коры
Космические	Электромагнитные, плазменные, ионосферные. Сейсмогенные деформации: сейсмодислокации (линеamentный анализ), сейсмдеформации

Каждый из предвестников может быть как долго-, так и средне- и краткосрочным. К настоящему времени во всем мире насчитывается несколько сотен различных по своей природе предвестников. Проведенный анализ позволил выделить общие характеристики существующих подходов к прогнозированию землетрясений и распределить методы прогнозирования по таким группам: геофизические, геологические, геохимические, гидродинамические, геодезические, космические.

Подготовка землетрясения отражается на геофизических полях: перед землетрясениями и во время их наблюдаются аномальные изменения электрических свойств горных пород, силы тяжести, скоростей сейсмических и акустических волн. Геофизические методы занимают большую часть сейсмологических исследований при прогнозировании землетрясений и считаются классическими. При соблюдении всех условий прогноз землетрясения по многим другим (не геофизическим) критериям возможен только в его эпицентре за короткий срок до основного толчка. Отсюда следует, что альтернативные способы прогнозирования землетрясений имеют важную научную значимость, являясь частью геолого-геофизического комплекса, но не могут быть использованы в качестве самостоятельных методов, что подтверждено многолетней практикой сейсмо-прогностических работ.

В то же время большая часть геофизических предвестников носит интуитивный характер: фиксируется факт наблюдения конкретной аномалии геофизического поля перед конкретным сильным землетрясением, но отсутствуют сведения о статистических характеристиках предвестников; не рассматриваются возможные физически обоснованные механизмы, приводящие к возникновению предвестников.

Исследования по прогнозированию землетрясений привлекли внимание специалистов различных областей, не обладающих достаточными знаниями сейсмологии. Поиск предвестников землетрясений зачастую осуществляется без всякой системы и не имеет физической и сейсмологической основы. Р. Геллер критикует многие работы, ставя под сомнение их научную ценность из-за отсутствия точных определений: много расплывчатых высказываний сейсмологов и геофизиков расцениваются ими как прогноз землетрясений, но мало систематических исследований, посвященных определениям и собственно прогнозу.

Часто к проблеме прогноза подходят как к задаче восстановления временного ряда моментов землетрясений, а не как к задаче о прогнозе катастрофы, т.е. о переходе через критический порог. Пренебрежение основами теории вероятностей и математической статистики нередко приводит к необоснованным выводам.

Другим одним из основных недостатков существующих методов и подходов является то, что в разных регионах различные предвестники работают по-разному, давая большой разброс в оценках места, времени и силы будущего землетрясения.

Значительную сложность при определении места очага будущего землетрясения по наблюдениям за предвестниками представляет собой большой ареал их распространения: расстояния, на которых наблюдаются предвестники, значительно превышают размеры разрыва в очаге. При этом краткосрочные предвестники наблюдаются на больших расстояниях, чем долгосрочные, что подтверждает более слабую их связь с очагом.

Существующие *программные продукты* обработки и анализа сейсмологических данных можно распределить по следующим группам в зависимости от типа решаемых задач.

1. Мониторы

К этой группе относятся такие продукты, как: проект «Экстремум» для прогнозирования последствий землетрясения (включает комплект программ, который дает возможность хранить, систематизировать и обрабатывать картографическую и семантическую информацию, прогнозировать последствия землетрясений в городах и населенных пунктах, а также обосновывать эффективные сценарии реагирования в пострадавшем регионе); Earth Alerts|Version 2011 – сейсмический монитор, работающий в режиме реального времени, может встраиваться в любой сайт и работать автономно. Большая часть продуктов работает и настраивается для конкретного района.

2. Программы чтения-записи

В настоящее время разработаны конверторы (программы-преобразователи) для следующих широко распространённых форматов представления сейсмических данных: XDATA – формата записей пунктами наблюдений радиотелеметрической системы NANOMETRICS; PCC-1 – формата записей пунктами наблюдений радиотелеметрической системы PCC; Css версий 2.8 и 3.0 – формата записей, разработанных для проведения технического эксперимента группой научных экспертов Конференции по разоружению в Женеве

(ТЭГНЭ-2 и ТЭГНЭ-3); DASS – формата записей для цифровых станций НПП «Геотех»; CM6 GSE2 – формата записей для сейсмических данных, распространяемых по запросам AutoDRM; PECATOS4 – преобразование файлов бюллетеней, получаемых HYPO71-91, в формат каталогов землетрясений.

Чтобы облегчить перемещение данных от одного обрабатывающего подрядчика к другому, международное общество прикладной геофизики SEG (Society of Exploration Geophysicist) сначала разработало формат SEG-X для обмена (eXchange) данными. Но он вскоре был заменен форматом SEG-Y, являющимся форматом с последовательно расположенными трассами (или демультимплексный), разработанным для хранения полностью или частично обработанных сейсмических данных.

ARCOPY – программный комплекс, предназначенный для перезаписи сейсмических данных различных форматов, применявшихся в СНГ (СЦС-3, СОС-ПС, СОС-Сайбер, SEG-B, SEG-D, ССЦ-2,3,4) и других (в том числе упакованных), с архивных магнитных лент (1/2 дюйма) на картриджи большой емкости, преобразования данных в формат SEG-Y.

Другие пакеты для работы с форматом SEG-Y: GSEGView, SegyMAT, SegyPY, segy-py, SeiSee (также форматы CWP/SU и CGG CST), SeisView.

WSG осуществляет импорт данных по TCP/IP-протоколу из систем NRTS и LISS (формат миниSEED – Stein 2).

3. Визуализация, интерпретация, анализ

Астропроцессор ZET GEO – универсальная астрологическая программа, которая может быть использована при краткосрочном прогнозировании землетрясений [4].

Методика и программа оптимального проектирования геодезических профилей для наблюдений за активными разломами земной коры, разработанная Институтом прикладной математики во Владивостоке, позволяет автоматически осуществить выбор оптимального числа и расположения в пространстве геодезических пунктов, предназначенных для GPS-наблюдений за активными разломами земной коры. Язык программирования – FORTRAN [5].

Тем же институтом разработана методика и программа строгой математической обработки свободных и несвободных плановых геодезических сетей, позволяющая уравнивать плановые линейно-угловые геодезические сети любой конфигурации с полной оценкой точности ее элементов, фиксировать систему координат любым

набором квазистабильных или стабильных пунктов, осуществлять поиск грубых ошибок измерений по поправкам из уравнивания. Язык программирования – FORTRAN [5].

Основные вычислительные процедуры WSG (Windows Seismic Grafer) предназначены для обработки сейсмических сигналов и получения оценок параметров гипоцентров сейсмических событий [6].

HYPO71-91 – консольная программа для расчёта основных параметров землетрясения по первичным данным обработки HYPO71-91 [7].

Математическое обеспечение анализа землетрясений – SEISAN для IBM PC и SUN компьютеров состоит из полного набора программ и простой базы данных для анализа данных о землетрясениях в аналоговой и цифровой форме [8].

Tesseral Technologies Inc. – программный пакет конечно-разностного моделирования сейсмического волнового поля [9].

Tatra Seis Inc. – программный пакет обработки сейсмических данных для поиска вертикальных границ с помощью технологии миграции дуплексных волн [9].

Madagaskar – многомерный анализ данных, включающий обработку сейсмичности [10].

CWP/SU (Seismic Un*x) – сейсмический пакет с открытым исходным кодом сейсмических утилит, наиболее широко использующийся в мире (69 стран и территорий) [11].

Другие бесплатные программы обработки: FreeUSP [12], FreeDDS (общие форматы – SEG-Y, SU, SEPlib, USP) [12], PSEIS_OSS (параллельная обработка сейсмических данных общих форматов – SEG-Y, SU, SEPlib, USP, Madagascar) [14], CPSeis [17], SPARC [12], SEPlib [16], BotoSeis [17] и GeBR (интерфейсы с SU, Madagascar и другие возможности) [18], SeaSeis (обработка SU-данных, SEG-D – ридер, SEG-Y ввод/вывод и другие возможности) [15], GPLib++ (сейсмические и магнитотеллурические обработки) [17], JavaSeis (предварительный стек для ввода/вывода сейсмических данных; обработка сейсмических данных в Eclipse) [17], Seismic Lab [19], CREWES [20], SW3D [21], PyShottab, Seismon [22], Seismic_Handler [23], STK [17].

Рассмотрим некоторые бесплатные анализаторы. В основе программного продукта Delivery лежит Байесовский метод инверсии, предусматривающий обеспечение оценки с погрешностью, которая обеспечивает преимущества для разработчиков моделей [24]. Геологический ресурс JRG является пакетом обработки и визуализации сейсмических данных на Java с качественной графикой (возможность 3-D

и 2-D), SEG-Y и звуковым файлом ввода/вывода [25]. Jive3D представляет собой пакет моделирования и томографической инверсии, способный моделировать сейсмические типы данных в широком временном диапазоне [26]. Программа WaveletExtractor позволяет извлекать оптимальные всплески от сейсмических, каротажных данных и каротажа (время–глубина) информации, может производить: оценки неопределенности для вейвлетов, информацию время–глубина, уровень шума для оценки сейсмической инверсии [28]. SLIMpy написана на чистом Python, предоставляет возможности обработки сейсмических данных пакетов, таких как Мадагаскар, через перегрузку операторов [27]. Возможности обработки данных проекта SEIZMO включают корреляционный анализ (correlation), свертки (convolution, deconvolution), удаление тренда (detrending), дифференцирование (differentiation), интегрирование (integration), интерполяцию (interpolation), изменение частоты дискретизации (resampling), фильтрацию (filtering), слияние (merging), ответ переноса (response transferring), вращение (rotation), укладку (stacking), спектральный анализ (spectral analysis), конусообразование (tapering) и оконный анализ (windowing) [28]. GeoSeis используется для расчета параметров из преломленного сейсмического сигнала [29]. Passeis позволяет проводить пассивный сейсмический анализ [30].

Программы отображения сейсмических сигналов: Mines JTK (обработка и отображение сигнала) [31], kogeо (включает в себя геофизический анализ и визуализацию) [32], SIOSEIS представляет собой программный пакет для расширения функциональности и управления, отражения морской сейсмичности и преломления данных [33], Landmark SeisWorks® 3D решает задачи визуализации, анализа и интерпретации сейсмических 3D данных, программный комплекс mech-Lander используется для построения (визуализации) очагов землетрясений.

4. Прогнозирование

Рассмотрим программные продукты, позволяющие выполнять прогнозирование сейсмической активности в контексте методов, лежащих в их основе.

Специалисты и ученые географического факультета Московского государственного университета на данный момент работают над новой задачей с применением нейросетей – так называемых самоорганизующихся карт Кохонена, предложенных финским ученым Теуво Кохоненом в 1984 году, на сегодняшний день являющихся самым популярным алгоритмом искусственных нейронных сетей в неконтролируемой категории

обучения; представляют собой соревновательную самообучающуюся нейронную сеть с обучением без учителя, выполняющую задачу визуализации и кластеризации с использованием следующих критериев для получения прогнозов: широтное распределение сейсмической активности, астрономические характеристики [34].

В 70-х годах В.Н. Вапником был предложен алгоритм «Фортран обобщенный портрет» (ФОП). Идея алгоритма заключается в восстановлении зависимостей по эмпирическим данным, в основе которых лежит задача распознавания – задача идентификации периодов повышенной сейсмоопасности с помощью построения гиперплоскости, разделяющей пространство заданных признаков. Программно-алгоритмический комплекс включает в себя программы для решения задач: обучения распознаванию образов, восстановления регрессии, интерпретации результатов косвенных экспериментов [35].

Система прогноза ионосферных данных на основе нейронной сети (разработана в Институте космических исследований и распространения радиоволн ДВО РАН Ю.А.Полозовым и О.В.Мандриковой) базируется на совместном применении конструкции вейвлет-преобразования и нейронных сетей. Для обучения сети использовались преобразованные с помощью вейвлет-разложения четвертого уровня сигналы сейсмически спокойных лет. В основу сети была положена трехслойная сеть прямой передачи сигнала с двумя сигмоидальными и одним линейным слоями [36].

Као Динь Чонг (работа выполнена в Российском университете дружбы народов) в своей работе применил нейросетевые технологии, используя функции Гуттенберга – Рихтера для определения сейсмической активности; в качестве сети выбрана Feedforward Back Propagation с обратным распространением ошибки для среднесрочного прогноза землетрясений на основе сравнительного анализа различных типов нейронных сетей [37].

В Институте космических исследований и распространения радиоволн ДВО РАН Ю.А.Полозовым и О.В.Мандриковой предложен алгоритм для вычисления функции принадлежности сейсмогенерирующих факторов к нечетким подмножествам факторов влияния на сейсмическую активность, позволяющий повысить точность решения задач прогнозирования; предложена модель прогнозирования, которая позволяет строить прогнозы для локального участка земной коры, разработанная на основе математического аппарата нечетких множеств и нечеткой логики [38].

Институтом космофизических исследований и распространения радиоволн разработана программа для вычисления вероятностей сейсмических событий, которая реализует вероятностную модель интерпретации каталога землетрясений. Позволяет представить сейсмический режим в виде интегральных функций распределения или функций плотности распределения вероятностей, на основе которых строится вероятностная модель сейсмичности [5].

Геоинформационная система Prediction (работа выполнена в Институте земной коры СО РАН в Иркутске Е.А.Лавиной) для прогноза землетрясений и горных ударов: разработка и примеры применения на конкретных территориях – в Байкальской рифтовой зоне и Норильском месторождении [39].

TRON – онлайн-технология системного прогнозирования землетрясений в режиме реального времени. Для математического и статистического анализа используется единая БД о поведении домашних птиц, рыб и зверей, наполняемая пользователями сети Интернет (лидер команды – Б.Д.Яровой, Центр информационных технологий, г. Владивосток) [13].

В 2009 году начала функционировать в режиме краткосрочного прогнозирования землетрясений и оперативной передачи этой информации странам – участникам Global Network for the Forecasting of Earthquakes (владельцы станций ATROPATENA). В основу торсионного детектора трехмерных гравитационных вариаций положен эффект трехмерного изменения гравитационного поля перед сильными землетрясениями на очень больших расстояниях от эпицентра землетрясений [40].

Выводы. Различные возможности и границы применения компьютерной техники и разработки компьютерных систем определяются уровнем формализации научно-технических знаний в конкретной отрасли. Чем глубже разработана теория, тем большие возможности существуют для создания и внедрения программного обеспечения обработки данных. Несмотря на большое количество и разнообразие методов и программных продуктов в области прогнозирования землетрясений, до недавнего времени не существовало математического обеспечения, позволяющего делать краткосрочный прогноз, так как не существовало теоретических методов, дающих возможность сделать такой прогноз, который бы с высокой долей вероятности реализовывался. На сегодняшний день проблема предсказания землетрясений остается открытой.

