Sentiment and Emotion Analysis in Social Media

Felipe Bravo-Marquez

Department of Computer Science, University of Chile
Millenium Institute Foundational Research on Data

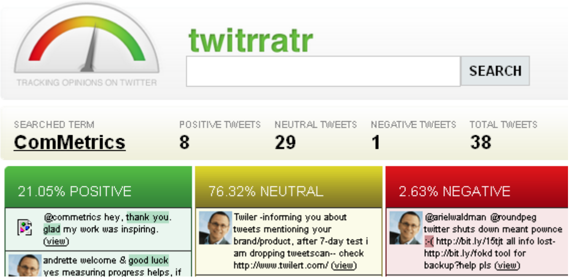
August 11, 2021

CIENCIAS DE LA COMPUTACIÓN
UNIVERSIDAD DE CHILE

Millennium Institute
Foundational
Research on Data

Sentiment Analysis and Social Media

- Opinions are provided **freely and voluntarily** by the users in social media.
- Analysing the sentiment underlying these opinions has important applications in product **marketing** and **politics**.
- Warning 1: Twitter and Facebook are not representative of the general population.
- Warning 2: Mechanisms provided by these plattforms to access posts from public accounts is limited.



Opinion Mining or Sentiment Analysis

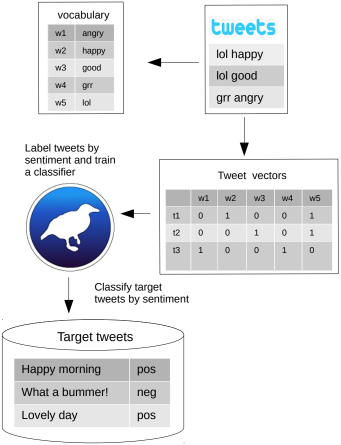
- Application of **NLP** and **machine learning** techniques to identify and extract subjective information from textual datasets.

Main Problem: Message-level Polarity Classification (MPC)

1. Automatically classify a sentence 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



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 data 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- Trained models can become outdated over time.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.

Data Annotation

Crowdsourcing

- Rely on services like **Amazon Mechanical Turk** or **Figure Eight** to ask the **crowds** to label a sample of the data.
- Studies show that crowdsourced annotations can be competitive to expert annotations [Snow et al., 2008].
- Achieving good annotation quality is challenging: how many annotators per sentence? how much to pay? how to consodolite disagreements?
- It is hard to ensure that the annotated sample will cover all the complexities of language usage.

Roadmap

- In the rest of this talk we will overview two dimensions of sentiment analysis.
- First: sentiment analysis is not a single well-defined problem. We will introduce many popular sentiment analysis **tasks**.
- Second: we will overview various techniques used to solve those tasks (e.g, **feature-based** machine learning, **deep learning**).

Fine-grained Sentiment Analysis

Calculate fine-grained sentiment labels for phrases in the parse trees of a sentences [Socher et al., 2013].

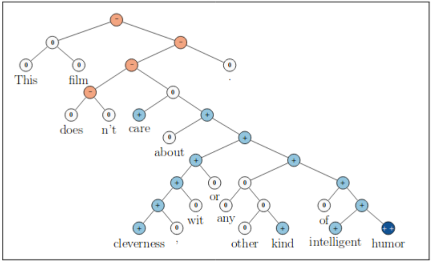


Figure 1: Example of the Recursive Neural Tensor Network accurately predicting 5 sentiment classes, very negative to very positive (– –, –, 0, +, + +), at every node of a parse tree and capturing the negation and its scope in this sentence.

Aspect-based Opinion Mining

Extract fine-grained information with respect to entities mentioned in user comments [Saeidi et al., 2016].

Sentence	Labels
The cheap parts of London are **Edmonton** and **Tottenham** and they are all poor, crime ridden and crowded with immigrants	(Edmonton,price,Positive) (Tottenham,price,Positive) (Edmonton,safety,Negative) (Tottenham,safety,Negative)
Hampstead area, more expensive but a better quality of living than in **Tufnell Park**	(Hampstead,price,Negative) (Hampstead,live,Positive)

Figure: Sample sentences taken from the Sentihood dataset.

Stance Detection

Detect if the author of a tweet is in favor, against or neutral regarding a given target (e.g., Donald Trump).

Emotion Intensity Detection

- Given a tweet and an emotion X (anger, fear, sadness, joy) , determine the intensity or degree of emotion X felt by the speaker – a real-valued score between 0 and 1.

% by Emotion

Emotion	
anger	23.97%
fear	31.73%
joy	22.70%
sadness	21.60%
Grand Total	100.00%

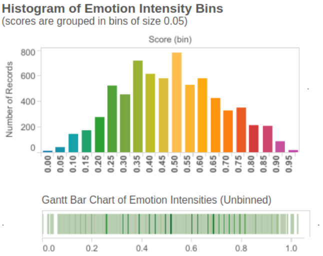
% by TDT

Testflag	
train	50.91%
dev	4.82%
test	44.27%
Grand Total	100.00%

% by Emotion, TDT
(TDT: train, dev, test sets)

Emotion	Testflag	
anger	train	50.38%
	dev	4.94%
	test	44.68%
fear	train	50.93%
	dev	4.88%
	test	44.18%
joy	train	51.09%
	dev	4.59%
	test	44.32%
sadness	train	51.27%
	dev	4.83%
	test	43.90%
Grand Total		100.00%

Histogram of Emotion Intensity Bins
(scores are grouped in bins of size 0.05)



Emotion
- (All)
- anger
- fear
- joy
- sadness

Tweets

Id	Tweet	Emotion	Intensity
10030	This fuck you is boiling up inside, its not gonna be good when I let it out.	anger	0.792
10031	@VodafoneUKhelp @VodafoneUK wow!! My bill is £44.77 and hav a text from u to prove that and you have taken £148!!!!..	anger	0.792
10032	I blame the whole season on Natalie! The season would have been so different had she not turned her back on her allianc..	anger	0.792
10033	Since the 'update' my @iPhone loses power nearly 40% faster. #furious	anger	0.792
10034	@bringyouhome2 I'm about to fly into a fit of rage it's not FAIR	anger	0.792
10035	Unbelievable takes 10 minutes to get through to @BarclaysUK then there's a fault and the call hangs up #fuming #treatcus	anger	0.792

Rule-based systems

- There are various commercial and free rule-based sentiment analysis systems: SentiStrength, Vader, LIWC [1].

