Explain sentiments using Conditional Random Field and a Huge Lexical Network

Mike Donald TAPI NZALI
I3M, Univ. Montpellier
Montpellier, France
LIRMM, Univ. Montpellier
Montpellier, France
mdtapinz@univ-montp2.fr
tapinzali@lirmm.fr

Joël MAÏZI
Pierre POMPIDOR
Sandra BRINGAY
LIRMM, Univ. Montpellier
Montpellier, France

Christian LAVERGNE
I3M, Univ. Montpellier
Montpellier, France
Caroline MOLLEVI
Biostatistics Unit, ICM
Montpellier, France

Abstract

In this paper, we focus on a particular task which consists in explaining the source and the target of sentiments expressed in social networks. We propose a method for French, which overcomes a fine syntactic parsing and successfully integrate the Conditional Random Field (CRF) method and a smart exploration of a very large lexical network. Quantitative and qualitative experiments were performed on real dataset to validate this approach.

1 Introduction

In this article, we focus on a particular task of sentiment analysis which consists in explaining the target and the source of sentiments. For example, in the sentence "*We like the green initiative*", the sentiment is expressed by the verb "*like*", the target is "*green initiative*" and the source is "*We*". In (Bringay et al., 2014), we have proposed a method based on syntactic roles for English texts. Experiments have shown that our method is robust, even on the texts that are difficult to process : messages in health forums that contain misspelling and slang. Indeed, the method is not based on fine syntactic parsing. However, it is not possible to transpose this method directly to French because, to our knowledge, there is no resource available to explain semantic roles.

In this context, we propose a new approach based on machine learning methods and a very large lexical network, in French, issued from a contributory game *JeuxDeMots*[1]. The challenge is twofold: 1) Instead of using fine syntactic parsing, we use a statistical modeling method called Conditional Random Field (CRF), to extract candidates for targets and sources in large volumes of texts issued from poorly written social web messages 2) we also exploit the huge French lexical network *JeuxDeMots* (more than $300,000$ nodes and $7,000,000$ relations) to choose the best sources and targets among the candidates identified with CRF. This new method has been successfully exploited to analyse the sentiments expressed in the French tweets dealing with environment and climate change.

The paper is organized as follows. In section 2, we briefly present the state of the art. In section 3, we provide a description of our method. In section 4, we provide all the detail of the experiments carried out and the prime results. Finally in section 5, we conclude this work by providing the main perspectives associated with this work.

2 State of the Art

Since the early 2000s, sentiment analysis, also called "opinion mining", has experienced growing interest. Many methods have been developed to extract emotional states expressed or implied in texts. To identify sentiments, many resources exist (e.g. list of words, phrases, idioms), which were built mostly for English and polarity (e.g. *Linguistic Inquiry and Word Count* (Tausczik and Pennebaker, 2010)) or emotions (e.g. *NRC* lexicon (Mohammad and Turney, 2010a)). Some methods extend these vocabularies to specific domains (Neviarouskaya et al., 2011). Others are not restricted to the use of lexicons as (Strapparava and Mihalcea, 2008) who implement learning approaches.

Two categories of approaches are used to link sentiments and potential target and source. 1) Methods that essentially implement syntactical aspects, represented by combinations of rules (Mudinas et al., 2012) as the polarity inverters (*do not, just, very...*), conjunctions (*but, or*), etc. The effectiveness of these methods is strongly linked to language style that impacts on the syntactic rules

[1] http://www.jeuxdemots.org/jdm-accueil.php

to take into account and are not adapted to social web texts. 2) Methods that are based on different distance computations between words denoting sentiments and potential targets and source (as the proximity (Hu and Liu, 2004) or the position in the syntactic tree (Wu et al., 2009)). There are also many hybrid methods (Ding and Liu, 2007). In (Bringay et al., 2014), we proposed an efficient approach for English texts that requires a resource *FrameNet*[2] and the *SEMAFOR* parser [3] for explaining the semantic roles. To our best knowledge, such a resource does not exist in French. Consequently, we have proposed a method combining learning approach to find targets and sources candidates and a smart exploration of a large French lexical network to select the best one.

To choose the best candidate, we use the *Games with a purpose JeuxDeMots*, created in 2007 (Lafourcade and Joubert, 2012), to build a huge lexical network for french. For example, the game asks the player ideas associated with term *climatic change*. The player freely associate terms such as *bear*. Other players have already faced the same term. The player wins credits if the proposed term has already been proposed by another player. The more the proposal is specific, the more points he obtains. The lexical network generated with this game is a directed graph, with terms (nodes) and typed and weighted relations (edges) between the terms. There are more than 50 types of relationships. To weight the edges, *JeuxDeMots* is based on crowdsourcing. Each relation is weighted by a *strength of association*, denoted C_{jdm} representing the number of players who have associated two terms by the same relation. A first challenge is to explore the network to link terms in the sentences (sentiment and target/source) and explain these relations. The second challenge will be to exploit this very large network that includes more than 300000 terms and more than 7000000 relations.

3 Methods

The method is organised into 3 steps :

Step 1: Corpus. The corpus we used and annotations have been made in the Ucomp project[4]. These tweets deal with climate change. Table 1 and 2 present detailed statistics on the corpus.

