Learning Domain-Specific Polarity Lexicons

Presenter: *Gizem Gezici*

Gulsen Demiroz, Berrin Yanikoglu, Dilek Tapucu, and Yücel Saygın

Faculty of Engineering and Natural Sciences,

Sabanci University

Istanbul, Turkey

Sabanci University

Sentiment Analysis Research Group

SentiWordNet

- A Word-Net based *domain independent* polarity lexicon
- It associates words with positivity, negativity and objectivity values
- Each word-sense pair is mapped to 3 values (positive, negative and objective)
- Word-sense pair examples:

word-sense	negative pol.	objective pol.	positive pol.
“good”-adv	0.000000	0.812500	0.187500
“good”-adj	0.005952	0.386904	0.607142
“good”-noun	0.000000	0.468750	0.531750

- *Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: Proceedings of the 5th Conference on Language Resources and Evaluation (LREC06. pp. 417-422 (2006)).*

Dominant Polarity

- While using the polarity values from SentiWordNet, dominant polarity is used for the corresponding word-sense pair.
- For a term *t*,

Dominant Pol(t) is

$$\begin{array}{ll} pol^{+} & \text{if } pol^{+} \geq pol^{-} \\ pol^{-} & \text{if } pol^{-} > pol^{+} \\ 0 & \text{otherwise} \end{array}$$

- There is no effect of objective polarity values in our formulation.
- We use the term *Pol (t)* in the remainder of the presentation.

Problem Definition

Hotel review:

- “*The hotel had really **small rooms***” (-)

Digital camera review:

- “*This camera is great as it has a **small size***” (+)

- However, *pol (“small”-adj)* which is the dominant polarity is 0.7250 (objective polarity).

Word	POSTag	Neg-Pol	Obj-Pol	Pos-Pol
small	Adjective	0. 2625	0.7250	0.0125

- Domain-independent lexicons (e.g. SentiWordNet) *cannot capture the context information*.

Motivation

- Observation: SentiWordNet has an assumption that a word-sense pair **always** has the same polarity in all circumstances.
- Goal: Adapt SentiWordNet polarities to a specific domain.

Background

- Yejin Choi, Claire Cardie, 2009: ***Adapting a Polarity Lexicon using Integer Linear Programming for Domain-Specific Sentiment Classification***
  - They start with an existing general-purpose polarity lexicon
  - Then adapt it into a domain-specific lexical usage
  - They use integer linear programming
  - Polarity of each word is one of: {positive, neutral, negative or negator}
  - They do expression-level polarity classification

Method

- For adapting the general purpose lexicon, we update the polarity of a word if its occurrence in labeled reviews strongly suggest one class, while SentiWordNet would suggest the other class.

Finding Domain Specific Words

- To determine the different occurrences of words between positive and negative class:

  - We first compute tf-idf scores of each word separately for positive and negative review classes.
  - There are a few variants of tf-idf computations and the one we use is computed as:

$$tf.idf\ (w_i, +) = log_e(tf\ (w_i, +) + 1) * log_e\ (N\ /\ df(w_i))$$

$$tf.idf\ (w_i, -) = log_e(tf\ (w_i, -) + 1) * log_e\ (N\ /\ df(w_i))$$

New Measure for Polarity Adaptation

- In order to determine the different occurrences of a word in positive vs. negative class, *Delta TFIDF metric* has already been proposed.
- Besides, we define a new measure, namely *(Δtf) idf*. It estimates whether the polarity of a word should be adjusted considering its occurrence in positive vs. negative class and computed as follows:

$$(\Delta tf)idf\,(w_i) = tf.idf\,(w_i,+) - tf.idf\,(w_i,-)$$

$$= [tf\,(w_i\,,+) - tf\,(w_i\,,-)\,]\,x\;idf\,(w_i)$$

- Although our new measure is very similar to *Delta TFIDF,* these two metrics *take into account of different things.*
- *Delta TFIDF considers the difference in the document frequencies; whereas our measure considers the term frequencies of the word in positive and negative reviews.*
- Then, by comparing *(Δtf) idf score* and *Pol (t)* of a word, we adjust its polarity.