LIWC SentiStrength

- These techniques use **opinion or emotion lexicons** together with aggregation rules.
- These rules deal with sentiment patterns such as negation and intensifiers (e.g., I really don't like onions).
- They are not as strong as machine-learning based systems.
- However, rule-based systems have the advantage of being easier to interpret and manipulate.

[1] `http://liwc.wpengine.com`

Feature-based Systems

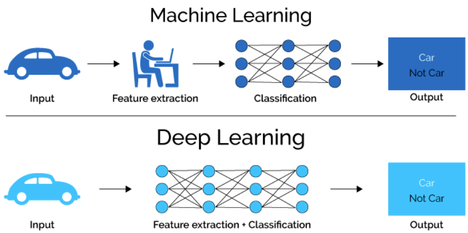
- In 2013, The Semantic Evaluation (SemEval) workshop organised the “Sentiment Analysis in Twitter task” [Nakov et al., 2013].
- The tasks involved the automatic classification of tweets into positive, negative and neutral classes.
- The organisers released training and testing datasets for both tasks. [Nakov et al., 2013]
- The team that achieved the highest performance in both tasks among 44 teams was the *NRC-Canada* team [Mohammad et al., 2013].
- The team proposed a supervised approach using a linear SVM classifier with the following hand-crafted features for representing tweets.

Feature-based Systems

1. Word *n*-grams.
2. Character *n*-grams.
3. Part-of-speech tags.
4. Word clusters trained with the Brown clustering method [Brown et al., 1992].
5. The number of elongated words (words with one character repeated more than two times).
6. The number of words with all characters in uppercase.
7. The presence of positive or negative emoticons.
8. The number of individual negations.
9. The number of contiguous sequences of dots, question marks and exclamation marks.
10. Features derived from polarity lexicons [Mohammad et al., 2013].

Feature Engineering and Deep Learning

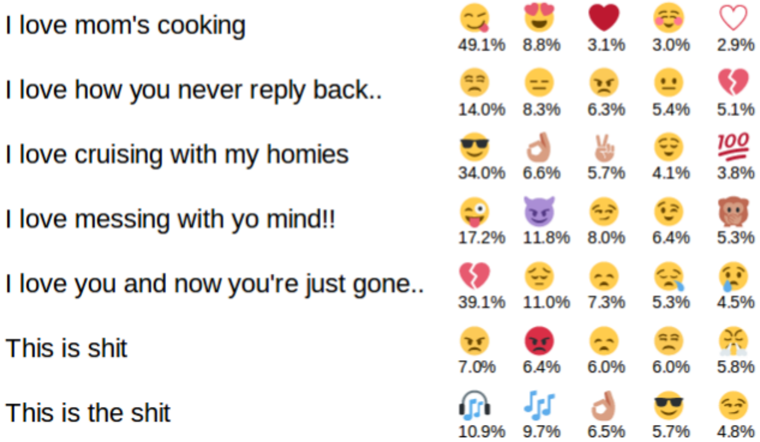
- Designing the features of a winning NLP system requires a lot of domain-specific knowledge.
- The NRC system was built before deep learning became popular in NLP.
- Deep Learning systems on the other hand rely on neural networks to automatically learn good representations.



Feature Engineering and Deep Learning

- Deep Learning yields state-of-the-art results in most NLP tasks.
- Large amounts of training data and faster multicore GPU machines are key in the success of deep learning.
- **Neural networks** and **word embeddings** play a key role in modern sentiment analysis models.

DeepEmoji


I love mom's cooking
49.1% 8.8% 3.1% 3.0% 2.9%
I love how you never reply back..
14.0% 8.3% 6.3% 5.4% 5.1%
I love cruising with my homies
34.0% 6.6% 5.7% 4.1% 3.8%
I love messing with yo mind!!
17.2% 11.8% 8.0% 6.4% 5.3%
I love you and now you're just gone..
39.1% 11.0% 7.3% 5.3% 4.5%
This is shit
7.0% 6.4% 6.0% 6.0% 5.8%
This is the shit
10.9% 9.7% 6.5% 5.7% 4.8%

Final Comments

- There is no single definition of sentiment analysis.
- It is hard to set universal criteria for the sentiment of a sentence.
- Training data is a bottleneck.
- Machine learning-based models reflect the distribution of the corpus on which they were trained.
- Models that work well on a dataset won't necessarily work well on another one.
- Models can be biased or incomplete.
- I do not recommend to trivially aggregate the output of sentiment analysis models deployed on social media data to monitor public opinion.

Questions?

Thanks for your Attention!

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.

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.

References II

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.

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

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.

References III

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.

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.

References IV

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.

Saeidi, M., Bouchard, G., Liakata, M., and Riedel, S. (2016).
SentiHood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods.
In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1546–1556, Osaka, Japan. The COLING 2016 Organizing Committee.

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.

References V

Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. (2008).
Cheap and fast – but is it good? evaluating non-expert annotations for natural language tasks.
In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 254–263, Honolulu, Hawaii. Association for Computational Linguistics.

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.

References VI

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.