

ОТЗЫВ ЗАРУБЕЖНОГО НАУЧНОГО КОНСУЛЬТАНТА
на диссертацию Акановой Акерке Сапаровны
на тему «Нейрокомпьютерная система семантического анализа текста»,
представленную на соискание степени PhD
по специальности 6D075100 – «Информатика, вычислительная техника и
управление»

Развитие научных исследований в области нейрокомпьютерных систем является на данный момент одним из приоритетных направлений в науке и технике. В наше время, называемое веком высоких технологий и больших данных, в решении вопроса автоматизации обработки текста с помощью вычислительных машин увеличиваются задачи высокой сложности, для их решения широко применяют искусственный интеллект, в частности машинное обучение.

Целью диссертационной работы является построение нейросетевой модели для глубокого обучения выходных данных тематической модели, построенной на базе казахскоязычного текста. В работе рассматриваются вопросы проведения автоматизированного стемминга текста на казахском языке путем адаптирования стеммера Портера, приведён семантический анализ текста на казахском языке с применением тематического моделирования и глубокого обучения нейронных сетей, освещены вопросы создания многослойной нейросетевой модели.

Актуальность темы исследования обусловлена возникновением проблем проведения автоматизированного стемминга текста на казахском языке, семантического анализа с применением тематического моделирования и глубокого обучения нейронных сетей, компьютерной обработки текста на казахском языке и решением перечисленных проблем при помощи моделей нейронных сетей.

В работе приведена классификация методов автоматизированной обработки текста, которая включает методы схожести, метод максимального правдоподобия, метод случайных блужданий, метод косинусного сходства, метод латентно-семантического анализа, методы машинного обучения, статистический метод, которые используются в целях автоматизированной обработки большого объема текста.

Для семантического анализа корпуса текстов на казахском языке требуются применение тематического моделирования и построение нейронной сети для глубокого обучения данных, полученных на основе латентного распределения Дирихле. Для реализации тематической модели корпуса текстов на казахском языке был адаптирован алгоритм Портера, который способствует созданию словаря словоформ на казахском языке. Адаптированный стеммер Портера выполняет функцию удаления аффиксов казахского языка, включенных в работу алгоритма. Ранее алгоритм Портера не применялся в автоматизации текста на казахском языке. Стеммер является неотъемлемой частью при создании словаря для LDA-модели. LDA-модель включает в себя распределение ключевых слов по темам, тем по документу, различных тем, каждая тема представлена, как комбинация ключевых слов, при этом каждое ключевое слово имеет свой определенный вес в теме. Созданная тематическая модель имеет матрицу весов ключевых слов в теме и весов тем в документе. Количество элементов полученной матрицы векторов входных данных дает количество нейронов. Над полученными нейронами производится глубокое обучение в построенной нейронной сети. Процедура глубокого обучения модели семантического анализа текста на казахском языке в многослойной нейронной сети состоит из комбинации слоев Embedding, SpatialDropout1D, LSTM, Dense.

В диссертационной работе автором получены следующие результаты:

- исследован и проведен анализ методов и моделей семантического анализа текста, применяемых при автоматической обработке;
- приведена классификация методов, применяемых при автоматизированной обработке текста в целом;

- проведен сравнительный анализ технологий распознавания объектов на основе нейронных сетей, где были описаны сравнительные характеристики 13 технологий: Apache Singa, Caffe, Deeplearning4j, Keras, Microsoft Cognitive Toolkit, MXNet, Neural Designer, OpenNN, Theano, Torch, Tensor Flow, NLTK, Gensim;

- адаптирован алгоритм работы стеммера Портера для создания эталонного словаря словоформ;

- построена тематическая модель при помощи LDA для обработки текста на казахском языке;

- построена нейросетевая модель для обработки данных, полученных путем тематического моделирования;

- разработана нейрокомпьютерная система обработки текстовых данных на казахском языке.

Апробация работы проходила на базе Павлодарского государственного университета имени С. Торайгырова, получен акт внедрения. Результаты диссертации опубликованы в 12 работах, среди которых 1 монография, 4 статьи в журналах, рекомендованных Комитетом по контролю в сфере образования и науки МОН РК, 1 статья в отечественном научном издании, 2 статьи в международных научных изданиях, индексируемых в базах данных Thomson Reuters и Scopus, 4 работы в материалах международных и республиканских конференций.

На основании вышеизложенного, считаю диссертационную работу Акановой Акерке Сапаровны на тему «Нейрокомпьютерная система семантического анализа текста» удовлетворяющей всем требованиям, предъявляемым к диссертационным работам на соискание учёной степени PhD, а её автора – Аканову Акерке Сапаровну – заслуживающей присуждения учёной степени PhD по специальности «6D075100 – Информатика, вычислительная техника и управление».

Зарубежный научный консультант:
кандидат технических наук,
доцент кафедры высшей
математики ФГБОУ ВО «Сибирский
государственный университет
геосистем и технологий

Е. Крылова

Крылова Е.М.

