

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра дискретной математики и алгоритмики

Аннотация к дипломной работе

«Алгоритмы нормализации текстов социальных сетей»

Юнусов Рустам Рустамович

Научный руководитель — старший преподаватель, Комаровский И.В

2017

РЕФЕРАТ

Дипломная работа, 52 с., 19 рис., 33 табл., 16 источников.

ТЕКСТЫ СОЦ. СЕТЕЙ, НОРМАЛИЗАЦИЯ ТЕКСТА, УСЛОВНЫЕ СЛУЧАЙНЫЕ ПОЛЯ, НАИВНЫЙ БАЙЕСОВСКИЙ КЛАССИФИКАТОР, СЕНТИМЕНТНЫЙ АНАЛИЗ ТЕКСТОВ, ЭМБЕДДИНГ, ЛИНЕЙНЫЙ КЛАССИФИКАТОР, МНОГОСЛОЙНЫЙ ПЕРСЕПТРОН, РЕКУРРЕНТНЫЕ НЕЙРОННЫЕ СЕТИ, БАТЧ НОРМАЛИЗАЦИЯ, ДРОПАУТ, РЕГУЛЯРИЗАЦИЯ, ВИЗУАЛИЗАЦИЯ, ОБРАБОТКА ЕСТЕСТВЕННЫХ ЯЗЫКОВ.

Объектом исследования являются тексты социальных сетей и алгоритмы их обработки.

Цель работы: разработать и реализовать модели для нормализации и сентиментного анализа текстов социальных сетей, продемонстрировать эффект нормализации текстов на качество работы методов сентиментного анализа.

Методы исследования: методы обработки естественных языков, язык программирования python.

В ходе работы получены следующие результаты:

- 1) Реализованы подходы на основе условных случайных полей и наивного байеса для нормализации текстов социальных сетей, проведён их сравнительный анализ.
- 2) Реализована демонстрационная система для нормализации текстов на фреймворке Django.
- 3) Реализованы подходы на основе линейного классификатора, многослойного персептрона, рекуррентных нейронных сетей и различных вариантов эмбединга для сентиментного анализа текстов, проведён их сравнительный анализ.
- 4) Реализованы подходы для нормализации текстов с предварительной нормализацией текста. Продемонстрировано влияние нормализации на качество работы алгоритмов анализа тональности текстов.

Область применения: машинное обучение, обработка естественных языков.

ABSTRACT

Graduate work, 52 p., 19 pic., 33 tablets, 16 sources.

TEXTS OF SOC. NETWORKS, NORMALIZATION OF TEXT, CONDITIONAL RANDOM FIELDS, NAIVE BAYES CLASSIFIER, SENTIMENT ANALYSIS, EMBEDDING, LINEAR CLASSIFIER, MULTILAYER PERSEPTRON, RECURRENT NEURAL NETWORKS, BATCH NORMALIZATION, DROPAUT, REGULARIZATION, VISUALIZATION, NATURAL LANGUAGE PROCESSING.

Object of research: texts of social networks and algorithms for their processing.

Purpose of the work: to develop and to implement models for normalization and sentimental analysis of social network texts. To demonstrate influence of text normalization on the quality of the sentiment analysis methods.

Methods of research: methods of processing natural languages, Python programming language.

In the course of the work the following results were achieved:

- 1) Models are implemented on the basis of conditional random fields and naive bayesian for, their comparative analysis is carried out.
- 2) Implemented a demonstration system for with Django framework.
- 3) Implement models based on the linear classifier, multilayer perceptron, recurrent neural networks and various options for the sentiment analysis of texts, their comparative analysis is carried out.
- 4) Implement models for the normalization of texts with preliminary normalization of the text. The effect of normalization on the quality of the algorithms for analyzing the tonality of texts.

Scope: machine learning, natural language processing.