

МЕТОДЫ ИЗВЛЕЧЕНИЯ АСПЕКТНЫХ ТЕРМИНОВ

Рой Дарья Александровна

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: dasha02.08@mail.ru

Одним из направлений задачи анализа тональности мнений (*sentiment analysis*) является аспектный анализ, который позволяет определить, какое мнение автор выражает не только по поводу сущности в целом, но и по ее аспектам [1]. В качестве сущности в данном случае может выступать объект, продукт, товар или услуга, по поводу которых автор выражает мнение. Под аспектом при этом понимаются атрибуты, качества и свойства, характеризующие анализируемую сущность. Например, для сущности *отель* оцениваемыми аспектами могут являться *сервис*, *интерьер*, *расположение* и *цена*.

В текстах мнений для обозначения аспектов используются аспектные термины. К примеру, для аспекта интерьер аспектными терминами являются потолок, обстановка, мебель.

Как правило, при решении задачи аспектного анализа мнений рассматриваются следующие подзадачи [2]:

- выделение аспектных терминов,
- объединение аспектных терминов в аспектные категории (т.е. отнесение аспектного термина к соответствующему аспекту),
- определение тональности по отношению к аспектам.

Для каждой из указанных подзадач существуют свои методы решения. В данной работе был проведен обзор методов выделения аспектных терминов.

Для извлечения аспектных терминов существуют три подхода [3]:

- статистический,
- лингвистический,
- машинное обучение.

Статистический подход рассматривает задачу извлечения аспектных терминов как задачу извлечения слов и словосочетаний, наиболее часто встречающихся в документах.

Идея лингвистического подхода состоит в том, что поскольку рядом с аспектными терминами часто употребляются оценочные слова, то для извлечения аспектов имеет смысл строить лингвистические правила и шаблоны, на основе которых в тексте будут находиться требуемые конструкции.

При использовании машинного обучения выделение аспектов рассматривается либо как проблема классификации слов и словосочетаний на два класса: аспект и неаспект, либо как задача последовательной разметки слов (*sequence labeling*).

Лучшие результаты для задачи извлечения аспектных терминов на данный момент показывают нейронные сети [4]. При этом, как правило, анализируются англоязычные мнения из одной области, т.е. рассматривающие одну определенную сущность.

По результатам проведенного обзора были выбраны и реализованы методы для решения подзадачи выделение аспектных терминов из русскоязычных мнений. Среди них: метод на основе статистики, метод на основе шаблонов, метод на основе машинного обучения (SVM) и различные архитектуры нейронных сетей. Для оценки качества работы методов использована коллекция отзывов про рестораны [2]. В дальнейшем планируется предложить метод, который позволит обрабатывать мнения из любой области.

В качестве дополнительной задачи ставится извлечение общей тональности из мнений о конкретной сущности, т.е. извлечение словосочетаний, наиболее ярко характеризующих ее.

Литература

1. Popescu A., Etzioni O. Extracting product features and opinions from reviews // Natural language processing and text mining. Springer: London. 2007. P. 9–28.
2. Loukachevitch N. V., Blinov P. D., Kotelnikov E. V., Rubtsova Y. V., Ivanov V. V., Tutubalina E. V. SentiRuEval: Testing Object-oriented Sentiment Analysis Systems in Russian // In Proceedings of International Conference Dialog. 2015. P. 3–9
3. Liu B., Zhang L. A survey of opinion mining and sentiment analysis // In Mining Text Data. Springer: US, 2012. P. 415–463.
4. Pontiki M., Galanis D., Papageorgiou H., Androutsopoulos I., Manandhar S., Al-Smadi M., Al-Ayyoub M., Zhao Y., Bing Q., De Clercq O., Hoste V., Apidianaki M., Tannier X., Loukachevitch N., Kotelnikov E., Bel N., Maria Jimenez-Zafra S., Eryigit G. // Proceedings of the 10th International Workshop on Semantic Evaluation, San Diego, California, 2016. P. 19–30