

Голові разової спеціалізованої вченої ради
Дніпровського національного університету
імені Олеся Гончара
доктору фізико-математичних наук, професору,
в.о. проректора ДНУ імені Олеся Гончара
з науково-педагогічної роботи
Гук Наталії Анатоліївни

РЕЦЕНЗІЯ

рецензента Когута Петра Ілліча,
доктора фізико-математичних наук, професора,
професора кафедри математичного аналізу та оптимізації
Дніпровського національного університету
імені Олеся Гончара
на дисертаційну роботу

Кошеля Євгена Васильовича

**«Моделі та методи синтезу мовлення на основі нейронних систем звичайних
диференціальних рівнянь»**

подану на здобуття ступеня доктора філософії
у галузі знань 11 Математика та статистика
зі спеціальності 113 Прикладна математика

Дисертаційна робота Кошеля Євгена Васильовича присвячена дослідженню проблеми синтезу мовлення за допомогою нового інструментарію, пов'язаного з використанням так званих систем нейронних звичайних диференціальних рівнянь.

В першому розділі дається огляд досліджень за обраною тематикою. Основна увага приділяється базовим поняттям та методам обробки аналогових та дискретних сигналів, а також їх застосуванню до моделювання голосового сигналу як динамічного процесу.

У другому розділі досліджується якісні властивості одного класу систем звичайних диференціальних рівнянь. Оскільки процес моделювання голосового сигналу людини є динамічним процесом, який протікає у реальному часі і має, як правило, хаотичну поведінку, то основна увага автором приділяється методам дослідження і прогнозування хаотичних процесів. За основу в роботі обирається метод часових рядів. Суть цього методу полягає в тому, що спочатку (в певні моменти часу) фіксуються спостережувані змінні, за допомогою яких описується

динаміка досліджуваного процесу. Поведінка в часі таких змінних утворює відповідні часові ряди.

Зазвичай дослідження поведінки часових рядів базується на методах теорії ймовірності та математичної статистики. Проте технології аналізу даних, які використовують статистичні підходи, мають ряд недоліків. Зокрема, такі методи, як правило, ґрунтуються на усереднених показниках, за якими важко з'ясувати реальну поведінку систем в окремо взяті моменти часу. Крім того, стандартні статистичні методи зазвичай нехтують спостереженнями, які виходять за рамки статистично надійних: піками, анормальними відхиленнями і т.п. Разом з тим, саме така поведінка часових рядів може бути вирішальною для прогнозування подальшої динаміки голосового сигналу. Виходячи з наведеного, автор надає перевагу залученню нових методів дослідження, які б базувалися на нелінійній динаміці, теорії катастроф, фрактальній геометрії та інформаційних технологіях інтелектуального аналізу даних (Data Mining). Власне основою таких технологій є перехід від класичних нейромережових моделей до так званих систем нейронних диференціальних рівнянь (NODE). Принципова сутність цього переходу полягає у тому, що у більшості задач моделювання нейромережа сприймається як дискретна система, в той час як NODE є системою звичайних нелінійних диференціальних рівнянь. Новизна такого підходу до процесу моделювання голосових сигналів не викликає сумнівів і полягає в тому, що голосові сигнали можна розглядати як специфічний клас розв'язків відповідних систем диференціальних рівнянь.

До основних результатів другого розділу варто віднести запропоновану автором математичну модель у формі певної системи звичайних диференціальних рівнянь, а також новий тип активаційної функції для моделювання процесу синтезу мовлення. Автором також отримано низку нових результатів стосовно дослідження якісної поведінки запропонованого класу динамічних систем.

У третьому розділі розглядається задача синтезу голосових сигналів на основі запропонованих моделей у формі систем нейронних диференціальних рівнянь.

На відміну від традиційних підходів до синтезу мови в часовій та частотній областях, в дисертації пропонується новий підхід, оснований на залученні методів нелінійного аналізу для відтворення основних інваріантів, що характеризують так звані геометричні властивості голосових сигналів. Власне самі ці інваріанти в подальшому служать основою до розв'язання проблеми синтезу мови. Ключовим результатом даного розділу є нова технологія, яку пропонується покласти в основу відтворення голосових сигналів.

