

Анализ тональности-2

Открытое тестирование систем анализа тональности

РОМИП-2011, 2012

- Анализ тональности

- Отзывы о фильмах,
- книгах,
-
- цифровых камерах



- Участие

- 2011: 12 участников, более 200 прс
- 2012: 17 участников, более 150 прогонов



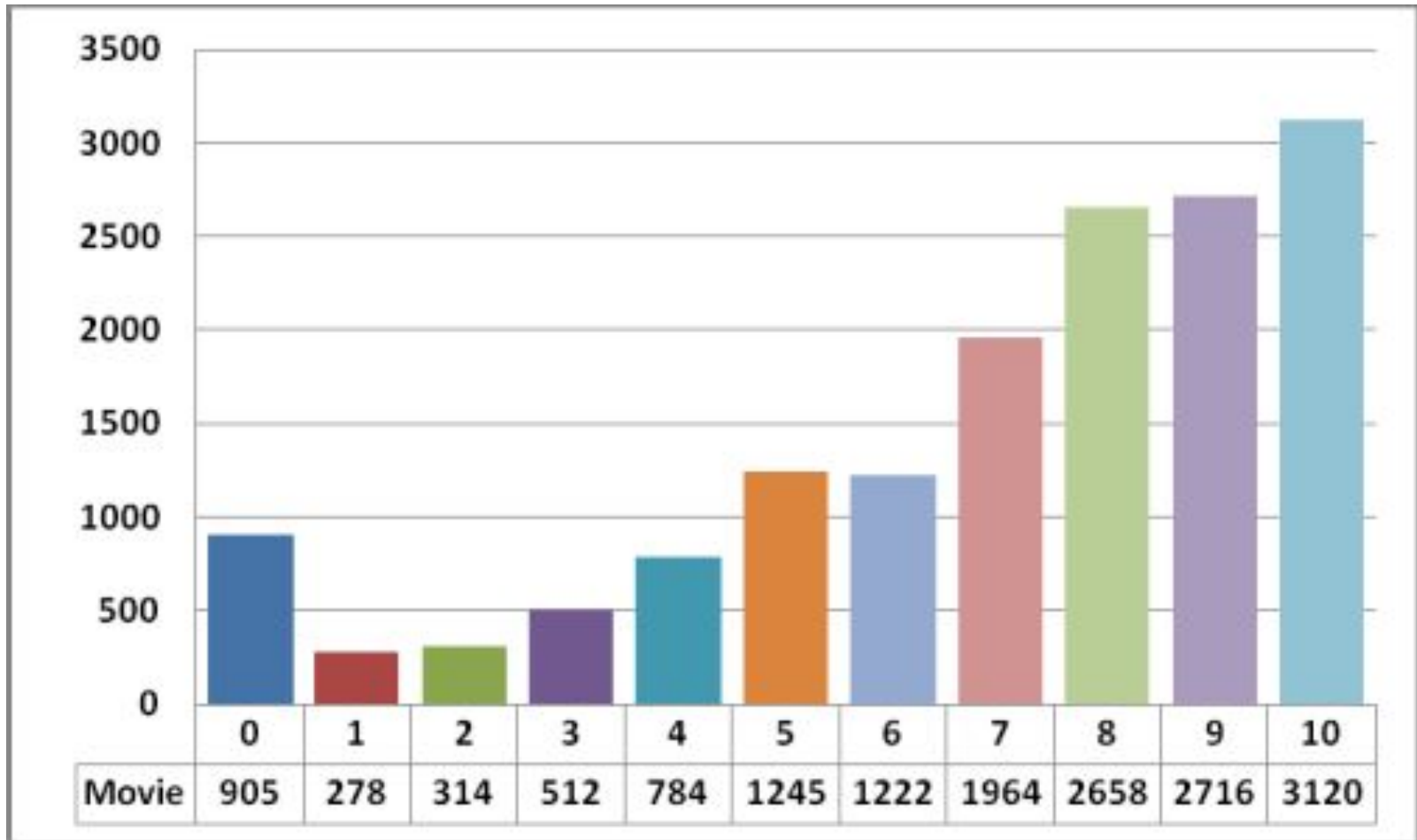
Задачи

- Классификация отзывов на 2 класса
– хорошо/плохо
- Классификация отзывов на три класса
– хорошо/нормально/плохо
- Классификация отзывов на 5 классов

