Affect Detection in Texts (ч2)

Федотова Мария

Recognizing emotions in texts

Проблема:

‘Секс’ ? ‘Университет’ ?

Recognizing emotions in texts

Проблема:

Что делать со словами, которые сами по себе эмоционально не окрашены? (aka generic words)

Recognizing emotions in texts

Решения:

WordNet-Affect Presence	LSA Single Word	LSA Emotional Synset	LSA All Emotion Words	NB Trained On Blogs

Knowledge-based Emotion Classification

Corpus-based Emotion Classification

WordNet-Affect Presence. Baseline

A-Labels	*Valence*	*Examples of word senses*
JOY	positive	noun **joy#1**, adjective **elated#2**, verb **gladden#2**, adverb **gleefully#1**
LOVE	positive	noun **love#1**, adjective **loving#1**, verb **love#1**, adverb **fondly#1**
APPREHENSION	negative	noun **apprehension#1**, adjective **apprehensive#3**, adverb **anxiously#1**
SADNESS	negative	noun **sadness#1**, adjective **unhappy#1**, verb **sadden#1**, adverb **deplorably#1**
SURPRISE	ambiguous	noun **surprise#1**, adjective **surprised#1**, verb **surprise#1**
APATHY	neutral	noun **apathy#1**, adjective **apathetic#1**, adverb **apathetically#1**
NEGATIVE-FEAR	negative	noun **scare#2**, adjective **afraid#1**, verb **frighten#1**, adverb **horryfyingly#1**
POSITIVE-FEAR	positive	noun **frisson#1**
POSITIVE-EXPECTATION	positive	noun **anticipation#1**, adjective **cliff-hanging#1**, verb **anticipate#1**

Входящий документ?

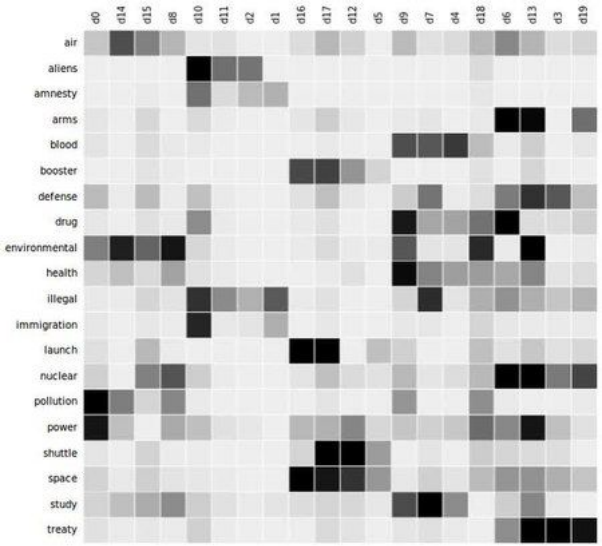
WordNet-Affect Presence. Baseline

A-Labels	*Valence*	*Examples of word senses*
JOY	positive	noun **joy#1**, adjective **elated#2**, verb **gladden#2**, adverb **gleefully#1**
LOVE	positive	noun **love#1**, adjective **loving#1**, verb **love#1**, adverb **fondly#1**
APPREHENSION	negative	noun **apprehension#1**, adjective **apprehensive#3**, adverb **anxiously#1**
SADNESS	negative	noun **sadness#1**, adjective **unhappy#1**, verb **sadden#1**, adverb **deplorably#1**
SURPRISE	ambiguous	noun **surprise#1**, adjective **surprised#1**, verb **surprise#1**
APATHY	neutral	noun **apathy#1**, adjective **apathetic#1**, adverb **apathetically#1**
NEGATIVE-FEAR	negative	noun **scare#2**, adjective **afraid#1**, verb **frighten#1**, adverb **horryfyingly#1**
POSITIVE-FEAR	positive	noun **frisson#1**
POSITIVE-EXPECTATION	positive	noun **anticipation#1**, adjective **cliff-hanging#1**, verb **anticipate#1**

Входящий документ?

