

## **МЕХАНИЗМЫ АНАЛИЗА ТОНАЛЬНОСТИ ТЕКСТА**

**Ю. С. Кузьменков, И. А. Мурашко**

*Учреждение образования «Гомельский государственный технический университет имени П. О. Сухого», Республика Беларусь*

Статья посвящена вопросу исследования методов анализа эмоциональной окраски текстов, а также разработке и реализации метода эмоциональной окраски на примере динамических интернет-ресурсов.

В ряде исследований по определению полярности текста высокую эффективность показали методы обучения с учителем. Эти методы использовались как в ранних работах по определению полярности документа, так и в современных работах, где анализируются предложения и короткие текстовые сообщения. Для решения поставленной задачи был выбран данный метод.

Для использования методов обучения с учителем требуется обучающая выборка. Обычно обучающее множество составляется из примеров той области, в которой будет применяться классификатор.

В качестве обучающей и проверочной выборки был составлен корпус, состоящий из 8000 предложений, для которых определена полярность. Часть этих предложений была извлечена из размеченного корпуса, предоставленного для свободного доступа. Другая часть была получена с помощью онлайн-системы Sentiment140 анализа эмоциональной окраски.

Все примеры полученного обучающего множества взяты из мнений об электронной технике, а именно о мобильных телефонах, планшетах, плеерах.

Эффективность алгоритмов извлечения аспектов формулируется в терминах точности и полноты. В контексте решаемой задачи эти метрики имеют следующий смысл. Алгоритм извлечения аспектов проверяет каждый термин документа на принадлежность множеству аспектов. Тогда точностью этого алгоритма называется отношение числа правильно определенных аспектов к числу всех терминов, отнесен-

ных к классу аспектов, а полностью – отношение числа правильно определенных аспектов к числу аспектов в документе.

Для тестирования используются предложения, взятые из размеченного корпуса. Из корпуса взяты 1700 предложений из отзывов о смартфонах и 1100 отзывов о ноутбуках. В каждом из этих предложений выделены аспекты. Результаты тестирования реализованного алгоритма показаны в таблице.

#### Результаты тестирования

Домен	Точность	Полнота
Смартфоны	0,67	0,79
Ноутбуки	0,61	0,71

Проанализировав полученные результаты, можно сделать следующие выводы: для решения задачи определения полярности предложений и коротких сообщений эффективны алгоритмы обучения с учителем. Проблемой обучения с учителем является составление тренировочного корпуса с примерами из предметной области, в которой будет использоваться классификатор. Однако схожей проблемой обладают и словарные методы: веса терминов словаря, составленного для одной предметной области, могут оказаться неадекватными для другой. Задача определения полярности текста успешно решается с помощью методов обучения с учителем. Для увеличения эффективности этих методов используются лингвистические и частотные фильтры, позволяющие отсеивать слова, не имеющие отношения к аспектам.