Библиографические ссылки

1. **Карцева М.В.** Анализ предвестников землетрясений / М.В. Карцева, А.Т. Зверев // Изв. ВУЗов. «Геодезия и аэрофотосъемка», МИИГАиК, Москва. – 2006. – №5. – С.125—133.
2. **Завьялов А.Д.** Среднесрочный прогноз землетрясений: основы, методика, реализация. / А.Д. Завьялов; Ин-т физики Земли им. О.Ю. Шмидта. – М.: Наука, 2006. – 254 с.
3. **Уломов В.И.** Методика поиска прогностических признаков землетрясений / В.И. Уломов // Информационное сообщение АН УзССР. – 1977. – №186. – 11 с.
4. [Электронный ресурс]. – Режим доступа: <http://zet.astrologer.ru/>.
5. **Кульчин Ю.Н.** Научно-технические разработки ДВО РАН для практического использования в социально-экономической сфере / Ю.Н.Кульчин. – Владивосток, 2011. – С. 63–64, 75.
6. Официальный сайт ГС РАН [Электронный ресурс]. – Режим доступа: <http://www.ceme.gsras.ru>.
7. Official site of School of EAS [Электронный ресурс]. – Режим доступа: <http://www.eas.gatech.edu/>.
8. [Электронный ресурс]. – Режим доступа: <http://seis.geus.net/software/seisan/seisan.html>.
9. Official site of Tesseral Technologies [Электронный ресурс]. – Режим доступа: <http://www.tesseral-geo.com/>.
10. [Электронный ресурс]. – Режим доступа: <http://reproducibility.org>.
11. Official site of Center for Wave Phenomena [Электронный ресурс]. – Режим доступа: <http://www.cwp.mines.edu/>.
12. Official download site for FreeUSP and FreeDDS [Электронный ресурс]. – Режим доступа: <http://freeusp.org/>.
13. [Электронный ресурс]. – Режим доступа: <http://tron.ru/AINUR/>.
14. Official site of Pseis-OSS [Электронный ресурс]. – Режим доступа: <http://pseis.org/>.
15. [Электронный ресурс]. – Режим доступа: <http://seaseis.com/>.
16. Official site of The Stanford Exploration Project [Электронный ресурс]. – Режим доступа: <http://sepwww.stanford.edu/>.
17. Official site of Source Forge [Электронный ресурс]. – Режим доступа: <http://sourceforge.net/>.
18. Official site of GêBR <http://www.gebrproject.com/> [Электронный ресурс]. – Режим доступа: <http://www.gebrproject.com/>.
19. Official site of University of Alberta [Электронный ресурс]. – Режим доступа: <http://www.ualberta.ca/>.
20. Official site of CREWES [Электронный ресурс]. – Режим доступа: <http://www.crewes.org/>.
21. Official site of Charles University in Prague [Электронный ресурс]. – Режим доступа: <http://cuni.cz/>. [Электронный ресурс]. – Режим доступа: <http://www.stefanmertl.com/science/software/seismon/>.

22. [Электронный ресурс]. – Режим доступа: <http://www.seismic-handler.org/>.
23. Official site of Commonwealth Scientific and Industrial Research Organization [Электронный ресурс]. – Режим доступа: <http://www.csiro.au/>.
24. Official site of University of Nevada, Reno [Электронный ресурс]. – Режим доступа: <http://www.unr.edu/>.
25. [Электронный ресурс]. – Режим доступа: <http://bullard.esc.cam.ac.uk/~hobro/3D/>.
26. [Электронный ресурс]. – Режим доступа: <http://slim.eos.ubc.ca/SLIMpy>.
27. [Электронный ресурс]. – Режим доступа: <http://epsc.wustl.edu/~ggeuler/codes/m/seismo/>.
28. GeoSeis Software 2011 © [Электронный ресурс]. – Режим доступа: <http://www.geoseis.tr.gg>.
29. **Haishan Z.** An Open-Source Passive Seismic Data Processing and Visualization Package: Passeis / Haishan Zheng, Leiv-J., Gelius // *Frontiers + Innovation – 2009 CSPG CSEG CWLS Convention*// P.408–411.
30. Official Site of the Colorado School of Mines [Электронный ресурс]. – Режим доступа: <http://www.mines.edu/>.
31. [Электронный ресурс]. – Режим доступа: <http://www.kogeo.de>.
32. [Электронный ресурс]. – Режим доступа: <http://sioseis.ucsd.edu/>.
33. **Kohonen T.** Self-Organizing Maps (Third Extended Edition), NY, 2001 – 501p..
34. **Вапник В.Н.** Алгоритмы и программы восстановления зависимостей. М.: ФИЗМАТЛИТ, 1984. - 816 с.
35. **Мандрикова О.В.** Анализ изменений, происходящих в ионосферных параметрах накануне сильных землетрясений на Камчатке, на основе вейвлет-преобразования сигналов критической частоты / О.В. Мандрикова, В.В. Богданов, Ю.А. Полозов // *Proc. of SCM'2007 (International Conference on Soft Computing and Measurements)* // Сб. докл. С.-Петербург, 27-29 июня 2007 г. / СПб: Изд-во СПбГЭТУ "ЛЭТИ", 2007.
36. **Чонг Као Динь** Исследование и применение нейросетевых технологий в задаче прогнозирования землетрясений (на примере северо-западного района Вьетнама): автореф. дис. на соискание ученой степени канд. техн. наук: 05.13.01 / Као Динь Чонг. – М., 2012
37. **Мандрикова О.В.** Автоматизированный способ обработки сигналов со сложной структурой / О. В. Мандрикова, Ю. А. Полозов // *Информационные технологии.* – 2008. – № 12. – С. 15–19.
38. **Левина Е. А.** ГИС для прогноза землетрясений и горных ударов: разработка и примеры применения в Байкальской рифтовой зоне и Норильском месторождении: автореф. дис. на соискание ученой степени канд. геол.-минерал. наук / Е.А Левина. – Иркутск, 2011.
39. Официальный сайт Глобальной Сети Прогнозирования Землетрясений [Электронный ресурс]. – Режим доступа: <http://seismonet.org/index.html>.

Надійшла до редколегії 12.07.2013 р.

УДК 519.688:336.76.066

О.П. Луценко, О.Г. Байбуз

Дніпропетровський національний університет імені Олеся Гончара

ВІДНОВЛЕННЯ ФОРМИ РОЗПОДІЛУ ТОЧОК РОЗЛАДНАНЬ НА ЧАСОВОМУ РЯДІ ВАЛЮТНИХ КОТИРУВАНЬ

Розглянуто форму розподілу, утвореного точками зміни статистичних характеристик ряду курсів валют. Проаналізовано можливості його відновлення з використанням параметричного та непараметричного підходів.

Ключові слова: статистика, визначення розладнань, біржова торгівля, котирування, часовий ряд, нестационарні процеси, відновлення розподілів.

Рассмотрена форма распределения, образованного точками изменения вероятностных характеристик временного ряда валютных котировок. Проанализированы возможности его восстановления с использованием параметрического и непараметрического подходов.

Ключевые слова: статистика, определение разладок, биржевая торговля, котировки, временной ряд, нестационарные процессы, восстановление распределений.

The distribution formed by change-points of financial quotes time series has been reviewed. Possibilities of its estimation using parametric and non-parametric approach have been analyzed.

Keywords: statistics, change-point detection, stock market, market quotations, time series, non-stationary processes, probability distribution estimation.

Постановка проблеми. Висока прибутковість торгівлі на Forex поєднується з ризиками, для мінімізації яких застосовується технічний аналіз ринку. Технічний аналіз керується твердженням «курс ураховує все» і ґрунтується тільки на поточній поведінці курсу валют, без урахування причин зміни.

Основою технічного аналізу є твердження, що ціни на ринку рухаються направлено, підкоряючись певній тенденції (тренду). Виділяють висхідну (бичачу), спадну (ведмежу) й бічну (флет) тенденції. Завданням технічного аналізу є визначення характеру

діючої тенденції та виявлення моментів її зміни шляхом аналізу історії зміни цін на ринку.

Серед методів, найбільш часто вживаних у технічному аналізі, можна виділити такі: допоміжні побудови на графіках курсів валют (лінії опору і підтримки, лінії тренда тощо), індикатори та осцилятори. Існує велика кількість різновидів і модифікацій цих методів, але для основної маси характерні загальні недоліки: суб'єктивність допоміжних графічних побудов, значна затримка сигналу про розворот тренда у часі, велика кількість хибних сигналів в умовах невираженого тренда.

У зв'язку з перерахованими недоліками, а також зі зростанням рівня обчислювальних можливостей персональних комп'ютерів відносно недавно сформувалася тенденція до переходу від індикаторів та осциляторів технічного аналізу до більш точних і складних методів. Додатковий вплив на розвиток ринку програмного забезпечення Forex має збільшення доступності торгівлі на цьому ринку (збільшення кількості компаній-посередників, поширення торгівлі через інтернет, реклама у ЗМІ), яке приводить до подальшого збільшення кількості учасників торгівлі й зростання попиту на технології оцінки та мінімізації ризиків. У зв'язку з цим розробка нових методів та уточнюючих моделей, що можуть бути використані у технічному аналізі, є актуальною задачею.

Відмінними рисами методу оцінки ризиків, пропонованого авторами даної статті, є:

1) імовірнісний підхід. Оскільки зміни валютних котирувань є значною мірою випадковим процесом, його доцільно розглядати з позицій імовірності різних гіпотез прогнозу. Детерміністичне прогнозування не дає уявлення про співвідношення таких імовірностей, показуючи лише найбільш імовірний варіант розвитку подій, який часто буває обраний із мінімальною перевагою щодо інших варіантів. Перехід до ймовірнісної моделі прогнозування є перспективним з огляду на прозорість даних про достовірність прогнозу;

2) розгляд часового ряду валютних котирувань не як безлічі окремих точок, а як сукупності відрізків стаціонарності, на які він розбивається точками розладнання. Знаходження точок розладнання ряду дозволяє розділити ряд на ділянки з подібними статистичними властивостями, що відповідає одній з основних парадигм технічного аналізу – поділу часового ряду на відрізки з постійною тенденцією.

Постановка цілей статті. Основу моделі становить припущення, що точки розладнання статистичних характеристик ряду котирувань

(зміни напрямку тренда) є випадковою величиною, функція розподілу якої повільно змінюється з часом. Відновивши щільність розподілу, отримаємо можливість на кожному кроці спостережень судити про ймовірність зміни напрямку ринкової тенденції.

Мета даної статті – опис форми розподілу, утвореного точками розладнань ряду валютних котирувань, а також аналіз ефективності його відновлення за допомогою параметричного (ЕМ-алгоритм) та непараметричного (сплайн-аппроксимація) підходів.

Основний матеріал. Перед тим як перейти до розгляду розподілу, опишемо методи, за допомогою яких було отримано дані.

Для вирішення завдання розбиття ряду на ділянки стаціонарності використовуються два методи: графічний метод пошуку перегинів графіка, призначення якого – знаходити точки розвороту апостериорно, і алгоритм кумулятивних сум (АКС), що належать до методів послідовного виявлення розладнань.

Графічний метод має високу точність і призначений для побудови основної вибірки ділянок стаціонарності, на основі якої відновлюється початкове значення щільності розподілу розладнань. Цей метод працює за таким принципом.

Зафіксуємо початковий і кінцевий моменти в процесі, всередині якого будемо шукати точку розладки: m – початковий момент часу, n – кінцевий момент часу. На цьому проміжку шукатимемо лише одну точку розладки.

Для цього проміжку побудуємо масив

$$\{d_i, i = \overline{n, m}\},$$
$$d_i = \frac{|kx_i - y_i + b|}{\sqrt{k^2 + 1}},$$

де k, b – знайдені коефіцієнти рівняння прямої ($y=kx+b$), що проходить через точки (m, x_m) та (n, x_n) .

Точку розладки знаходимо в такий момент часу r , що

$$x_r = \max_{i=m, n} d_i, \quad r \in (m, n), \quad x_r > bar,$$

де bar – бар'єр чутливості методу.

Якщо точка розладки була знайдена, то момент часу r розбиває проміжок $[m, n]$ на два: $[m, r]$ та $[r, n]$, які розглядаємо аналогічним способом, переходячи до пункту 1.

Процес закінчується, коли жодне значення d_i на проміжку не перевищує встановленого бар'єра.

У процесі використання моделі часовий ряд котирувань поповнюється новими даними і виникає необхідність знаходити розладнання в найновіших отриманих даних. Для цієї мети використовуватимемо алгоритм кумулятивних сум (АКС). АКС полягає в аналізі поведінки величини [1]:

$$S_t = S_{t-1} + \ln(\omega(x_t/\theta_2)\omega(x_t/\theta_1)).$$

На кожному кроці сума порівнюється із заданим порогом h . Якщо на кроці t $S > h$, подається сигнал про розладнання, накопичення суми починається з нуля. Відбиваючий екран встановлено в точці 0.

$$S_t = \max(0, S_{t-1} + \Delta S),$$

$$\Delta S = \ln(\omega(x_t/\theta_2)/\omega(x_t/\theta_1)).$$

Знаючи тип розподілу та визначаючи одну з імовірнісних характеристик розподілу як параметр θ , можна отримати рекурентні формули накопичення кумулятивної суми. Так, для випадку виявлення зміни середнього значення нормального розподілу формула для припущення матиме вигляд

$$g_t = \frac{m_2 - m_1}{\sigma^2} \left(x - \frac{m_1 + m_2}{2} \right),$$

де m_1 – математичне очікування величини до розладнання;

m_2 – передбачуване математичне очікування величини після розладнання;

σ – середнє квадратичне відхилення;

x – значення спостереження у момент часу t .

У даному випадку зручно застосовувати різновид АКС, у якому на кожному кроці із заданим порогом порівнюється різниця

$$g_t = S_t - \min_{k < t} S_k. \quad (1)$$

Якщо необхідно виявляти зміни θ у бік зменшення, різниця приймає вигляд

$$g_t = \max_{k < t} S_k - S_t \quad (2)$$

Сигнал про розладку подається в момент часу

$$\tau = \inf(t \geq 1 : g_t > h).$$

Розділення умов подання сигналу про розладнання «вгору» і «вниз», що забезпечується рівняннями (1) і (2), є важливим для даної задачі, оскільки зміни характеристик у бік посилення чинного тренда не призводять до розвороту тренда і не повинні враховуватися.

Оскільки стабільний тренд сам по собі характеризується зміною середнього значення спостережень, у даному випадку доцільно застосовувати АКС не до початкової кривої спостережень, а до її першої похідної за часом. Фізичний сенс такої похідної – швидкість зміни значень функції. Різке змінення значення першої похідної означатиме розладнання процесу. У разі, коли потрібно визначити зміну діючої тенденції, точка розвороту графіка буде характеризуватися тим, що значення першої похідної перетне нульову позначку. Відсіваючи розладнання, зміна спостережень у яких направлена у протилежний від нульової позначки бік (посилення тренда), а також розладнання, що лежать надто далеко від нульової позначки (уповільнення тренда, яке ще далеко від розвороту), отримуємо точки, що розділяють області дії трьох основних ринкових трендів.

Нехай після розбиття графіка маємо n відрізків стаціонарності. Кожен із відрізків характеризується вектором параметрів v , що включає в себе його тривалість у часі (довжину відрізка стаціонарності по осі $x - x_S$) і різницю між цінами на початку і в кінці відрізка (різниця по осі $y - y_S$). Таким чином, процес описується вектором випадкових величин (x_S, y_S) , типова гістограма розподілу якого показана на рис. 1.

Обчислення критеріїв χ^2 і точного критерію Фішера вказують на наявність імовірнісної залежності між функціями розподілу x_S та y_S ; отже, спрощення форми розподілу вектора до добутку двох одновимірних неможливе.

Отримана функція розподілу має складну форму з двома піками і не може бути апроксимована з прийнятним ступенем точності жодним зі стандартних двовимірних розподілів.

Наблизимо отриману функцію розподілу сумішшю двовимірних нормальних розподілів за допомогою ЕМ-алгоритму.

ЕМ-алгоритм припускає, що дані можуть бути кластеризовані і підпорядковуються лінійній комбінації (суміші) нормальних (гауссових) розподілів:

$$p(x) = \sum_{j=1}^k w_j p_j$$

$$\sum_{j=1}^k w_j = 1,$$

$$w_j \geq 0,$$

де $p_j(x)$ – функція правдоподібності j -ї компоненти суміші, w_j – її апіорна ймовірність.

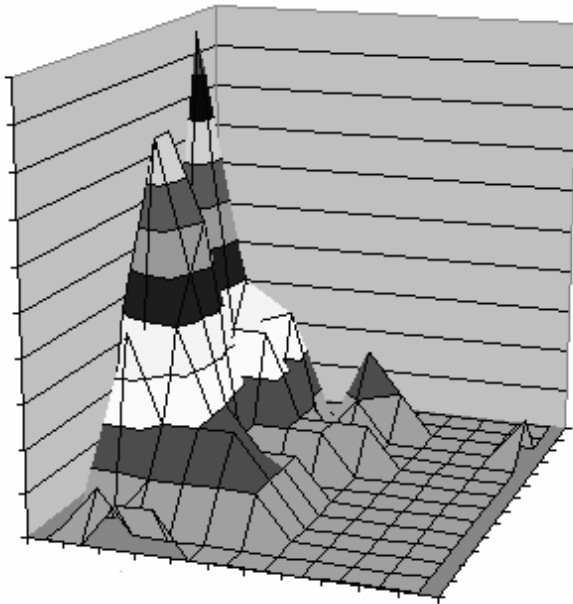


Рис. 1. Гістограма розподілу вектора (x_s, y_s)

Нехай функція правдоподібності належить до параметричного сімейства $\varphi(x; \theta)$ і відрізняється лише значеннями параметра θ .

