Sentiment analysis track at ROMIP 2012

Ilia Chetviorkin
CMC MSU

Loukachevitch Natalia
Research Computing Center MSU

Sentiment analysis motivation

- Large amount of sentiment-oriented data in social media

- Various useful applications
  - Reputation management for brands
  - Review aggregation for products and services
  - Monitoring of political events and attitudes connected with them

ROMIP 2012 motivation

- Evaluation framework for sentiment analysis in Russian

- ROMIP 2011 statistics:
  - 23 groups/persons registered
  - 12 groups/persons submitted results
  - 6 wrote papers
  - More than 200 runs…

ROMIP 2012 tracks

- Sentiment classification of user reviews
  - The same task as in ROMIP 2011. New test collection
- Query-based extraction of opinionated blog posts,
  - Three domains: Movies, Books, Digital Cameras
- Sentiment classification of news-based opinions.
  - News-based opinions are fragments of direct or indirect speech extracted from news articles.

- Partial support by RFBR grant N11-07-00588-a (head of the project – Loukachevitch Natalia)

ROMIP 2012 participants

- In all 16 groups took part in 5 tasks

- Number of runs
  - 94 runs in two-class classification
  - 46 runs in three-class classification
  - 15 runs in five-class classification
  - 16 runs in news-based opinion classification
  - 2 participants in query-based opinion retrieval with 33 runs

Sentiment classification

Procedure

- Training collection
  - Participants receive training collections with users' scores and can adapt or train their methods
- Evaluation
  - Testing collections without users' scores
  - Participants can apply different methods to process reviews and can send many "runs"

Training collections

- Training collection for every domain
  - 15 thousand movie reviews (Imhonet.ru)
  - 24 thousand book reviews (Imhonet.ru)
  - 10 thousand digital camera reviews (Yandex.market)

- Users' scores
  - Imhonet: 10-point scale,
  - Yandex/Market: 5-point scale
  - Adaptation of scores to chosen task is necessary

Testing collection

- !! Testing collections do not have users' scores
- Blog posts from Livejournal.ru
  - Yandex Blog Search
  - 60 thousand «posts»
  - Special queries
- We had to select only opinionated texts
  - Text should express sentiment about any entity from the chosen domains

Expert assessment of texts

- Numbers of selected reviews: 129 book reviews, 408 movie reviews, 411 camera reviews
- Each review was annotated by experts using three scales:
  - 2 classes: positive or negative
  - 3 classes: positive, negative or mediocre
  - 5 classes
- Highly skewed test set: 96% of positive reviews for cameras, 87% for books and 81% for movies

Results two-class classification

Run_ID	*Object*	*Macro_P*	*Macro_R*	*Macro_F*	*Accuracy*
xxx-17	book	0.749	0.684	0.715	0.884
xxx-1	book	0.666	0.748	0.705	0.821
Baseline	book	0.434	0.500	0.465	0.868
yyy-12	camera	0.589	0.734	0.669	0.895
yyy-13	camera	0.688	0.635	0.660	0.961
Baseline	camera	0.483	0.500	0.491	0.966
zzz-19	film	0.695	0.719	0.707	0.806
zzz-23	film	0.731	0.641	0.683	0.831
zzz-12	film	0.759	0.586	0.661	0.828
Baseline	film	0.404	0.500	0.447	0.809

Analysis of the results

- All best methods were based on application of machine learning
  - Rule-based systems performance was lower in the considered domains due to large amount of labeled training data
  - Various linguistic features could improve the quality of sentiment classification
- Created collections with manual annotation can be received by researchers for the experiments
- Further experiments ..?

News-based opinion classification

Task description

- The first step for sentiment analysis of whole news articles
  - Opinion classification of news quotes
  - The example of quote is: *“Посредством этих структур десяткам тысяч избирателей предлагают деньги в обмен на паспортные данные и подписи за какого-либо кандидата”*

- The task was to separate all text fragments into 3 classes: Positive, Negative or Neutral (without expression of any opinion)

Collections

- Training collection
  - 4260 text fragments with direct or indirect speech extracted from news articles
  - Balanced collection: 41% of quotes were negative, 32% positive and 27% neutral
- Test collection contains 124, 647 direct and indirect speech fragments from news articles
  - Random fraction of 5500 text fragments for system evaluation

Results

Run_ID	*Macro_P*	*Macro_R*	*Macro_F*	*Accuracy*
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

- The leaders in the news-based task were knowledge-based (dictionary + rules) approaches
  - Broad scope of quotation topics
  - Absence of a large training collection

Query-based sentiment extraction

Task description

В декабре выходит **CANON EOS 6D** FF...

