Leveraging Personalized Sentiment Lexicons for Sentiment Analysis

Dominic Seyler (dseyler2@Illinois.edu)

Jiaming Shen (js2@illinois.edu)

Jinfeng Xiao (jxiao13@illinois.edu)

Yiren Wang (yiren@illinois.edu)

ChengXiang Zhai (czhai@Illinois.edu)

University of Illinois at Urbana-Champaign

Motivation

- In sentiment analysis, the system must predict the sentiment of an input text.
- User reviews, such as Yelp reviews can be used as a proxy for sentiment analysis.
- User ratings are highly subjective: Incorporate personalization into the model.
- Leverage user profile: documents previously written by the same author.
- User profile can be utilized to infer user traits related to her writing style.

Main Ideas

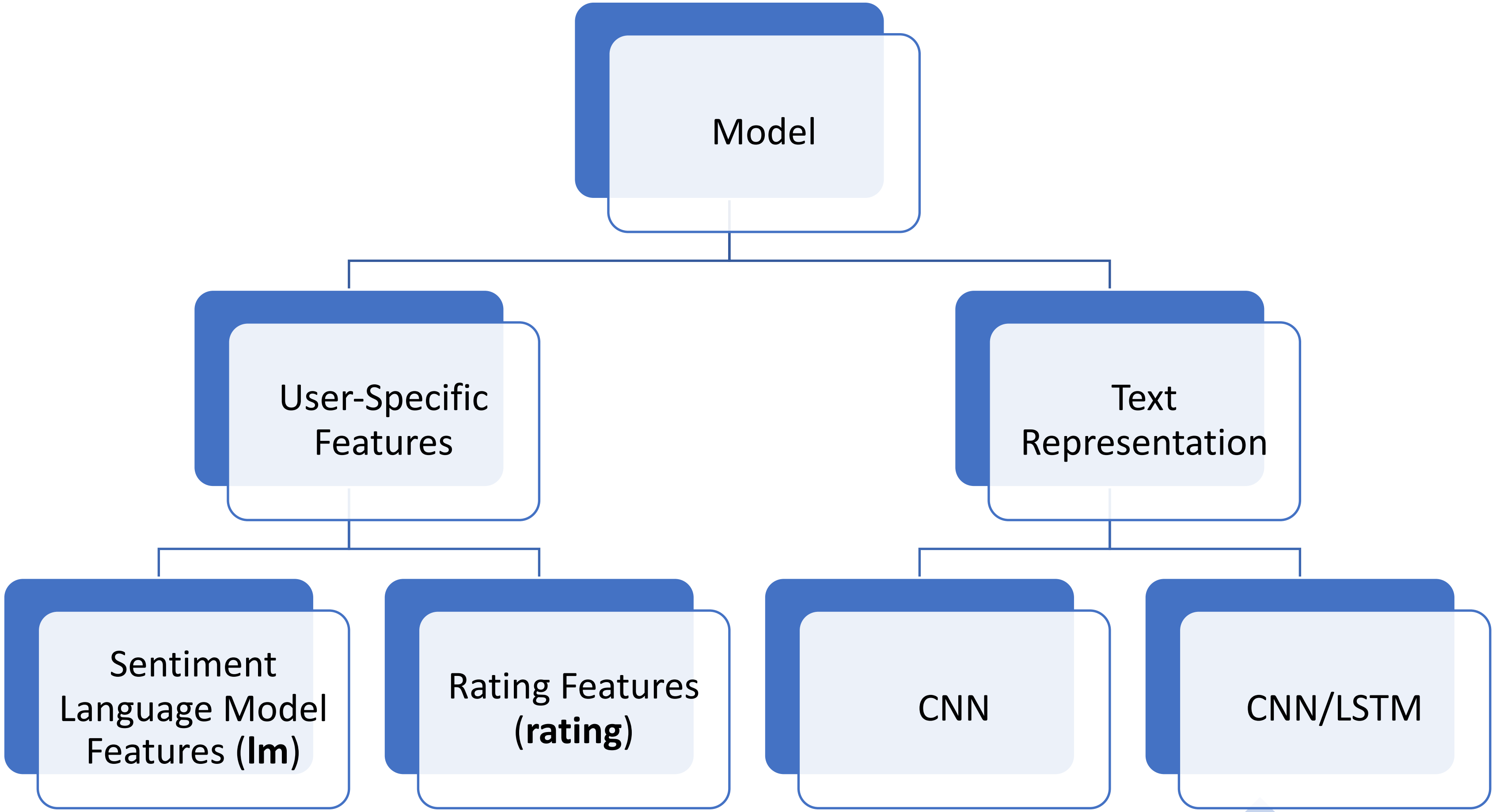
- Sentiment words carry different sentiment weight for different users.
- Model is personalized by learning a user-specific unigram language model over the most prominent words in a sentiment lexicon.
- Exploit rating history in addition to sentiment lexicons.

