

Варламов А.Д.

*Муромский институт (филиал) федерального государственного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирская обл., ул. Орловская, 23
E-mail: varlamov_aleks@mail.ru*

Учет эмоциональных квантификаторов для оценки тональности текстов

Под тональностью текста понимают эмоциональный окрас и эмоциональную оценку автора по отношению к объектам, речь о которых идёт в тексте [2]. Авторские высказывания могут иметь тон от позитивного до негативного. Чаще всего выделяют классы документов с положительной и с отрицательной тональностью текста, а также класс, в котором нет явного превосходства одного из них.

Анализ тональности (Sentiment Analysis) относится к типу задач классификации текстов. Чаще всего документы делят на два класса (позитивный и негативный тон) и на три класса (с учетом нейтрально-тональных текстов).

Практически все подходы применяют методы обучения на составленной экспертами коллекции из высказываний, разделенных по группам тональности. Это весьма обосновано, так как, имея выборку, можно легко вычислить меру соответствия оценок тональностей, данных экспертами и системой, и по ней же сравнить исследуемый алгоритм с другими [3].

На эмоциональный оттенок каждого слова влияет общий контекст предложения, в котором оно используется. В разных сочетаниях с другими словами одно и то же слово может в разной степени отражать положительную или отрицательную эмоцию автора. Наиболее сильное такое воздействие оказывают слова и выражения, которые в теории нечетких множеств называются квантификаторами. Их можно разделить на следующие группы:

1. усиливающие эмоциональное влияние слов (например, “слишком”, “чересчур”, “очень”);
2. ослабляющие эмоциональное влияние слов (например, “почти”, “возможно”, “вероятно”);
3. изменяющие эмоциональное влияние слов на противоположное (например, “не”, “без”).

В докладе приведены результаты исследования алгоритма классификации текста по тональности, основанного на машинном обучении, и его модификации, учитывающей воздействие квантификаторов на ключевые слова [1]. Исследование проводилось с целью получения ответов на следующие вопросы:

1. Какова степень влияния наличия отдельных квантификаторов на эмоциональный окрас текста?
2. Как изменится качество оценки тональности текста, если система станет учитывать влияние эмоциональных квантификаторов на тон высказываний?

По результатам исследования были сделаны следующие выводы:

1. В большинстве случаев при сентимент-анализе квантификаторы оказывают воздействие на отдачу ключевыми словами эмоционального тона всему анализируемому тексту.
2. Учет влияния квантификаторами на тональность слов улучшил качество работы алгоритма на 4,5 % по метрике "accuracy".

Таким образом, учет эмоциональных квантификаторов в алгоритмах и методах оценки тональности текстов можно считать оправданным.

Литература

1. Варламов А.Д. Проблемы практического применения наивного байесовского классификатора для оценки тональности текстов // Алгоритмы, методы и системы обработки данных. 2017. № 36. С. 8-16.
2. Российский семинар по Оценке Методов Информационного Поиска (РОМИП) [Электронный ресурс]. — Режим доступа: <http://www.romip.ru>.
3. Lewis D.D. Naive (Bayes) at forty: The independence assumption in information retrieval. Proceedings of 10th European Conference on Machine Learning, Chemnitz, Germany, 1998, pp. 4–15.