English Opinion Analysis for NTCIR7 at POSTECH

Dec. 19th, 2008

Jungi Kim Hun-Young Jung Sang-Hyeob Nam Yeha Lee Jong-Hyeok Lee
{yangpa, blesshy, namsang, sion, jhlee}@postech.ac.kr

Knowledge and Language Engineering Laboratory
Department of Computer Science and Engineering
Pohang University of Science and Technology

Hypotheses on Opinion Sentence Analysis

- *Opinionativeness* of a word consists of a *sentiment* aspect and an *informative* aspect
  - Sentiment: polarity {Pos/Neg}, Strength {Strong/Weak}
  - Informative: discriminative {Common/Rare}, {Significant/Trivial}, {Meaningful/Useless}

- A sentence of a document with many opinionated sentences is more likely to be opinionated. Similarly, a document tends to contain either mostly of positive sentences or mostly of negative sentences.
  - Prior probability of a sentence from its document

Sentiment Weight

- Resources
  - SentiWordNet [Esuli and Sebastiani, LREC'06]
    - WordNet Synsets with {Pos/Neg/Neu} scores
  - Appraisal verbs [Whitelaw et al., CIKM'05]
    - Levin's Verb Classes [Levin, 1993]

$$W_{Sentiment}(w) = SWN_{Pos}(w) + SWN_{Neg}(w) + Appraisal(w)$$

$$SWN(w) = Max(S(w_{s_1}), S(w_{s_2}), \ldots, S(w_{s_n}))$$

$$Appraisal(w) = \begin{cases} 1.5 & \text{if w is an appraisal verb} \\ 0 & \text{otherwise} \end{cases}$$

Term Weight

- A good opinionated term is discriminable, prominent, relevant to topic
- Topic-Independent (rareness, importance, significance)
  - Global Knowledge (from the collection of documents)
    - Inverse Document Frequency
  - Local Knowledge (from the document, sentence, phrase the word belongs)
    - Term Frequency in a sentence
    - Depth of a word in a dependency parse tree
- Topic-Dependent (correspondence, association, relevance)
  - Global Knowledge (co-occurrence, term dependence, ~domain-dependent)
    - Point-wise Mutual Information
  - Local Knowledge (modifying/modified by topical words)
    - Minimum Distance to a topical word in a parsed tree
- Resources
  - Document Collection: English Newspaper Articles from NTCIR CLIR Corpus
  - Sentence Analysis: Stanford Parser

Term Weight

$$W_{term}(w) = W_{BM25}(w) \cdot W_{TreeDepth}(w) \cdot W_{TopicProximity}(w)$$

$$W_{BM25}(w) = \log \frac{N - df + 0.5}{df + 0.5} \cdot \frac{tf \cdot (k_1 + 1)}{tf + k_1 \cdot (1 - b + b \cdot \frac{sl}{avgsl})}$$

idf tf

$$W_{TreeDepth}(w) = DepTreeDepth(w)^{0.9}$$

$$W_{TopicProximity}(w) = \begin{cases} 1.5 & \text{if distance to topic in dependency tree} <= 2 \\ 1.0 & \text{otherwise} \end{cases}$$

