Semi-supervised emotion lexicon expansion with label propagation

Mario Giulianelli[1] Daniël de Kok[2]

[1]University of Amsterdam

[2]Seminar für Sprachwissenschaft
University of Tübingen

CLIN, 2018

Emotion and sentiment

- **Sentiment analysis** commonly refers to the task of polarity annotation. A piece of text is positioned on a value scale from negative to positive.
- **Emotion analysis** replaces the value scale with a set of m basic emotions. A text is assigned to an emotion class or it is mapped onto an m-dimensional space.
- Our work: document-based

Challenges of emotion analysis

- **Lack of contextual information**: judgements on affective orientation are subjective and susceptible to cross-cultural differences.
  - *Time to start this research paper*
  - *Am not gonna watch Barcelona match today*
- **Inter-annotator agreement**: trained annotators agree with a simple-average Pearson correlation of 53.67, and with a frequency-based average correlation of 43 (Strapparava and Mihalcea, 2007).
- **Insufficient lexical coverage**

[1] for anger, disgust, fear, joy, sadness, and surprise.

Challenges of emotion analysis

- **Lack of contextual information**: judgements on affective orientation are subjective and susceptible to cross-cultural differences.
  - *Time to start this research paper*
  - *Am not gonna watch Barcelona match today*
- **Inter-annotator agreement**: trained annotators agree with a simple-average Pearson correlation of 53.67, and with a frequency-based average correlation of 43 (Strapparava and Mihalcea, 2007).
- **Insufficient lexical coverage**:
  - only 3,462[1] emotion words in the NRC Emotion Lexicon
  - one third of the Hashtag Corpus contains no lexicon words
  - a tweet contains on average 1.09 lexicon words

[1] for anger, disgust, fear, joy, sadness, and surprise.

Approaches to emotion analysis

Corpus-based

Emotion analysis as a supervised classification problem.

- Datasets: news (SemEval-2007 Affective Text), tweets (**Hashtag Emotion Corpus**), blog posts, fairy tales
- Features: Weighted PMI, SentiWordNet scores, synonyms and lexical contrast from WordNet

Lexicon-based

Relies on labeled dictionaries to calculate the emotional orientation of a text from the words and phrases that constitute it.

- Lexica: WordNet Affect, Hashtag Emotion Lexicon, **NRC Emotion Lexicon**

Approaches to emotion analysis

Lexica and corpora are **complementary** sources of information and can be used jointly (Strapparava and Mihalcea, 2008).

- Corpus-based approaches learn to use contextual information.
- Lexicon-based approaches typically have a wider coverage of emotion-bearing words but are context-independent.

Problems

Narrow coverage

- *Saddened by the terrifying events in Virginia.*

Affective content but emotionally neutral words

- *I want cake. I bet we don't have any.*

Indirect affective words

- *I am going to have a monster year.*

Compositionality

- *Beating poverty in a small way.*

Implicatures

- *I'm not actually writing a physics exam today.*

Solution

Assumption: all terms in a text contribute to its affective content.

Use **transductive learning** to extend the coverage of an existing emotion lexicon, thereby:

- addressing the disproportion between lexicon words and unseen types
- leveraging latent information within the (semantic) space of lexicon words

Label propagation (Zhu and Ghahramani, 2002)

Construct a fully connected graph:

- **labeled and unlabeled** words are vertices
- edges are weighted by the distances between distributional word representations

  $w_{ij} = \exp\left(-\frac{dist(x_i,x_j)^2}{\sigma^2}\right)$

Compute a probabilistic transition matrix T

$$T_{ij} = P(i \rightarrow j) = \frac{w_{ij}}{\sum_k w_{kj}}$$

and a label matrix Y that stores, for each word,
its probability distribution over labels.

Label propagation (Zhu and Ghahramani, 2002)

Iterative algorithm

1. Propagate $Y \leftarrow TY$
2. Row-normalise Y
3. Repeat until convergence

Closed-form solution

- Partition the transition matrix
  $$T = \begin{bmatrix} T_{ll} & T_{lu} \\ T_{ul} & T_{uu} \end{bmatrix}$$
- Compute solution directly
  $$Y_u = \left(I - T_{uu}^{-1}\right) T_{ul} Y_l$$

Label propagation with word embeddings

Use cosine similarity to weight edges:

$$w_{ij} = \sigma\left(a\left(\frac{x_i \cdot x_j}{||x_i||_2\ ||x_j||_2}\right) + b\right)$$

Replace $a \in \mathbb{R}$ with $\alpha \in \mathbb{R}^d$ parameters that control edge weights along the d dimensions of the chosen word representation:

$$w_{ij} = \sigma\left(\alpha\left(\frac{x_i}{||x_i||_2} \odot \frac{x_j}{||x_j||_2}\right) + b\right)$$

Parameter optimisation
Minimise $H = -\sum_{ij} Y_{ij} \log Y_{ij}$ using gradient descent.

