Using Sentiment Analysis as a Case Study for Introducing Modern NLP Concepts

Felipe Bravo-Marquez

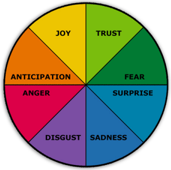
Department of Computer Science, University of Chile
IMDF

January 12, 2018

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

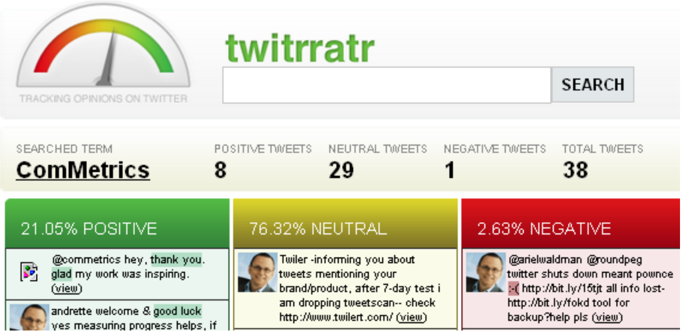
Social Media

- Microblogging services are increasingly being adopted by people in order to access and publish information.
- **Twitter**: Massively used Microblogging platform where users post messages limited to 140 characters referred to as **tweets**.
- Tweets can be used to convey emotions, opinions, and stance.



Sentiment Analysis and Social Media

- Opinions are provided **freely and voluntarily** by the users in Twitter.
- Analysing the sentiment underlying these opinions has important applications in product **marketing** and **politics**.



Opinion Mining or Sentiment Analysis

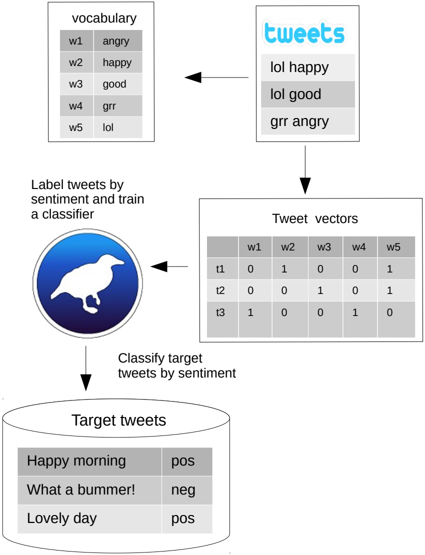
- Application of **NLP** and **text mining** techniques to identify and extract subjective information from textual datasets.

Main Problem: Message-level Polarity Classification (MPC)

1. Automatically classify a tweet to classes **positive**, **negative**, or **neutral**.

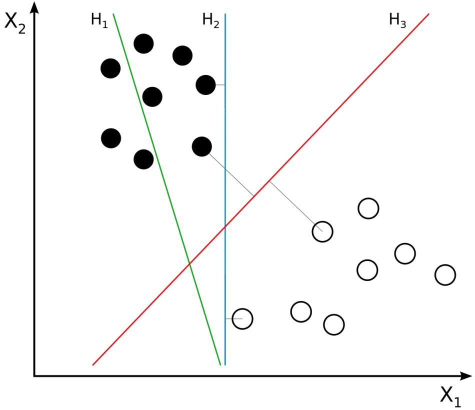
2. State-of-the-art solutions use **supervised** machine learning models trained from **manually** annotated examples [Mohammad et al., 2013].

Sentiment Classification via Supervised Learning



Supervised Learning: Support Vector Machines (SVMs)

- Idea: Find a hyperplane that separates the classes with the maximum margin (largest separation).



- H_3 separates the classes with the maximum margin.