Для 6 базовых эмоций собрали synset’ы слов и просто считают для входящего документа, сколько слов из этих synset’ов в нем встретилось. Слов какой эмоции больше, ту эмоцию и присваиваем

Affective Semantic Similarity



occurrence matrix

LSA →

векторные представления слов

Affective Semantic Similarity

векторные представления слов
усредняются с весами

документ можно представить
одним вектором

Affective Semantic Similarity

векторные представления слов усредняются с весами → документ можно представить одним вектором

✨ И слова, и документы можно представить в виде векторов, которые можно сравнивать друг с другом ✨

Knowledge-based Emotion Classification

LSA Single Word	LSA Emotional Synset	LSA All Emotion Words
Считаем близость (от -1 до 1) между вектором входящего текста и вектором эмоции, в котором по сути одно слово (например, «радость»)	Считаем близость (от -1 до 1) между вектором входящего слова и вектором, в котором усреднены несколько слов – само слово, обозначающее эмоцию + synset, относящийся к этому слову	Считаем близость (от -1 до 1) между вектором входящего слова и вектором, в котором усреднены несколько слов – само слово, обозначающее эмоцию + synset, относящийся к этому слову + synset'ы слов, к которым относится данная эмоция

Knowledge-based Emotion Classification

A-Labels	*Valence*	*Examples of word senses*
JOY	positive	noun **joy#1**, adjective **elated#2**, verb **gladden#2**, adverb **gleefully#1**
LOVE	positive	noun **love#1**, adjective **loving#1**, verb **love#1**, adverb **fondly#1**
APPREHENSION	negative	noun **apprehension#1**, adjective **apprehensive#3**, adverb **anxiously#1**
SADNESS	negative	noun **sadness#1**, adjective **unhappy#1**, verb **sadden#1**, adverb **deplorably#1**
SURPRISE	ambiguous	noun **surprise#1**, adjective **surprised#1**, verb **surprise#1**
APATHY	neutral	noun **apathy#1**, adjective **apathetic#1**, adverb **apathetically#1**
NEGATIVE-FEAR	negative	noun **scare#2**, adjective **afraid#1**, verb **frighten#1**, adverb **horryfyingly#1**
POSITIVE-FEAR	positive	noun **frisson#1**
POSITIVE-EXPECTATION	positive	noun **anticipation#1**, adjective **cliff-hanging#1**, verb **anticipate#1**

LSA Single Word	LSA Emotional Synset	LSA All Emotion Words
JOY	noun **joy#1**, adjective **elated#2**, verb **gladden#2**, adverb **gleefully#1**	noun **joy#1**, adjective **elated#2**, verb **gladden#2**, adverb **gleefully#1** + [noun **joy#1**, ...]

Corpus-based Emotion Classification

1. Берем посты с LifeJournal. Токенизируем. Оставляем только короткие посты
2. Обучаем на этом простой классификатор (Naïve Bayes classifier), который будет нам предсказывать вероятность того, что текст относится к одной из 6 базовых эмоций
3. ✨✨✨

Emotion	LiveJournal mood	Number of blogposts
ANGER	angry	951
DISGUST	disgusted	72
FEAR	scared	637
JOY	happy	4,856
SADNESS	sad	1,794
SURPRISE	surprised	451

Evaluation on SemEval 2007 task

	Fine	Coarse		
	r	Prec.	Rec.	F1
ANGER				
WordNet-Affect presence	12.08	**33.33**	3.33	6.06
LSA single word	8.32	6.28	63.33	11.43
LSA emotion synset	17.80	7.29	86.67	13.45
LSA all emotion words	5.77	6.20	**88.33**	11.58
NB trained on blogs	**19.78**	13.68	21.67	**16.77**
DISGUST				
WordNet-Affect presence	-1.59	0	0	-
LSA single word	**13.54**	**2.41**	70.59	**4.68**
LSA emotion synset	7.41	1.53	64.71	3.00
LSA all emotion words	8.25	1.98	**94.12**	3.87
NB trained on blogs	4.77	0	0	-
Fear				
WordNet-Affect presence	24.86	**100.00**	1.69	3.33
LSA single word	**29.56**	12.93	**96.61**	**22.80**
LSA emotion synset	18.11	12.44	94.92	22.00
LSA all emotion words	10.28	12.55	86.44	21.91
NB trained on blogs	7.41	16.67	3.39	5.63
JOY				
WordNet-Affect presence	10.32	**50.00**	0.56	1.10
LSA single word	4.92	17.81	47.22	25.88
LSA emotion synset	6.34	19.37	72.22	30.55
LSA all emotion words	7.00	18.60	**90.00**	30.83
NB trained on blogs	**13.81**	22.71	59.44	**32.87**
SADNESS				
WordNet-Affect presence	8.56	**33.33**	3.67	6.61
LSA single word	8.13	13.13	55.05	21.20
LSA emotion synset	13.27	14.35	58.71	**23.06**
LSA all emotion words	10.71	11.69	**87.16**	20.61
NB trained on blogs	**16.01**	20.87	22.02	21.43
SURPRISE				
WordNet-Affect presence	3.06	**13.04**	4.68	6.90
LSA single word	9.71	6.73	67.19	12.23
LSA emotion synset	12.07	7.23	89.06	13.38
LSA all emotion words	**12.35**	7.62	**95.31**	**14.10**
NB trained on blogs	3.08	8.33	1.56	2.63