Обучающие коллекции

- Для каждой области была сформирована обучающая коллекция
 - 15,718 отзывов о фильмах с портала Имхонет
 - 24,159 отзывов о книгах с портала Имхонет
 - 10,370 отзывов о камерах с Яндекс.Маркета (есть мета информация о достоинствах, недостатках и полезности отзыва)
- Каждый отзыв с Имхонета имел оценку по 10 балльной шкале, с Яндекс.Маркета по 5 балльной шкале

Распределение отзывов о фильмах



Коллекция для тестирования

- Коллекция сформирована из блогов Livejournal.ru
 - 16,821 текст входит в коллекцию
- Для выгрузки использовался Поиск по Блогам Яндекса
 - 61 запрос о книгах
 - 112 запросов о фильмах
 - 922 запроса о цифровых камерах

*Впечатления от книга «Виктор Пелевин» -
список*

Оценка экспертами

- Тестовая коллекция содержит много «мусора» для нашей задачи
 - Как релевантные так и нерелевантные
 - Может идти описание нескольких сущностей
 - Может не быть никакой оценки
- **Необходимо отобрать** только
 - Субъективные тексты
 - Об одной сущности из выбранных трех областей

Оценка экспертами

- Таким образом было отобрано
 - 275 отзывов о книгах
 - 329 отзывов о фильмах
 - 270 отзывов о цифровых камерах
- Каждый отзыв был размечен двумя экспертами по трем шкалам:
 - 2 класса, положительный или отрицательный
 - 3 класса, добавили «средне», смесь + и –
 - 5 классов, по аналогии с пятью звездами

Статистика по разметке

- Для двух классовой задачи
 - 84% положительных отзывов для камер, 92% для книг и 85% для фильмов
 - Аналогичная ситуация для 3 и 5 классов
- Оценка согласованности экспертов:

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

Согласованность экспертов

Kappa	2 classes	3 classes	5 classes
Movies	0.818	0.615	0.429
Books	0.812	0.674	0.545
Digital Cameras	0.808	0.602	0.398

Доля ОТЗЫВОВ в схеме AND

	2 classes	3 classes	5 classes
Movies	0.948	0.799	0.590
Books	0.967	0.829	0.684
Digital Cameras	0.944	0.766	0.548

Основные метрики

- Precision, recall, F1-measure, accuracy and average Euclidian distance
- Традиционный вариант по классам и макро усреднение

	actual class	
predicted class	tp_x (true positive) Correct result	fp_x (false positive) Unexpected result
	fn_x (false negative) Missing result	tn_x (true negative) Correct absence of result

Основные метрики

$$Precision = \frac{tp_x}{tp_x + fp_x}$$

$$Macro_Prec = \frac{1}{|S|} \cdot \sum_{x \in S} \frac{tp_x}{tp_x + fp_x}$$

$$Recall = \frac{tp_x}{tp_x + fn_x}$$

$$Macro_Rec = \frac{1}{|S|} \cdot \sum_{x \in S} \frac{tp_x}{tp_x + fn_x}$$

Класификация на 2 класса OR

<i>Run_ID</i>	<i>Object</i>	<i>Macro_Prec</i>	<i>Macro_Re</i> <i>c</i>	<i>Macro_F1</i>	<i>Accuracy</i>
xxx-40	book	0.714	0.804	0.747	0.895
xxx-0	book	0.751	0.721	0.735	0.924
xxx-24 (46)	book	0.968	0.630	0.690	0.938*
xxx-19	book	0.790	0.651	0.694	0.931
Baseline	book	0.460	0.500	0.479	0.920
yyy-24	camera	0.918	0.940	0.929*	0.959*
yyy-16	camera	0.944	0.898	0.919	0.956
Baseline	camera	0.426	0.500	0.460	0.852
zzz-23	film	0.776	0.797	0.786	0.881
zzz-9	film	0.706	0.794	0.730	0.812
zzz-14	film	0.743	0.597	0.623	0.860
Baseline	film	0.427	0.500	0.461	0.854