[1] Image source: Wikipedia

Related Tasks

- Stance Detection: detect if the author of a tweet is in favor, against or neutral regarding a given target (e.g., Donald Trump).
- Irony/Sarcasm Detection: detecting sarcasm in tweets.
- Emotions classification: classify tweets according to multiple emotions (e.g., anger, fear, sadness, joy)
- Infer emotion intensities (numerical values) in tweets (e.g., degree of anger).
- Affective[2] Lexicon Induction: classification of words into affective dimensions.
- Many of these probems have been evaluated in **SemEval** tasks.

[2] We will use the term "**affect**" to encompass sentiment, emotions, and other related concepts.

Challenges

- **Label sparsity (LS)**: manual annotation is **labour-intensive** and **time-consuming**.
- **Concept drift**: the sentiment pattern can vary from one collection to another (domain-drift, temporal-drift).
- A classifier trained from tweets annotated for one domain will **not necessarily** work on another one!
- Trained models can become outdated over time.

Examples of domain-Drift

1. For me the queue was pretty **small** and it was only a 20 minute wait I think but was so worth it!!! :D @raynwise
2. Odd spatiality in Stuttgart. Hotel room is so **small** I can barely turn around but surroundings are inhumanly vast & long under construction.

Label Sparsity

- A possible approach to **overcome** the sentiment-drift problem is to **constantly update** the sentiment classifier with **recent labelled data** [Bifet and Frank, 2010, Silva et al., 2011].
- The high arrival rates of social streams make the continuous acquirement of sentiment labels **infeasible** [Silva et al., 2011, Calais Guerra et al., 2011, Guerra et al., 2014].

Approaches to overcome label sparsity

Distant Supervision

- Automatically **label** unlabelled data (**Twitter API**) using a heuristic method.
- **Emoticon-Annotation Approach (EAA)**: tweets with positive **:)** or negative **:(** emoticons are labelled according to the polarity indicated by the emoticon [Read, 2005].
- The emoticon is **removed** from the content.
- The same approach has been extended using hashtags #anger, and emojis.
- Drawback: emoticons can induce noisy and incomplete information. Moreover they are not necessarily used in all domains (e.g., politics).

Crowdsourcing

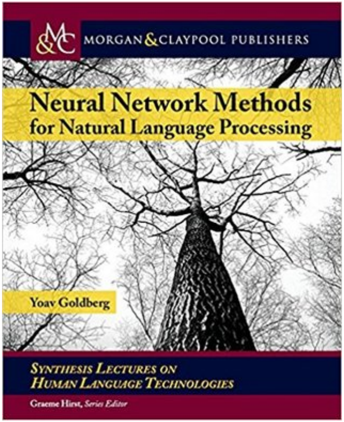
- Rely on services like **Amazon Mechanical Turk** or **Crowdflower** to ask the **crowds** to label a sample of the data on a demand-driven basis.
- This can be expensive for online sentiment analysis (label sparsity problem).

Roadmap

- In this talk we will overview various approaches tackling the main sentiment analysis problems.
- We will also introduce modern concepts in natural language processing based on **neural networks** such as **word embeddings**, **convolutional neural networks** (CNNs), and **Long short-term memory networks** (LSTMs).

Disclaimer

The presentation of neural network models in this talk is heavily based on this book:



Recursive Neural Networks over Sentiment Treebank

- A recursive neural tensor network for learning the sentiment of pieces of texts of different granularities, such as words, phrases, and sentences, was proposed in [Socher et al., 2013].
- The network was trained on a sentiment annotated treebank `http://nlp.stanford.edu/sentiment/treebank.html` of parsed sentences for learning compositional vectors of words and phrases.
- Every node in the parse tree receives a vector, and there is a matrix capturing how the meaning of adjacent nodes changes.
- The network is trained using a variation of backpropagation called Backprop through Structure.
- The main drawback of this model is that it relies on parsing.