Future Directions: Humor Recognition

Спойлер:

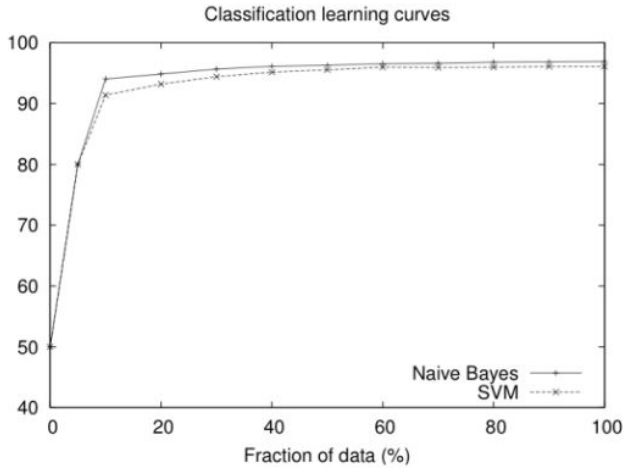
The results obtained in the automatic classification experiments reveal the fact that ***computational approaches represent a viable solution for the task of humor-recognition****, and good performance can be achieved using classification techniques based on stylistic and content features*

Future Directions: Humor Recognition

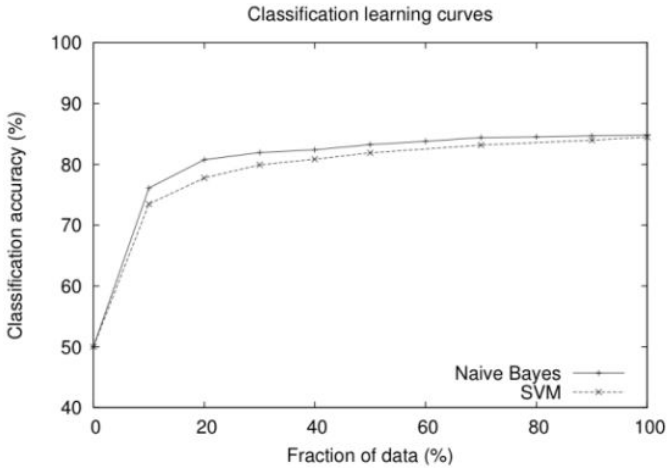
One-liners
Take my advice; I don't use it anyway.
I get enough exercise just pushing my luck.
I just got lost in thought, it was unfamiliar territory.
Beauty is in the eye of the beer holder.
I took an IQ test and the results were negative.
Reuters titles
Trocadero expects tripling of revenues.
Silver fixes at two-month high, but gold lags.
Oil prices slip as refiners shop for bargains.
Japanese prime minister arrives in Mexico.
Chains may raise prices after minimum wage hike.
Proverbs
Creativity is more important than knowledge.
Beauty is in the eye of the beholder.
I believe no tales from an enemy's tongue.
Do not look at the coat, but at what is under the coat.
A man is known by the company he keeps.

Шутки? Не шутки?