Задача розділення суміші розподілів полягає в тому, щоб, маючи вибірку X^m випадкових спостережень із суміші, знаючи число k та функцію $\varphi(x; \theta)$, оцінити вектор параметрів $\Theta = (\omega_1, \dots, \omega_k, \theta_1, \dots, \theta_k)$.

Щільність імовірності нормального розподілу має вигляд

$$\varphi(x; \theta) = p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\},$$

де μ – математичне очікування, σ^2 – дисперсія.

Для двовимірного випадку

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \\ \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - \rho\frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2}\right]\right\}.$$

ЕМ-алгоритм складається з ітераційного повторення двох кроків [2]:

1. На Е-кроці обчислюється значення вектора прихованих змінних G .

$$g_{ij} \equiv P(\theta_j | x_i),$$

$$\sum_{j=1}^k g_{ij} = 1,$$

$$g_{ij} = \frac{\omega_j p_j(x_1, x_2)}{\sum_{s=1}^k \omega_s p_j(x_1, x_2)},$$

де $p_j(x_1, x_2)$ – значення функції щільності двовимірного нормального розподілу;

g_{ij} – приховані змінні, фізичний зміст яких – апостеріорна ймовірність того, що об'єкт x_i належить до j -ї компоненти суміші.

На М-кроці вирішується задача максимізації логарифма правдоподібності і знаходиться наступне значення вектора Θ за поточними значеннями вектора прихованих змінних.

$$Q(\Theta) = \ln \prod_{i=1}^m p(x_1, x_2) \rightarrow \max_{\Theta}.$$

Маємо

$$w_j = \frac{1}{m} \sum_{i=1}^m g_{ij}, \quad j = 1, \dots, k,$$

$$\theta_j = \arg \max_{\theta} \sum_{i=1}^m g_{ij} \ln p(x_1, x_2), \quad j = 1, \dots, k.$$

Метод не є стійким до початкового наближення, і кінцевий результат залежить від даних, якими ініціалізується алгоритм. Під час рішення задачі ініціалізація проводилася двома типами даних: результатом кластерізації вибірки за методом k середніх, а також заданням наближення в ручному режимі. На жаль, в обох випадках алгоритм виявився нечутливим до спаду щільності розподілу біля нульових значень по y , яке ясно видно на гістограмі незалежно від вибору відрізка вхідних даних і розміру вибірки. Результат роботи наведений на рис. 2.

Як бачимо, спроби наблизити форму функцій сумою гауссіан за допомогою ЕМ-алгоритму дають досить низьку точність апроксимації. Набагато кращий результат був отриманий при використанні

непараметричної апроксимації функції розподілу за допомогою В-сплайна. Вибір В-сплайн-апроксимації серед інших методів пояснюється тим, що:

1) гладкість одержуваної на виході поверхні відповідає природним формам функцій розподілу;

2) В-сплайни швидкі у побудові, а їхня властивість локальності дозволяє з меншими витратами оновлювати розподіл при отриманні нових даних про розкладання.

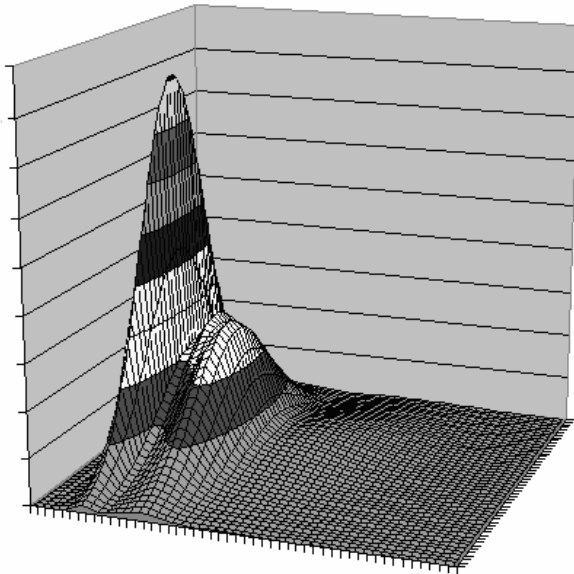


Рис. 2. Результат апроксимації розподілу сумою гауссіан за допомогою ЕМ-алгоритму

Розділимо графік розподілу функції (x_s, y_s) на прямокутні ділянки розміром dx на dy . Підрахуємо кількість попадань у кожну ділянку і на підставі отриманих значень побудуємо гістограму. Точки гістограми будуть служити вузлами сплайна. Отримуємо двовимірний масив вузлових точок величиною $m \times n$. Двовимірну В-сплайн-функцію запишемо у вигляді тензорного добутку одновимірних В-сплайнів, побудованих на кожній із двох осей координат. У такому випадку значення В-сплайн-поверхні в довільній точці буде рівним [3]:

$$x(s, t) = \sum_{i=0}^n \sum_{j=0}^m P_{i,j} B_i(s) B_j(t),$$

$$B_{k,1} = \begin{cases} 1, & t_k < t < t_{k+1} \\ 0, & \text{інакше} \end{cases}$$

$$B_{k,d} = \left(\frac{t - t_k}{t_{k+d-1} - t_k} \right) B_{k,d-1}(t) + \left(\frac{t_{k+d} - t}{t_{k+d} - t_{k+1}} \right) B_{k+1,d-1}(t),$$

де s, t – координати точки,

$P_{i,j}$ – значення функції в точці i, j ;

$B_{k,d}(t)$ – значення стикувальної функції порядку d в точці t , що належить до інтервалу $(t_k; t_{k+1})$;

$t_k; t_{k+1}$ – пари сусідніх вузлових точок сплайна.

При рішенні задачі використовувалися кубічні стикувальні функції і відкриті однорідні вузлові вектори

$$U = (-3, -2, -1, 0, 1, \dots, n+1),$$

$$V = (-3, -2, -1, 0, 1, \dots, m+1).$$

Результат відновлення щільності розподілу (x_s, y_s) наведений на рис. 3.

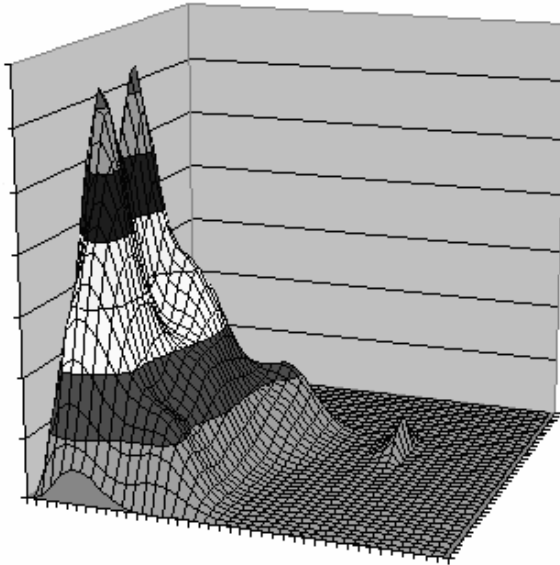


Рис. 3. Результат апроксимації щільності розподілу кубічним В-сплайном

Можемо бачити, що отримана поверхня є значно ближчою за формою до гістограми експериментального розподілу, за якою відновлювалася щільність. Крім того, на відміну від ЕМ-алгоритму, ступінню гладкості поверхні легко керувати, змінюючи розмір комірок гістограми та вводячи кратні вершини при побудові сплайна.

Висновки. Ураховуючи складність форми розподілу бачимо, що використання непараметричних методів відновлення щільності є більш доцільним, тому що на виході дає результат, значно ближчий до форми вихідної гістограми. Відновлена щільність може бути використана для отримання функції ризиків, яка виражає ймовірність зміни на пряму діючої тенденції у певний відрізок часу в майбутньому. Така функція може використовуватись як самостійно, так і як уточнююча величина при побудові торговельних стратегій.

Перспективи подальших досліджень: побудова функцій ризику, встановлення оптимальних налаштувань алгоритмів залежно від масштабу та форми вхідних даних.

Бібліографічні посилання

1. **Brodsky B.E.** Nonparametric Methods in Change-Point Problems / B.E.Brodsky, B.S. Darkhovsky. – Dordrecht: Kluwer Academic Publishings, 1993. – 210 p.
2. **Королёв В.Ю.** ЕМ-алгоритм, его модификации и их применение к задаче разделения смесей вероятностных распределений: теоретический обзор / В.Ю. Королев. – М.: ИПИРАН, 2007. – 94 с.
3. **Rogers D.F.** An Introduction to NURBS with Historical Perspective / D.F. Rogers. – San Francisco: Morgan Kaufmann Publishers, 2000. – 344 p.

Надійшла до редколегії 12.09.2013 р.

УДК 004.942:378.14

О.Г. Байбуз, Н.І. Харченко

Дніпропетровський національний університет імені Олеся Гончара

СИСТЕМИ УПРАВЛІННЯ ДИСТАНЦІЙНИМ НАВЧАННЯМ

Розглянуто структурні компоненти та основні властивості найпоширеніших зарубіжних та вітчизняних систем управління дистанційним навчанням. Виконано порівняльний аналіз систем навчання Moodle, Lotus Learning Space, «Веб-клас-ХПІ», ATutor.

Ключові слова: *дистанційне навчання, системи управління дистанційним навчанням, стандартизація систем дистанційного навчання, SCORM.*

Рассмотрены структурные компоненты и основные свойства наиболее распространенных зарубежных и отечественных систем управления дистанционным обучением. Проведен сравнительный анализ систем обучения Moodle, Lotus Learning Space, «Веб-класс-ХПИ», ATutor.

Ключевые слова: *дистанционное обучение, системы управления дистанционным обучением, стандартизация систем дистанционного обучения, SCORM.*

It considers structural components and the basic properties of the most common foreign and national systems of distance learning. The comparative analysis of learning Moodle, Lotus Learning Space, «Web-Class KPI», ATutor is made.

Keywords: *distance learning, distance learning management system, standardization of e-learning, SCORM.*

Постановка проблеми. Основними напрямками модернізації освіти є забезпечення можливості продовжувати навчання в межах концепції безперервної освіти на базі телекомунікаційних засобів.

Існування людини в інформаційному просторі вимагає її компетентності в тому, що стосується інформаційного співтовариства, форматів і стандартів реалізації інформаційних програм, умінь і навичок використання інформаційного ресурсу.

Розвиток педагогічних систем є одним із визначальних напрямів розв'язання проблем підвищення якості освіти. Відповідно, відбувається активний пошук нових форм організації освіти, які б були засновані на сучасних інформаційних і комунікаційних технологіях навчання й підвищення кваліфікації. Однією з таких форм є дистанційне навчання.

Постановка цілей статті. Стаття спрямована на дослідження характеристик програмних засобів для реалізації дистанційного навчання та аналіз існуючих систем управління дистанційним навчальним процесом, які відповідають загальноприйнятим умовам стандартизації.

Основний матеріал. Під дистанційним навчанням розуміється індивідуалізований процес передавання і засвоєння знань, умінь, навичок і способів пізнавальної діяльності людини, що відбувається за опосередкованої взаємодії віддалених один від одного учасників навчання у спеціалізованому середовищі, яке створене на основі сучасних психолого-педагогічних та інформаційно-комунікаційних технологій.

Дистанційне навчання є оперативним і таким, яке завдяки інформаційним та комунікаційним технологіям повністю відповідає вимогам часу, має низку позитивних рис та потужний потенціал для свого ефективного впровадження в систему підвищення кваліфікації працівників, перш за все за умов незалежності від фізичного розташування слухачів та організації, що долучені до подібної системи.

Актуальність і доцільність упровадження дистанційного навчання в освітню діяльність інститутів післядипломної освіти обумовлюється низкою причин, серед яких важливими є такі:

- дистанційне навчання спроможне виконувати роль каталізатора процесу реформування освітньої діяльності в інституті щодо надання їй рис, характеристик і показників систем відкритої освіти (доступність, реальна неперервність, особистісна орієнтованість, демократичність, варіативність тощо);

- дистанційне навчання активізує застосування у навчальному процесі інформаційних, комп'ютерних, телекомунікаційних технологій, підвищує творчу складову навчання, ініціює формування в Інтернеті вітчизняного веб-середовища післядипломної педагогічної освіти [3].

Важливою складовою частиною дистанційного навчання є його реалізація за допомогою використання інформаційних технологій, а саме системи управління навчанням, створеної для розробки, управління та поширення навчальних матеріалів онлайн із забезпеченням спільного доступу багатьох користувачів.

В англomовній літературі можна зустріти таку аббревіатуру систем управління навчанням:

- LMS – Learning Management System (система управління навчанням);
- CMS – Course Management System (система управління курсами);
- LCMS – Learning Content Management System (система управління навчальним матеріалом);
- MLE – Managed Learning Environment (оболонка для управління навчанням);
- LSS – Learning Support System (система підтримки навчання);
- LP – Learning Platform (освітня платформа);
- VLE – Virtual Learning Environments (віртуальні навчальні середовища).

Найбільш поширеними системами управління є LMS та CMS.

У зв'язку з поширенням дистанційного навчання у багатьох освітніх галузях питання стандартизації технологій дистанційного навчання вирішуються на національному та міжнародному рівнях. Стандартизація систем дистанційного навчання (СДН) взагалі та електронних дисциплін зокрема почалася з середини 80-х років XX століття. Стандарт AICC (Aviation Industry Computer-Based Training Committee) [1] був створений у 1988 році і впорядковував основи обміну текстових файлів у СДН для пілотів літаків. Із часом цей стандарт уже не відображав нові можливості Інтернет-технологій і тому для створення нового стандарту був організований консорціум, до складу учасників якого увійшли Apple, IBM, Oracle, Sun Microsystems, Microsoft, University of California – Berkley. Консорціум було названо IMS Global Learning Consortium, і основним його завданням стала розробка сучасного стандарту обміну навчальними матеріалами на основі XML. У результаті було розроблено стандарт SCORM [2] – для електронного навчання через Веб.

Наявність стандартів є важливою для будь-якого користувача інформаційних технологій, оскільки саме завдяки стандартизації кожен користувач може комбінувати обладнання та програми різних виробників відповідно до своїх індивідуальних потреб. Якщо єдиний стандарт відсутній, то користувач повинен обмежуватися пристроями та програмами лише одного виробника. Стандартизації підлягають як обладнання, так і програмне забезпечення, зокрема програми, що використовуються в електронному навчанні.

До найбільш поширених стандартів у сфері електронного навчання належать такі:

IEEE – Institute of Electrical and Electronic Engineers (Інститут електротехніки та електроніки), Комітет технології освітніх стандартів (LTSC – Learning Technology Standards Committee) (<http://ltsc.ieee.org/>);

– AICC – Airline Industry Computer Based Training Committee (Міжнародний комітет по комп'ютерному навчанню в авіації) (<http://www.aicc.org/>);

– IMS – Instructional Management Systems (Системи організації навчання), Консорціум всесвітньої освіти – Специфікація IMS –XML – базування стандарт описує структуру курсу (<http://www.imspj.org/>);

– ADL – Advanced Distributed Learning (Просунуте розподілене навчання) і створений ADL стандарт SCORM – Sharable Content Object Reference Model (Модель обміну навчальними матеріалами) (<http://www.adlnet.org/>);

– ARIADNE – Alliance of Remote Instructional Authoring & Distribution Networks for Europe (Консорціум АРІАДНА) (<http://www.ariadne-eu.org/>) стандартизація обміну навчальним контентом для Європейського Союзу;

– PROMETEUS <http://www.prometeus.org/>;

– The Dublin Core Metadata Initiative <http://dublincore.org/>.