Updating Word Polarities

- If there is a disagreement between the dominant SentiWordNet polarity, namely *Pol (t)* and *(Δtf) idf* score of a word, we consider changing its polarity.
- As seen In Table I, the polarity of the words '*comfy*' and '*joke*' should be updated.

w_i	$\Delta(tf)\,(w_i)$	$Pol\,(w_i)$	$Result$
comfy	6.01	-0.75	Disagreement
joke	-8.25	0.53	Disagreement
dirty	-6.7	-0.47	Agreement

Table I.

- Yet, ***how*** we will update the polarities?
- We have some updating method alternatives for these words which will be discussed next.

Updating Method Alternatives

- When there is a mismatch between SentiWordNet's dominant polarity and *(Δtf) idf* score of a word, for update process, we have several alternatives:
  - *Flip: Using the opposite polarity of the word (e.g. if the negative polarity of a word was dominant, we switch to its positive polarity and vice versa).*
  - *ObjectiveFlip: Switching the objective polarity to either negative or positive of a word; similarly switching the negative or positive to objective instead of its opposite polarity as done in **Flip**.*
  - *Shift:* Shifting the polarity of a word toward the other pole.
  - *DeltaScore: Computing the new polarity based on the (Δtf) idf score of the word.*

Extent Alternatives of the Updates

- We decided how we update the polarities of the determined words.
- Then a new question comes: ***How many words*** will be affected by our updating methods?
- Again, we have several alternatives:
  - *Top-k%*: *Changing the polarity of the top-k% of the words showing a mismatch. For this option, we sorted the determined words in descending order with respect to |(Δtf) idf| scores and examined the top-k% of the list.*
  - *Threshold*: *Changing the polarity of all the words below/above a fixed threshold where a disagreement occurs.*
  - *Iterative*: *Changing the polarity of a word one at a time using hill-climbing.*

Feature

- We have one feature to be used for classification which is the average review polarity:

$$\text{Average review polarity}(R) = \frac{1}{|R|} \sum_{w_i \in R} Pol(w_i)$$

- Feature Computation Steps:
  - Apply Stanford NLP tool in order to extract POS Tags.
  - Compute the average polarity of the review using $Pol\,(w_i)$.
  - In this process, only words with POS Tags JJ*(Adjective), RB*(Adverb), NN*(Noun), VB*(Verb) which have dominant polarity positive or negative.
  - We don't count the objective polarity words as their dominant polarity is 0.

Sentiment Classification

- We applied *Flip* and *DeltaScore* approaches among four updating method alternatives and reported them.
- Moreover, we tried all of three updating method alternatives: *Top-k%*, *Threshold* and *Iterative*.
- However, we report first two approaches since *Iterative* approach is too slow and not better than others.
- For *Top-k%* selection, we tried top-5% and top-10%.
- For *Threshold* selection, we did two runs with different positive and negative threshold values ranges that will enable a good number of words to be picked.
- After all of these steps, average review polarity is computed and reviews are classified as follows:

$$\mathtt{AverageReviewPol(R)} = \begin{cases} \mathit{Positive}, \text{ if average word polarity} > 0 \\ \mathit{Negative}, \text{ if average word polarity} <= 0 \end{cases}$$

Experimental Evaluation

Dataset I:

- TripAdvisor corpus
- Around 250.000 customer-supplied reviews
- About 1850 hotels
- Each review has a star rating (1* to 5*)
- Our dataset :
  - 6000 randomly chosen reviews (3000 positive, 3000 negative reviews).
  - These reviews were shuffled and splitted into train and test sets.
  - Each contains 1500 positive, 1500 negative reviews.
  - The reviews with star rating bigger than 2 are positive reviews, the rest are negative. (binary classification).

Dataset II:

- Pang&Lee (2004) Movie Corpus
- 2000 reviews (1000 positive, 1000 negative reviews).
- These reviews were shuffled and splitted into train and test sets.
- Each contains 500 positive, 500 negative reviews.
- Reviews are already marked as positive vs. negative (“+” for positive, “-” for negative reviews).