Contributions

- Novel personalized sentiment features derived from a user-specific unigram language model over an external sentiment lexicon.
- Show that personalized sentiment features improve deep learning framework.
- Our model outperforms other personalized baselines.
- Create dataset with user history based on the Yelp 2018 dataset.

Approach

Approach Overview



Sentiment Language Model Features
(lm)

- Sentiment words are important triggers for review rating classification.
- Capture user-specific language features: Use unigram language model on positive and negative sentiment lexicons.
- Build sentiment word candidate list derived from SentiWordNet [1], filter by positive and negative score according to:
  - Sentiment score threshold:

    $$score^{+}(word) > \epsilon \quad \text{OR} \quad score^{-}(word) > \varepsilon$$

  - Absolute difference:

    $$|score^{+}(word) - score^{-}(word)| > \varepsilon^{diff}$$

[1] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In Proceedings of the International Conference on Language Resources and Evaluation.

Sentiment Language Model Features
(lm)

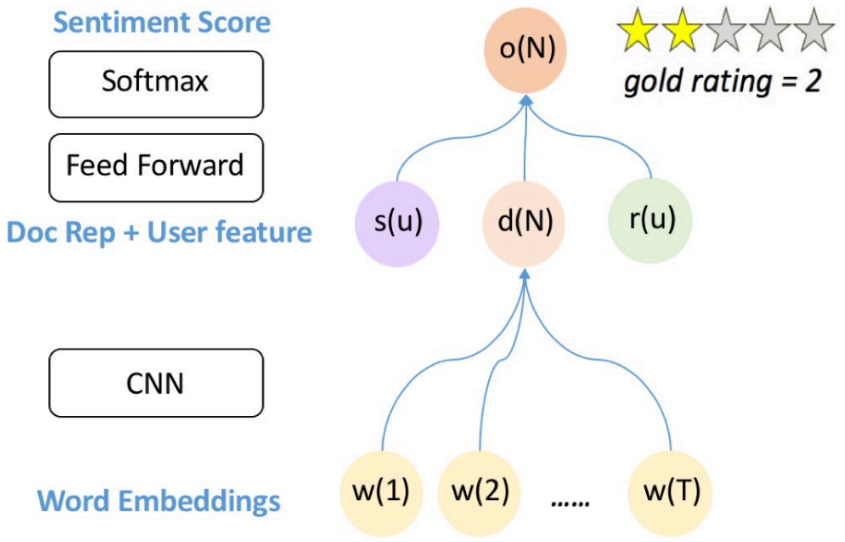
- Learn user-specific language model using maximum-likelihood.
- Use truncated-SVD to compress the user-specific language model into a low-dimensional vector $s(u)$.
  - $s(u) \in \mathbb{R}^k$ is a dense vector.
  - k is the number of largest singular values selected in the truncated SVD.

[1] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In Proceedings of the International Conference on Language Resources and Evaluation.

Rating Features

(rating)

- Rating history is a useful signal for sentiment analysis.
- Additional user-specific features based on rating statistics: maximum, minimum, mean score, standard deviation and histogram.
- Denote user-specific feature vector as $r(u)$.



Text Representations:

CNN

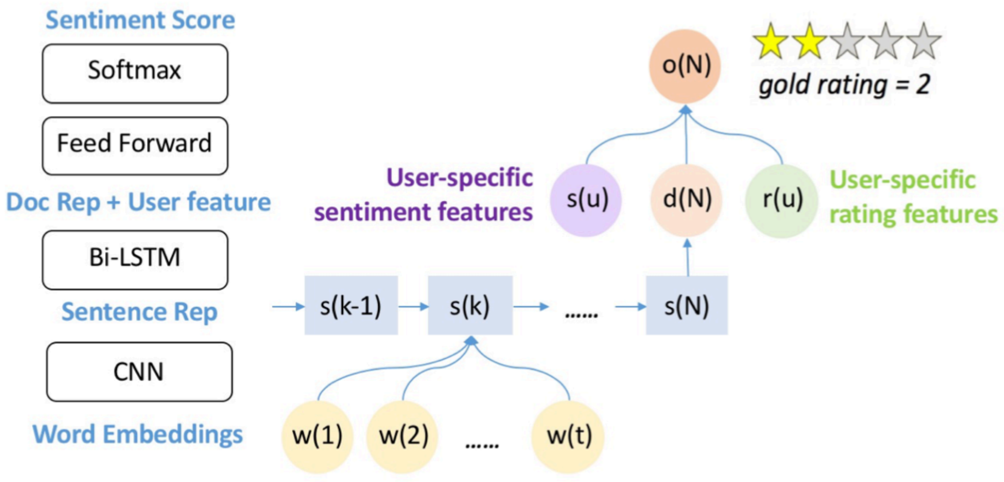
- Similar to Kim [1]: CNN model obtains the review document representation directly from the word level:

$$d(N) = CNN(w_1, w_2, ..., w_T)$$

[1] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In Proceedings of the Conference on Empirical Methods in Natural Language Processing. 1746–1751.