Выводы по проведенному тестированию

- 9 дорожек
 - Во всех лучшим методом был метод SVM
 - Многие участники экспериментировали со списками оценочных слова, правилами, другими методами обучения, но...
- Анализ выдачи SVM:
 - Хорошая (отличная) идея (замысел, сценарий) - в плохих отзывах
 - Имена известных актеров обычно вносят положительный вклад

РОМИП 2012. Задание 2.

Классификация цитат из новостей

- Классификация новостных цитат: прямой или косвенной речи
 - *Секретарь президиума генсовета «Единой России», зампреда Госдумы Сергей Неверов в субботу заявил, что партия **не боится раскола** в связи с появлением в ней разных идеологических платформ .”*
- Нужно классифицировать цитаты на 3 класса: позитивный, негативный или нейтральный (нет мнения)
- Задание похоже
 - NTCIR-6 (2007) – извлечение оценочных предложений из новостных статей
 - Анализ тональности политических цитат (Awadallah et al., 2012)

Коллекции для классификации цитат

- Коллекция для обучения
 - 4260 цитат, размеченных по классам экспертами
 - Сбалансированная коллекция: 41% of негативные, 32% позитивные, 27% нейтральные
- Тестовая коллекция содержит 124, 647 цитат
 - Случайные доля 5500 цитат для оценки систем

Результаты

<i>Run_ID</i>	<i>P</i>	<i>R</i>	<i>F</i>	<i>Accuracy</i>
xxx-4	0.626	0.616	0.621	0.616
xxx-11	0.606	0.579	0.592	0.571
xxx-15	0.563	0.560	0.562	0.582
Baseline	0.138	0.333	0.195	0.413

- Лидеры в этой задаче инженерно-лингвистические подходы (словарь+правила)
 - Широкая тематика цитат
 - Отсутствие большой обучающей коллекции

Оценочная лексика vs. Машинное обучение

- Машинное обучение требует больших размеченных коллекций, и дает относительно высокие результаты, если такие коллекции имеются
- Машинное обучение неустойчиво к переносу на другие предметные области, во времени
- Результаты машинного обучения плохо объяснимы

SentiRuEval 2014-2015:

Анализ репутации по твитам

- «Репутационный» твит может выразить
 - положительное или отрицательное отношение к компании
 - сообщать положительный или отрицательный факт
- Обучающая выборка
 - 5000 твитов про банки и 5000 твитов про ТКК
- Участники
 - 10 участников
 - 19 прогонов

Пример твита и формат разметки

- `<table name="bank">`
- `<column name="id">71</column>`
- `<column name="twitid">492547326574360000</column>`
- `<column name="text">Сбербанк России не будет работать в
Крыму и Севастополе </column>`
- `<column name="sberbank">0</column>`
- `<column name="vtb">NULL</column>`
- `<column name="gazprom">NULL</column>`
- `<column name="alfabank">NULL</column>`
- `<column name="bankmoskvy">NULL</column>`
- `<column name="raiffeisen">NULL</column>`
- `<column name="uralsib">NULL</column>`
- `<column name="rshb">NULL</column>`

Проблема разметки

- Несовпадение в предоставленных оценках
 - *я сегодня ходил в сбербанк за картой, там оч милая девушка работала*
- Тестовая выборка размечалась голосованием трех разметчиков
 - Были включены только те твиты, в которых более двух из трех участников разметили твит одинаково
- Размеры тестовых выборок
 - Банки – 4549 твитов
 - ТКК – 3845 твитов

Разметка тональности твитов

[Инструкции по разметке](#)

[Разметить текст](#)

[О проекте](#)

Выберете отношение автора твита к указанным организациям:

@svintuss @MegaFonHelp @ru_mts нормально))) Т.е. заблокировали, а ты потом ищи-свищи что это и из-за чего...