Серед усіх продуктів стандартизації електронного навчання, що з'явилися останнім часом, SCORM отримав найширше визнання. Ця модель використовується при створенні систем навчання, що спираються на ресурси інтернету. Еталонна модель SCORM складається із трьох частин:

– Вступ, або оглядова частина (the Overview);

– опис моделі інтеграції змісту (the Content Aggregate Model);

– опис робочого середовища, або середовища виконання програм (the Run-Time Environment - RTE).

У першій частині описуються стандарти ADL (Advanced Distributed Learning) і дається логічне обґрунтування створення еталонної моделі. Друга частина містить практичні поради щодо виявлення ресурсів і перетворення їх на структурований навчальний матеріал. В останній частині даються практичні поради щодо здійснення зв'язку з веб-середовищем і відстеження його вмісту.

В ідеальній ситуації, відповідної еталону SCORM, усі елементи навчальних програм функціонально сумісні з усіма системами LMS та середовищами VLE. Будь-яку відповідну стандарту навчальну комп'ютерну програму можна ввести в наявну систему організації

навчання або віртуальне середовище, і між ними буде можливий обмін даними.

SCORM – це, скоріше, не стандарт, а еталон, за допомогою якого перевіряються ефективність і практична застосовність набору окремих специфікацій та стандартів. Цей еталон використовується такими розробниками стандартів, як IEEE і IMS, для об'єднання створених ними специфікацій.

Потрібно мати на увазі, що SCORM поки ще остаточно не утвердився як стандарт і що процедуру незалежного сертифікування для нього ще навіть не розпочато. Тому стосовно SCORM правомірно вживати термінологічний вислів «претендує на відповідність стандарту». Тим не менше для ефективної роботи у системах навчання, що використовують ресурси Інтернету, дотримання вимог SCORM необхідне.

Згідно з вимогами SCORM навчальні програми повинні містити три основних компоненти:

1) мова взаємодії програм (run-time communications) – іншими словами, стандартна мова, на якій навчальна програма «спілкується» із системою організації навчання (LMS) або з віртуальним середовищем навчання (VLE). Наявність такої мови важлива насамперед тому, що вона дозволяє запуснути й завершити програму навчання, перебуваючи в LMS або VLE. Крім того, ця мова робить можливою передачу даних про оцінки з навчальної програми в LMS;

2) файл-маніфест, або пакет утримання (Content package). Цей файл містить повний опис курсу навчання та його складових;

3) метадані про курс. Кожен фрагмент курсу – зображення, сторінка HTML або відеокліп – асоціюється з певним файлом метаданих, у якому містяться вказівки на те, що цей фрагмент собою являє і де знаходиться.

SCORM – це зібрання специфікацій і стандартів, які були зібрані в декілька «технічних книг». Кожна може розглядатися як окрема книга. Майже всі специфікації та основні принципи взяті від інших організацій. Ці технічні книги стосуються трьох головних тем:

- «Content Aggregation Model (CAM)» (Модель нагромадження змісту);

- «Run-time Environment (RTE)» (Середовище виконання);

- «Sequencing and Navigation (SN)» (Впорядкування та навігація).

ADL буде оновлювати ці книги або додавати нові у міру необхідності.

Оскільки передбачається, що книги можуть використовуватися окремо, у них існують загальні повторювані розділи. Так, наприклад, книга RTE, яка стосується механізмів взаємодії, відтворення й запуску навчальних матеріалів у середовищі виконання на основі стандартного інтерфейсу і моделі даних, часто звертається до об'єктів Sharable Content Objects (SCOs). Більш докладно про об'єкти Sharable Content Objects (SCOs) (об'єднані одиниці змісту) йдеться в книзі Content Aggregation Model (CAM) .

Впровадження дистанційної форми навчання реалізується за допомогою програмних засобів, побудованих на сучасних інформаційно-комунікаційних технологіях, які одержали загальну назву «системи дистанційного навчання» (СДН).

При виборі програмного забезпечення для систем навчання доречно враховувати такі характеристики:

- надійність в експлуатації;
- безпека;
- сумісність (відповідність стандартам);
- зручність використання та адміністрування;
- модульність;
- забезпечення доступу;
- вартість ПЗ, супроводження та апаратної частини.

Розглянемо вищезазначені характеристики окремо.

Надійність в експлуатації. Цей параметр характеризує зручність адміністрування й простоту оновлення контенту за допомогою вже існуючих шаблонів. Обираючи програмне забезпечення, необхідно звернути увагу на те , щоб зміст навчального курсу та структура сайту були розділені, для того щоб при оновленні контенту ви не могли випадково видалити важливі позиції меню. Крім того, якщо в систему важко додавати нових користувачів, виключати старих, додавати контент, якщо виникають проблеми з оновленням сайту та інше, викладачі швидко відмовляться від її використання.

Сумісність. Системи мають бути сумісні з іншими. Хоча «універсального» програмного рішення, відповідного всім можливим стандартам, не існує, все ж можна вибрати систему, що підтримує хоча б один широко розповсюджений стандарт.

Сумісність корисна при обміні (переміщенні) контенту з однієї системи управління навчання в іншу, при використанні розроблених курсів, при введенні в курс справи нових співробітників.

Один зі способів гарантувати сумісність – шукати програмне забезпечення, що підтримує певні стандарти, прийняті в індустрії. В ідеальному випадку воно повинне дозволяти використання одних і тих же навчальних матеріалів у різних системах управління навчання та управління контентом.

Сумісність – це можливість взяти один і той же навчальний матеріал і, не вносячи до нього змін, використовувати його в різних системах управління навчанням. У даний час стандарти є тільки загальним напрямом для досягнення сумісності. Не варто заздалегідь вважати, що навчальний курс, що відповідає стандарту SCORM, автоматично можна використовувати в системі управління навчанням на основі SCORM.

Зручність використання та адміністрування. При виборі нової системи необхідно забезпечити зручність її використання. Це важливий параметр, оскільки потенційні учні ніколи не стануть використовувати технологію, яка здається громіздкою або створює труднощі при навігації. Технологія навчання має бути інтуїтивно зрозумілою. У навчальному курсі необхідно знайти меню допомоги, нескладно переходити від одного розділу до іншого й спілкуватися з викладачем. Програмне забезпечення має бути простим і відкритим.

Модульність. У сучасних системах дистанційного навчання (ДН) можуть використовуватися взаємозамінні об'єкти знань – невеликі елементи навчального контенту. Це невеликі самодостатні інформаційні блоки, які можуть бути повторно використані для навчальних цілей. Об'єкти знань при цьому просто переносяться з одного курсу або уроку в інший, абсолютно відмінний від нього курс. Мета створення цих об'єктів – скорочення часу розробки курсів, оскільки, створивши один об'єкт, його можна повторно використовувати знову й знову. Такі блоки з'єднуються, роз'єднуються і розміщуються в різному порядку незалежно від їхнього розміру або кольору.

Забезпечення доступу. Це питання має два аспекти. Перший: ті, кого навчають, не повинні мати перешкод для доступу до навчальної програми. Наприклад, він має бути сумісним зі screen-readers-програмами, що забезпечують зчитування слів на екрані для тих, у кого ослаблений зір. Другий аспект – необхідно переконатися, що придбана технологія придатна для всіх можливих користувачів. Наприклад, якщо деякі з учнів не мають останнього варіанта

Macromedia Flash, вони не побачать анімації, створеної вами в цій технології.

Програмне забезпечення (ПЗ) має бути протестоване тими браузерями, які використовуватимуться учнями. Аби переконатися, що навчальна програма працює на тій платформі, на якій необхідно здійснити тестування за кількома сценаріями, слід провести тестування на декількох комп'ютерах із різними варіантами браузерів і програмами або дати жорсткі рекомендації щодо конфігурації обладнання.

Огляд та порівняння систем управління дистанційним навчанням

Сьогодні широко використовується велика номенклатура систем дистанційного навчання та управління дистанційним навчанням як із відкритим кодом (умовно безплатних), так і платних, широкоживаних та вузькоорієнтованих.

Moodle (Modular Object-Oriented Dynamic Learning Environment) – пакет модульного програмного забезпечення з відкритим кодом (ліцензія GNU GPL), який призначений для створення курсів дистанційного навчання та web-сайтів. Ця програма управління дистанційним навчанням, орієнтована на взаємодію між викладачем і студентом, також використовується для підтримки очних курсів. Moodle може бути встановлений на будь-який комп'ютер, що підтримує PHP та роботу із СУБД MySQL, PostgreSQL, Microsoft SQL Server; програмне забезпечення є кросплатформним [4].

До основи проекту покладено п'ять принципів, які об'єднані спільною назвою «соціальний конструктивізм»:

- у сучасному навчальному середовищі ми всі одночасно є потенційними вчителями та учнями;
- ми успішні у навчанні, особливо тоді, коли намагаємось створити щось, чи пояснювати щось людям;
- розуміння інших дає змогу вивчити їх більш індивідуально;
- великий внесок у навчання робить спостереження за роботою більш досвідчених педагогів;
- навчальне середовище має бути гнучким, забезпечувати учасникам навчального процесу простий інструмент для реалізації їхніх навчальних потреб [6].

При проектуванні Moodle особлива увага приділялася таким принципам:

- просування педагогіки соціального конструкціонізму (співпраця, активне навчання, критична рефлексія тощо);
- підтримка різних підходів до навчання: дистанційне, змішане, очне;
- простий, інтуїтивно зрозумілий, ефективний, крос-платформенний інтерфейс у вікні браузера;
- проста установка на більшість платформ, що підтримують PHP;
- сумісність із більшістю широко використовуваних баз даних;
- список курсів містить описи і є доступним будь-якому користувачеві;
- курси структуруються за категоріями;
- істотна увага приділяється питанням безпеки;
- для більшості текстових областей (ресурси, повідомлення форумів тощо) використовується вбудований WYSIWYG HTML-редактор.

Lotus Learning Space розроблена компанією IBM, надає можливість учитися й викладати в асинхронному режимі, звертаючись до матеріалів курсів у зручний час, брати участь в онлайн-заняттях у режимі реального часу. Викладач може створювати зміст курсу в будь-яких програмах і потім розміщувати створений матеріал у Learning Space [7]. Програма має гнучку систему редагування й адміністрування курсу, дозволяє обирати різні режими викладання й відстежувати поточні результати студентів.

Курси організовані у вигляді послідовних занять, які можуть бути самостійними, інтерактивними або колективними. Самостійні заняття зазвичай містять матеріали для читання і тести, які необхідно виконати після вивчення матеріалу. Інтерактивні заняття включають лекції у віртуальному класі, участь в онлайнівій дискусії або чаті, роботу з віртуальною дошкою (Whiteboard) і системою сумісного перегляду Web-сайтів (Follow me). Інтерактивні заняття плануються на певну дату і часто проводяться викладачем у віртуальному класі в режимі реального часу. Поточні результати студентів (етап проходження курсу, оцінки, витрачений час, кількість звернень та ін.) зберігаються в базі даних. Ця інформація доступна викладачеві у будь-який час у вигляді звітів різної форми. Колективні заняття передбачають заняття в офлайновій та онлайнівій дискусіях, чаті [7].

Віртуальне навчальне середовище – «**Веб-клас-ХІІ**». Ця СДН, розроблена проблемною лабораторією дистанційного навчання

НТУ«ХП», призначена для створення динамічного інформаційного простору, який має на меті забезпечувати продуктивну навчальну діяльність і враховувати всі пізнавальні потреби слухачів, а саме: презентацію структурованих і мотивованих навчальних матеріалів; підтримку пізнавальної та діяльної активності користувачів, необхідну комунікацію й співробітництво учасників навчального процесу у різних формах; засоби адміністрування навчального процесу та його активного супроводження; система має динамічно настроювану мову інтерфейсу (українську, російську, англійську). Перша версія системи «Веб-Клас-ХП» була використана у травні – вересні 2001 року для організації та проведення дистанційного курсу «Практичний курс дистанційного навчання». Із курсами, які реалізовані у даному середовищі, можна ознайомитися на сайті університету (<http://dl.kpi.kharkov.ua>) [8].

ATutor – модульна система дистанційного керування навчанням із відкритим кодом. Поширюється на основі GNU General Public License. Для установки необхідно мати комп'ютер із веб-сервером Apache 1.3.x, PHP версії 4.2.0 та MySQL версій 3.23.x і 4.0.12 (версії 4.1.x і 5.x офіційно не підтримуються). Система розроблена з урахуванням доступності та можливістю адаптації за бажанням користувача. Щодо операційної системи сервера обмежень немає – система є крос-платформеною [5].

Використання системи ATutor дозволяє викладачам легко організовувати різні курси навчання. Студенти ж отримують адаптивне і просте середовище навчання. Адміністратору ця система також особливого клопоту не завдасть. Зовнішній вигляд можна змінити буквально за кілька клацань мишки, доступність вихідного коду й відкриті інструменти, застосовувані для побудови сервера курсів, дозволяють у разі крайньої необхідності внести і більш серйозні зміни. Серед необхідного для створення курсів та управління ними і процесом навчання є й засоби обміну повідомленнями. Особлива увага приділяється й безпеці. За допомогою додаткових модулів можна наростити функціональність.

Отже, описані системи дистанційного навчання – програмне забезпечення процесу дистанційного навчання, проте мають різні параметри й можливості. Наведемо основні відомості про описані системи управління у таблиці.

Спільними зусиллями програмістів і педагогів розроблено достатню кількість систем дистанційного навчання для організації

дистанційного навчання у навчальних закладах різного типу. Передбачити, що буде в майбутньому, складно, але вже зараз можна підібрати необхідне програмне забезпечення для автоматизації навчання, і не тільки дистанційного, але також очного й заочного.

Таблиця

Характеристики Система управління	Виробник	Умови поширення	Стандартизація	Підтримка української мови	Система перевірки знань
Moodle	Австралія	Вільно	IMS, SCORM	Так	Тести, завдання, семінари, активність на форумах, контроль активності
Lotus Learning Space	США	Платно	AICC, SCORM	Так	Тести, інтерактивні заняття, контроль активності
Веб-клас-ХІІІ	Україна	Умовно безплатно	Відсутня	Так	Тести, активність на форумах, контроль активності
ATutor	Канада	Вільно	IMS, SCORM	Так	Тести, контроль активності

Основні недоліки багатьох систем:

- відсутність у багатьох системах української «вузівської» специфіки організації навчального процесу, тобто документообігу між керівництвом ВНЗ, деканатами факультетів, кафедрами тощо;
- відсутність українських сертифікатів із захисту інформації.

Можливо, у майбутньому доведеться перейти на іншу систему, але найголовніше – залишаться навички роботи з подібними системами, тобто буде створена інфраструктура їх підтримки. А якщо використані системи відповідатимуть стандартам, то й дані перенести буде нескладно. **Висновки.** Інформація, наведена у роботі, містить інформацію щодо можливостей існуючих систем управління

дистанційним навчанням та сьогодні є актуальною для розвитку й розробки цих систем. Інформаційно-комунікаційні технології постійно розвиваються, і нові версії розглянутих платформ дистанційної освіти матимуть нові можливості для вдосконалення навчального процесу.

Перспективи подальших досліджень. Розглянувши основні принципи створення систем управління дистанційним навчанням, вважаємо, що наступним кроком буде використання статистичних методів для розвитку цього напрямку.