Prior with Document Smoothing

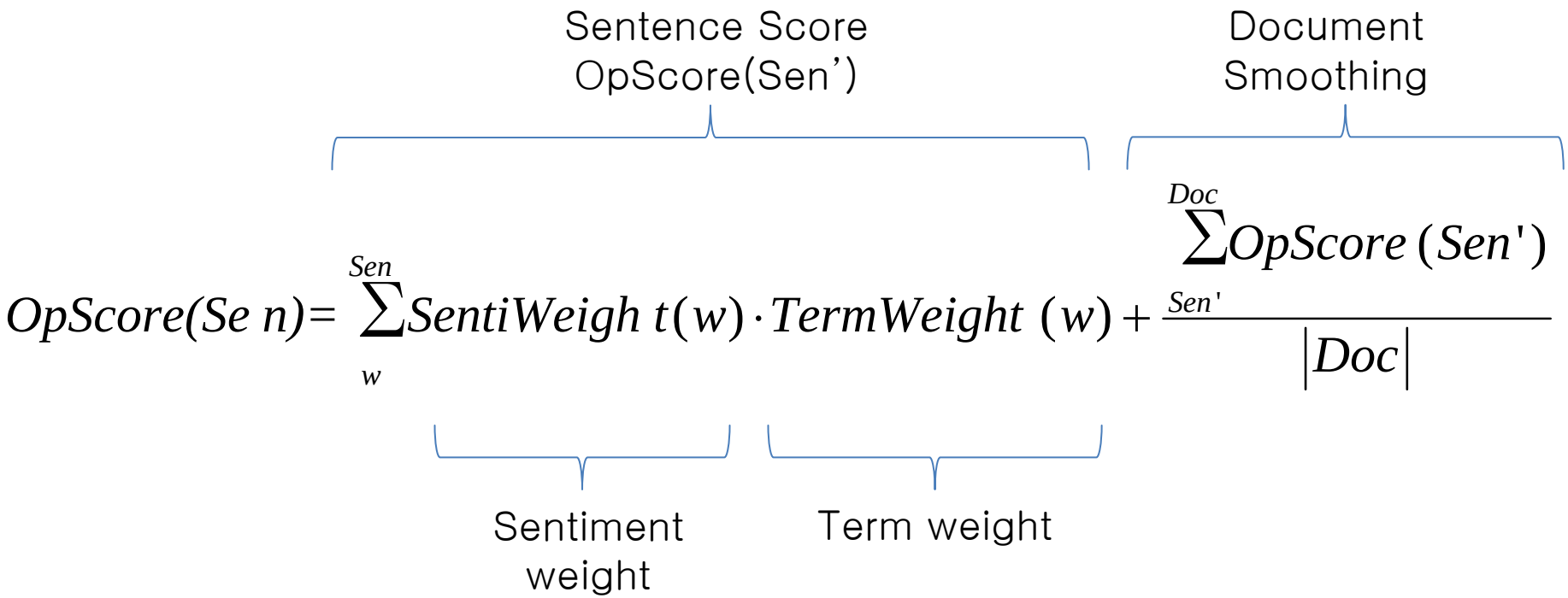
- Prior opinion or polarity scores of a sentence is presumed from the opinion or polarity score of the document the sentence belongs.
  - Jelinek-Mercer Smoothing

Sentence Score

Document Score (Average Sentence Score)

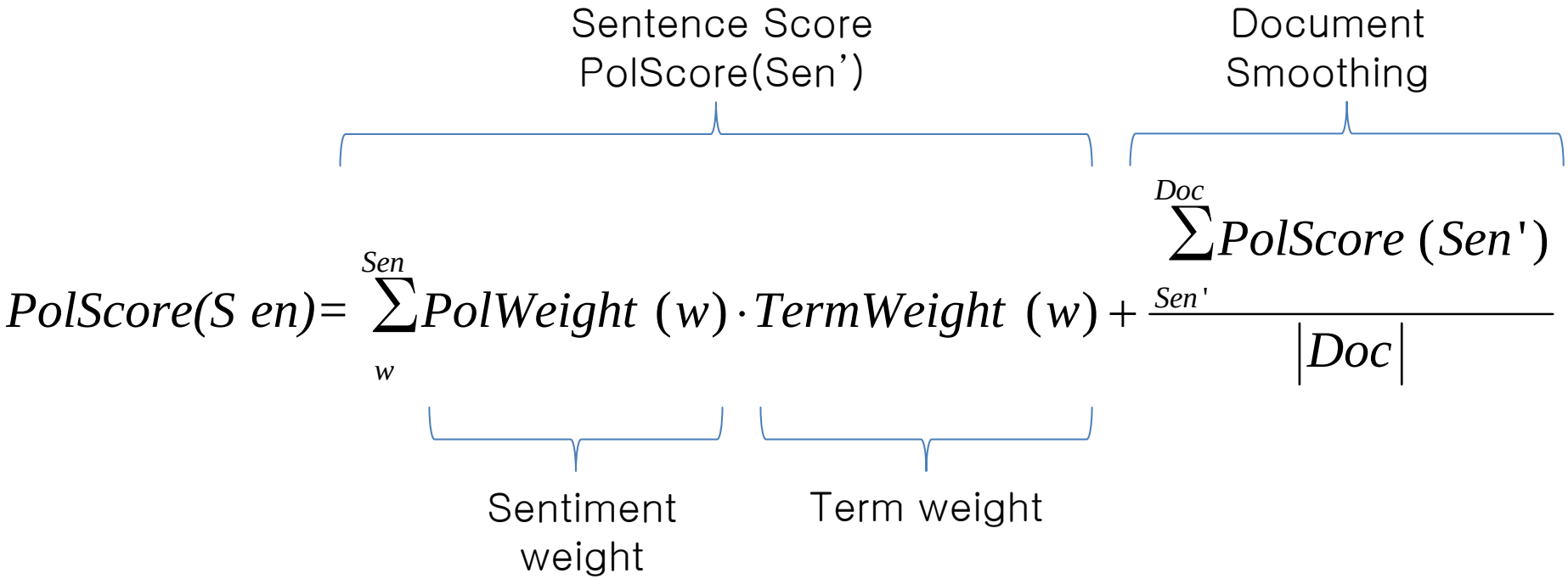
$$Score(Sen) = \lambda \cdot \sum_{w}^{Sen} Score(w) + (1-\lambda) \cdot \frac{\sum_{Sen'}^{Doc} Score(Sen')}{|Doc|}$$

