Методы распознавания сарказма в тексте

Кибитова Валерия

Спецсеминар "Алегбра над алгоритмами и эвристический поиск закономерностей"

11 апреля 2016 г.

Определение сарказма

Сарказм – это способ использования слов таким образом, что буквальное и истинное значение текста являются противоположными. Как правило, используется с целью обидеть кого-то или посмеяться над кем-то.

Постановка задачи

Дан текст, необходимо определить присутствует или нет в нем сарказм.

Основные методы решения

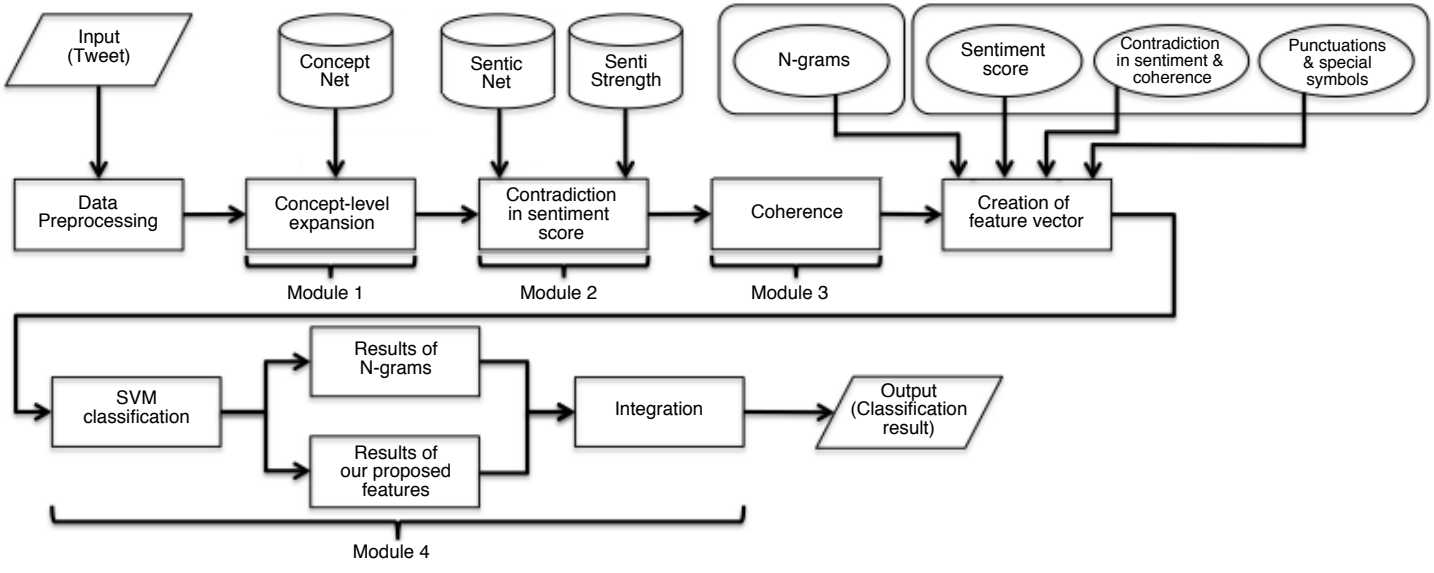
- Методы, основанные на машинном обучении
- Методы, основанные на лингвистической структуре сарказма

Recognition of Sarcasm in Tweets Based on Concept Level Sentiment Analysis and Supervised Learning Approaches

Структура алгоритма

Данный алгоритм состоит из 4 частей:

- Анализ тональности
- Анализ концептов
- Идентификация согласованности
- Классификация



Для оценки тональности текста использовалась:

- SentiStrength – предоставляет оценку тональности для каждого слова в пределах от $[-5, 5]$
- SenticNet – предоставляет оценку тональности для каждого слова в пределах от $[-1, 1]$
- ConceptNet – позволяет получить список концептов, связанных с данным словом

Оценка эмоциональной окраски слова

$$w\_score(w) = \begin{cases} polarity\_score(w), & if\ w \in SS\ or\ SN \\ average\_polarity\_score(w), & if\ w \in SS\ and\ SN \\ \frac{1}{|C|} \sum_{c \in C} polarity\_score(c), & otherwise \end{cases}$$

$$sum\_pos\_score = \sum_{pos\_w \in TW} w\_score(pos\_w)$$

$$sum\_neg\_score = \sum_{neg\_w \in TW} w\_score(neg\_w)$$

Предложения $s1$ и $s2$ согласованы:

