

Министерство образования Республики Беларусь  
Учреждение образования  
«Белорусский государственный университет  
информатики и радиоэлектроники»

УДК 004.934.2

Талецкий  
Антон Игоревич

**Программная реализация классификатора тональности речевых сигналов  
на основе нейросетевых алгоритмов**

**АВТОРЕФЕРАТ**

диссертации на соискание степени  
магистра информатики и вычислительной техники

по специальности 1-40 81 04 – Обработка больших объемов информации

Научный руководитель  
Калугина М.А.  
к.физ-мат.н., доцент

Минск 2020

Работа выполнена на кафедре информатики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Научный руководитель: **КАЛУГИНА Марина Алексеевна**,  
кандидат физико-математических наук, доцент  
кафедры информатики учреждения образования  
«Белорусский государственный университет  
информатики и радиоэлектроники»

Рецензент: **ШЕШОЛКО Владимир Константинович**,  
кандидат физико-математических наук, доцент  
кафедры управления информационными  
ресурсами учреждения образования «Академия  
управления при Президенте РБ»

Защита диссертации состоится «23» июня 2020 г. года в 10<sup>00</sup> часов на заседании Государственной экзаменационной комиссии по защите магистерских диссертаций в учреждении образования «Белорусский государственный университет информатики и радиоэлектроники» по адресу: 220013, Минск, ул. Гикало, 9, копр. 4, ауд. 112, тел. 293-85-91, e-mail: [inform@bsuir.by](mailto:inform@bsuir.by)

С диссертацией можно ознакомиться в библиотеке учреждения образования «Белорусский государственный университет информатики и радиоэлектроники».

## ВВЕДЕНИЕ

Искусственный интеллект – одно из самых бурно развивающихся направлений развития в современной науке и информационных технологиях. Они образуются на пересечении нейронаук, математики и информатики.

Круг задач, которые можно решить с помощью ИИ, постоянно растет. Применение для ИИ находят везде – от прогнозирования рака в медицине до управления автомобилем в автомобилестроении.

В основе скачка развития ИИ в последних десятилетиях лежат три основные причины:

- наличие большого количества данных для анализа и их доступность;
- развитие математического аппарата для моделирования ИИ;
- рост вычислительных мощностей ЭВМ.

Одна из наиболее популярных областей развития ИИ – анализ тональности некоторого канала передачи информации. Например, sentiment-анализ занимается автоматизированным выявлением эмоциональной окраски текстов. Аналогичный анализ можно осуществлять с человеческой мимикой или речевыми сигналами. Основной идеей при проектировании таких алгоритмов является поиск некоего подмножества признаков, извлекаемых из данного канала информации, которые позволят модели ИИ с наибольшей точностью определить его тональность.

Темой данной диссертации является улучшение реализации автоматизированного анализа тональности речевых сигналов человека. В рамках нее были улучшены показатели качества работы системы в сравнении с результатами достигнутыми в рамках дипломной работе по аналогичной теме. Как и планировалось, были применены модели искусственных нейронных сетей.

Отдельно стоит отметить, что в рамках данной работы не планируется осуществлять анализ смыслового содержания аудиофайлов, так как это, вообще говоря, сводится к задаче анализа тональности текста.

# **ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

## **Цели и задачи исследования**

Целью данной работы являлось исследование эффективности применения искусственных нейронных сетей для решения задачи классификации тональности речевых сигналов человека, а также сравнение полученных показателей качества работы системы с аналогичными результатами, полученными в предыдущей работе. Основной упор в работе делается на исследование искусственных нейронных сетей базирующихся на сверточных и рекуррентных архитектурах.

Список основных задач которые нужно выполнить в рамках данной работы:

- В первую очередь необходимо выполнить анализ актуальных литературных источников по теме анализа тональности речевых сигналов;
- Выполнить краткий анализ результатов достигнутых в предыдущей работе, а также установить бейзлайн, относительно которого можно будет оцениваться результаты, которые будут достигнуты средствами искусственных нейронных сетей;
- Составить краткий теоретический обзор нейросетевых архитектур, используемых для анализа последовательных данных. В том числе, рекуррентных нейронных сетей и их основных компонентов (GRU, LSTM);
- Выбрать наилучшую архитектуру нейронной сети. Сравнить её с бейзлайном.

Основной гипотезой, положенной в основу данной диссертационной работы, является предположение о том, что искусственные нейронные сети могут справляться с задачей классификации тональность речевых сигналов человека быстрее и качественнее в сравнении с классическими методами, представляющими из себя pipeline из генерации фичей и последующей мета-классификацией.

## **Область исследования**

Содержание диссертации соответствует образовательному стандарту высшего образования второй ступени (магистратуры) ОСВО 1-40 81 04-2020 специальности 1-40 81 04 – «Обработка больших объемов информации».

## **Апробация результатов диссертации**

Основные положения диссертационной работы докладывались на 56 научной конференции аспирантов, магистрантов и студентов БГУИР.

## **Опубликованность рецептов диссертации**

По теме диссертации опубликованы 2 печатные работы в сборнике научной конференции аспирантов, магистрантов и студентов БГУИР

### **Структура и объем диссертации**

Диссертация состоит из введения, общей характеристики работы, четырех глав, заключения, списка использованных источников и приложений. В первой главе представлен анализ предметной области и актуальных источников по теме. Вторая глава посвящена обзору результатов достигнутых в предыдущей работе и формулированию ориентира для дальнейшей оценки качества нового подхода. В третьей главе выполнен обзор теоретической базы лежащей в основе анализа последовательностей данных нейронными сетями. В частности рекуррентными нейронными сетями. В четвертой главе проводится обзор лучших архитектур и их результатов среди разных концепций нейронных сетей (сверточные и рекуррентные) с последующим выбором лучшей из них, а также обоснованием этого выбора.