МТС

Негативное Позитивное Нейтральное Спам Содержит обе эмоции

Мегафон

Негативное Позитивное Нейтральное Спам Содержит обе эмоции

[Далее](#)

Положительная тональность – если сообщается положительное отношение автора к организации, или факт, который свидетельствует об успехах организации (увеличение прибыли, увеличение числа клиентов).

Отрицательная тональность – если сообщается отрицательное отношение автора к организации, или факт, который свидетельствует о проблемах организации (снижение прибыли, уменьшение числа клиентов).

нейтральными – факты, которые относятся к стандартной деятельности организаций.

Результаты

- Ручная разметка участника для ТКК
 - Macro-F – 0.703
 - Абсолютный возможный максимум качества для систем
- Лучшие результаты участников далеки от экспертных
 - Банки – 0.36
 - ТКК – 0.488

Наиболее трудные твиты: ошиблись все участники

1. Проблемы из-за небольшой обучающей выборки
 - Самый **безалаберный** банк по отношению к клиентам - Сбербанк
 - В столице произошло дерзкое **ограбление** Сбербанка
 - **Гребаный** сбербанк

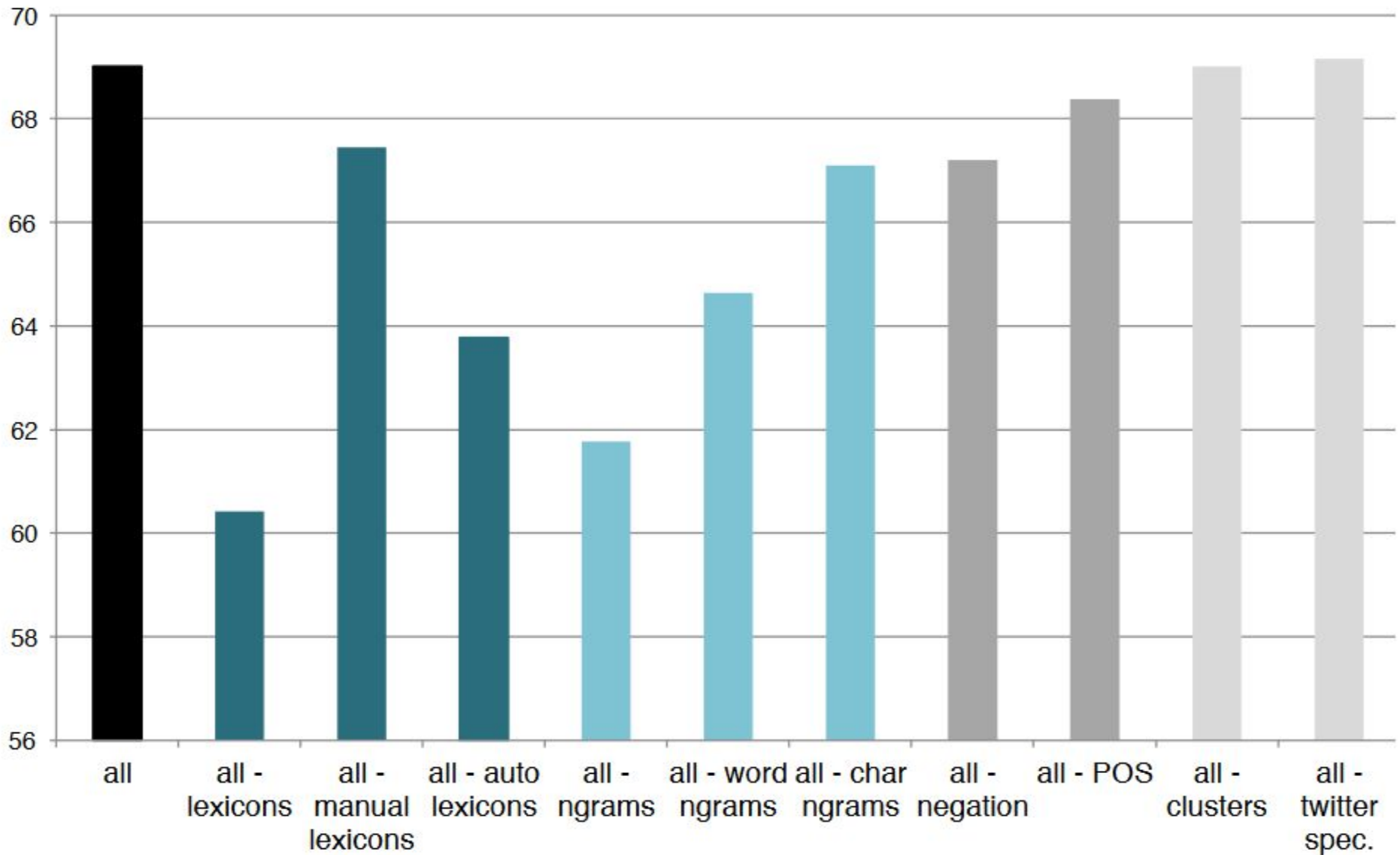
 2. Действительно сложные твиты: ирония
 - Сбербанк России – лучший в мире производитель пластиковых карточек для отскабливания льда от автомобиля
 - Нормально @sberbank зарабатывает - размен 5% от суммы
- Твитов первой группы значительно больше