если существует такое слово $w1$ в предложении $s1$ и слово $w2$ в предложении $s2$, что выполняется одно из условий:

- w1 и w2 идентичные местоимения
- w1 и w2 идентичны как строки(стоп-слова не учитываются)
- w2 начинается с *the*
- w2 начинается с *this*, *that*, *these*, *those*
- w1 и w2 именованные сущности

Recognition of Sarcasm in Tweets Based on Concept Level Sentiment Analysis and Supervised Learning Approaches

Используемые признаки:

- N-граммы($N = 1, 2, 3$)
- Два бинарных признака: *contra* и *contra* + *coher*, которые определяют присутсвует ли в тексте противоречие тональностей.
- Признаки, определяющие, степень позитивности и негативности твита:
  $pos\_low\ if\ sum\_pos\_score <= 1$
  $pos\_medium\ if\ 1 < sum\_pos\_score <= 2$
  $pos\_high\ if\ sum\_pos\_score > 2$

Recognition of Sarcasm in Tweets Based on Concept Level Sentiment Analysis and Supervised Learning Approaches

Используемые признаки:

- Число смайликов
- Число последовательностей, в которых пунктуационные символы повторяются
- Число последовательностей, в которых буквы повторяются
- Число слов, написанных большими буквами
- Число слов сленговых слов и слов-усилителей
- Число восклицательных знаков
- Число идиом

Recognition of Sarcasm in Tweets Based on Concept Level Sentiment Analysis and Supervised Learning Approaches

Результаты:

Table 1: The result of contradiction in sentiment score approach

Methods	Recall	Precision	F-measure	Accuracy
Contradiction in sentiment score (Baseline 1)	0.55	0.56	0.56	57.14%

Table 2: The result of SVM classification based on various features

Methods	Recall	Precision	F-measure	Accuracy
Our proposed features	0.64	0.63	0.63	63.42%
Uni-gram features (Baseline 2)	0.72	0.73	0.73	73.81%
Uni-gram, bi-gram and tri-gram features (Baseline 2)	**0.76**	**0.76**	**0.76**	**76.40%**

Table 3: The result of marjority vote and margin based SVM classification

Methods	Recall	Precision	F-measure	Accuracy
uni-gram and contradiction	0.72	0.72	0.72	72.83%
uni-gram and sentiment score	0.75	0.75	0.75	75.64%
uni-gram and punctuations + special symbols	0.72	0.73	0.73	73.91%
uni-gram and our proposed features without coherence	0.75	0.75	0.75	75.72%
uni-gram and our proposed features without concept level knowledge generation	0.74	0.75	0.75	75.48%
uni-gram and all our proposed features	0.76	0.77	0.76	76.35%
uni-gram, bi-gram, tri-gram and all our proposed features	**0.79**	**0.78**	**0.79**	**79.43%**

Sarcasm Detection on Twitter: A Behavioral Modeling Approach

Ставится следующая задача:

Дан твит t от пользователя U, вместе с историей твитера пользователя. Решением задачи обнаружения сарказма является автоматическое обнаружение является ли твит саркастичным или нет.

Следующие факторы влияют на саркастичность текста:

- Контраст настроений в тексте
- Когнитвные способности пользователя
- Текущее эмоциональное состояние пользователя
- Грамматические знания пользователя
- Нетрадиционный стиль написания

$A = \{affect(w)|w \in t\}$
$S = \{sentiment(w)|w \in t\}$
$\Delta affect = max(A) - min(A)$
$\Delta sentiment = max(S) - min(S)$

affect(w) – оценка для слова из Warriner([1-9])
sentiment(w) – оценка для слова из SentiStrength.

Оценка для n-граммов:

$$\frac{POS(b) - NEG(b)}{POS(b) + NEG(b)}$$

Признаки: число положительных n-граммов, число отрицательных n-граммов, сумма оцеок для положительных n-граммов, сумма оценок для отрицательных n-граммов.