Opinion Sentence Analysis



$$IsOpinionated(Sen) = \delta\ [\ opScore(Sen) > \Theta_{op}\ ]$$

Sentence Polarity Analysis



IsNeu(Sen) = δ [| PosScore(Sen) − NegScore(Sen) | < Θ_{pol}]

IsPos(Sen) = δ [PosScore(Sen) > NegScore(Sen)]

IsNeg(Sen) = δ [PosScore(Sen) < NegScore(Sen)]

Opinion Holder Extraction

- Opinion Holder Candidates
  - Speaker of a quotated remark
  - NOMINAL_SUBJECT of a simple clause of the most opinionated word
  - "AUTHOR"
- Most Opinionated Word
  - Communication Verb: 0.9
  - Appraisal Words: 0.7
  - SentiWordNet
- Resources
  - Sentiment Resources
    - Communication and Appraisal Words [Whitelaw et al., CIKM'05]
    - SentiWordNet [Esuli and Sebastiani, LREC'06]
    - Non-Named Entity Opinion Holder Candidates (pronouns, professions)
  - NLP tools
    - Named Entity Recognizer [Finkel et al., ACL'05]
    - Syntactic Parser [Klien and Manning, ACL'03]

Experimental Results: Opinion Sentence Analysis

Optimized on NTCIR6 (Lenient Evaluation)

System	PRECISION	RECALL	F-MEASURE
BASELINE (SentiWN)	0.285	0.809	0.422
BASELINE+APPRAISAL	0.305	0.707	0.426 (+0.95%)
BASELINE+OKAPI	0.317	0.776	0.450 (+6.64%)
BASELINE+TREEHEIGHT	0.299	0.741	0.426 (+0.95%)
BASELINE+TOPICPROXIMITY	0.281	0.835	0.421 (-0.24%)
BASELINE+SMOOTHING	0.296	0.783	0.430 (+1.90%)
ALL	0.345	0.717	0.466 (+10.4%)

Official Submissions to NTCIR7 (Optimized on Precision, Recall, F-measure)

System	L/S	PRECISION	RECALL	F-MEASURE
KLE1	L	0.353	0.727	0.475
KLE2	L	0.375	0.541	0.443
KLE3	L	0.274	0.933	0.423
KLE1	S	0.111	0.768	0.194
KLE2	S	0.119	0.579	0.198
KLE3	S	0.081	0.926	0.149

Experimental Results: Sentence Polarity Analysis

NTCIR6 (Lenient Evaluation)

System	PRECISION	RECALL	F-MEASURE
BASELINE (SentiWN)	0.092	0.353	0.147
BASELINE+APPRAISAL	0.096	0.365	0.152 (+3.40%)
BASELINE+OKAPI	0.104	0.344	0.160 (+8.84%)
BASELINE+TREEHEIGHT	0.100	0.324	0.152 (+3.40%)
BASELINE+TOPICPROXIMITY	0.097	0.323	0.149 (+1.36%)
BASELINE+SMOOTHING	0.101	0.357	0.157 (+6.80%)
ALL	0.145	0.395	0.212 (+44.2%)