Recursive Neural Networks over Sentiment Treebank

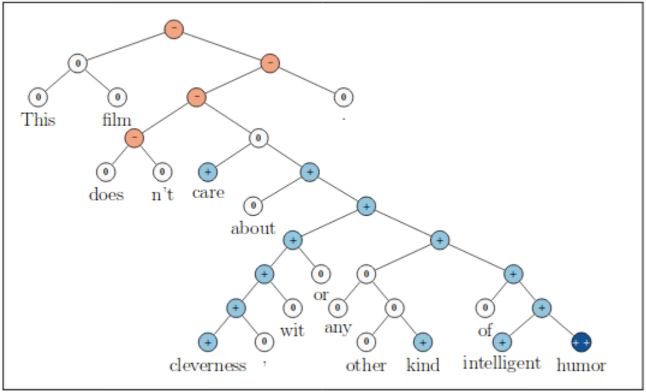


Figure 1: Example of the Recursive Neural Tensor Network accurately predicting 5 sentiment classes, very negative to very positive (– –, –, 0, +, + +), at every node of a parse tree and capturing the negation and its scope in this sentence.

Recursive Neural Networks over Sentiment Treebank

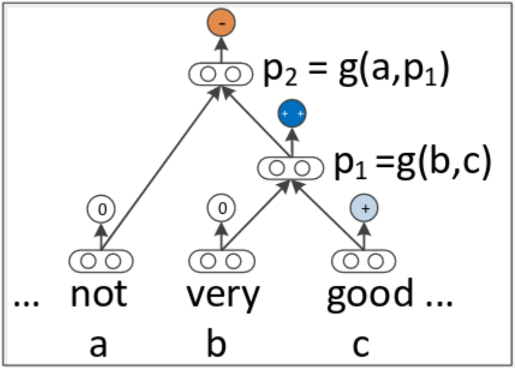


Figure 4: Approach of Recursive Neural Network models for sentiment: Compute parent vectors in a bottom up fashion using a compositionality function g and use node vectors as features for a classifier at that node. This function varies for the different models.

Paragraph vector

- A paragraph vector-embedding model that learns vectors for sequences of words of arbitrary length (e.g, sentences, paragraphs, or documents) without relying on parsing was proposed in [Le and Mikolov, 2014].
- The paragraph vectors are obtained by training a similar network as the one used for training the CBOW embeddings.
- The words surrounding a centre word in a window are used as input together with a paragraph-level vector for predict the centre word.
- The paragraph-vector acts as a memory token that is used for all the centre words in the paragraph during the training the phase.
- The recursive neural tensor network and the paragraph-vector embedding were evaluated on the same movie review dataset used in [Pang et al., 2002], obtaining an accuracy of 85.4% and 87.8%, respectively.
- Both models outperformed the results obtained by classifiers trained on representations based on bag-of-words features.
- Many researchers have have struggled to reproduce these paragraph vectors [Lau and Baldwin, 2016].

Paragraph vector

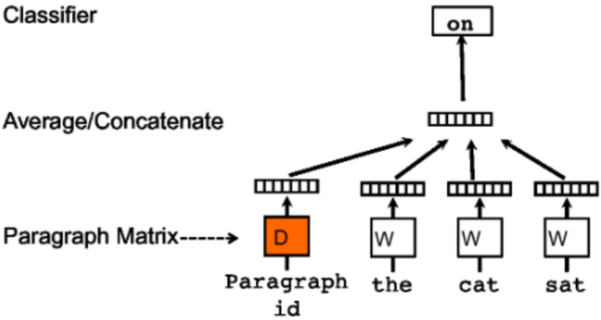


Figure 2. A framework for learning paragraph vector. This framework is similar to the framework presented in Figure 1; the only change is the additional paragraph token that is mapped to a vector via matrix D. In this model, the concatenation or average of this vector with a context of three words is used to predict the fourth word. The paragraph vector represents the missing information from the current context and can act as a memory of the topic of the paragraph.