Признаки, связанные с распределением длин слов в тексте

$$< E[l_w], med[l_w], mode[l_w], \sigma[l_w], max\{l_w\} >$$

$L = \{l_i\}$ – распределение длин слов в твите

$$JS(D1||D2) = \frac{1}{2}KL(D_1||M) + \frac{1}{2}KL(D_2||M)$$

$$M = \frac{D_1 + D_2}{2}$$

$$KL(T_1||T_2) = \sum ln(\frac{T_1(i)}{T_2(i)})T_1(i)$$

D1 – Распределение длин слов в текущем твите
D2 – Распределение длин слов в в предыдущих твитах пользователя

Признаки, связанные настроением пользователя

Все предыдущие твиты пользователя разделяются на корзины состоящие из n твитов($n \in \{1, 2, 5, 10, 20, 40, 80\}$).

$$< \overset{+}{\sum}, \overset{-}{\sum}, P, max(\overset{+}{\sum}, \overset{-}{\sum}) >$$

$$\overset{+}{\sum} = \sum pos(t)$$
$$\overset{-}{\sum} = \sum neg(t)$$

$$< n_+, n_-, n_0, Q, max(n_+, n_-, n_0) >$$

$n_{+(-)}$ – число положительных(отрицательных) твитов

$$< E[ad_w], med[ad_w], mode[ad_w], \sigma[ad_w], max\{ad_w\} >$$
$$< E[sd_w], med[sd_w], mode[sd_w], \sigma[sd_w], max\{sd_w\} >$$

AD – распределение *affect_score* в твите
SD – распределение *sentiment_score*

Признаки, которые использовались для оценки настроения пользователя:

- Сравнение распределений оценок в данном твите с распределением оценок в предыдущих твитах.
- Вероятность появления каждой оценки *sentiment_score* в твите.
- Оценка вероятности написания твита в данный промежуток времени.
- Промежуток времени между предыдущим твитом и текущим.
- Присутствие бранных слов.

Sarcasm Detection on Twitter: A Behavioral Modeling Approach

Признаки связанные с оценкой знания пользователем используемого языка:

- Общее число, написанных пользователем слов; число различных слов, написанных пользователем; отношение различных слов к общему числу.
- Вероятность появления каждой части речи в твите(TweetNLP).
- Правильное использование *your*(*you're*)and *its*(*itis*).
- Число предыдущих хештогов *#sarcasm*, использованных пользователем.

Признаки, определяющие опытность пользователя:

- Число дней с момента регистрации.
- Число твитов, среднее число ежедневных твитов.
- Число ретвитов; присутствие слов, содержащих цифры; содержащие только согласные; процент слов, которые содержат только слова, которые встречаются в словаре.
- Число подписчиков и подписок.

Sarcasm Detection on Twitter: A Behavioral Modeling Approach

Признаки связанные со способом написания текста:

- Присутствие повторяющихся символов(3 или больше) во всех словах и в словах, выражающих эмоции.
- Число символов, число различных символов, число слов с большой буквы.
- Распределение пунктуации в текущем твите.
- Теги частей речи первых трех слов в твите.
- Позиция первого эмоционального слова в твите.
- Число существительных, глаголов, прилагательных и наречий, используемых в твите, число стоп-слов в твите.
- Лексическая плотность, число используемых слов-усилителей.

Technique	**Dataset Distribution**					
	1:1		20:80		10:90	
	Acc.	AUC	Acc.	AUC	Acc.	AUC
SCUBA	83.46	0.83	88.10	0.76	92.24	0.60
Contrast Approach	56.50	0.56	78.98	0.57	86.59	0.57
SCUBA++	86.08	0.86	89.81	0.80	92.94	0.70
Hybrid Approach	77.26	0.77	78.40	0.75	83.87	0.67
SCUBA - `#sarcasm`	83.41	0.83	87.53	0.74	91.87	0.63
n-gram Classifier	78.56	0.78	81.63	0.76	87.89	0.65
Majority Classifier	50.00	0.50	80.00	0.50	90.00	0.50
Random Classifier	49.17	0.50	50.41	0.50	49.78	0.50

Modelling Sarcasm in Twitter, a Novel Approach

Основная идея:

Избегать слов слов или паттернов слов, как признаков, формировать признаковое пространство, на основе структуры предложений.

Набор данных:

Данные твиттера, которые одинакова разделены на 6 тем: сарказм, ирония, образование. юмор, политика и новости.
American National Corpus(ANC) – содержит частоту встречаемости слов, используемых в письменном и устном языке.

