



РАЗРАБОТКА АЛГОРИТМА ВЫДЕЛЕНИЯ ЭЛЕМЕНТОВ ЭМОЦИОНАЛЬНОСТИ

Медведская А.В.

*Национальный технический университет
«Харьковский политехнический институт»
61002, Харьков, ул. Фрунзе, 21, тел. (057) 707-64-60
e-mail: brovka90@inbox.ru*

С развитием информационных технологий и всемирной паутины Интернет формируется Интернет язык, который, по форме существования является письменным языком, фактически близок к разговорному, поскольку обладает ее основными признаками: непосредственностью и неподготовленностью общения, преобладанием диалога над монологом, эмоциональностью, экспрессивностью и логической непоследовательностью высказываний.

Таким образом, если большая часть коммуникации переходит в область письма, неизбежно встает вопрос о недостаточности письменной речи. Во время речи мы используем различные средства передачи эмоций: мимику, интонацию, жестикуляцию. На письме мы лишены этих возможностей. Но в языке интернета для передачи эмоций существуют некоторые инструменты, которые помогают понять эмоции пользователя. Следовательно, происходит обогащение языка [1].

Эти средства можно классифицировать следующим образом:

1. Фонетические:

- многократное повторение звуков (пролонгация);
- написание по слогам (скандирование или произношение);
- искривление стандартов орфографии;
- словесное ударение с помощью большой гласной буквы;
- фразовый ударение;
- интонационная окраска.

2. Графические. Основные графические средства выражения эмоций – это графические улыбки (смайлики). Широкой популярности в форумах, чатах они достигли из-за удобства вставки в текст, понятности, возможности выражения эмоций.

3. Для выражения эмоций существуют различные лексические средства, которые уже по своему содержанию эмоциональные, выражающие то или иное чувство [2, 3].

Целью данной работы является решение задачи определения эмоциональной лексики в электронных сообщениях. Эта задача зачастую является одной из подзадач анализа тональности текста.

Для решения задачи определения эмоциональных элементов используются следующие методы: методы, основанные на знаниях (правилах, словарях); Machine Learning и скрытая Марковская модель.



Для построения алгоритма извлечения эмоциональных элементов из электронных сообщений был выбран метод Machine Learning с учителем.

Общий алгоритм Machine Learning выглядит следующим образом:

1) Необходимо собрать коллекцию документов для обучения классификатора.

2) Каждый документ учебной коллекции нужно представить в виде вектора признаков. Сообщение с учебной коллекции нужно разметить тегами.

1. Многократное повторение звуков (проголгация).

`<prolong>aaaaaaaaa</prolong>`

2. Написание по слогам (скандирование или произношение).

`<scans> Поз-дра-вл яю</scans>`

3. Искривление стандартов орфографии.

`<fault>Аффтар</fault>`

4. Словесное ударение с помощью большой гласной буквы.

`<wstress> КрасавЕц </wstress>`

5. Фразовое ударение.

`<phstress> ДА </phstress>`

6. Интонационная окраска.

`<tone>!!!!</tone>`

7. Смайлик.

`<smile>:</smile>`

3) Для каждого документа надо указать «правильный ответ», то есть, содержит эмоциональные элементы или не содержит, за этими ответами и будет обучаться классификатор.

4) С помощью классификатора система начинает сравнивать два класса, к которым мы отнесли сообщения на предыдущем этапе, и обнаруживает характерные признаки эмоциональности в тексте. Так система учится на тренировочной коллекции.

5) Используя тестовую выборку, тестируем обученный классификатор. Документ подается на вход классификатора, и полученные ответы сравниваются с экспертными оценками [4].

В ходе работы был разработан алгоритм определения эмоциональных элементов в электронных сообщениях с использованием Machine Learning, с помощью которого система будет обучаться распознавать и выделять эмоциональные элементы в тексте.

Список литературы

1. Валгина Н. С. Активные процессы в современном русском языке: учебное пособие для студентов вузов /Н. С. Валгина. – Москва: Логос, 2001. – 210 с.

2. Голуб И. Б. Стилистика русского языка. /И. Б. Голубь – М.: Айрис-пресс, 2003. – 155 с.

3. Максимов В. И. Стилистика и литературное редактирование: учебник /В. И. Максимов. – М.: Гарда-рики, 2007. – 274 с.

4. Пак А. Обучаем компьютер чувствам (sentiment analysis по-русски) [Электронный ресурс] – Режим доступа: <http://habrahabr.ru/post/149605/>