Text Representations:

CNN-LSTM



- CNN to embed each sentence into a dense vector $s(k)$.
- Sentence representations are passed into a LSTM layer to generate document level representations $d(N)$.
- Max pooling over N sentence vectors for constructing the final document representation.

Experiments

Dataset	#Reviews	#Users
Total	89,150	1,950
Train. (2011-2015)	50,018	1,950
Dev. (2016)	21,120	1,950
Test (2017)	18,012	1,950

Dataset

- Yelp Reviews from 2018.
- Restrict our analysis to users with 20-200 reviews.
- Ensure all users in the development and test datasets have been observed in the training dataset.
- The total number of reviews in the training set should be close to 50k.
- Training: 2011-15, development: 2016, testing: 2017.

Experimental Setup

- Research questions:
  1. Measure the effectiveness of our personalization features by performing an ablation study.
  2. Compare our model's performance with other personalized methods:
- Baselines:
  - Semantic Representations for Users and Products (UPNN) [1]
  - Neural Sentiment Classification (NSC) [2]

[1] Duyu Tang, Bing Qin, and Ting Liu. 2015. Learning semantic representations of users and products for document level sentiment classification. In Proceedings of the Association for Computational Linguistics. 1014–1023.
[2] Huimin Chen, Maosong Sun, Cunchao Tu, Yankai Lin, and Zhiyuan Liu. 2016. Neural sentiment classification with user and product attention. In Proceedings of the Conference on Empirical Methods in Natural Language Processing. 1650–1659.

Model	Accuracy
CNN	66.37%
P-CNN	67.14% (+1.2%)
– rating feature (lm-only)	66.77% (+0.6%)
– lm feature (rating-only)	67.09% (+1.1%)
CNN-LSTM	66.15%
P-CNN-LSTM	67.26% (+1.7%)
– rating feature (lm-only)	66.04% (-0.2%)
– lm feature (rating-only)	67.04% (+1.3%)

Feature Ablation Analysis

- CNN and CNN-LSTM have no personalization component and perform worse than the personalized models.
- Personalized CNN-LSTM model outperforms the personalized CNN model.
- When lm is added to the rating features it is still able to improve overall.
- Personalized features are effective, and the highest additive benefit can be achieved when combined.

Model	Accuracy
UPNN (Tang et al. [22])	59.11%
NSC (Chen et al. [7])	66.06%
NSC+LA (Chen et al. [7])	66.17%
NSC+UPA (Chen et al. [7])	66.42%
CNN	66.37% (+0.5%)
P-CNN	67.14% (+1.1%)
CNN-LSTM	66.15% (+0.1%)
P-CNN-LSTM	**67.26%** (+1.3%)

Comparison to Baselines

- CNN and CNN-LSTM models achieve comparable results with the NSC model.
- Adding user features improves performance for the baseline NSC+LA/UPA models and our P-CNN(-LSTM) models.
- Both of our personalized models outperform the best personalized baseline model.
- P-CNN-LSTM model can slightly better leverage the user features (1.3% improvement), compared to the P-CNN model (1.1% improvement).

Tang et al. 2015. Learning semantic representations of users and products for document level sentiment classification. ACL. 1014–1023.
Chen et al.. 2016. Neural sentiment classification with user and product attention. EMNLP. 1650–1659.

Conclusions and Future Work

- We presented a novel personalized approach for sentiment analysis that:
  - Captures a user's specific use of sentiment words in a language model.
  - Correlates them with her rating behavior.
- We showed that incorporating personalized sentiment lexicons can improve overall performance and beat other personalized baselines.
- Future Work
  - Explore more rigorous ways to derive sentiment words matching the informal language used in social media.
  - More sophisticated incorporation of personalized features could achieve larger performance improvement.

Thank You!

ILLINOIS

Leveraging Personalized Sentiment Lexicons for Sentiment Analysis

Dominic Seyler (dseyler2@Illinois.edu)

Jiaming Shen (js2@illinois.edu)

Jinfeng Xiao (jxiao13@illinois.edu)

Yiren Wang (yiren@illinois.edu)

ChengXiang Zhai (czhai@Illinois.edu)

University of Illinois at Urbana-Champaign

more info at: https://dominicseyler.com