

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 Байбуза Олега Григоровича,
доктора технічних наук, професора,
завідувача кафедри математичного забезпечення ЕОМ
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 сучасних методів синтезу мовлення, в основу яких покладені нові підходи до побудови архітектури нейромережових моделей, які потім використовуються в алгоритмах синтезу. Ця архітектура створюється як права частина деякої системи нелінійних диференціальних рівнянь. Таким чином, до дослідження властивостей нейромережі залучається потужний апарат сучасної теорії диференціальних рівнянь.

В першому розділі наведено огляд базових понять та методів обробки аналогових та дискретних сигналів у контексті роботи зі звуком, розглянуто голосовий сигнал як об'єкт дослідження, викладено загальні теоретичні відомості про його аналіз та синтез, інформацію про його структуру та головні характеристики, розглянуто та обґрунтовано моделювання голосового сигналу як динамічного процесу.

У другому розділі досліджується нова модель керування системою звичайних диференціальних рівнянь у латентному просторі.

Синтезом мови на сьогоднішній день називається технологія, яка здатна перетворювати текстову інформацію на звичайну мову. З розвитком "розумних машин" ця технологія стає все більш актуальною, і з кожним днем потребує все більшої досконалості. Власне, на сьогоднішній день розроблено цілу низку методів синтезу мови. І моя задача полягає в тому, щоб показати чим відрізняються методи синтезу, про які йдеться в дисертації, від аналогічних методів розроблених раніше.

Насамперед слід сказати, що у цій роботі процес моделювання голосового сигналу людини досліджується як динамічний процес у реальному часі. Модель, яка описує цей процес, представлена 10-вимірною системою нелінійних диференціальних рівнянь (2.28). Її відмінність від класичних систем полягає в наявності лінійної частини, яка дозволяє моделювати періодичні або квазіперіодичні процеси. Крім того, відомо, що успіх (або невдача) моделювання істотно залежить від типу активаційної функції, що застосовується. У цій роботі автор вперше застосував узагальнену степеневу функцію для моделювання процесу синтезу мови. При цьому йому довелося серйозно обґрунтувати корисність застосування таких активаційних функцій порівняно з класичними, такими як гіперболічний тангенс, сигмоїда та ReLU.

У третьому розділі розглядається задача моделювання голосового сигналу, як динамічного процесу. Розроблено програмний продукт, який реалізує методи аналізу мовлення та виділення з нього визначних характеристик, а також проведено порівняння використання нової реалізації цих алгоритмів з іншими. Нова реалізація показує кращі результати в обраних здобувачем метриках. На основі розроблених методів будується практична реалізація запропонованої у другому розділі моделі для синтезу мовлення. Наведено сигнатури функцій, що реалізують модель, та пояснено метод використання моделі для синтезу потоку звукового сигналу в реальному часі.

З появою останніми роками нелінійного аналізу динамічних систем стало зрозуміло, що традиційні підходи до синтезу мови в часовій та частотній областях можуть бути далекі від оптимальних. У цій дисертації представлений ряд інструментів нелінійного аналізу, які потім застосовуються до бази даних голосних, щоб дістати звідти основний інваріант, що описує геометричні властивості голосу, що моделюється. Потім результати цього аналізу застосовуються (з використанням прийнятих ідей нелінійної динаміки) до проблеми синтезу мови.

Інструменти, що застосовуються до аналізу мовного сигналу досить широкі: вбудовування тимчасової затримки, розкладання за сингулярними значеннями, кореляційна розмірність, локальний сингулярний аналіз, спектри Ляпунова, біфуркаційні діаграми. Хоча ці інструменти та запропоновані алгоритми в даний час не є загальноприйнятими, проте вони є досить ефективними, особливо для розрахунку спектрів Ляпунова у присутності шуму та моделювання при обмеженнях на довжину даних. Нові надійні методи застосовуються до великої бази даних голосних звуків показують, що перевірені голосні демонструють докази нелінійної, низькорозмірної, не хаотичної поведінки. Ці методи використовуються в останньому розділі дисертації, в якому і розглядається нова техніка синтезу. Ця техніка показує, що вона здатна відтворювати високоякісну природну мову.

Головними досягненнями роботи я вважаю такі результати

1. У другому розділі була вперше 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. При цьому були доведені теореми про обмеженість розв'язків таких систем та

проведений біфуркаційний аналіз, який дозволив встановити границі параметрів при яких в системі виникає хаотична поведінка.

2. У третьому розділі розроблено новий програмний продукт та нову модель для синтезу мовлення, описано метод її навчання та продемонстровано здатність моделі синтезувати мовлення з визначних характеристик.

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 по дисертації є деякі зауваження.

1. Як відомо, при побудові будь-якої моделі, адекватність цієї моделі і процесу суттєво залежить від розмірності простору вкладення. Але в роботі я не помітив, що таке дослідження розмірності де-небудь проводилось. (Єдине, про що згадує автор, так це те, що ця розмірність задовольняє умовам теореми Такенса. Але ця теорема дає значення нижньої границі розмірності і нічого не говорить про реальну розмірність простору вкладення для даного процесу.)
2. При вирішенні проблеми синтезу мови автор ніде не розрізняє синтез жіночого або чоловічого голосу. Можливо з точки зору ІТ технологій таке розрізнення не має значення. Але в такому разі автор повинен був сказати про це конкретно.

Незважаючи на зроблені зауваження, я вважаю, що дисертація має високий науковий рівень і заслуговує на високу оцінку, а її автор, Кошель Є.В., заслуговує на присудження йому ступеня доктора філософії.

Рецензент,
доктор технічних наук, професор,
завідувач кафедри інженерії програмного
забезпечення та інформаційних технологій
Дніпровського національного університету
імені Олеся Гончара

Олег БАЙБУЗ

О'бз

Підпис проф. Байбуза Олега Григоровича засвідчую:



Тетяна ХОДАНЕН