Бібліографічні посилання

1. [Електронний ресурс]. – Режим доступу: http://www.aicc.org/joomla/dev/index.php?option=com_content&view=article&id=7&Itemid=9.
2. [Електронний ресурс]. – Режим доступу: <http://www.adlnet.gov/scorm/scorm-2004-4th/#tab-resources>.
3. **Олійник В.В.** Підвищення кваліфікації керівників освіти за дистанційною формою навчання / В.В. Олійник, В.Ю. Биков. – К.: Логос, 2006. – 408 с.
4. **Белозубов А.В.** Система дистанционного обучения Moodle : учеб.-метод. пособие / А.В.Белозубов, Д.Г.Николаев. – СПб., 2007.
5. ATutor User Documentation [Електронний ресурс]. – Режим доступу : <http://help.atutor.ca/general/>.
6. **Андреев А.В.** Практика электронного обучения с использованием Moodle / А.В.Андреев, С.В.Андреева, И.Б.Доценко. – Таганрог: ТТИ ЮФУ, 2008. – 146 с.
7. **Bowling E.** The evolution of Lotus e-Learning Software.[Електронний ресурс]. – Режим доступу: http://www.ibm.com/developerworks/lotus/library/ls-elearning_evolution.
8. Віртуальне навчальне середовище «Веб-клас ХІІІ» [Електронний ресурс]. – Режим доступу: <http://dl.kpi.kharkov.ua/WebCl/index.htm>.

Надійшла до редколегії 30.09.2013 р.

ВИЗНАЧЕННЯ ЗНАЧУЩИХ ПАРАМЕТРІВ МОВНОГО СИГНАЛУ ЗА ДОПОМОГОЮ КОЛОКОЛОПОДІБНИХ ФУНКЦІЙ

Наведений один із можливих варіантів технології визначення функціонального стану людини за його мовою на підставі отримання різниці спектрально-часових представлень, а також їх відношень.

Ключові слова: *функціональний стан, людина, мова, настрій, розподіл частот, значущі параметри, апроксимація, колоколоподібні функції, функція Аньєзі.*

Приведен один из возможных вариантов технологии определения функционального состояния человека по его речи на основании получения разности спектрально-временных представлений, а также их отношений.

Ключевые слова: *функциональное состояние, человек, речь, настроение, распределение частот, аппроксимация, колоколообразные функции, функция Аньези.*

This work tells us how to determine the functional state of a person according to his speech. Which is based on the finding difference between two wavelength-time representations.

Keywords: functional state, human, speech, mood, frequency distribution, witch of agnesy, approximation.

Постановка задачі. Авторами був виконаний ряд досліджень при реалізації пристроїв розпізнавання промови.

Задача побудови пристроїв розпізнавання для сучасних персональних ЕОМ складається із двох частин:

1) вибір параметрів опису об'єктів (фонем, складів, слів, фраз), а також методів інтерпретації описів і проектування програмних систем, побудови програмних засобів реалізації описів і розпізнавання;

2) вбудовування розроблених проектів і програмних реалізацій у системні середовища (оболонки) ПЕОМ.

Останніми роками рівень розвитку обчислювальних засобів значно виріс, що дозволило вирішувати задачі побудови систем розпізнавання із залученням великого числа параметрів і методів, орієнтованих на глибинну сутність мовного сигналу.

Однак залишилася необхідність узгодження (вирішення) суперечливих вимог: забезпечення високої надійності розпізнавання для великих словників об'єктів, що потребує залучення великого числа параметрів (ознак) і, відповідно, великого часу їхньої обробки; виконання обробки в реальному масштабі часу (мінімальні витрати часу).

У цілому задача проектування пристрою розпізнавання мовлення може бути сформульована в такий спосіб:

а) вибрати таку систему параметрів, що забезпечувала б досить повний опис деякої заданої кількості об'єктів мовних сигналів та їхню відмінність один від одного в цій системі з деякою заданою надійністю; параметри повинні обчислюватися зі швидкістю, що має реальний час функціонування пристрою розпізнавання мовлення; сукупність параметрів кожного об'єкта повинна займати такий обсяг пам'яті, щоб забезпечувати можливість розміщення в ній еталонних параметрів досить великих словників (наприклад, одночасне розміщення не менш 500–1000 слів із параметрами);

б) здійснити сегментацію мовлення і групування інформації про параметри для формування структури слова і завдання відповідності між символьним і параметричним представленнями;

в) побудувати стратегію розпізнавання, що забезпечує мінімальні витрати часу на пошук найбільш відповідного еталона з максимальною надійністю в заданій системі параметрів;

г) розробити й реалізувати проект пристрою розпізнавання мовлення, що має можливість подальшого розвитку (розширення або звуження) із мінімальними витратами на його модифікацію.

У результаті теоретичних та експериментальних досліджень була розроблена загальна схема проектування пристроїв розпізнавання мовлення [1; 2], що містить складові частини:

- блок аналізу мовних сигналів;
- вибір типів мовних одиниць (сегменти, склади, фонemi, слова, сегментно-складове представлення);
- сегментація;
- вибір метрики та критеріїв порівняння;
- вибір методів порівняння.

При проектуванні складних програмних систем необхідно вирішити такі практичні задачі:

а) вибір універсального мовного середовища для написання модулів будь-якої складності і будь-якого функціонального призначення;

б) забезпечення мовних засобів для маніпулювання текстами до їхньої трансляції;

в) розробка технології проектування програмних систем, що містять параметри їхньої перебудови по вихідних текстах програм.

Алгоритм апроксимації на основі колоколоподібних функцій спектрально-часового представлення мовних сигналів

Задача апроксимації спектрально-часового представлення в класі колоколоподібних функцій полягає в розбивці заданого початкового опису на ділянки, що відповідають деяким мовним одиницям, і наступний їхній опис у класі заданих функцій. Вихідне спектрально-часове представлення розглядається у прямокутній області $R = [\omega_a, \omega_b] \times [T_c, T_d]$, в області R задана таблична спектрально-часова функція $S(\omega_k, t_l)$, де ω_k – дискретно задана частота, t_l – дискретно заданий час: $\omega_a = \omega_0 < \omega_1 < \dots < \omega_m = \omega_b$,

$$T_c = t_0 < t_1 < \dots < t_n = T_d.$$

Визначено класи функцій $\{W_i(\omega_k)\}$, $\{h_i(t_l)\}$, що мають такі властивості:

- для представлення резонансу функція повинна бути колоколоподібною (мати максимум на заданій частоті);
- форма функції повинна визначатися деякими параметрами;
- функції асимптотично наближаються до площини області R у будь-якому напрямку від максимуму;
- гілки функції повинні монотонно убувати (зростати), якщо максимум колоколу знаходиться за межами області R.

Функція $S(\omega_k, t_l)$ в області R містить довільну кількість сплесків спектральної енергії, розміщених довільним чином у заданій області. Необхідно функцію $S(\omega_k, t_l)$ сегментувати і апроксимувати у класі функцій $\{W_i(\omega_k)\}$, для цього визначити параметри сплесків функції

Спектрально-часова складова шукається у класі функцій модифікованого локона Аньєзі

$$(\omega^2 + r^2)W(\omega) - a^3 = 0, \quad W(\omega_k) = \frac{a_{(i)}^3}{c_{(i)}^2 + (\omega_k - \beta_{(i)})^2},$$

$$h(t_l) = \frac{b_{(j)}^3}{d_{(j)}^2 + (t_l - T_{(j)})^2},$$

$$\begin{aligned} & Z_{(i,j)}(k, q_{(i,j)}, a_{(i,j)}, c_{(i,j)}, \beta_{(i,j)}, b_{(i,j)}, d_{(i,j)}, T_{(i,j)}, \omega_k, t_l) = \\ &= S_{\mathcal{G}(j)}(\omega_k) W_{(i)}(\omega_k) h_{(j)}(t_l) = \\ &= \frac{q_{(i)}^{\text{lnk}}}{k} \frac{a_{(i)}^3}{c_{(i)}^2 + (\omega_k - \beta_{(i)})^2} \frac{b_{(j)}^3}{d_{(i)}^2 + (t_l - T_{(i)})^2}. \end{aligned}$$

Різницева схема вилучення складових – СЕТ [2] для спектральної функції $S(\omega_k, t_l)$:

[illegible]

Вилучення складових $S_{ij}(\omega_k, t_l)$ здійснюється для кожної клітки сітки $M \times N$, $j = \overline{0, M-1}$, $i = \overline{0, N-1}$. За методом найменших квадратів визначаються параметри $k, q_{(ij)}, a_{(ij)}, c_{(ij)}, \beta_{(ij)}, b_{(ij)}, d_{(ij)}, T_{(ij)}$ для

$$\sigma_{(i,j)}^2 = \sum_{\omega_k} \sum_{t_l} [S_{i-1,j-1}(\omega_k, t_l) -$$

$$-Z_{(i,j)}(k, q_{(i,j)}, a_{(i,j)}, c_{(i,j)}, \beta_{(i,j)}, b_{(i,j)}, d_{(i,j)}, T_{(i,j)}, \omega_k, t_l) \Gamma^2.$$

На першому кроці параметри визначаються для

$$\begin{aligned} q_{(1,1)}^2 &= \sum_{\omega_k} \sum_{t_l} [S(\omega_k, t_l) - \\ &- Z_{(1,1)}(q_{(1,1)}, a_{(1,1)}, c_{(1,1)}, \beta_{(1,1)}, b_{(1,1)}, d_{(1,1)}, T_{(1,1)}, \omega_k, t_l)]^2 = \\ &= \sum_{\omega_k} \sum_{t_l} \left\{ S(\omega_k, t_l) - \frac{q_{(1)}^{\ln k}}{k} \frac{a_{(1)}^3}{c_{(1)}^2 + (\omega_k - \beta_{(1)})^2} \frac{b_{(1)}^3}{d_{(1)}^2 + (t_l - T_{(1)})^2} \right\}^2, \\ \sigma &= S(\omega_k, t_l) - \frac{q_{(1)}^{\ln k}}{k} \frac{a_{(1)}^3}{c_{(1)}^2 + (\omega_k - \beta_{(1)})^2} \frac{b_{(1)}^3}{d_{(1)}^2 + (t_l - T_{(1)})^2}. \end{aligned} \quad (2)$$

Параметри $q_{(i,j)}, a_{(i,j)}, c_{(i,j)}, \beta_{(i,j)}, b_{(i,j)}, d_{(i,j)}, T_{(i,j)}$ визначаються на наступних кроках.

Для рішення задачі мінімізації (2) реалізований метод покоординатного спуску, що полягає у виборі напрямку зміни параметрів $k, q_{(i,j)}, a_{(i,j)}, c_{(i,j)}, \beta_{(i,j)}, b_{(i,j)}, d_{(i,j)}, T_{(i,j)}$ за правилом

$$\begin{aligned} q_{(i,j)} &= q_{0(i,j)} \pm \Delta_q, \quad a_{(i,j)} = a_{0(i,j)} \pm \Delta_a, \quad c_{(i,j)} = c_{0(i,j)} \pm \Delta_c, \quad \beta_{(i,j)} = \beta_{0(i,j)} \pm \Delta_\beta, \\ b_{(i,j)} &= b_{0(i,j)} \pm \Delta_b, \quad d_{(i,j)} = d_{0(i,j)} \pm \Delta_d, \quad T_{(i,j)} = T_{0(i,j)} \pm \Delta_T \end{aligned}$$

із заданою точністю EPS.

Для кожного виду колоколоподібних функцій існує свій параметр, що керує шириною відповідної функції щодо положення її максимуму. Цей параметр є аналогом добротності резонансних ланок. Для локона Аньєзі такими параметрами є: $c_{(i,j)}$ – за частотою, $d_{(i,j)}$ – за часом.

Базовий алгоритм розпізнавання на основі екстремальних функцій

У результаті рішення задачі сегментації формується деяка тимчасова послідовність, що відображає структуру мовного висловлення. В основу алгоритму розпізнавання покладене рішення задачі синтезу спектрально-часового представлення в базисі аналітичних функцій двовимірного локона Аньєзі, Гауса та ін. Для цього в частотній області діапазон частот від 100 Гц до 6 КГц представляється у вигляді 15 шарів із граничними частотами: (100÷400) Гц, (400÷800) Гц, (800÷1200) Гц, (1200÷1600) Гц, (1600÷2000) Гц, (2000÷2400) Гц, (2400÷2800) Гц, (2800÷3200) Гц,

(3200÷3600) Гц, (3600÷4000) Гц, (4000÷4400) Гц, (4400÷4800) Гц, (4800÷5200) Гц, (5200÷5600) Гц, (5600÷6000) Гц. Для кожного шару на сегменті в заданих частотних границях обчислюються параметри функцій Аньєзі, Гауса або ін., вирішуючи задачу апроксимації вихідного спектрально-часового представлення як суму локонів

$$Z(\omega_k, T_{IN}) = \sum_{i=1}^n G_i[W_{gij}(q_{ij}, \omega_k), W_{ij}(\alpha_{ij}, \beta_{ij}, c_{ij}, \omega_k), h_{ij}(\tau_{ij}, T_{ij}, d_{ij}, t_{IN})] \quad (4)$$

при $Z(\omega_k, T_{IN}) \rightarrow S(\omega_k, T_{IN})$.

За обчисленими значеннями параметрів локонів

$q_{(1,1)}, a_{(1,1)}, c_{(1,1)}, \beta_{(1,1)}, b_{(1,1)}, d_{(1,1)}, T_{(1,1)}$ будується п'ять груп описів

еталонних послідовностей кожного слова (складу):

- 1) максимуми функцій локона;
- 2) синтезоване спектрально-часове представлення;
- 3) бітове представлення із синтезованого спектрально-часового представлення;
- 4) двовимірний сплайн-опис спектрально-часового представлення на основі двовимірного фільтра [2];
- 5) вихідне спектрально-часове представлення $S(\omega_k, t_l)$.

При розпізнаванні реалізована ієрархія розпізнавання (3). Проста схема розпізнавання для словників до 1000 слів оперує з представленнями

Вид 2) → вид 5) ,

для яких задана внутрішня ієрархія, утворена послідовністю шарів у граничних частотах (100÷2000) Гц → (100÷3000) Гц → (100÷4000) Гц → (100÷4000) Гц → (100÷5000) Гц → (100÷6000) Гц.

На кожному кроці внутрішньої ієрархії реалізується розподіл обсягу розпізнаваної підмножини на дві.

Повна схема містить у собі всі параметричні представлення:

Вид 1) → вид 2) → вид 3) → вид 4) → вид 5).

Зіставлення за представленням «вид 1)» має найбільшу швидкодію, зіставлення по представленнях «вид 2)», «вид 5)» має найбільшу надійність. Кожен рівень зіставлення поставляє наступному половину еталонних послідовностей, звужуючи розпізнавану підмножину. На останньому етапі приймається остаточне рішення. Експериментальні дослідження на словнику у 500 слів виявили високу стійкість одержуваних правильних відповідей розпізнавання. У момент

навчання надійність розпізнавання становила не менш 98%. У процесі експлуатації на окремі слова виконується донавчання. Етап навчання полягає в накопиченні еталонних параметрів слів (слів-складів), вимовлених по одному разу, донавчання – додавання одного еталона ненадійно розпізнаваного слова.

Задача проектування бази знань аналізу і розпізнавання мовлення

Аналіз результатів у діючій системі розпізнавання виявив значну обмеженість алгоритмів і застосовуваних параметрів із причини жорсткості їхньої структури й неможливості підстроювання системи на кроці виконання з метою підвищення надійності розпізнавання та адаптації системи до ситуаційної варіативності мовлення. Подібну задачу можна і потрібно вирішувати в середовищі баз знань, керуючи системою розпізнавання на кроці виконання.

Задача побудови бази знань (БЗ) визначається цілями, що мають бути досягнуті в результаті її реалізації. Цілі створення БЗ – керування об'єктом та процесами, які в ньому відбуваються, що забезпечує досягнення деякого найкращого співвідношення між параметрами, які характеризують об'єкт і процеси, виявляють фактори, що викликають відхилення цих параметрів.

Об'єкт керування характеризується складом і призначенням як його частин, так і об'єкта в цілому, що визначає його статичний стан у вигляді інформаційного вектора $S(p_1, p_2, \dots, p_n)$ взаємозв'язку його параметрів p_l .

Процеси, що характеризують об'єкт, описуються функціями часу і параметрів динаміки об'єктів $f(t, p_1, p_2, \dots, p_n)$.

