CIS-positive: Combining Convolutional Neural Networks and SVMs for Sentiment Analysis in Twitter

Sebastian Ebert, Ngoc Thang Vu, Hinrich Schütze

Center for Information and Language Processing, University of Munich, Germany

ebert@cis.lmu.de

1. Introduction

- non-standard language makes tweet normalization necessary
  - informal language
  - different grammar
  - intentional misspellings
  - abbreviations
  - $\rightarrow$ high OOV rate
  - e.g., “read some fiction,or text bk,sure can zzz:p”
- polarity classification difficult
- linguistic knowledge, such as sentiment lexicons, is crucial

1.1 Contributions

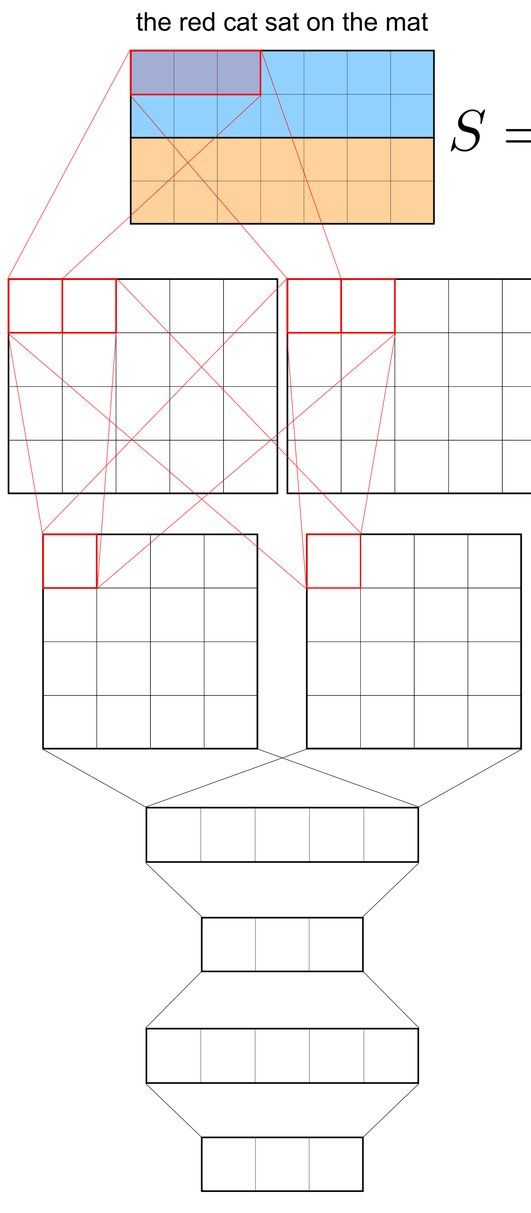
1. incorporate linguistic features into Convolutional Neural Network (CNN)
2. combination of SVM with tweet-specific features and CNN

2. Preprocessing

- tokenization and POS tagging using CMU tokenizer and tagger [Owoputi et al., 2013]
- replace user mentions by “<user>”
- replace urls by “<web>”
- keep hashtags (e.g., #happy)
- normalize “!?!?” to “[!?]” to reduce OOV rate
- normalize elongated words
  - $>$ 3 equal characters in a row
  - create candidate set by step-wise removing one repeated character (e.g., “cooolll” $\rightarrow$ {coolll, colll, cooll, coll, cool, col})
  - search in sentiment lexicon
  - take shortest when multiple matches exist
  - use more reliable lexicons first
- lowercase
- vocabulary size: 18k
- OOV rate tokens: 10.62% (test progress), 9.32% (test)
- OOV rate types: 57.06% (test progress), 40.57% (test)

3. Convolutional Neural Network

- capture sequential phenomena, i.e., keep word order
- consider words in their contexts
- capture long-distance effects
- goal of CNN: conflate the input sequence into a meaningful representation by finding salient features that indicate polarity



$$S = \begin{bmatrix} P_{:,w_1} & \cdots & P_{:,w_n} \\ Q_{:,w_1} & \cdots & Q_{:,w_n} \end{bmatrix}$$

$$C_{i,j} = \mathbf{m}^T S_{i,j:j+m-1}$$

$$A = max(0, C + \mathbf{b})$$

$$\mathbf{z} = max(0, W\mathbf{x} + b)$$

$$\mathbf{o}_i = \frac{\exp(\mathbf{z}_i)}{\sum_j \exp(\mathbf{z}_j)}$$

- trainable parameters: P, $\mathbf{m}$, W, $\mathbf{b}$, b
- training hyper-parameters: mini-batch stochastic gradient descent with 200 batch size, AdaGrad with initial $lr = 0.001$, ℓ_2 with $\lambda = 0.001$
- CNN hyper-parameters: $d_p = 50$, 30 filters with $m = 5$ in 1st conv layer, 10 filters with $m = 3$ in 2nd conv layer, hidden layers with 200, 40, 200 neurons