Summary

- Neural networks are making improvements across many NLP tasks (e.g., sentiment analysis).
- Deep Learning ! $=$ Feature Engineering.
- Word embeddings provide a practical framework for semi-supervised learning (i.e., leveraging unlabelled data).
- Character-level embeddings are worth paying attention to!
- Convolutional neural networks can capture useful features (e.g., n-grams) regardless of the position.
- Recurrent Neural Networks are very useful for learning temporal patterns, especially for long dependencies.
- We just touched the surface!!

Other projects

- WASSA 2017 Shared Task in Emotion Intensity: given a tweet an emotion X (anger, fear, joy, or sadness) determine the intensity or degree of emotion X felt by the speaker—a real-valued score between 0 and 1.
- English tweets were annotated using Best-Worst scaling.
- Twenty-two teams participated. Best system: ensemble of deep learning models (r = 0.74).
- SemEval 2018 Task 1: Affect in Tweets. Extension of previous task including VAD emotions and two more languages: Spanish and Arabic.
- AffectiveTweets Weka Package

Questions?

Thanks for your Attention!

Acknowledgments

- Machine Learning Group at the University of Waikato

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

References I

Amir, S., Ling, W., Astudillo, R., Martins, B., Silva, M. J., and Trancoso, I. (2015).
Inesc-id: A regression model for large scale twitter sentiment lexicon induction.
In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 613–618, Denver, Colorado. Association for Computational Linguistics.

Baroni, M., Dinu, G., and Kruszewski, G. (2014).
Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors.
In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 238–247. Association for Computational Linguistics.

Becker, L., Erhart, G., Skiba, D., and Matula, V. (2013).
Avaya: Sentiment analysis on twitter with self-training and polarity lexicon expansion.
Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013).

Bifet, A. and Frank, E. (2010).
Sentiment knowledge discovery in twitter streaming data.
In *Proceedings of the 13th international conference on Discovery science*, DS'10, pages 1–15, Berlin, Heidelberg. Springer-Verlag.

References II

Bravo-Marquez, F., Frank, E., and Pfahringer, B. (2015).
From unlabelled tweets to twitter-specific opinion words.
In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 743–746, New York, NY, USA. ACM.

Brown, P. F., Desouza, P. V., Mercer, R. L., Pietra, V. J. D., and Lai, J. C. (1992).
Class-based n-gram models of natural language.
Computational linguistics, 18(4):467–479.

Calais Guerra, P. H., Veloso, A., Meira Jr, W., and Almeida, V. (2011).
From bias to opinion: a transfer-learning approach to real-time sentiment analysis.
In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 150–158. ACM.

Conneau, A., Schwenk, H., Barrault, L., and Lecun, Y. (2017).
Very deep convolutional networks for text classification.
In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, volume 1, pages 1107–1116.

References III

Esuli, A. and Sebastiani, F. (2005).
Determining the semantic orientation of terms through gloss classification.
In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, CIKM '05, pages 617–624, New York, NY, USA. ACM.

Esuli, A. and Sebastiani, F. (2006).
Sentiwordnet: A publicly available lexical resource for opinion mining.
In *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06*, pages 417–422.

Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., and Lehmann, S. (2017).
Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm.
In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 1615–1625.

Goldberg, Y. (2016).
A primer on neural network models for natural language processing.
J. Artif. Intell. Res.(JAIR), 57:345–420.

References IV

Goldberg, Y. and Levy, O. (2014).
word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method.
arXiv preprint arXiv:1402.3722.

Guerra, P. C., Meira, Jr., W., and Cardie, C. (2014).
Sentiment analysis on evolving social streams: How self-report imbalances can help.
In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, pages 443–452. ACM.

Guo, W., Li, H., Ji, H., and Diab, M. T. (2013).
Linking tweets to news: A framework to enrich short text data in social media.
In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 239–249. Association for Computational Linguistics.

Harris, Z. (1954).
Distributional structure.
Word, 10(23):146–162.

References V

Hatzivassiloglou, V. and McKeown, K. R. (1997).
Predicting the semantic orientation of adjectives.
In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, ACL '98, pages 174–181, Stroudsburg, PA, USA. Association for Computational Linguistics.