Задача керування може бути вирішена як задача моделювання передатної функції W об'єкта у виді сукупності його частин зі своїми передатними функціями W_i і як моделювання процесів у вигляді інформаційних потоків, що характеризують об'єкт у деякі фіксовані моменти часу $Y_{ki}(t_j, p_1, p_2, \dots, p_l, \dots, p_n)$, де k – номер частини об'єкта, i – номер інформаційного потоку в k -й частині об'єкта, p_l – l -й параметр як функція часу i -го інформаційного потоку в k -й частині об'єкта.

Задача практичного порівняння сигналів

Розглянемо задачу зіставлення двох сигналів на практиці. Порівняння будемо проводити на двох реалізаціях слова «чотири».

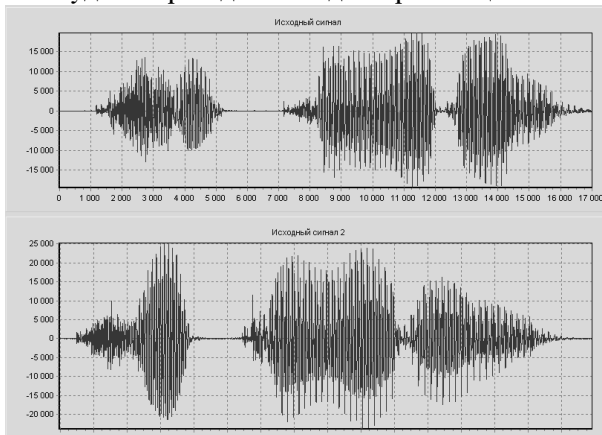


Рис. 1. Вид сигналу слова «чотири» у двох варіаціях

Після деяких стандартних перетворень із сигналом до отриманої матриці можна застосувати модифікований локон Аньєзі. Це перетворення у деякому сенсі подібне до перетворення Фур'є із тією лише різницею, що за базисну функцію обирається нетригонометрична функція.

Після цього можна подивитись на дві реалізації слова «чотири» у спектрально-часовому представленні.

Залежно від вимог до швидкості обчислення існують декілька підходів до отримання необхідних значень.

Перший із них полягає у пошуку максимального піку на кожній ітерації і відліку колоколу Аньєзі від цієї точки (де значення $S(\omega, t)$ набуває максимуму).

Інший підхід полягає в розбитті координатної площини на невеликі сегменти і підрахунку функції Аньєзі в кожному з них послідовно. Такий підхід є дещо швидший.

Наступним логічним кроком є порівняння отриманих величин. Спочатку розглянемо результат операції віднімання.

Як можна бачити з рис. 4, різниця є доволі невеликою і сигнали схожі. Іншою картина буде в разі порівняння діленням.

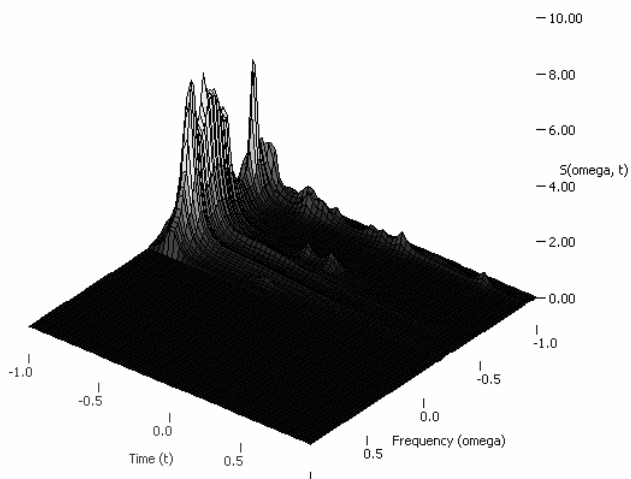


Рис. 2. Спектрально-часове представлення першої реалізації слова «чотири»

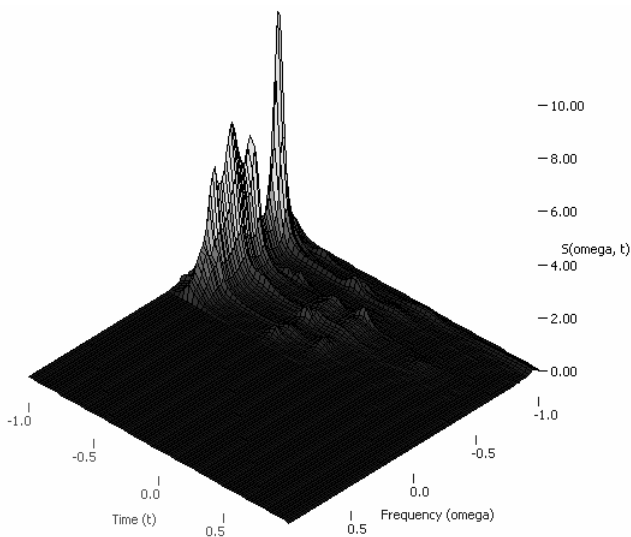


Рис. 3. Спектрально-часове представлення другої реалізації слова «чотири»

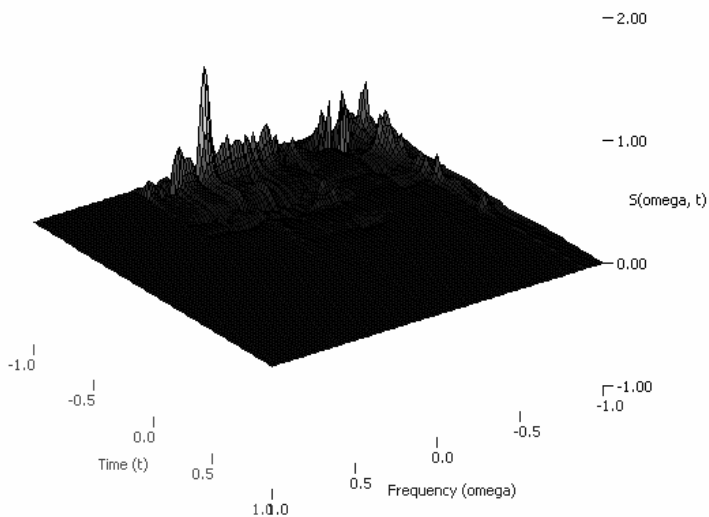


Рис. 4. Різниця спектрально-часового представлення (результат операції віднімання у збільшеному масштабі)

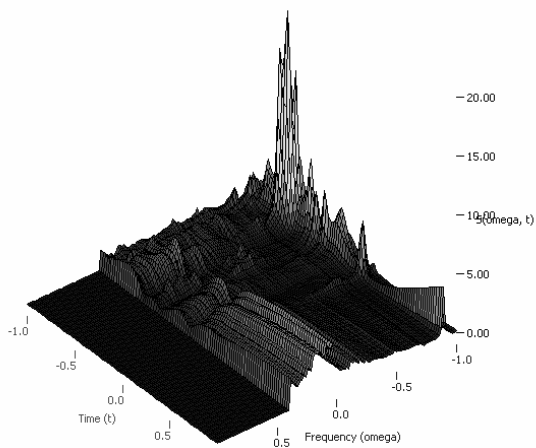


Рис. 5. Відношення спектрально-часових представлень (результат операції ділення)

Таким чином, ми отримали матрицю – передаточну функцію, що ідентифікує людину унікальним чином. Тобто виділили параметричне представлення для конкретної людини, що виражає певну емоцію.

Висновки. Відповідно до функції розпізнавання, технології проектування і загальної гіпотези про мовний сигнал як ієрархічну послідовність станів з обмеженими відхиленнями і необхідності його багаторівневої апроксимації як нелінійної функції побудована система виділення параметричного представлення мови людини.

Бібліографічні посилання

1. **Карпов О.М.** Вступ до курсу «Технологія проектування пристроїв розпізнавання промови»: навч. посіб. / О.М. Карпов. –Д.: РВВ ДДУ, 2000. – 64с.
2. **Карпов О.Н.** Технология построения устройств распознавания речи: монография / О.Н. Карпов. – Днепропетровск: Изд-во ДНУ, 2001. – 190 с.
3. **Карманов В.Г.** Математическое программирование / В.Г.Карманов. – М: Наука, 1975. – 272 с.
4. **Попов Э.В.** Экспертные системы. Решение неформализованных задач в диалоге с ЭВМ / Э.В.Попов. – М.: Наука, 1987. –288 с.
5. Компьютерные технологии распознавания речевых сигналов / [О.Н.Карпов, А.Г.Габович, Б.Г.Марченко, В.А.Хорошко и др.]. –К.: ООО «ПолиграфКонсалтинг », 2005, – 138 с.
6. **Douglas-Cowie** A New Emotion Database: Considerations, Sources and Scope / E.Douglas-Cowie, R.Cowie, M.Schroder. – Newcastle, 2000.

Надійшла до редколегії 12.10.2013 р.

ПОБУДОВА ПОВЕРХОНЬ ЗА ДОПОМОГОЮ ТЕХНОЛОГІЇ Т-СПЛАЙНІВ

Проведено огляд та порівняльний аналіз сучасних методів побудови поверхонь, а саме Т-сплайнів і NURBS. Розглянуто основні риси Т-сплайнів, їхні особливості, переваги та недоліки.

Ключові слова: *NURBS поверхні, Т-сплайни, локальне уточнення.*

Проведен обзор и сравнительный анализ современных методов построения поверхностей, а именно Т-сплайнов и NURBS. Рассмотрены основные черты Т-сплайнов, их особенности, преимущества и недостатки.

Ключевые слова: *NURBS-поверхности, Т-сплайны, локальное уточнение.*

It was made a review and comparative analysis of modern methods of building of surfaces, namely, T-Splines and NURBS. Also was considered the main features, characteristics, advantages and disadvantages of T-Splines.

Keywords: *NURBS surfaces, T-splines, local refinement.*

Постановка проблеми. Важливу роль в аналізі даних екологічного моніторингу відіграють методи візуалізації, а саме методи побудови поверхонь, за допомогою яких можна зробити оцінку потенційної небезпеки для навколишнього середовища. Залежно від галузі застосування використовуються різні підходи – такі як трикутні сітки [3–5], параметричні сплайн-поверхні [6; 7], скалярні сплайн-функції [8], функції з радіальним базисом [9]. Серед численної кількості методів постає проблема вибору найбільш ефективного методу побудови поверхні для даних гідрогеохімічного моніторингу підземних вод.

Аналіз останніх досліджень і публікацій. Сьогодні будь-який інженерний проект складається із двох частин: проектування та аналізу. У проектуванні найбільш широко застосовується технологія NURBS (неоднорідні раціональні В-сплайни), які, у свою чергу, є також і інженерним стандартом. Перевагами технології NURBS є те, що її зручно використовувати для побудови поверхонь будь-якої форми; більш того, вже існує багато ефективних алгоритмів для побудови

NURBS-об'єктів. У системах САПР майже всюди використовується технологія NURBS. Але основний недолік NURBS полягає в розривах та дублюваннях у місцях перетину поверхонь. Крім того, більшість поверхонь неможливо зобразити, використовуючи для цього лише одну NURBS-поверхню. Для усунення таких недоліків нещодавно було розроблено нову технологію, яка є узагальненням NURBS-технології. Ця технологія називається Т-сплайни [1]. Т-сплайни – вид поверхні вільної форми, подібної до неоднорідного раціонального В-сплайна (NURBS). Ключова відмінність Т-сплайнів від NURBS полягає у тому, що контрольні точки NURBS-поверхні повинні утворювати топологічну прямокутну сітку, у той час як у Т-сплайнів допустимі так звані внутрішні Т-точки (контрольна точка з трьома, а не чотирма сусідами).

Мета статті. Подати огляд технології побудови поверхні на основі Т-сплайнів; здійснити аналіз основних відмінностей побудови поверхонь на основі NURBS та на основі Т-сплайнів.

Основний матеріал. Т-сплайни були винайдені доктором Томасом Седербергом (Dr. Thomas Sederberg) у 2003 р. Заснована у 2004 р. компанія T-Splines, Inc. (Прово, штат Юта, США) розробляє програмне забезпечення для моделювання поверхонь вільної форми на основі запатентованої технології Т-сплайнів. Продукти компанії (Plug-in для Rhino і Maya, а також набір бібліотек T-Tools для вбудовування в інші додатки) знаходять застосування у промисловому дизайні архітектурних об'єктів, яхт, ювелірних виробів і товарів широкого вжитку.

Контрольна сітка для Т-сплайн-поверхні називається Т-сіткою. Якщо Т-сітка утворює прямокутну сітку, то Т-сплайн вироджений у В-сплайн-поверхню. Інформація про вузол для Т-сплайнів виражається за допомогою вузлового інтервалу – невід'ємне число, яке є різницею між двома вузлами. Вузловий інтервал зв'язується з кожним ребром у Т-сітці. На рис.1 показаний прообраз частини Т-сітки у (s,t) -параметричному просторі, d_i та e_i – вузлові інтервали. Вузлові інтервали мають задовольняти умові: сума всіх вузлових інтервалів з одного боку будь-якої фігури повинна дорівнювати сумі вузлових інтервалів на протилежному боці фігури. Наприклад, на рис. 1 на фігурі F_1 $e_1 + e_2 + e_4 = e_6 + e_7$ і на фігурі F_2 $d_6 + d_7 = d_8$.

Щоб визначити систему координат, спочатку треба визначити їхній початок. Нариклад, на рис.1 позначимо (s_0, t_0) початком координат. Далі можна знайти координати кожного вузла сітки. Отже, ґрунтуючись на виборі початку координат, маємо

$$s_0 = t_0 = 0, s_1 = d_1, s_2 = d_1 + d_2, s_3 = d_1 + d_2 + d_3, t_1 = e_1.$$

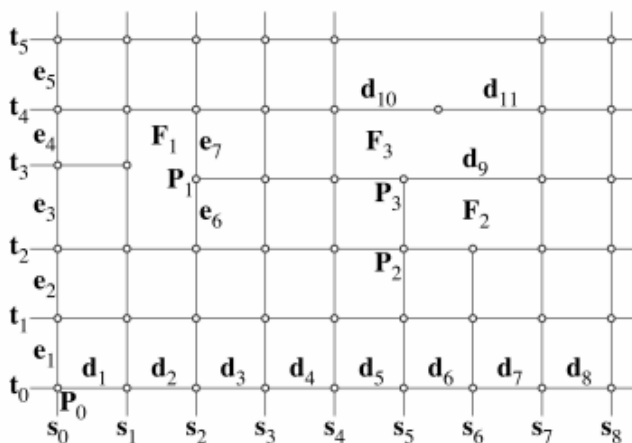


Рис. 1. Контрольна сітка для Т-сплайнів

Крім того, кожна контрольна точка має вузлові координати: наприклад, для P_1 ($s_5, t_2 + e_6$).

Ще одне правило для Т-сітки: щоб Т- подібне перехрестя з одного боку фігури могло з'єднуватись із Т-подібним перехрестям на протилежному боці фігури, треба, щоб сума вузлових векторів з обох боків була однаковою. Наприклад, для фігури F1 необхідно, щоб $e_3=e_6$ та $e_4=e_7$.

Т-сплайн-поверхня визначається так:

$$P(s, t) = (x(s, t), y(s, t), z(s, t), w(s, t)) = \sum_{i=1}^n P_i B_i(s, t), \quad (1)$$

де $P_i(x_i, y_i, z_i)$ – контрольні точки, w_i – їхні ваги відповідно, чий декартові координати визначаються $\frac{1}{w_i}(x_i, y_i, z_i)$.