Batched-based label propagation

The size of the transition matrix can cause memory issues:

- $a \in \mathbb{R} \quad \rightarrow \quad T \in \mathbb{R}^{V \times V} \quad \rightarrow \quad \sim 2GB$
- $\alpha \in \mathbb{R}^d \quad \rightarrow \quad T \in \mathbb{R}^{V \times V \times d} \quad \rightarrow \quad \sim 600GB$

(V=32,930; half-precision; 300-dimensional vectors)

Label Propagation in batches:

- Randomly select a subset of the vocabulary of size $U < V$
  - possibly fix the distribution of labeled and unlabeled instances to be equal to the proportion that they have in the original transition probability matrix
- Compute the submatrix $T' \in R^{U \times U \times d}$
- Propagate labels within submatrix
- Repeat M times for each submatrix
- Repeat for N submatrices

Representing ~~words~~ linguistic units

Specialised word embeddings

- Learn emotion-specific word vectors directly from a large annotated corpus by extending an existing general purpose embedding algorithm (e.g. Collobert and Weston, Skipgram)
- Use pretrained embeddings as weights for an emotion classifier and update them during training (our approach)

Other features

- Character-level models of emotion intensity (Lakomkin et al., 2017)
- Additional lexical resources: WordNet, SentiWordNet

Experiments

We compare four emotion classifiers:

- One-vs-all SVM (Mohammad and Kiritchenko, 2015)
- Bidirectional LSTM
- Bidirectional LSTM model with an emotion lexicon (NRC Lexicon)
- Bidirectional LSTM model with the extended emotion lexicon obtained through label propagation

Results emotion classification

Classification on the Hashtag Emotion Corpus

Classifier	P	R	F_1
Mohammad and Kiritchenko, 2015	55.1	45.6	49.9
Bidirectional LSTM	55.0	55.0	55.0
Bidirectional LSTM + emotion lexicon	55.2	55.2	55.2
Bidirectional LSTM + expanded lexicon[2]	**56.2**	**56.2**	**56.2**

Domain adaptivity: classification on the SemEval-2007 headlines

Classifier	P	R	F_1
Mohammad and Kiritchenko, 2015	**46.7**	38.6	42.2
Bidirectional LSTM	38.8	50.3	43.8
Bidirectional LSTM + emotion lexicon	39.2	**50.9**	44.3
Bidirectional LSTM + expanded lexicon[2]	43.1	48.9	**45.9**

[2]with scalar parameter *a*

Survey

Test the *classification accuracy* of an untrained person with respect to an emotion-annotated dataset, the Hashtag Corpus.

- 33 participants: undergraduate and graduate students
- task: read 25 tweets and assign each to one emotion class
- 825 unique tweets

Dropped my phone in coffee shitty day

anger	disgust	fear	joy	sadness	surprise

Three main types of tweets:

I'm so excited for starting gift shopping early! #joy

Grateful for the sudden ability to make amazing omelettes! #surprise

Dropped my phone in coffee shitty day #joy

Humans as classifiers

	Survey			Our model		
Emotion class	**P**	**R**	**F_1**	**P**	**R**	**F_1**
anger	25	50	33	38	27	32
disgust	18	70	29	40	18	25
fear	48	22	30	58	52	55
joy	52	46	49	66	76	71
sadness	50	52	51	40	44	42
surprise	40	23	29	53	46	49
average	40.9	40.4	40.6	56.2	56.2	56.2

- Assigning an emotion to a short paragraph is a hard task for both a human and a statistical classifier.
- More contextual information is required than it is available in the paragraph itself.

Summary

Conclusions

- Label propagation can be used to extend the coverage of an existing emotion lexicon.
- Access to an expanded emotion lexicon can improve emotion classification as it combines context-sensitivity with wide, context-independent lexical coverage.

Outlook

- Can character-level models of emotion intensity (Lakomin et al., 2017) be used for label propagation?
- Enrich word representations with lexical contrast information.

Thank you!

Intrinsic evaluation

Average Kullback–Leibler divergence for 10-fold cross-validation on the NRC Emotion Lexicon.

Lexicon expansion	KL divergence
Uniform distribution	1.34
Majority class (Hashtag Corpus)	21.32
Prior class distribution (Hashtag Corpus)	1.53
Label propagation ($a \in \mathbb{R}$)	1.31
Batch label propagation ($a \in \mathbb{R}$)	1.31
Batch label propagation[3] ($\alpha \in \mathbb{R}^{300}$)	14.37

[3] 500 batches of size 3,000; 5 epochs per batch