Official Submissions to NTCIR7

System	L/S	PRECISION	RECALL	F-MEASURE
KLE1	L	0.092	0.353	0.147
KLE2	L	0.096	0.365	0.152
KLE3	L	0.104	0.344	0.160
KLE1	S	0.041	0.500	0.075
KLE2	S	0.042	0.357	0.074
KLE3	S	0.033	0.670	0.063

Experimental Results: Opinion Holder Extraction

NTCIR6 (Lenient Evaluation)

System	PRECISION	RECALL	F-MEASURE
BASELINE (SentiWN+NER)	0.122	0.482	0.194
BASELINE+COM.VERB	0.140	0.552	0.223 (+15.0%)
BASELINE+MAN.NE	0.125	0.492	0.199 (+2.58%)
ALL	0.145	0.575	0.231 (+19.1%)

Official Submissions to NTCIR7

System	L/S	PRECISION	RECALL	F-MEASURE
KLE1	L	0.400	0.508	0.447
KLE1	S	0.133	0.532	0.213

Conclusion

- High performance achieved in opinion, polarity judgments, and holder extraction

- Term weighting scheme has proven to be very effective in sentiment analysis
  - Empirical study with
    - different collections (NTCIR, movie review data)
    - various methods (Lexicon-based, Machine Learning)

- On-going work with theoretically-motivated interpretations and experiments
  - Formal study with probabilistic and language modeling
  - TREC Blog06 Collection with 06~08

References

[1] Yohei Seki, David Kirk Evans, Lun-Wei Ku, Hsin-His Chen, Noriko kando, and Chin-Yew Lin. Overview of Opinion Analysis Pilot Task at NTCIR-6. In Proceedings of the sixth NTCIR Workshop.

[2] Yohei Seki, David Kirk Evans, Lun-Wei Ku, Le Sun, Hsin-Hsi Chen, and Noriko kando. Overview of Multilingual Opinion Analysis Task at NTCIR-7. In Proceedings of the seventh NTCIR Workshop.

[3] Andrea Esuli and Fabrizio Sebastiani. 2006. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In Proceedings of the LERC 2006, pages 417–422, Geneva, IT.

[4] Yi Hu, Jianyong Duan, Xiaoming Chen, Bingzhen Pei, and Ruzhan Lu. 2005. A new method for sentiment classification in text retrieval. In Proceedings of the IJCNLP 2005.

[5] Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In Proceedings of the COLING 2004, pages 1367–1373, Geneva, CH.

[6] Soo-Min Kim and Eduard Hovy. 2006. Identifying and analyzing judgment opinions. In Proceedings of HLT/NAACL, 2006.

[7] Rada Mihalcea, Carmen Banea, and Janyce Wiebe. 2007. Learning Multilingual Subjective Language via Cross-Lingual Projections. In Proceedings of the ACL 2007.

[8] Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, and Daniel M. Ogilvie. 1966. The General Inquirer A Computer Approach to Content Analysis., MIT Press, Cambridge, MA.

[9] Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting emotional polarity of words using spin model. In Proceedings of the ACL 2005, pages 133–140, Ann Arbor, US.

References

[10] Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2006. Latent variable models for semantic orientations of phrases. In Proceedings of the EACL 2006.

[11] Peter D. Turney and Michael L. Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems, 21(4):315–346.

[12] Casey Whitelaw, Navendu Garg, and Shlomo Argamon. 2005. Using Appraisal Groups for Sentiment Analysis. In Proceedings of the CIKM 2005, pages 625–631, Bremen, DE.

[13] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In Proceedings of HLT-EMNLP- 2005.

[14] Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In Proceedings of the EMNLP 2003.

[15] Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. In Proceedings of the ACL 2003, pp. 423-430.

[16] Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. In Proceedings of the ACL 2005, pp. 363-370.

[17] Beth Levin. English Verb Classes and Alternations: a preliminary investigation. University of Chicago Press, Chicago and London, 1993.