До основних результатів даної дисертаційної роботи я би відніс такі:

1. Запропонована нова модель для апроксимації та прогнозування одновимірних нестационарних часових рядів за допомогою системи нейронних диференціальних рівнянь, у якій вперше була застосована узагальнена степенева активаційна функція, а також теореми про обмеженість розв'язків таких систем та області значень на основні параметри, при яких в системі виникає хаотична поведінка.
2. Проведено порівняльний аналіз нейромережових моделей, які застосовують класичні активаційні функції (сигмоїду, гіперболічний тангенс та ReLU) та степеневу активаційну функцію. Показано, що моделювання із степеневою активаційною функцією має більшу степінь схожості з реальним процесом в порівнянні з моделями, які побудовані за допомогою класичних активаційних функцій.
3. Досліджено запропоновану систему нейронних диференціальних рівнянь і показано її суттєву відмінність від класичних моделей. Саме ця обставина дозволила виявити існування досі невідомих хаотичних атракторів, які не могли бути згенерованими за класичними моделями.
4. У третьому розділі автор запропонував процедуру сегментації голосового сигналу (якісні характеристики таких сигналів зазнають постійних змін під час мовлення), сутність цієї процедури полягає в тому, щоб в межах одного сегменту певні характеристики такого сигналу (інваріанти) мають залишатися незмінними. Це дозволило суттєво знизити обсяги пам'яті та часові затрати при комп'ютерному моделюванні синтезу таких сигналів.

Вважаю, що всі отримані в роботі результати є новими та оригінальними. Вони можуть знайти своє застосування не лише в проблемі синтезу мови, але і в багатьох інших задачах (наприклад, при моделюванні зміни погоди чи процесів еволюції біологічних популяцій), де аналіз та дослідження часових рядів є першочерговим завданням. Головні результати дисертаційної роботи були опубліковані у фаховому математичному журналі «Journal of Optimization, Differential Equations and Their Applications» (квартиль Q3). Результати роботи неодноразово доповідались на семінарах і конференціях. Дисертація оформлена у відповідності до вимог, які висуваються до робіт на здобуття ступеня доктора філософії. Порушень академічної доброчесності у дисертації та наукових працях Є.В. Кошеля я не виявив.

Разом з тим за матеріалами дисертації є певні зауваження та побажання.

1. При розробці нейромережових моделей вагому роль відіграє швидкість налаштування (або тренування) таких моделей. Проте в дисертаційній роботі автор ніде не наводить жодних даних щодо

швидкості налаштування моделей з різними активаційними функціями.

2. Як відомо, при побудові будь-якої моделі, визначальним фактором є її перевірка на адекватність та точність відтворення бажаного сигналу. Для цього існує багато напрацьованих метрик таких як нарахування коефіцієнта Пірсона, метрика PSNR чи популярна нині метрика структурної подібності SSIM. У випадку моделі, яку пропонує автор даної роботи, її адекватність суттєво залежить від багатьох факторів, включаючи вибір параметрів впливу, їх налаштування, коректність розбиття звукового сигналу на так звані зони інваріантності і що є мабуть самим основним — якою має бути кількість таких інваріантів (або атракторів) для коректного відтворення обраного звукового сигналу. Проте дослідженню цієї проблеми, включаючи аналіз чутливості моделі до обраних параметрів впливу, приділено недостатньо уваги.
3. У другому розділі автор досліджує якісні властивості розв'язків одного класу диференціальних рівнянь, праві частини яких не містять лінійної складової. Разом з тим з тексту дисертації незрозуміло де в задачі синтезу мовлення автор застосовує ці результати.
4. Хотілося б побачити в роботі більш детальне обґрунтування принципу вибору визначних характеристик голосу. До таких характеристик автор відносить мел-кепстральні коефіцієнти, які отримуються зі спектра сигналу, знайденого за допомогою перетворення Фур'є. Проте добре відомо, що спектр сигналу, який утворено коефіцієнтами вейвлет розвинення, дозволяє точніше передавати локальну поведінку звукових сигналів. Вважаю, що було б доцільним в подальших дослідженнях цієї проблеми залучити апарат вейвлет-перетворень.

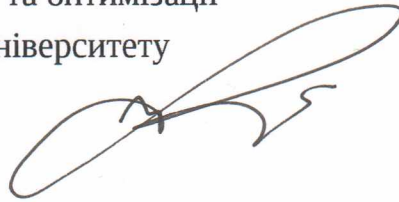
Попри наведені зауваження, я вважаю, що дана дисертація відповідає всім вимогам, які затверджені наказом Міністерства освіти і науки України від 12 січня 2017 р. № 40. Основні результати пройшли достатню апробацію, а отримані в дисертаційній роботі наукові результати є новими, та мають важливе практичне значення.

Виходячи з наведеного, вважаю, що дисертаційне дослідження Кошеля Є.В. відповідає вимогам "Порядку присудження ступеня доктора філософії та

скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії", затвердженого Постановою Кабінету Міністрів України від 12 січня 2022 р., № 44, а її автор --- Кошель Є.В. заслуговує на присудження наукового ступеня доктора філософії.

Рецензент,

доктор фізико-математичних наук, професор
кафедри математичного аналізу та оптимізації
Дніпровського національного університету
ім. О. Гончара



Петро КОГУТ

Підпис проф. Когути Петра Ілліча засвідчую:



Тетяна ХОДАНЕН