

РАЗРАБОТКА АЛГОРИТМА АНАЛИЗА ТОНАЛЬНОСТИ ТЕКСТОВЫХ СООБЩЕНИЙ О ЗАДАННОЙ БАНКОВСКОЙ ОРГАНИЗАЦИИ, РАЗМЕЩЕННЫХ НА ВЕБ-САЙТАХ СОЦИАЛЬНЫХ СЕТЕЙ И СОДЕРЖАЩИХ ИНФОРМАЦИЮ О ТЕХНИЧЕСКИХ СБОЯХ И БАНКОВСКОМ МОШЕННИЧЕСТВЕ

Чудакова А. А. (Университет ИТМО)

Научный руководитель – канд. техн. наук, доцент Воробьева А. А.
(Университет ИТМО)

В работе представлен алгоритм анализа тональности текстовых сообщений о заданной банковской организации, размещенных на веб-сайтах социальных сетей и содержащих информацию о технических сбоях и банковском мошенничестве, позволяющий выделять негативные отзывы для сокращения времени реагирования на них.

Введение

Социальные сети являются распространенным, открытым источником данных для анализа и интересны тем, что сообщения о каких-либо значимых событиях практически моментально публикуются в социальных сетях. Стоит отметить то, что существуют специальные узконаправленные сайты, на которых пользователи обсуждают или делятся отзывами по поводу какой-либо сферы своей жизни.

Определение тональности сообщений, размещенных на веб-сайтах социальных сетей, позволяет выявлять события, которые требуют реакции со стороны банковской организации, и сообщения, релевантные для банка. Так сообщения позитивной тональности в большинстве случаев отражают ситуации, которые не требуют анализа службами банка. Определение эмоциональной окраски позволяет определять тональность сообщения более точно.

Разработка технологии автоматического выявления оценки мнений на информационных порталах и в социальных сетях является важной задачей, решение которой позволит сократить время реагирования на инциденты и упростить работу сотрудников. Для банковского сектора извлечение такой информации является критичным, в силу высокой конкуренции в этой сфере, а также упущенное время может повлечь большие денежные и репутационные потери. В связи с этим тема данной исследовательской работы является актуальной и находит практическое применение.

Целью работы является исследование и разработка метода анализа тональности текстовых сообщений о заданной банковской организации, размещенных на веб-сайтах социальных сетей и содержащих информацию о технических сбоях и банковском мошенничестве.

Анализ предметной области, в том числе анализ научной литературы и патентные исследования, показали, что автоматизированных решений на мировом и российском рынке не представлено.

Основная часть

Разработанный алгоритм использует методы машинного обучения и включает в себя следующие этапы:

- получение отзывов из открытых источников;
- определение степени эмоциональности отзыва;
- предобработка текста отзыва;
- анализ тональности текста отзыва.

В качестве этапов предварительной обработки текста в данной НИР выбраны следующие методы:

- замена URL-адресов и упоминаний пользователей,
- замена сокращений,

- удаление чисел;
- замена повторений пунктуации;
- лемматизация.

Необходимо отметить, что негативные отзывы с высокой эмоциональной окраской будут иметь наибольший приоритет.

Выводы

В результате работы был разработан алгоритм, выполняющий анализ тональности текстовых сообщений, позволяющий выделять негативные отзывы для сокращения времени реагирования на них. В дальнейших исследованиях планируется проведение экспериментов, подтверждающих эффективность разработанного алгоритма.

Чудакова А.А. (автор)

Воробьева А.А. (научный руководитель)