Hu, M. and Liu, B. (2004).
Mining and summarizing customer reviews.
In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '04, pages 168–177, New York, NY, USA. ACM.

Jurafsky, D. and Martin, J. H. (2008).
Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.
Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition.

Kamps, J., Marx, M., Mokken, R. J., and De Rijke, M. (2004).
Using WordNet to Measure Semantic Orientation of Adjectives.
In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, volume 4, pages 1115–1118. European Language Resources Association (ELRA).

References VI

Kim, S.-M. and Hovy, E. (2004).
Determining the sentiment of opinions.
In *Proceedings of the 20th international conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

Lau, J. H. and Baldwin, T. (2016).
An empirical evaluation of doc2vec with practical insights into document embedding generation.
arXiv preprint arXiv:1607.05368.

Le, Q. V. and Mikolov, T. (2014).
Distributed representations of sentences and documents.
In *Proceedings of the 31th International Conference on Machine Learning*, pages 1188–1196.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998).
Gradient-based learning applied to document recognition.
Proceedings of the IEEE, 86(11):2278–2324.

Levy, O. and Goldberg, Y. (2014).
Neural word embedding as implicit matrix factorization.
In *Advances in neural information processing systems*, pages 2177–2185.

References VII

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013).
Distributed representations of words and phrases and their compositionality.
In Burges, C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K., editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.

Mohammad, S. M., Kiritchenko, S., and Zhu, X. (2013).
Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets.
Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013).

Nakov, P., Rosenthal, S., Kozareva, Z., Stoyanov, V., Ritter, A., and Wilson, T. (2013).
Semeval-2013 task 2: Sentiment analysis in twitter.
In *Proceedings of the seventh international workshop on Semantic Evaluation Exercises*, pages 312–320, Atlanta, Georgia, USA. Association for Computational Linguistics.

Pang, B., Lee, L., and Vaithyanathan, S. (2002).
Thumbs up? Sentiment classification using machine learning techniques.
In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, pages 79–86. Association for Computational Linguistics.

References VIII

Pennington, J., Socher, R., and Manning, C. D. (2014).
Glove: Global vectors for word representation.
In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1532–1543.

Read, J. (2005).
Using emoticons to reduce dependency in machine learning techniques for sentiment classification.
In *Proceedings of the ACL Student Research Workshop*, ACLstudent '05, pages 43–48, Stroudsburg, PA, USA. Association for Computational Linguistics.

Severyn, A. and Moschitti, A. (2015).
Twitter sentiment analysis with deep convolutional neural networks.
In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 959–962, New York, NY, USA. ACM.

Silva, I. S., Gomide, J., Veloso, A., Meira, Jr., W., and Ferreira, R. (2011).
Effective sentiment stream analysis with self-augmenting training and demand-driven projection.
In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 475–484, New York, NY, USA. ACM.

References IX

Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., and Potts, C. (2013).
Recursive deep models for semantic compositionality over a sentiment treebank.
In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642. Association for Computational Linguistics.

Tang, D., Wei, F., Qin, B., Zhou, M., and Liu, T. (2014).
Building large-scale twitter-specific sentiment lexicon : A representation learning approach.
In *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*, pages 172–182.

Turian, J., Ratinov, L., and Bengio, Y. (2010).
Word representations: a simple and general method for semi-supervised learning.
In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 384–394. Association for Computational Linguistics.

References X

Turney, P. D. and Littman, M. L. (2003).
Measuring praise and criticism: Inference of semantic orientation from association.
ACM Transactions on Information Systems (TOIS), 21(4):315–346.

Zhou, Z., Zhang, X., and Sanderson, M. (2014).
Sentiment analysis on twitter through topic-based lexicon expansion.
In Wang, H. and Sharaf, M., editors, *Databases Theory and Applications*, volume 8506 of *Lecture Notes in Computer Science*, pages 98–109. Springer International Publishing.