Future Directions: Humor Recognition

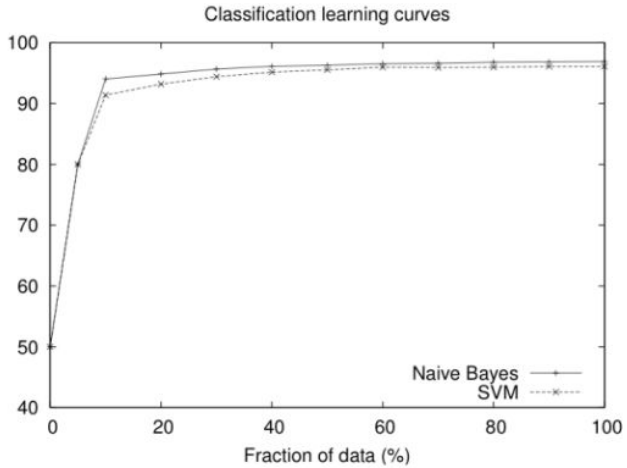


(a) One-liners / Reuters

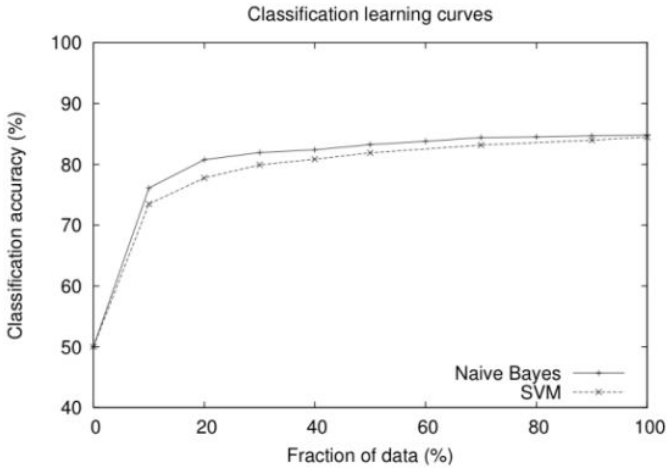


(b) One-liners / Proverbs

Future Directions: Humor Recognition



(a) One-liners / Reuters



(b) One-liners / Proverbs

Но! Плато ближе к концу обучения говорит нам о том, что просто содержания недостаточно для эффективной детекции юмора. В дальнейшем предлагается обращать внимание также на **стилистические фичи** юмористических текстов: **аллитерацию, антонимию, грубые слова** (alliteration, antonymy, and adult slang) (Ruch, 2002; Bucaria, 2004).

Future Directions: Exploiting Extra-linguistic Features

CORPS corpus (CORpus of tagged Political Speeches)

contains political speeches tagged with audience reactions (e.g. applause, standing-ovation, booing).

The collected texts come from various Web sources (e.g. politicians' official sites, News web sites) to create a specific resource useful for the study of persuasive language (Guerini, Strapparava, & Stock, 2008)

Future Directions: Music and Lyrics

```
<song filename=AHARDDAY.m2a>
<key time=0>G major</key>
<line pvers=1 raising=3 anger=1.5 disgust=0.7 sadness=2.5 surprise=0.8 >
<token time=5040 orig−note=B degree=3 duration=210>IT</token>
<token time=5050 orig−note=B degree=3 duration=210>'S </token>
<token time=5280 orig−note=C′ degree=4 duration=210>BEEN </token>
<token time=5520 orig−note=B degree=3 duration=210>A </token>
<token time=5760 orig−note=D′ degree=5 duration=810>HARD </token>
<token time=6720 orig−note=D′ degree=5 duration=570>DAY</token>
<token time=6730 orig−note=D′ degree=5 duration=570>'S </token>
<token time=7440 orig−note=D′ degree=5 duration=690>NIGHT</token>
</line>
<line pvers=2 raising=5 anger=3.5 disgust=2 sadness=1.2 surprise=0.2 >
<token time=8880 orig−note=C′ degree=4 duration=212>AND </token>
<token time=9120 orig−note=D′ degree=5 duration=210>I</token>
<token time=9130 orig−note=D′ degree=5 duration=210>'VE </token>
<token time=9360 orig−note=C′ degree=4 duration=210>BEEN </token>
<token time=9600 orig−note=D′ degree=5 duration=210>WOR</token>
<token time=9840 orig−note=F′ degree=7− duration=930>KING </token>
<token time=10800 orig−note=D′ degree=5 duration=210>LI</token>
<token time=11040 orig−note=C′ degree=4 duration=210>KE </token>
<token time=11050 orig−note=C′ degree=4 duration=210>A </token>
<token time=11280 orig−note=D′ degree=5 duration=330>D</token>
<token time=11640 orig−note=C′ degree=4 duration=90>O</token>
<token time=11760 orig−note=B degree=3 duration=330>G</token>
</line>
```

Figure 2: Two lines of a song in the corpus: *It-'s been a hard day-'s night, And I-'ve been wor-king li-ke a d-o-g*

Emotion	Baseline	Textual	Musical	Textual and Musical
ANGER	89.27%	91.14%	89.63%	92.40%
DISGUST	93.85%	94.67%	93.85%	94.77%
FEAR	93.58%	93.87%	93.58%	93.87%
JOY	50.26%	70.92%	61.95%	75.64%
SADNESS	67.40%	75.84%	70.65%	79.42%
SURPRISE	94.83%	94.83%	94.83%	94.83%
AVERAGE	81.53%	86.87%	84.08%	88.49%

Table 16: Evaluations using a coarse-grained binary classification.