Experimental Results

- *For below results:*
  - *Flip* and *DeltaScore* updating methods with *top-5%* and *top-10%* of all of the words were carried out.
  - Furthermore, *Threshold* update with different threshold values were applied for picking the words to flip.

Word	POS Tag	SentiWordNet	Flip	DeltaScore
joke	NN	0.53	-0.19	-0.41
ludicrous	JJ	0.56	-0.125	-0.36
implausible	JJ	0.44	-0.25	-0.27
sufficient	JJ	-0.75	0.125	0.50
complicated	JJ	-0.625	0.125	0.32
courage	NN	-0.5	0.375	0.22

Table II. Polarity Scores: Before and After Update

Hotel Domain Example Polarity Updates

- sufficient JJ 0.75 0.125 0.125

  Word: **sufficient** was Negative (-0.75),

  i. Flip Approach: now Positive (0.125)

  ii. DeltaScore Approach: now Positive (0.49)

  Ideal and very very friendly. Just about everything you read on tripadvisor about the Castle Inn is true. It is15mins walk down to F/Wharf or 15mins up to Union Square / Chinatown area. ***Simple but sufficient complimentary breakfast (coffee, good orange juice, yoghurt, fruit, pastry, cereal bars) left us satisfied (including 2 teenagers!)****. …* Overall Rating: 4

- joke NN 0.1875 0.2812 0.5312

  Word: **joke** was Positive (0. 5312), now Negative (-0.1875)

  i. Flip Approach: now Negative (-0.1875)

  ii. DeltaScore Approach: now Negative (-0.41)

  Terrible Terrible Terrible ***Check in was a joke, our room wasn't ready until 5:00 pm.****Only one elevator was working which left us waiting for approx. 20 minutes every time we wanted to use it (should have left when we got there). ...* Overall Rating: 1

Movie Domain Example Polarity Updates

- complicated JJ 0.625 0.25 0.125

  Word: **complicated** was Negative (-0.625),

  i. Flip Approach: now Positive (0.125)

  ii. DeltaScore Approach: now Positive (0.32)

  ... this is an insightful , haunting exploration of the last days of the frankenstein and bride of frankenstein director , and ***it is notable for introducing one of the first complicated gay characters in a hollywood movie*** *. ...*

  Review Label: Positive

- ludicrious JJ 0.125 0.3125 0.5625

  Word: **ludicrious** was Positive (0.5625),

  i. Flip Approach: now Negative (-0.125)

  ii. DeltaScore Approach: now Negative (-0.36)

  ... ***the action in armageddon are so over the top , nonstop , and too ludicrous for words , i had to sigh and hit my head with my notebook a couple of times*** *. ...*

  Review Label: Negative

Classification Results on Hotel Dataset

Update Method	Training	Testing	Training (no 3-stars)	Testing (no 3-stars)
None (using SentiWordNet)	76.03	75.13	78.10	77.25
After 5% *Flip*	77.33	75.87	79.15	77.76
After 10% *Flip*	78.23	76.53	80.94	79.32
After 5% *DeltaScore*	80.40	78.03	82.16	80.12
After 5% *DeltaScore*	82.37	**80.27**	84.85	**82.72**
After Threshold (≥5 or ≤-10) *Flip*	77.80	76.33	79.93	78.30
After Threshold (≥5 or ≤-5) *Flip*	78.27	76.53	80.94	79.32

Table III. Classification Accuracies on TripAdvisor Dataset

Classification Results on Movie Dataset

Update Method	Training	Testing
None (using SentiWordNet)	60.00	61.30
After 5% *Flip*	60.80	62.60
After 10% *Flip*	62.70	63.90
After 5% *DeltaScore*	68.90	64.10
After 5% *DeltaScore*	73.00	**65.80**
After Threshold (≥10 or ≤-5) *Flip*	60.50	62.00
After Threshold (≥5 or ≤-5) *Flip*	61.60	63.10

Table IV. Classification Accuracies on MovieDataset

Conclusions & Future Work

Conclusions:

- In this work, we aimed at finding out how we can adapt an existing general-purpose lexicon.
- New polarity orientations for the words were captured by looking at how they are used in a particular domain.
- Although the proposed method is very simple yet efficient, it increased the review sentiment classification accuracy in both of the domains.
- Our work is comparable to *Choi et al (2009)*, where around 2% improvement in accuracy had been obtained using an adaptation done by linear programming; whereas we obtained around 5% improvement in accuracy in both hotel and movie domains.