Новые подходы: комбинирование методов

- Предварительное создание оценочных словарей
 - Вручную
 - Автоматически по большим текстovým коллекциям
- Признаки для машинного обучения
 - Позитивное слово из словаря
 - Негативное слово из словаря
 - Общая количество оценок в твите и др.
 - Максимальный позитивный и негативный вес слова в твите
 - Средняя сумма весов слов в твите
 - => Снижение зависимости от обучающей выборки



Анализ тональности твитов (SemEval 2013)



Другой подход к преодолению ограниченности обучающей выборки

- Обучение без учителя (unsupervised learning):
представление слов как точек в вещественном пространстве низкой размерности на основе сходства контекстов слов (word embedding)
- Используется большой корпус текстов

new		1		2	6			9		3	...
old	1	1			2	1		4		2	...
good	1		6	3		1		7	1		...
bad	2	1	4			2				3	...
...											...



new	-0.03	0.5	0
old	-0.04	0.3	0
good	1.4	0	2.5
bad	1.3	0	3.6
...			

Результаты SemEval-2016

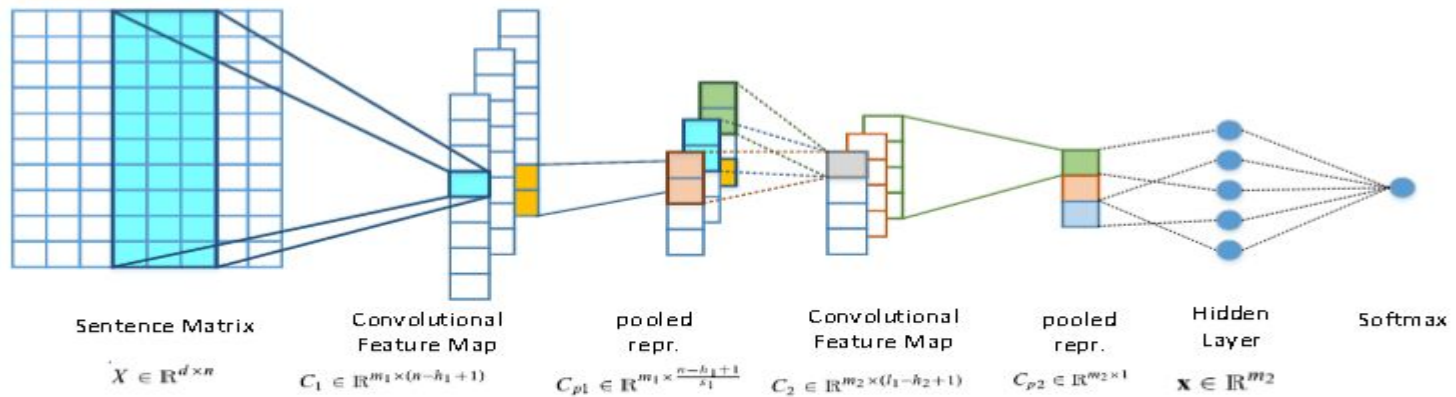
New Approaches as winners