Декартові координати точок на площині:

$$\frac{\sum_{i=1}^n (x_i, y_i, z_i) B_i(s, t)}{\sum_{i=1}^n w_i B_i(s, t)} \quad (2)$$

$$B_i(s, t) = N[s_{i0}, s_{i1}, s_{i2}, s_{i3}, s_{i4}](s)N[t_{i0}, t_{i1}, t_{i2}, t_{i3}, t_{i4}](t), \quad (3)$$

де $N[s_{i0}, s_{i1}, s_{i2}, s_{i3}, s_{i4}](s)$ – це кубічна базисна функція В-сплайна, яка зв'язана із вузловим вектором

$$s_i = [s_{i0}, s_{i1}, s_{i2}, s_{i3}, s_{i4}]. \quad (4)$$

Аналогічно $N[t_{i0}, t_{i1}, t_{i2}, t_{i3}, t_{i4}](t)$ зв'язана з вузловим вектором

$$t_i = [t_{i0}, t_{i1}, t_{i2}, t_{i3}, t_{i4}]. \quad (5)$$

Рівняння Т-сплайн-поверхні дуже схоже з рівнянням тензорного добутку для раціональної В-сплайн-поверхні. Різниця між Т-сплайн-рівнянням і В-сплайн-рівнянням у тому, як вузлові вектори s_i, t_i визначаються для кожної функції $B_i(s, t)$.

NURBS-поверхні визначаються за допомогою набору контрольних точок, які лежать топологічно у прямокутній сітці, як показано на рис. 2,а. Це означає, що велика частина контрольних точок у NURBS є зайвими, тому що вони не містять значної кількості геометричної інформації, а необхідні лише для задоволення топологічних обмежень. З іншого боку, Т-сплайн може мати лише частину контрольних точок у кожному рядку мережі, як показано на рис. 2,б. Частковий ряд контрольних точок закінчується Т-подібним перехрестям – звідси й назва Т-сплайнів. Для проектування чим менша кількість контрольних точок, тим швидше час моделювання. Більш того, для того щоб зробити уточнення для Т-сплайнів, достатньо лише зробити вставку однієї контрольної точки. У випадку з NURBS знадобиться вставка точок уздовж усього рядка. Наступним обмеженням NURBS є те, що в більшості випадків об'єкт буде складатися з декількох NURBS-поверхонь і після «зклеювання» частин, можливо, будуть присутні розриви. Т-сплайни дозволяють конструювати гладкі без розривів (або їх ще називають «водонепронекні») поверхні.

Локальне уточнення кривих і поверхонь NURBS здійснюється через вузол вставки, у процесі чого не змінюється форма кривої або поверхні. Існує декілька алгоритмів для вставки вузлів. Алгоритм вставки вузла є фундаментальним, його можна використати і як математичний інструмент для розуміння та аналізу В-сплайнів, і як практичний інструмент для управління та побудови В-сплайн-кривих і поверхонь. Вставка вузла добре відпрацьовує для локального уточнення В-сплайн-кривої: вставка одного вузла породжує оновлення лише кількох контрольних точок у локальному регіоні.

Тим не менш для тензорного добутку В-сплайн-поверхні локальне уточнення неможливе, оскільки вставка одного вузла в один із

вузлових векторів поверхні спричиняє вставку нових контрольних точок у цілий рядок або стовпець.

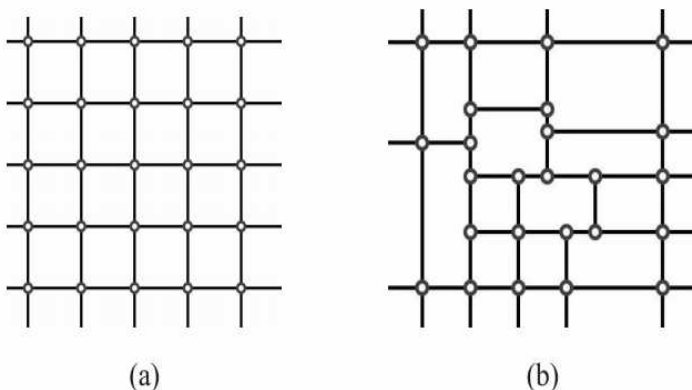


Рис. 2. Полігональна мережа для: а – NURBS; б – Т-сплайнів

Алгоритм локального уточнення Т-сплайна. Локальне уточнення означає, що треба зробити вставку однієї або декількох контрольних точок у Т-сітку без зміни форми Т-сплайн-поверхні. Ця процедура також може називатися локальною вставкою вузла, тому що додаткові контрольні точки для Т-сітки повинні впливати на вставку додаткових вузлів при побудові базисних функцій.

Локальний алгоритм буде складатися із двох етапів: топологічний та геометричний. На першому етапі визначається, які контрольні точки треба буде вставити додатково. Після того як усі необхідні нові контрольні точки визначені, обчислюються декартові координати і ваги для шуканої Т-сітки.

З'ясуємо тепер алгоритм топологічного етапу. Важливим ключем до розуміння рішення проблеми є розуміння, як у Т-сплайні щільно пов'язані між собою базисні функції та Т-сітка: кожній контрольній точці відповідає базисна функція, і для кожної базисної функції за правилом 1 визначається вузловий вектор.

Правило 1: визначимо вузлові вектори $\mathbf{s}_i$ (4) та $\mathbf{t}_i$ (5) для точки P_i . Тоді вузловий вектор $\mathbf{s}_i$ складатиметься з перших двох вузлів, які будуть спільними для променя, що виходить з вузла $(\mathbf{s}_i, \mathbf{t}_i)$ уздовж осі

s , та першими двома ребрами, з якими промінь перетинається. Аналогічно визначаються координати вузлового вектора t_1 .

Наприклад, на рис.1 для точки P_1 вузлові вектори мають координати $s_1 = (s_0, s_1, s_2, s_3, s_4)$, $t_1 = (t_1, t_2, t_2 + a_6, +t_4)$.

Припустимо, що на деякий час можуть існувати базисні функції, які будуть порушувати правило 1. Розглянемо три можливих порушення, які можуть виникнути в алгоритмі.

- *Порушення 1:* базисна функція припускає наявність вузла, який задовольняє правилу 1 для поточної Т-сітки.

- *Порушення 2:* базисна функція має вузол, який не задовольняє правилу 1 для поточної Т-сітки.

- *Порушення 3:* контрольна точка не має базисної функції, яка з нею пов'язана.

Якщо порушень не існує, Т-сплайн побудований без порушень. Якщо порушення є, алгоритм буде вирішувати їх по одному, поки ніяких подальших порушень не буде існувати. Топологічний етап нашого алгоритму локального уточнення складається з наступних дій:

1. Вставити всі необхідні контрольні точки у Т-сітку.
2. Якщо базисні функції винні в порушенні 1, виконати необхідні вставки вузла для таких базисних функцій.
3. Якщо змішувані функції винні в порушенні 2, додати відповідні контрольні точки у Т-сітку.
4. Повторити кроки 2 і 3, поки не буде виявлено більше порушень.

Вирішення всіх випадків порушення 1 і 2 автоматично вирішить усі випадки порушення 3.

Проілюструємо алгоритм на прикладі. На рис 3,а зображено початкову Т-сітку, в яку ми хочемо вставити одну контрольну точку P_2 . Оскільки Т-сітка на рис. 3,а правильна, то жодних порушень не існує. Але якщо ми просто вставимо P_2 у Т-сітку (рис. 3,б) без зміни базових функцій, то одразу отримаємо декілька порушень. Оскільки P_2 має координати вузлів (s_3, t_2) , то чотири базисні функції стають винними в порушенні 1: із центром у точках (s_1, t_2) , (s_2, t_2) , (s_4, t_2) та (s_5, t_2) . Щоб усунути ці порушення, ми повинні вставити вузол s_3 у кожній із цих функцій. Базисні функції із центром у точці (s_2, t_2) :

$$N[s_0, s_1, s_2, s_3, s_4](s)N[t_0, t_1, t_2, t_3, t_4](t).$$

Вставка вузла s_3 у вектор s розбиває базисну функцію на дві функції:

$$c_1 N[s_0, s_1, s_2, s_3, s_4](s)N[t_0, t_1, t_2, t_3, t_4](t),$$

$$d_2 N[s_0, s_1, s_2, s_3, s_4](s)N[t_0, t_1, t_2, t_3, t_4](t).$$

Перша із функцій задовольняє правилу 1. Крім того, функції з центром у точках (s_1, t_2) , (s_4, t_2) та (s_3, t_2) задовольняють правилу 1.

Тим не менш вузловий вектор t функцій $d_2 N[s_1, s_2, s_3, s_4, s_5](s) N[t_1, t_2, t_3, t_4, t_5](t)$ має порушення 2, оскільки для вектора $t = [t_0, t_1, t_2, t_3, t_4]$ правило 1 не виконується для вузла t_3 . Ця проблема не може бути усунута шляхом переробки функцій. Слід додати додаткову точку до контрольної Т-сітки. Необхідна контрольна точка P_3 на рис. 3,е. Вставка контрольної точки усуває порушення 2, але виникає новий випадок порушення 1. Як показано на рис. 3,ф, функція із центром у (s_2, t_3) має вузловий вектор s , який не включає в себе вузол s_3 , що необхідно для правила 1. Вставка s_3 у вузловий вектор вирішує проблему, і немає подальших порушень правила 1.

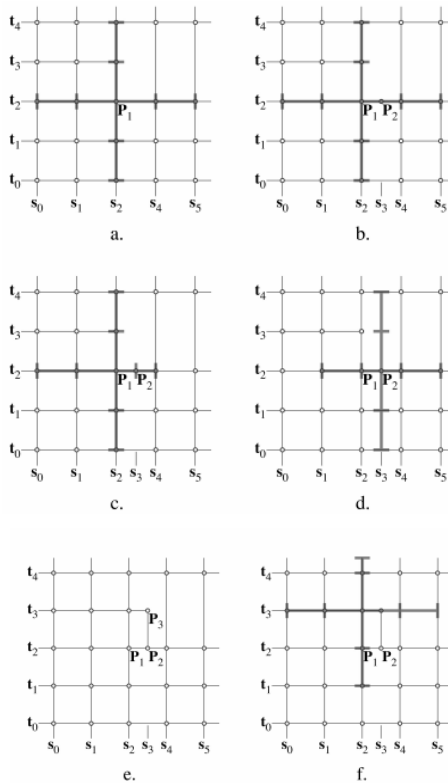


Рис. 3. Алгоритм

Цей алгоритм завжди гарантовано зупиниться. У гіршому випадку алгоритм буде розповсюджувати скрізь ряди контрольних точок, щоб перетнути всю поверхню. У практиці, як правило, алгоритм вимагає кількох додаткових нових контрольних точок, крім тих, що користувач хоче вставити. Таким чином, це уточнення алгоритму має два істотних плюси: алгоритм завжди відпрацьовує і мінімізує кількість додаткових вузлів для вставки.

Висновки. У статті було розглянуто основні положення побудови поверхні за допомогою Т-сплайнів. Розглянуто алгоритм локального уточнення. У наступному дослідженні вивчатиметься якість побудованих Т-сплайн-поверхонь для даних екологічного моніторингу.

Бібліографічні посилання

1. Isogeometric Analysis using T-splines / Y. Bazilevs, V.M. Calo, J.A. Cottrell [and others], 2008
2. T-spline Simplification and Local Refinement / W. S. Thomas, L. C. David, G. T. Finnigan [and others] // Brigham Young University, 2008.
3. **Amenta N.** A simple algorithm for homeomorphic surface reconstruction / S. Choi, T. K. Dey, N. Leekha. // Proceedings of the sixteenth annual symposium on Computational geometry, 2000. – P. 213–222.
4. **Bernardini F.** The ball-pivoting algorithm for surface re-construction / J. Mittleman, H. Rushmeier, C. Silva, and G. Taubin. // IEEE Transactions on Visualization and Computer Graphics, 2005, 11(2). – P. 181–192.
5. **Dey T.K.** Tight cocone: a water-tight surface reconstructor / S. Goswami // In Proc. SMI'03, 2003. – P. 127–134.
6. **Eck M.** Automatic reconstruction of b-spline surfaces of arbitrary topological type / H. Hoppe // SIGGRAPH'96, 1996. – P. 325–334.
7. **Gu X.** Manifold splines / Y. He, and H. Qin // Graphical Models, 2006, 68(3). – P. 23–254.
8. **Ettler B.J.** Least squares fitting of algebraic spline surfaces / A. Felis // Adv. Comput. Math., 17, 2002. – P. 135–152.
9. **Carr J.C.** Reconstruction and representation of 3D objects with radial basis functions. / R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, T. R. Evans // Proc. SIGGRAPH'01, 2001. – P. 67–76.

Надійшла до редколегії 15.10.2013 р.

УДК 519.816:519.237.8

О.Г. Байбуз*, В.Д. Рублевський**, Н.І. Рублевська***

**Дніпропетровський національний університет імені Олеся Гончара*

***ДЗ “Дніпропетровська медична академія МОЗ України”*

****ОКЗ “Дніпропетровське медичне училище”*

ДЕЯКІ АСПЕКТИ ВИКОРИСТАННЯ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ПРИ ОЦІНЦІ ВПЛИВУ ЗАБРУДНЕННЯ ДОВКІЛЛЯ ВАЖКИМИ МЕТАЛАМИ НА АДАПТАЦІЙНІ МОЖЛИВОСТІ ДІТЕЙ

Застосування сучасних інформаційних технологій дозволило обґрунтувати величини аерогенного надходження металів до організму (мг/кг×добу) і рівні вмісту металів у біосередовищі дітей – мешканців промислового регіону, які вірогідно ($p < 0,05-0,001$) призводять до зниження адаптаційних можливостей дитячого організму.

Ключові слова: *інформаційні технології, дитяче населення, важкі метали, адаптація, ризик.*

Использование современных информационных технологий позволило обосновать величины аерогенного поступления металлов в организм (мг/кг×сутки) и уровни содержания металлов в биосреде детей – жителей промышленного региона, которые достоверно ($p < 0,05-0,001$) приводят к снижению адаптационных возможностей детского организма.

Ключевые слова: *информационные технологии, детское население, тяжёлые металлы, адаптация, риск.*

The use of modern information technologies has to prove the value of aerogenic flow of metal into the body (mg / kg × day), and levels of metals in biological media of child residents of the industrial region, which was significantly ($p < 0.05-0.001$) lead to lower adaptive capacity of the child's body.

Keywords: *information technology, child population, heavy metals, adaptation, risk.*

Вступ. Важливе значення для досягнення кінцевої мети досліджень має правильний вибір способу обробки інформації. Вибір адекватних способів обробки інформації у сукупності з необхідними математичними методами дозволяє отримати важливі наукові результати на порівняно невеликій кількості спостережень [1].

Спостереження щодо вмісту ВМ в об'єктах навколишнього середовища свідчать про постійну присутність свинцю, кадмію, хрому, нікелю, марганцю, міді, цинку у повітрі, воді, продуктах харчування в умовах промислового регіону [2], що визначає актуальність досліджень, їх впливу на формування здоров'я населення.

Мета роботи: за допомогою використання інформаційних технологій обґрунтувати чинники ризику зниження адаптаційних можливостей дітей – мешканців промислових територій у зв'язку із впливом важких металів.

Матеріали і методи. В атмосферному повітрі проаналізовано вміст заліза, марганцю, цинку, міді за шестирічний період спостереження за результатами досліджень Держкомгідромету. Гігієнічна оцінка стану атмосферного повітря проведена згідно із [3]. Розрахунок середньодобових доз аерогенного надходження металів до організму проведено відповідно до [4].

Для вивчення ранніх змін у стані здоров'я населення, було обстежено 284 здорові дитини – мешканці міст Дніпропетровська та Дніпродзержинська на вміст заліза, марганцю, цинку, міді у біосубстраті (волосі) методом атомно-абсорбційної спектрофотометрії на спектрофотометрі ААС-1N згідно із [5].

Адаптаційні можливості дітей визначали за адаптаційним потенціалом (АП), який оцінювали згідно з [6].

Для визначення ризику зниження адаптаційних можливостей дитини використовували показник «відношення шансів» (ВШ). Розрахунок ВШ та 95% вірогідних інтервалів (95% ВІ) для показника ВШ проводили відповідно до [7]. Оцінювали ВШ таким чином: якщо значення ВШ < 1 – ризик зниження адаптаційних можливостей знижується, при показниках ВШ, рівних 1, ефект відсутній, при ВШ, вищих за 1, ризик зниження адаптаційних можливостей підвищується.