Future work:

- We are going to test the proposed methods on a larger dataset in different domains and with more lexicons.
- We also plan to incorporate this polarity adaptation approach to our open source sentiment analysis tool SARE.

- Y. Choi and C. Cardie, "Adapting a polarity lexicon using integer linear programming for domain specific sentiment classification," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, 2009, pp. 590–598.*

SARE – Sentiment Analysis Research Environment

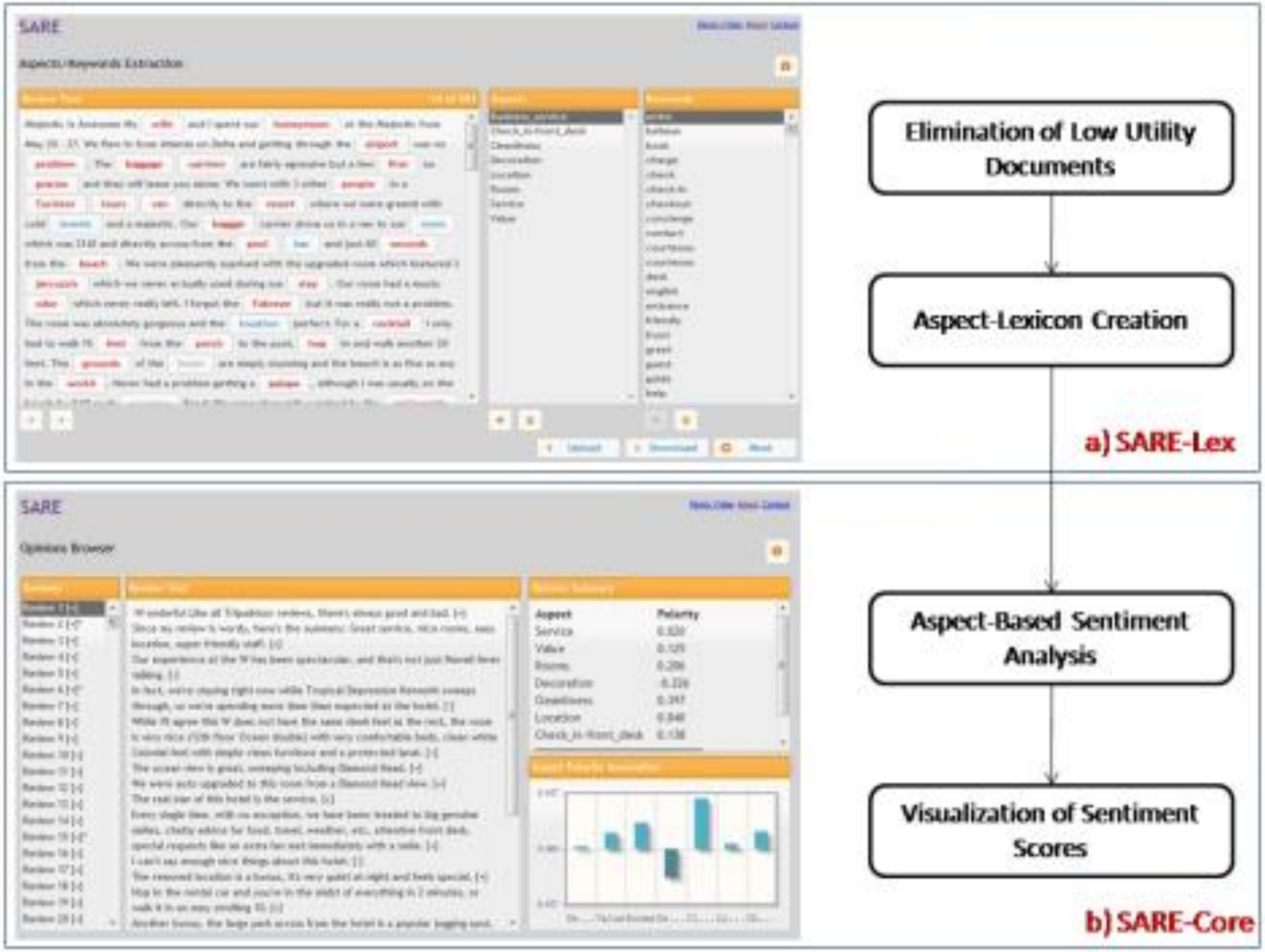


Fig. 1. SARE: (a) SARE-Lex module (b) SARE-Core module

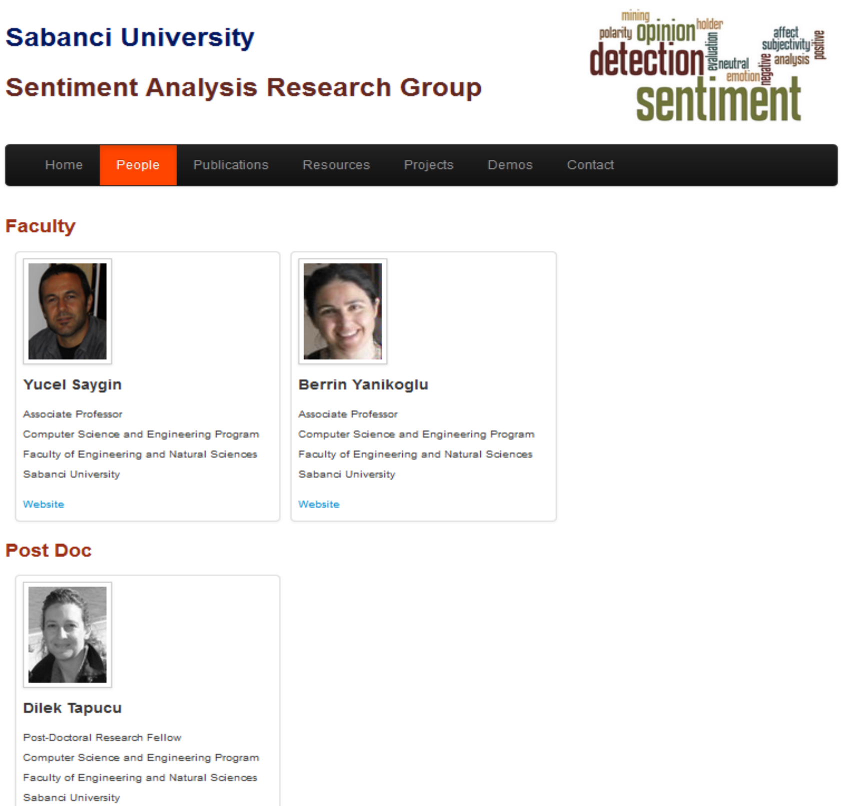
References

1. Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: Proceedings of the 5th Conference on Language Resources and Evaluation (LREC06. pp. 417-422 (2006)).
2. Y. Choi and C. Cardie, “Adapting a polarity lexicon using integer linear programming for domain specific sentiment classification,” in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, 2009, pp. 590–598.*
3. G. Paltoglou, S. Gobron, M. Skowron, M. Thelwall, and D. Thalmann, “Sentiment analysis of informal textual communication in cyberspace,” 2010.
4. J. Martineau and T. Finin, “Delta TFIDF: An improved feature space for sentiment analysis,” in *ICWSM, 2009.*
5. Y. Lu, and C. Zhai, “Latent aspect rating analysis on review text data: A rating regression approach,” *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 783–792, 2010.*
6. Pang and L. Lee, “A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts,” in *Proceedings of the ACL, 2004, pp. 271–278.*
7. M. Hussaini, A. Kocyigit, D. Tapucu, B. Yanikoglu, and Y. Saygin, “An aspect-lexicon creation and evaluation tool for sentiment analysis researchers,” in *ECMLPKDD, 2012.*

Thanks for your attention!

For any questions, contact:

http://sentilab.sabanciuniv.edu



Email :

- berrin@sabanciuniv.edu
- ysaygin@sabanciuniv.edu