Используемые признаки

- Частота: средняя частота слов в твите, самое редкое слово(его частота), разница первых двух признаков
- Стиль написания: средняя частота слов, написанных в письменном стиле, средняя частота слов в устном стиле, разница первых двух признаков.
- Структура предложения: число символов, из которых состоит текст; число слов в тексте; средняя длина слов в тексте; число глаголов, существительных, прилагательных и наречий; доля глаголов, существительных, наречий и прилагательных в тексте; число всех пунктационных символов в тексте; признаки, связанные с количесством каждого отдельного пунктуционного символа; наличие слов, обозначающих смех; число смайликов;

Используемые признаки

- Интенсивность: суммарная интенсивность прилагательных(наречий), средняя интенсивность, максимальная интенсивность, разность между максимальной и средней интенсивностью
- Синонимы:

$$sl_{w_i} = |syn_i : f(syn_i) < f(w_i)|$$

$$mean\{sl_{w_i}\}$$

$$wls_t = max_{w_i}\{|syn_i : f(syn_i) < f(w_i)|\}$$

$$wgs_t = max_{w_i}\{|syn_i : f(syn_i) > f(w_i)|\}$$

$$sg_{w_i} = |syn_{w_i} : f(syn_{w_i}) > f(w_i)|$$

$$mean\{sg_{w_i}\}$$

$$abs(wls_t - mean\{sl_{w_i}\})\ abs(wgs_t - mean\{sg_{w_i}\})$$

Используемые признаки

- Неоднозначность: среднее числа значений слов в тексте; максимальное число значение слова; разность предыдущих двух
- Эмоциональная окраска(SentiWordNet): сумма всех положительных оценок; сумма всех отрицательных оценок; разность между предыдущими двумя признаками; разность между максимальной положительной оценкой и средней, разность между минимальной негативной оценкой и средней

Результаты:

	Prec.	Recall	F1
Education	.87	.90	.88
Humour	.88	.87	.88
Irony	.62	.62	.62
Newspaper	.98	.96	.97
Politics	.90	.90	.90

Sarcasm as Contrast between a Positive Sentiment and Negative Situation

Структура предложения, содержащего сарказм:

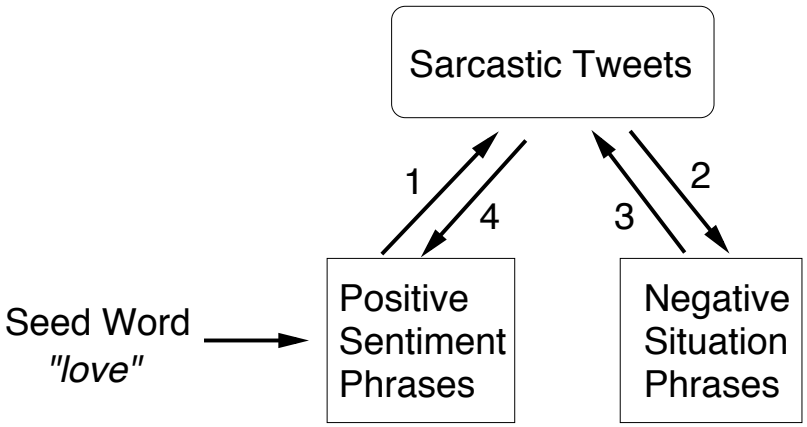
$$[+\mathit{VERB\ PHRASE}][-\mathit{SITUATION\ PHRASE}]$$

Пример: "I love waiting forever for the doctor"

Ключевая задача:

Идентифицировать стереотипные негативные ситуации или состояния

Sarcasm as Contrast between a Positive Sentiment and Negative Situation
Алгоритм



$$\frac{|follows(-candidate, +sentiment)\&sarcasm|}{|follows(-candidate, +sentiment)|}$$

$$\frac{|precedes(+candidate, -situation)\&sarcasm|}{|follows(+candidate, -situation)|}$$

$$\frac{|near(+candidatePRED, -situation)\&sarcasm|}{|near(+candidatePRED, -situation)|}$$

Sarcasm as Contrast between a Positive Sentiment and Negative Situation

Результаты

System	Recall	Precision	F score
Supervised SVM Classifiers			
1grams	.35	.64	.46
1+2grams	.39	.64	.48
Positive Sentiment Only			
Liu05	.77	.34	.47
MPQA05	**.78**	.30	.43
AFINN11	.75	.32	.44
Negative Sentiment Only			
Liu05	.26	.23	.24
MPQA05	.34	.24	.28
AFINN11	.24	.22	.23
Positive and Negative Sentiment, Unordered			
Liu05	.19	.37	.25
MPQA05	.27	.30	.29
AFINN11	.17	.30	.22
Positive and Negative Sentiment, Ordered			
Liu05	.09	.40	.14
MPQA05	.13	.30	.18
AFINN11	.09	.35	.14
Our Bootstrapped Lexicons			
Positive VPs	.28	.45	.35
Negative Situations	.29	.38	.33
Contrast(+VPs, –Situations), Unordered	.11	.56	.18
Contrast(+VPs, –Situations), Ordered	.09	**.70**	.15
& Contrast(+Preds, –Situations)	.13	.63	.22
Our Bootstrapped Lexicons ∪ SVM Classifier			
Contrast(+VPs, –Situations), Ordered	.42	.63	.50
& Contrast(+Preds, –Situations)	.44	.62	**.51**