- SemEval Twitter polarity classification results (Nakov et al., 2016) – 34 teams
- Out of 10 top-ranked systems,
 - 7 teams used word embeddings
 - 5 teams used deep neural networks

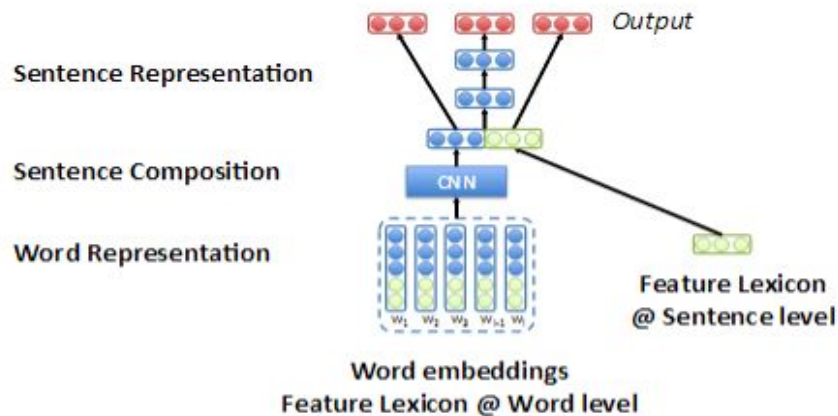
#	System	F_1^{PN}	ρ^{PN}	Acc
1	SwissCheese	0.633 ₁	0.667 ₂	0.646 ₁
2	SENSEI-LIF	0.630 ₂	0.670 ₁	0.617 ₇
3	UNIMELB	0.617 ₃	0.641 ₅	0.616 ₈
4	INESC-ID	0.610 ₄	0.662 ₃	0.600 ₁₀
5	aueb.twitter.sentiment	0.605 ₅	0.644 ₄	0.629 ₆
6	SentiSys	0.598 ₆	0.641 ₅	0.609 ₉
7	I2RNTU	0.596 ₇	0.637 ₇	0.593 ₁₂
8	INSIGHT-1	0.593 ₈	0.616 ₁₁	0.635 ₅
9	TwISE	0.586 ₉	0.598 ₁₆	0.528 ₂₄
10	ECNU (*)	0.585 ₁₀	0.617 ₁₀	0.571 ₁₆

Winning Neural Networks: very different architectures

Swiss Cheese – the first place – convolutional NN



SENSEI-LIF – the second place – another convolutional NN



SentiRuEval: Лучший подход по классификации твитов в 2016 году

- Первый подход
 - Кластеризация слов: нейронные сети и word2vec по большой текстовой коллекции постов и комментариев в социальной сети
- Второй-четвертый подходы
 - использование словарей для порождения признаков

