

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

УДК 681.3

Гордеенко
Валерий Игоревич

НЕЙРОННАЯ СЕТЬ ДЛЯ
ОПРЕДЕЛЕНИЯ ТОНАЛЬНОСТИ ТЕКСТА

АВТОРЕФЕРАТ

на соискание степени магистра технических наук
по специальности 1-40 80 02

«Системный анализ, управление и обработка информации»

Научный руководитель
Герман Олег Витольдович
кандидат технических наук, доцент

Минск 2020

ВВЕДЕНИЕ

Цель и задачи исследования. Цель диссертационной работы - разработка алгоритма, который должен выполнить следующие задачи: обработку входного текста, анализ эмоциональной окраски текста некоторым методом и вывод пользователю решения задачи в виде сообщения. Также требованием к программе будет максимально высокая скорость работы и высокая точность определения эмоциональной окраски текста. Исходя из целей данной работы, следует выделить несколько этапов работы над проектом. Первый этап – считывание, обработка и адаптация текста для нейросетевого анализатора эмоциональной окраски. Второй этап - создание обработчика полученного текста (ниже более подробно описаны его методы): стемминг (выделение базы однокоренных слов для приведения их к определенному виду), таблица матриц (большая база данных корней слов, каждому из которых присвоен статус положительной или отрицательной эмоциональной окраски на основе решения задачи нейросети), выделение наиболее значимых слов и группировка их по рейтингу значимости (данный шаг необходим для ускорения работы программы с таблицей матриц) и анализ текста на основе логит-регрессии.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Задачи исследования. Разработка алгоритма для определения тональности текста.

Объект исследования. Входящие текстовые данные.

Предмет исследования. Модели, алгоритмы и структуры нейронной сети для определения тональности текста.

Цель и задачи исследований. Целью является исследование задач о существующих алгоритмов эмоциональной оценки текста, а также разработка методов для определения тональности текста.

Структура и объем диссертации. Диссертационная работа состоит из введения, трёх глав, заключения, списка использованных источников. Работа включает в себя 40 страниц основного текста, 16 рисунков и 4 таблиц. Список использованных источников содержит 5 наименований.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В главе 1 «Обзор существующих решений» был проведен анализ и сравнение существующих методов и алгоритмов для решения данной задачи.

В главе 2 «Цель работы» были сформулированы цель магистерской диссертации и задачи, которые нужно решить для её достижения. Также дано предварительное краткое описание методов, выбранных для решения поставленной задачи. Описываются методы выделения значимых слов, стемминг, сбор данных для обучения нейросети, логит-регрессия, метод анализа иерархий Томаса Саати, обучение нейросети и блок-схема работы алгоритмы.

В заключительной главе 3 «Программная реализация обучения сети логистическому распознаванию» был продемонстрирован пример решения задачи с помощью Python и пакета Sklearn.

Диссертация выполнена самостоятельно, проверена в системе «Антиплагиат». Процент оригинальности соответствует норме, установленной кафедрой. Цитирования обозначены ссылками на публикации, указанные в «Списке использованных источников».

ЗАКЛЮЧЕНИЕ

Цель работы можно считать достигнутой - разработан алгоритм, который выполняет следующие задачи: обработку входного текста, анализ эмоциональной окраски текста некоторым методом и вывод пользователю решения задачи. Выполнены считывание, обработка и адаптация текста для нейросетевого анализатора эмоциональной окраски, создан обработчик полученного текста с использованием следующих методов: стемминг, таблица матриц, выделение наиболее значимых слов и группировка их по рейтингу значимости и анализ текста на основе логит-регрессии.

СПИСОК ПУБЛИКАЦИЙ АВТОРА

[1–А.] Лагойский М.В., Гордеенко В.И., Герман О.В. Вариант задачи линейного программирования / Интернаука: электронный научный журнал. – № 22(151). – М. : Изд. «Интернаука», 2020.