

Научно-исследовательская работа

Анализ тональностей текстов на естественном языке

Выполнил Муратов Ф.
Руководитель Толмачев А.

Академический Университет, осень 2014

Введение

Что такое анализ тональностей?

Где можно применить?

Мотивация

Задача хорошо исследована для текстов на английском языке.

Однако, очень мало исследований решения данной задачи для русского языка.

Задача

Исследование применимости существующих методов бинарной классификации текстов для анализа тональности текстов на русском языке.

Цели

1. Обзор методов решения данной задачи
2. Выбор методов
3. Оценка качества работы методов
4. Анализ результатов

Подход к решению

1. Реализация трех методов анализа: словарный, наивный байесовский классификатор, SVM
2. Оценка качества методов для определенных выборок
3. Анализ результатов

Данные

Две тестовые выборки предоставленные
<http://romip.ru/>.

1. Набор текстов из блогов
2. Коллекция цитат из новостного потока

Результаты

Отзывы о товарах

Метод	Ассурасу
Random - base line	~0.5
Словарный метод	0.84
Naive Bayes	0.87
SVM	0.88

Новостная лента

Метод	Ассурасу
Random - base line	~0.5
Словарный метод	0.79
Naive Bayes	0.79
SVM	0.81

Выводы

О предметной области

- Сильная зависимость результатов от предметной области.

О методах

- Словарный метод хорошо подходит, если есть словарь под предметную область.
- Наивный байесовский классификатор несмотря на свою простоту показал хорошие результаты.

Полученный опыт

Познакомился с новой для себя областью - обработкой текстов на естественном языке. Узнал некоторые подходы к решению задач машинного обучения.

Реализовал несколько методов бинарной классификации.

Контакты

mail:

muratovfedor@yandex.ru

code:

https://github.com/muratovv/Sentiment_Analysis