Общий объем работы составляет 54 страниц, из которых основного текста – 42 страниц, 35 рисунков на 15 страницах, 7 таблиц на 5 страницах, список использованных источников из 20 наименований на 2 страницах и 1 приложение на 7 страницах.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** содержится краткий рассказ о целях и логике построения исследования в работе, частично обоснованно. Также описано актуальность темы и методологии.

В **общей характеристике работы** сформулированы цели и задачи диссертации, обозначена область исследований, научная и практическая значимость. Приведено количество опубликованных работ по теме диссертации, а также описана ее структура.

В **первой** проведен анализ литературных источников, а также наиболее актуальных готовых инструментов анализа речевых сигналов. Была выявлена новизна и целесообразность проводимого исследования.

Во **второй** главе проводится краткий обзор результатов, достигнутых в предыдущей работе. Проводится краткий обзор методологии тестирования качества, с помощью которой было получено качество работы системы, разработанной в предыдущей работе. Далее, на эти показатели качества используются как базовые для оценки результатов диссертационной работы.

В **третьей** главе проводится достаточно масштабное изложение теории анализа последовательных данных с помощью нейронных сетей. Рассматриваются и разбираются наиболее эффективные и популярные в научном сообществе подходы и архитектуры. В частности большой упор делается на разбор классических на данный момент рекуррентных архитектур, таких как LSTM и GRU. Обозреваются их особенности, а также сильные и слабые стороны. Формулируются рекомендации по их использованию. Также в главе упоминаются альтернативные способы анализа последовательных данных, основанных на идее выявления позиционно инвариантных паттернов в последовательных данных. Если точнее, речь идет об одномерных сверточных архитектурах.

В последней **четвертой** главе приводятся результаты практических исследований. Обозреваются и сравниваются архитектуры, показавшие наилучшее качество и базирующиеся на разных концепциях (сравниваются GRU-based и Conv-based архитектуры). Сравнение производится по ряду характеристик включающих финальную оценку качества работы сети, скорость обработки данных, размер и вес моделей. Далее на основе выполненного сравнения делаются финальные выводы и рекомендации по выбору подходов для решения задач схожего типа.

**Заключение** содержит более выводы более общего характера. О результативности, эффективности и успешности проведенных исследований. Причем делается это на основе сразу двух работ по этой теме, исследующих разные области или подходы к решению одной задачи.

## ЗАКЛЮЧЕНИЕ

В рамках данного диссертационного исследования был проведен дополнительный, по отношению к диплому, анализ литературных источников. Большинство статей посвященных теме анализа тональности речевых сигналов используют комбинированные подходы: параллельно анализируют тональные характеристики речи и смысловое содержание сказанного через анализ текста. Однако, даже при этом, результаты не поднимаются выше 85-90% качества.

Диссертация является логическим продолжением исследований по теме дипломной работы. Таким образом, результат, достигнутый в дипломной работе (67% ассигасу и f1-score), изначально являлся ориентиром для понимания успешности результатов и целесообразности применения нейросетевых алгоритмов для решения данной задачи.

После проведенных исследований можно с уверенностью сказать, что применение нейронных сетей для анализа тональности речевых сигналов является эффективным и результативным. Также применение такого подхода достойно внимания при решении схожих задач или той же задачи на других данных. Обосновываются эти утверждения достигнутыми с помощью нейросетевых архитектур результатах в примерно 95% качества. Что является приростом в 30% относительно диплома (подхода с фичер инженерингом через спектральный анализ и классическим мета-классификатором).

Также в рамках данной работы сформулированы и обоснованы рекомендации по выбору наиболее подходящей архитектуры нейронной сети. А именно речь идет об одномерной сверточной нейронной сети. Которая показала незначительное отличие в качестве и значительное увеличение скорости обработки данных в сравнении с рекуррентной сетью.

Стоит отметить, что в данном случае обе архитектуры спроектированы таким образом, что размер входной последовательности не фиксирован (в сверточной сети благодаря глобал-пулингу). Вследствие чего была устранена необходимость в алгоритме анализа больших файлов средствами скользящего окна с трапециевидным усреднением.

Для дальнейших улучшений и исследований можно попробовать улучшить архитектуры с помощью слоя Attention (предварительные эксперименты не дали результата). Также имеет смысл поиск и сбор более репрезентативных и объемный датасетов.

## СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

1-А. Талецкий А. И. Оценка эффективности применения моделей классического машинного обучения для анализа тональности речевых сигналов человека / А. И. Талецкий // Компьютерные системы и сети: 56-я научная конференция аспирантов, магистрантов и студентов, Минск / Белорусский государственный университет информатики и радиоэлектроники. – Минск, 2020 (В печати).

2-А. Талецкий А. И. Оценка эффективности применения искусственных нейронных сетей для анализа тональности речевых сигналов человека / А. И. Талецкий // Компьютерные системы и сети: 56-я научная конференция аспирантов, магистрантов и студентов, Минск / Белорусский государственный университет информатики и радиоэлектроники. – Минск, 2020 (В печати).

Библиотека БГУИР