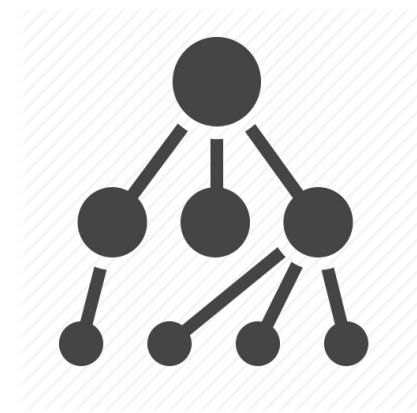
Заключение

- Автоматический анализ тональности текстов – активно развивающееся направление
 - Различные типы задач
- Методы для решения многих задач
 - Машинное обучение
 - Работает лучше, когда есть большая обучающая выборка
 - Проблемы с переносимостью на другую область, ухудшение качества со временем
 - Лингвистико-инженерный метод
 - Словари и правила
 - Комбинированный подход
 - Машинное обучение, использующее словари
 - Общей оценочной лексики
 - Разговорной лексики, извлеченной из социальных сетей
 - Использование словарей позволяет снизить зависимость от недостаточной обучающей выборки

Словарь РуСентиЛекс

РусентиЛекс: источники-1

- Списки оценочных слов из нескольких проектов по анализу тональности в конкретных предметных областях (инженерный подход)
- Автоматическое сопоставление с тезаурусом русского языка РуТез, нахождение соответствующих понятий
 - Возможность просмотра синонимов,
 - родовых и видовых понятий
- Далее просмотр и внесение в словарь экспертом



Оценочные выражения в тезаурусе RuТез

[Главная](#)

[О проекте](#)

[Справка](#)

[А](#) [Б](#) [В](#) [Г](#) [Д](#) [Е](#) [Ж](#) [З](#) [И](#) [Й](#) [К](#) [Л](#) [М](#) [Н](#) [О](#) [П](#) [Р](#) [С](#) [Т](#) [У](#) [Ф](#) [Х](#) [Ц](#) [Ч](#) [Ш](#) [Щ](#) [Ы](#) [Э](#) [Ю](#) [Я](#)

БЕССМЫСЛЕННЫЙ

([БЕССМЫСЛЕННОСТЬ](#), [БЕССМЫСЛЕННЫЙ](#), [ЛИШЕННЫЙ СМЫСЛА](#), [МАЛООСМЫСЛЕННОСТЬ](#), [МАЛООСМЫСЛЕННЫЙ](#), [НЕОСМЫСЛЕННЫЙ](#))

ВЫШЕ [НЕЛЕПЫЙ](#)

ЦЕЛОЕ [БЕССМЫСЛИЦА](#), [БЕССМЫСЛЕННОЕ ВЫСКАЗЫВАНИЕ](#)

НИЖЕ [АБСУРДНЫЙ \(НЕЛЕПЫЙ, БЕССМЫСЛЕННЫЙ\)](#)

АССОЦ [РАЗУМНЫЙ](#), [ЗДРАВΟΣМЫСЛЯЩИЙ](#)

РуСентиЛекс: источники-2

- Извлечение слов с негативными и позитивными ассоциациями (из новостей):
 - безработица, инфляция, и др.
- Применение шаблонов (2 млн. документов)
 - Позитивные (18 шаблонов)
 - Защищать вин.
 - Бороться за вин.
 - Охранять вин.
 - Ценить вин.
 - Негативные (36 шаблонов)
 - Противостоять вин.
 - Бороться с творит.
 - Обвинить в творит.
 - Не допустить вин.

Извлеченные слова с коннотациями

- Список слов с негативными коннотациями
 - коррупция 14869
 - терроризм 6876
 - преступность 2631
 - преступление 2618
 - экстремизм 2483
 - оборот 2166
 - распространение 1939
 - последствие 1930
 - экономический преступление 1791
 - ...
- Добавлены экспертом в словарь

РуСентиЛекс: источники-3

- Извлечение оценочных слов из Твиттера
- Модель извлечения оценочных слов,
 - основана на машинном обучении
 - несколько текстовых коллекций
 - Слова с высоким содержанием оценочной лексики
 - Низким содержанием оценочной лексики
 - Новости
- Просмотр и добавления в словарь экспертом

Новые слова с высоким весом, извлеченные из Твиттера (без мата)

- Офигенный
- Упоротый
- Афигенный
- Дибильный
- Дерьмовый
- Стремный
- Класный
- Халявный
- Симпотичный
- Бесбашенный
- Безбашенный
- Дебильный...