Parsing-based Sarcasm Sentiment Recognition in Twitter Data

Схема распознавания сарказма состоит из 2-х частей:

- Идентификации сарказма на данных твиттера основанная на парсинге. Распознает сарказм в случаях, сочетания положительной оценки и негативной ситуации и негативной оценки и положительной ситуации.(PBLGA)
- Алгоритм, который распознает сарказм в твитах, которые начинаются с междометий.(IWS)

Parsing-based Sarcasm Sentiment Recognition in Twitter Data
Алгоритм:

PBLGA:

$SF = \emptyset, sf = \emptyset, PSF = \emptyset, NSF = \emptyset, psf = \emptyset, nsf = \emptyset$
for T in C **do**
    $k = find\_parse(T)$
    $PF = PF \cup k$
end for
for TWP in PF **do**
    $k = find\_subset(TWP)$
    if $k == NP||ADVP||(NP + VP)$ **then**
        $SF = SF \cup k$
    else if $k == VP||(ADVP + VP)||(VP + ADVP)||(ADJP + VP)||$
$(VP+NP)||(VP+ADVP+ADJP)||(VP+ADJP+NP)||(ADVP+ADJP+NP)$ **then**
        $sf = sf \cup k$
    end if
end for

Parsing-based Sarcasm Sentiment Recognition in Twitter Data

Алгоритм:

```
for P in SF do
    SC = sentiment_score(P)
    if SC > 0.0 then
        PSF = PSF ∪ P
    else if SC < 0.0 then
        NSF = NSF ∪ P
    else
        Neutral Sentiment Phrase
    end if
end for
for P in sf do
    SC = sentiment_score(P)
    if SC > 0.0 then
        psf = psf ∪ P
    else if SC < 0.0 then
        nsf = nsf ∪ P
    else
        Neutral Situation Phrase
    end if
end for
```

$$PR = \frac{PWP}{TWP}$$

$$NR = \frac{NWP}{TWP}$$

$$SentimentScore = PR - NR$$

Методы, основанные на лингвистической структуре сарказма
Parsing-based Sarcasm Sentiment Recognition in Twitter Data

IWS:

for T in C **do**
  $k = find\_postag(T)$
  $TF = TF \cup k$
end for
for TWT in TF **do**
  $t = find\_subset(TWT)$
  $FT = find\_first\_tag(t)$
  $INT = find\_immediate\_next\_tag(t)$
  $NT = find\_next\_tag(t)$
  if $(FT == UH)\&\&(INT == ADJ||ADV)$ **then**
    Tweet is sarcastic
  else if (FT==UH)&&(NT==(ADV+ADJ)) || (ADJ +N) || (ADV +V)) **then**
    Tweet is sarcastic
  else if $FT \neq UH$ **then**
    Invalid tweet.
  else
    Tweet is not sarcastic
  end if
end for

Примеры:

"Wow, that's a huge discount, I'm not buying anything!!"

"Aha, great night"

Методы, основанные на лингвистической структуре сарказма
Parsing-based Sarcasm Sentiment Recognition in Twitter Data

Approach	*Precision*	*Recall*	$F-score$
Barbieri *et al.* system	0.88	0.87	0.88
Tungthamthiti *et al.* system	0.76	0.76	0.76
Riloff *et al.* system with positive verb	0.28	0.45	0.35
with negative situation	0.29	0.38	0.33
Contrast (+VPs, -situation)unordered	0.11	0.56	0.18
Contrast (+VPs, -situation)ordered	0.09	0.70	0.15
Contrast (+preds, -situation)	0.13	0.63	0.22
Liebrecht *et al.* system with 50/50	0.75	-	-
with 25/75 neg, pos ratio	0.56	-	-
PBLGA with sar tweets	**0.89**	0.81	0.84
PBLGA without sar tweets	0.64	0.75	0.69
IWS sarcastic tweets	0.85	**0.96**	**0.90**
IWS without sarcastic tweets	0.77	0.73	0.74