3.1 Features

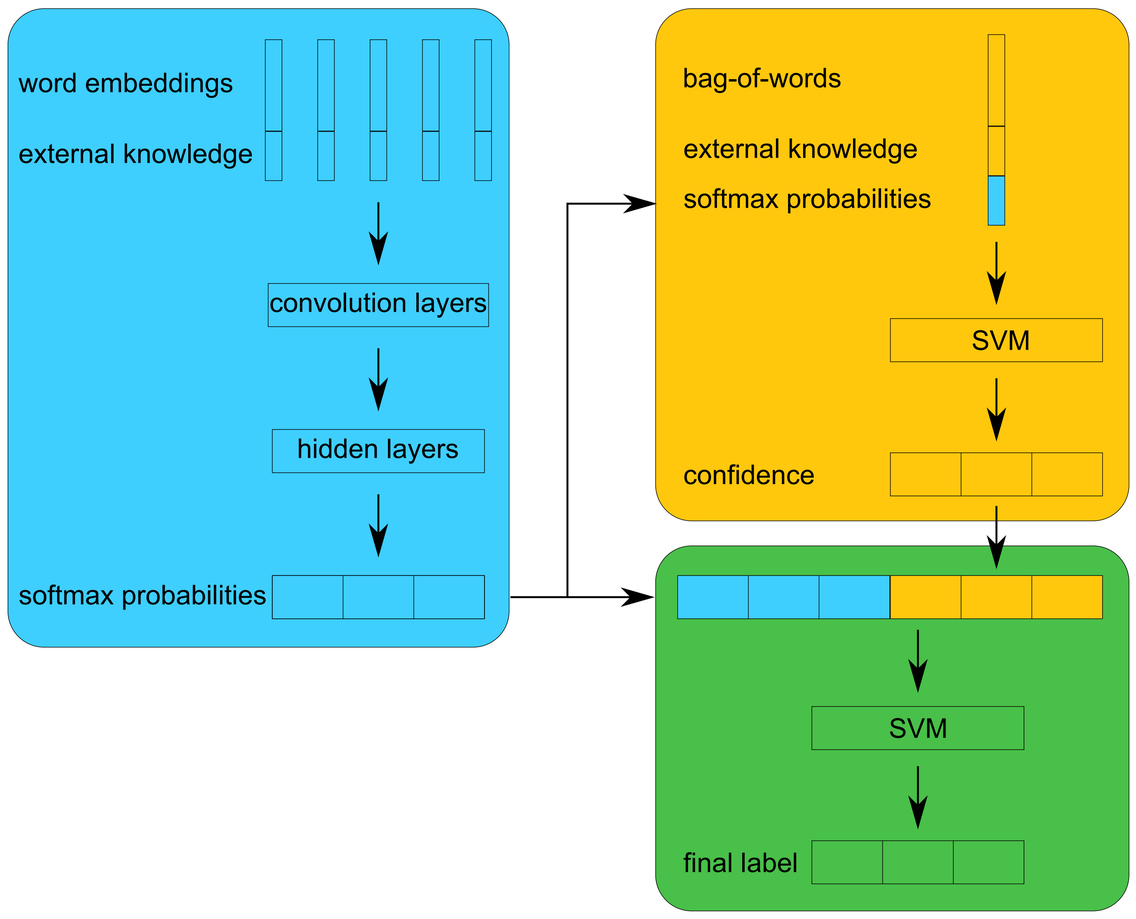
word embeddings randomly initialized or pre-trained with word2vec[1] on unlabeled Twitter data

binary sentiment indicators binary polarity label per token; lexicons: MPQA [Wilson et al., 2005], Opinion lexicon [Hu and Liu, 2004], NRCC Emotion lexicon [Mohammad and Turney, 2013]

sentiment scores sentiment score per token (or bigram); lexicons: sentiment 140 lexicon, hashtag lexicon [Mohammad et al., 2013]

binary negation indicator if token is between a negation word and the next punctuation[2]

4. Architecture



5. SVM

- following winner system from SemEval 2013 and SemEval 2014 [Mohammad et al., 2013]
- SVM 1: LIBLINEAR [Fan et al., 2008], C tuned on development set

5.1 Features

binary bag-of-words binary bag-of-words features of uni- and bigrams and character trigrams

sentiment features for every tweet and every lexicon: number of tokens in the tweet that occur in the lexicon, sum of all sentiment scores in the tweet, maximum sentiment score, sentiment score of the last token in the tweet

CNN output softmax output of the CNN informs the SVM about the CNN’s decision

6. Model Combination

- SVM 2: LIBLINEAR [Fan et al., 2008], C tuned on development set
- combine softmax output of the CNN and SVM 1 confidences

7. Results

	#pos	#neg	#neu	$F_{1,positive}$	$F_{1,negative}$	$F_{1,neutral}$	$F_{1,macro}$
SemEval 2013 Twitter	1572	601	1640	71.32	58.31	72.53	64.82
SemEval 2013 SMS	492	394	1207	66.94	63.34	80.33	65.14
SemEval 2014 LiveJournal	427	304	411	71.09	71.84	69.04	71.47
SemEval 2014 Twitter	982	202	669	73.63	58.47	67.14	66.05
SemEval 2014 Twitter sarcasm	33	40	13	60.00	38.46	53.33	49.23
SemEval 2015 Twitter	1038	365	987	65.32	53.82	68.06	59.57

- lowest performance on negative class: class is under-represented
- macro F_1 score of 59.57 leads to rank 20 out of 40 participants
- much better performance on LiveJournal $\rightarrow$ Twitter is difficult

Acknowledgments

This work was supported by DFG (grant SCHU 2246/10).

[1] https://code.google.com/p/word2vec/

[2] http://sentiment.christopherpotts.net/lingstruc.html