показать полный текст

В декабре выходит **CANON EOS 6D** FF...

2 ч. 16 мин. назад · Вяче$Love · blogs.mail.ru/mail/slawikim

Canon EOS 6D: размышления о бюджетном полном кадре | Цифровое фото и видео - 3DN... vk.cc/Y8kNv

10 ч. 48 мин. назад · xo6otgoose

Canon EOS 6D: самая легкая полнокадровая зеркальная камера j.mp/SA3eTq

вчера, 17:10 · fotomeridian

Canon EOS 6D: самая легкая полнокадровая зеркальная камера

показать полный текст 8 комментариев

Canon EOS 6D

вчера, 17:10 · fotomeridian · fotomeridian.livejournal.com

Canon представляет новую цифровую зеркальную камеру **EOS 6D** 17.09.2012 fb.me/yfAZamTO

вчера, 11:00 · olesyasukhomlin

Canon анонсировали **EOS 6D**

показать полный текст

Canon анонсировали **EOS 6D**

вчера, 09:59 · Твой DSLR · youdslr.blogspot.com

Canon EOS 6D - полнокадровая зеркалка с Wi-Fi и GPS-модулем. prophotos.ru/news/14779-can... Пойду убьюсь, задолбали высосанные из пальца "фичи".

вчера, 09:42 · pingwin87

Collections

- Three domains: books, movies, digital cameras
- Training dataset
  - 874 blog posts about various products with sentiment markup (it was used during ROMIP 2011 evaluation)
- Test collection
  - For each domain a list of search queries was manually compiled: 2,713 book queries, 1,412 camera queries, and 281 movie queries
  - For each query we obtained a set of blog posts using Yandex Blog Search Engine.
  - The resulting collection size 60,737 blog posts

Text markup

- Random selection of queries set:
  - 221 book queries
  - 235 movie queries
  - 301 queries about digital cameras
- For each document-query pair the assessor should decide:
  - If the document is relevant to a specific query
  - What sentiment is expressed about the object in the query.

Official metrics

- **Precision@n** indicates the number of correct (relevant) objects in the first ***n*** objects in the result set

$$P@n = \sum_{i=1}^{n} rel(i)$$

- **NDCG@n** measures the usefulness, or gain, of a document based on its position in the result list, where *IDCG@n* is *DCG@n* of perfect ranking algorithm

$$NDCG@n = \frac{DCG@n}{IDCG@n} \qquad DCG@n = rel(1) + \sum_{i=2}^{n} \frac{rel(i)}{\log_2(i)}$$

Simple approach

- Only one participant submitted his result before the deadline
  - To conduct the track we built our own very simple approach of base of TFIDF and opinion word list
- Ranking formula

$$Weight = \alpha \cdot (\sum_{w \in q} tfidf_w + \sum_{w \in q} tfidf_w^{header}) + (1 - \alpha) \cdot SentiWeight$$

SentiWeight – is the fraction words in the blog post from the *ProductSentiRus* word list

ProductSentiRus

- The general meta-domain sentiment lexicon for products and services
  - Intrinsic quality P@1000 = 91.4%
- Useful for various sentiment analysis tasks
  - Already used in cross-domain classification task and opinion extraction task
- Freely available!
  - http://www.cir.ru/SentiLexicon/ProductSentiRus.txt

Results

Run_ID	*Object*	*P@1*	*P@5*	*P@10*	*NDCG @10*
xxx-0	book	0.3	0.32	0.286	**0.305**
xxx-9	book	0.3	0.31	0.323	**0.304**
xxx-8	book	0.25	0.31	**0.332**	0.298
xxx-6	book	0.25	0.31	**0.327**	0.302
yyy-9	camera	0.402	0.313	0.302	**0.305**
yyy-7	camera	0.427	0.319	0.300	**0.303**
yyy-1	camera	0.402	0.328	**0.325**	0.226
yyy-2	camera	0.440	0.325	**0.311**	0.303
zzz-3	film	0.494	0.449	**0.438**	**0.338**
zzz-8	film	0. 494	0.448	**0.444**	**0.332**

Conclusion

- Huge interest to the sentiment analysis task in Russian both from industrial companies and academia
- Several new tracks were conducted: news-based sentiment classification, query-based sentiment extraction
- Investigation of Machine learning vs. rule-based and lexicon-based approaches

Thank you!
Questions?