Для кількісної оцінки ймовірності зниження рівня адаптації дітей залежно від факторів ризику використовували логістичну регресію [8]. У загальному вигляді рівняння логістичної регресії має вигляд

$$P = 1 + e^{\frac{z}{2}}$$
$$z = b_0 + b_1 * x_1 + \dots + b_i * x_i$$

де b_0 – константа, x_i – значення факторів, b_i – коефіцієнти регресії, p – ймовірність зниження рівня адаптації.

Рівень значущості коефіцієнтів регресії оцінювали за критерієм Вальда. Адекватність моделі оцінювали за критерієм χ^2 Пірсона.

Статистичну обробку матеріалів досліджень проводили з використанням методів біостатистики, реалізованих у пакеті програм статаналізу STATISTICA v. 6.1 (ліц. № AJAR909E415822FA).

Результати та їх обговорення. Аналіз результатів досліджень свідчить, що в атмосферному повітрі міст Дніпропетровська та Дніпродзержинська постійно реєструються марганець, залізо, мідь, цинк. Їх концентрації (у середньому за період спостереження) не перевищують ГДК с.д. (табл. 1).

Таблиця 1

Вміст металів в атмосфері міст Дніпропетровськ та Дніпродзержинськ

Метали	Концентрації металів (у середньому за період спостереження), мкг/м ³ (M±m), min/max (с.р.)		ГДК с.д., мкг/м ³
	м. Дніпропетровськ n=7280	м. Дніпродзержинськ n=3200	
Марганець	0,122±0,02 0,030/0,320	0,092±0,015 0,024/0,160	1,0
Залізо	1,899±0,412 0,306/3,850	2,333±0,432 0,890/4,340	4,0
Мідь	1,287±0,346 0,050/4,070	1,503±0,360 0,330/3,700	2,0
Цинк	0,314±0,071 0,078/0,970	0,205±0,044 0,032/0,420	50,0

Примітка: n – кількість спостережень.

Однак слід відзначити більш значний вміст у повітрі Дніпропетровська порівняно з Дніпродзержинськом марганцю (в 1,32 раза) та цинку (в 1,53 раза). Вміст заліза (в 1,23 раза), міді (в 1,17 раза) вище у Дніпродзержинську, що, деякою мірою, пов'язано із пріоритетністю тих чи інших видів промисловості у різних регіонах, а також умовами розсіювання викидів від джерел забруднення повітря металами.

Аналіз середньодобових доз надходження окремих металів з атмосферним повітрям (табл. 2) показав, що дитяче населення міст спостереження зазнає впливу більш значного порівняно з референтними величинами [9] аерогенного навантаження марганцем та міддю. Встановлена вірогідна різниця ($p<0,05$) АН доз. між містами: в умовах Дніпродзержинська до організму дітей надходження заліза вище, ніж у Дніпропетровську.

Таблиця 2

Аерогенне надходження металів до організму дітей

Метали	Місто, мг/кг*добу, $M \pm m$		Референтні величини [9]
	Дніпропетровськ	Дніпродзержинськ	
Заліза оксид	0,00068±0,00015	0,00083±0,00015*	0,010000
Марганець	0,00004±0,00001	0,000033±0,000005	0,000014
Міді оксид	0,00046±0,00001	0,00053±0,00001	0,000006
Цинку оксид	0,00001±0,000005	0,000007±0,000017	-

Примітка: * $p < 0,05$.

Таблиця 3

Вміст металів у волоссі дітей із районів спостереження

Місто, район	Метали, мкг/г, $M \pm m$			
	Zn	Cu	Mn	Fe
м. Дніпропетровськ Промисловий район №1, n=30	103,14±15,99	3,45±1,00	9,83±4,46 *	274,51± ±49,10
Район з інтенсивним транспортним рухом, n=29	129,56±24,69	15,42±3,36*	1,25±0,68	235,40± ±45,32
Промисловий район №2, n=34	123,57±11,75	9,52±1,85	2,09±1,01	110,51± ±19,05
Район порівняння, n=28	125,28±8,78	6,52±1,22	0,34±0,12	117,58± ±14,02
м. Дніпродзержинськ Промисловий район №1, n=29	106,21±10,24	7,05 ± 0,84	0,22±0,08	72,51± ±17,25*
Промисловий район №2, n=29	102,33±7,68	6,13 ±0,99*	0,54±0,17	47,23± ±18,74*
Район порівняння, n=29	96,07±13,10	8,30±1,54	0,15±0,05	60,35± ±10,85
Дані літератури [5; 10]	130,38±9,5	6,96±0,66	0,04-0,27	13-27

Примітки: n – кількість спостережень; * - достовірна відмінність від району порівняння ($p < 0,05$).

Вміст цинку у дітей із більшості районів спостереження достовірно ($p < 0,05$) є нижчими від фізіологічних величин [5; 10]. Зазначимо, що на тлі недостатнього вмісту цинку має місце високий (більше за фізіологічний у 4,1–10,2 раза) вміст заліза в організмі, якому притаманні

властивості імундепресанту, що може зумовлювати зниження імунної резистентності та підвищення захворюваності [10].

Привертає увагу в 1,3–2 рази нижчий ($p<0,05$) (порівняно з фізіологічними величинами [5; 10]) вміст міді у волоссі дітей із промислових районів №1 та №2 і в 2 рази вищий ($p<0,05$) у волоссі дітей із району з інтенсивним транспортним рухом. При порівнянні отриманих нами результатів з літературними даними [11] можна зазначити, що у волоссі дітей із промислових районів Дніпропетровська марганцю міститься більше в 1,7–7,9 раза, ніж у дітей з інших промислових регіонів.

Отримані результати з оцінки адаптаційних можливостей показали, що у промислових районах більша ($p<0,05$) питома вага дітей із напруженням, незадовільним рівнем адаптаційних можливостей та зривом адаптації (табл. 4).

Таблиця 4

Питома вага дітей з різним рівнем адаптації у районах спостереження

Місто, район	Рівень адаптації			
	задовільний	напруження	незадовільний	зрив
м. Дніпропетровськ Промисловий район №1, n=30	75,7±4,6*	18,9±1,2*	2,7±0,7*	2,7±0,7*
Район з інтенсивним транспортним рухом, n=29	81,5±4,9	14,8±1,5*	3,7±0,7*	0,0±0,0
Промисловий район №2, n=34	77,5±3,6*	17,5±1,8*	2,5±0,6*	2,5±0,6*
Район порівняння, n=28	88,6±3,2	11,4±0,7	0,0±0,0	0,0±0,0
м. Дніпродзержинськ Промисловий район №1, n=29	73,4±3,8*	20,0±1,4*	3,3±0,8*	3,3±0,8*
Промисловий район №2, n=29	80,0±4,2	13,3±1,5	6,7±0,9*	0,0±0,0
Район порівняння, n=29	90,0±5,1	10,0±1,1	0,0±0,0	0,0±0,0

Примітки: n – кількість спостережень; * - достовірна відмінність із районом порівняння ($p<0,05$).

Аналіз ВІШ зниження рівня адаптації залежно від величини аерогенного навантаження окремими ксенобіотиками дозволив визначити метали та такі їх величини аерогенного дозового

надходження (мг/кг×добу) до організму дітей, із якими пов'язаний ризик зниження рівня адаптації (табл. 5).

Таблиця 5

Ризик зниження рівня адаптації дітей залежно від величини аерогенного навантаження металами, $p < 0,05$

Метали, АН доз., мг/кг×добу	ВШ	95% ВІ	χ^2	p
Марганець, $< 0,000044$	0,36	0,14 - 0,92	4,760	0,029
Марганець, $> 0,000044$	2,77	1,09 - 7,03	4,760	0,029
Мідь, $< 0,000164$	0,24	0,1 - 0,57	10,696	0,001
Мідь, $> 0,000164$	4,24	1,74 - 10,34	10,696	0,001
Цинк, $< 0,0000077$	3,99	1,45 - 10,92	7,790	0,005
Цинк, $> 0,0000077$	0,25	0,09 - 0,69	7,790	0,005

Для прогнозування індивідуального ризику зниження адаптаційних можливостей побудовані рівняння логістичної регресії за граничними рівнями показників аерогенного дозового навантаження, а також за фактичними значеннями вмісту ксенобіотиків у біосередовищі дітей. Аналіз логістичних регресій показав, що ризик зниження рівня адаптації дітей достовірно ($p < 0,05-0,001$) асоційований із такими факторами: величиною аерогенного надходження міді (ВШ 4,24; 95% ВІ 1,74-10,34; $p < 0,001$), цинку (ВШ 3,99; 95% ВІ 1,45-10,92; $p < 0,001$), марганцю (ВШ 2,77; 95% ВІ 1,09-7,03; $p < 0,05$), підвищеним вмістом у волоссі Mn (ВШ=2,07), Fe (ВШ=2,01), а також зі зниженим вмістом цинку у біосередовищі дітей (ВШ=2,01). Побудовані регресійні рівняння дозволяють розрахувати ймовірність зниження адаптаційних можливостей дітей залежно від величини аерогенного дозового навантаження металами (табл. 6), а також при різних рівнях вмісту металів у біосередовищі дітей – мешканців промислових територій (табл. 7). Отримані моделі рекомендовані для впровадження в діяльність закладів санітарно-епідеміологічної служби з метою підвищення ефективності первинної профілактики несприятливих змін у стані здоров'я дитячого населення [12].

Таблиця 6

**Показники логістичних регресій для оцінки ймовірності зниження
адаптаційних можливостей дітей залежно від вмісту металів
у біосередовищі**

Метали	Коефіцієнти		ВШ (95% ВІ)	Адекватність моделі за χ^2
	b0	b1		
Марганець	-1,46	0,0667	2,069 (2,006 - 2,136)	$\chi^2=7,2829$; $p<0,001$
Залізо	-2,37	0,0071	2,007 (2,004 - 2,010)	$\chi^2=27,474$; $p<0,001$

Таблиця 7

**Показники мультиваріантних логістичних регресій для оцінки
ймовірності зниження адаптаційних можливостей дітей залежно від
вмісту металів у біосередовищі**

Параметри показника			Адекватність моделі за χ^2
Найменування	Коефіцієнт регресії (b)	ВШ (95% ВІ)	
Константа	-1,761	-	$\chi^2=29,777$; $p<0,001$
Zn	-0,006	0,994 (0,985 - 1,002)	
Fe	0,008	2,008 (2,004 - 2,011)	
Константа	-2,636	-	$\chi^2=29,978$; $p<0,001$
Mn	0,011	2,011 (2,004 - 2,082)	
Fe	0,007	2,007 (2,004 - 2,011)	

Висновки. Використання інформаційних технологій дозволило обґрунтувати чинники ризику зниження адаптаційних можливостей дітей – мешканців промислових територій: величини аерогенного надходження міді (ВШ 4,24; 95% ВІ 1,74-10,34; $p<0,001$), цинку (ВШ 3,99; 95% ВІ 1,45-10,92; $p<0,001$), марганцю (ВШ 2,77; 95% ВІ 1,09-7,03; $p<0,05$), а також рівень вмісту у біосередовищі Mn (ВШ=2,07), Fe (ВШ=2,01), Zn (ВШ=2,01).

Бібліографічні посилання.

1. **Антомонов Ю.М.** Математическая обработка и анализ медико-биологических данных / М.Ю. Антомонов. – К., 2006. – 558 с.
2. **Сердюк А.М.** Тяжелые металлы внешней среды и их влияние на репродуктивную функцию женщин: монография / [А.М. Сердюк, Э.Н.

Белицкая, Н.М. Паранько и др.]. – Днепропетровск: АРТ – ПРЕСС, 2004. – 148 с.

3. ДСП–201-97. Державні санітарні правила охорони атмосферного повітря населених місць (від забруднення хімічними і біологічними речовинами). – К., 1997. – 57 с.
 4. **Рахманин Ю.А.** Руководство по оценке риска для здоровья населения при воздействии химических веществ, загрязняющих окружающую среду / [Ю.А. Рахманин, С.М. Новиков, Т.А. Шапина и др.]. – М.: Федеральный центр Госсанэпиднадзора Минздрава России, 2004. – 143 с.
 5. **Деркачов Е.А.** Донозологічна оцінка стану здоров'я населення у зв'язку з впливом факторів навколишнього середовища / [Е.А. Деркачов, Л.Б. Огір, Т.Є. Дрозд та ін.]: метод. рекомендації МР 2.2.12.068–2000. – К., 2000. – 42 с.
 6. **Рублевська Н.І.** Нормативні значення адаптаційного потенціалу для оцінки рівня адаптації дітей допільного віку до екологічно несприятливих умов: інформ. лист / Н.І. Рублевська. – К.: Укрмедпатентінформ, 2010. – №19. – 3 с.
 7. **Бабищ П.Н.** Применение современных статистических методов в практике клинических исследований. Сообщение третье. Отношение шансов: понятие, вычисление и интерпретация / П.Н. Бабищ, А.В. Чубенко, С.Н. Лапач // Укр. мед. часопис. – 2005. – № 2 (46). – С. 113–119.
 8. **Юнкеров В.И.** Математико-статистическая обработка данных медицинских исследований / В.И. Юнкеров, С.Г. Григорьев. – СПб: СПб ВМедА, 2002. – 266 с.
 9. **Петросян А.А.** Аналіз дозового інгалаційного навантаження від забруднення атмосферного повітря хімічними речовинами / А.А. Петросян, О.І. Турос, О.М. Картавцев // Довкілля та здоров'я. – 2009. – № 2. – С. 25–28.
 10. **Корчина Т.Я.** Содержание тяжелых металлов в волосах детей севера Тюменской области / Т.Я. Корчина // Гигиена и санитария. – 2007. – №4. – С. 27–29.
 11. **Жданов В.В.** Оценка влияния выбросов теплоэлектростанций на содержание металлов в волосах детей / В.В. Жданов // Гігієна населених місць. – 2008. – Вип. 52. – С. 46–52.
 12. Пат. 49113 Україна, МПК А 61 В 10/00 G 01 N 33/50. Спосіб прогнозування адаптаційних можливостей дітей / Н.І. Рублевська; заявник і власник Рублевська Н.І. – u 201002335; заявл. 02.03.2010; опубл. 12.04.2010, Бюл. № 7.
- Надійшла до редколегії 18.10.2013 р.*

ЗМІСТ

Говоруха В.Б., Дубель О.В. Механізація і автоматизація ділових процесів	3
Карпов А.А., Карпов О.М., Журба Д.С. Алгоритми моделювання біомеханіки артикуляції губ людини на основі В-сплайн-поверхонь.....	9
Івахнік Г.В., Земляна С.В. Інформаційна технологія перевірки гіпотези про параметри нормального розподілу на основі послідовного аналізу.....	15
Сидорова М.Г. Застосування ансамблів алгоритмів для підвищення стійкості результатів кластеризації.....	22
Мацуга О.М., Дроздова І.В. Інформаційна технологія короткострокового прогнозування первинної інвалідності в Україні.....	30
Приставка П.О., Чолишкіна О.Г. Дослідження В-сплайна шостого порядку та їх лінійної комбінації.....	42
Волченко Ю.К., Ємел'яненко Т.Г. Інформаційна технологія та програмне забезпечення розпізнавання облич.....	52
Колдунова А.А., Емельяненко Т.Г. Обзор методов и программных продуктов прогнозирования землетрясений.....	59
Луценко О.П., Байбуз О.Г. Відновлення форми розподілу точок розладнань на часовому ряді валютних котирувань.....	71
Байбуз О.Г., Харченко Н.І. Системи управління дистанційним навчанням.....	81
Карпов О.М., Глушак К.М. Визначення значущих параметрів мовного сигналу за допомогою колоколоподібних функцій.....	93
Пирогова Я.М., Байбуз О.О., Мащенко Л.В. Побудова поверхонь за допомогою технології t-сплайнів.....	104
Байбуз О.Г., Рублевський В.Д., Рублевська Н.І. Деякі аспекти використання інформаційних технологій при оцінці впливу забруднення довкілля важкими металами на адаптаційні можливості дітей.....	112
Відомості про авторів	120