РусСентиЛекс: структура

- Текстовый файл
 - слово или фраза,
 - часть речи,
 - слово или фраза, в которой каждое слово стоит в лемматизированной форме (*бросаться деньги*)
 - тональность: позитивная (positive), негативная (negative), нейтральная (neutral) или двойная, positive/negative (*верноподанный*).
 - источник тональности (явно выраженная оценка, эмоция, или факт);
 - Для многозначных слов указание на понятия тезауруса РусТез, если слово в разных значениях имеет разные оценочные характеристике

Примеры описания в RuСентиЛекс

- **Пресный**

- пресный, Adj, пресный, negative, feeling, "НЕВКУСНЫЙ"
- пресный, Adj, пресный, negative, opinion, "НЕИНТЕРЕСНЫЙ"
- пресный, Adj, пресный, positive, fact, "ПРЕСНАЯ ВОДА"

- **Грязный, грязь**

- грязный, Adj, грязный, negative, opinion
- грязь, Noun, грязь, negative, opinion

Словарь RuSentiFrames

Оценочные фреймы

- Описывают позитивные и негативные отношения между участниками
- Позитивные и негативные эффекты на участников
- Позитивное/негативное настроение
- Отношение автора к участникам

Примеры фреймов

- Хотеть (want) (X, Y)
- Polarity (X, Y, pos, 1)

- Добиться (Achieve) (X, Y)
- Polarity (A1, A2, pos, 1)
- Effect (A2, +, 0.7)
- State (A1, pos, 1)
- Effect (A1, +, 1)
- Polarity (author, A1, pos, 0.7)
- Polarity (author, A2, pos, 0.7)

- Изоляция (Isolation) (A1, from A2)
- Polarity (A1, A2, pos, 0.7)
- Effect (A1, -, 1)
- State (A1, neg, 1)

Правила вывода

- $\text{polarity}(X, Y, \text{pos}) \ \& \ \text{effect}(Y, Z, +) \Rightarrow \text{polarity}(X, Z, \text{pos})$
- $\text{polarity}(X, Y, \text{neg}) \ \& \ \text{effect}(Y, Z, +) \Rightarrow \text{polarity}(X, Z, \text{neg})$
- $\text{polarity}(X, Z, \text{neg}) \ \& \ \text{effect}(Y, Z, +) \Rightarrow \text{polarity}(X, Y, \text{neg})$
- $\text{polarity}(X, Z, \text{neg}) \ \& \ \text{effect}(Y, Z, -) \Rightarrow \text{polarity}(X, Y, \text{pos})$

Извлечение отношений

- Янукович призвал инициировать создание **специальной комиссии**
- Yanukovich called to initiate the creation of a **special commission**
- Евросоюз хотел добиться изоляции **Ирана**
- The European Union wanted to isolate **Iran**

Отношения **Финляндии** и **Швеции** можно считать хорошими. Ведь входили же страны в состав одного королевства до 1809 года. Страны объединяет также и то, что они не входят в **НАТО**, но являются партнерами альянса. Кроме того, **Финляндия** и **Швеция** укрепляют двустороннее сотрудничество в области обороны.

Несмотря на все это, в ходе обсуждений в **Култаранте** возникли разногласия. Бывший министр обороны Швеции **Карин Энстрём** возмутилась тем, что президент **Финляндии** встречается в следующем месяце с президентом **России** в **Хельсинки**. По мнению второго шведского участника дискуссий, **Владимиру Путину** не были бы рады в **Швеции**.