Step 2: candidates generation with CRF. The CRF model was developed with domain independent surface and lexical features for the text tokens:

Class	**Learning step**		**Test**	
	#	%	#	%
Source	2448	31	1057	31
target	1875	24	804	24
Total	**7867**	**55**	**1861**	**55**

Table 1: Distribution of *source* and *target* in the corpus used

- The original token from the text (word form);
- Surface features: capitalization of the token (all in upper/lower case, combination of both), and punctuation mark in the token (PUNCT, NO_PUNCT);
- Lexical features: *n*-grams, number of consecutive repeats. Token frequency was computed based on the entire training corpus.
- Brown clustering: we used Percy Liang's implementation of Brown clustering (Brown et al., 1992), which is an HMM-based algorithm. In our work, we partition words into a base set of 100 clusters, and induces a hierarchy among those 100 clusters.
- Emotion lexicon: We built semi-automatically a new lexicon of French sentiments (Amine et al., 2014) by translating and expanding the English NRC lexicon (Mohammad and Turney, 2010b). This lexicon is free to download [5]. For each tokens, the corresponding feature takes the value "Yes" if the token appear in the lexicon and "No" otherwise. As a *source* and a *target* are usually surrounded by a sentiment token, we also consider the apparition of the sentiment in the neighborhood of the token (e.g. two tokens before or after the current token).

We experimented with standard tokenization (provided by TreeTagger) and custom tokenization (Tapi-Nzali et al., 2015) of French TreeTagger by

[2] https://FrameNet.icsi.berkeley.edu/fndrupal/home

[3] http://demo.ark.cs.cmu.edu/parse

[4] http://www.ucomp.eu/

[5] http://www.lirmm.fr/~abdaoui/FEEL.html

adding some segmentation rules (e.g. apostrophe : *l'image* is segmented into *l'* and *image*).

Step 3: Lexical network construction of each sentence. The purpose of this step is to extract a part of the lexical network *JeuxDeMots* representing the relationship between the meaning of the word and the candidates identified in Step 2. The intuition of the algorithm is the following one. We cross the lexical network from node to node. We stop when we no longer encounter new words or if we reach a maximum depth. Two other constraints are used to limit the expansion of the graph.

Constraint 1. To consider only the parts of the network related to our topic (environment and sentiment), we expand a node to another if the new one belong to these two predefined lexical fields chosen via Larousse thesaurus[6]. If there is no node in the lexical field, we expand to all neighboring nodes.

Constraint 2. We use the *association strength* C_{jdm} weighting the edges and consider only the relations frequently instantiated by players. A threshold is set by default.

Step 4: Identification of shortest paths. The objective of this step is to identify in the graph generated in step 3, the paths that must correspond to a compromise between the shortest paths, with a little depth, most reliable according to *strength of association* C_{jdm}. We have therefore redefined weights w_{rt} to foster some relationships like synonymy or significant semantic roles such as *patient* and *agent*. To identify the paths, we have adapted the shortest path algorithm and used the weights computed according to formulas 1 and 2. The weight w_i foster relationships that interest us with w_{rt} while taking into account the *strength of association* C_{jdm}. In equation 1, we verify a balance between w_{rt} and C_{jdm} terms. In equation 2, the term $(n-1)^2$ enables to penalize depth. We only consider paths which contain at least one *agent* or *patient* relationship. The path with the best score is proposed to the user to explain the link between the sentiment and the target or source.

$$w_1 = \frac{1}{1 - \frac{1}{C_{jdm}}} + w_{rt} \tag{1}$$

[6]http://www.larousse.fr/dictionnaires/francais/thesaurus/77857

$$w_n = (max(w_{n-1}) * \frac{1}{1 - \frac{1}{C_{jdm}}} + w_{rt}) * (n-1)^2 \tag{2}$$

4 Experiments

Our experiments were carried out using 10-fold cross-validation. To do this, the training corpus was divided into 10 folds. To build our model, we need a training, development and test corpus. Cross-validation has been distributed as follows: The model is built on 8 folds, the optimization of the construction is performed on the ninth part (development) and the model evaluation performed on the last fold (test).

To perform our experiments, we use Wapiti[7](Lavergne et al., 2010). It is a very fast toolkit for segmenting and labelling sequences with discriminative models. For the iterative estimation of the model parameters, we used the algorithm RPROP (Riedmiller and Braun, 1992).

Table 3 presents the results obtained by different CRF models on training set by cross validation. The features of four bests configurations are :

- **Configuration 1** : Part Of Speech tagging + lemmatization + lowercase
- **Configuration 2** : Part Of Speech tagging + lemmatization + lowercase + brown clustering
- **Configuration 3** : All (Part Of Speech tagging + lemmatization + lowercase + brown clustering + emotion lexicon
- **Configuration 4** : Part Of Speech tagging + lemmatization + lowercase + emotion lexicon

	Training	**Test**	**All**
Tweets	3,001	1,783	4,784
Tokens	78,771	48,612	127,383
Source	1,131	604	1,735
Target	3,954	2,251	6,205

Table 2: Description of the corpus

The results of the evaluation are reported in terms of precision (the number of *source* and *target* correctly extracted over the total number of *source* and *target* extracted), recall (the number

[7]http://wapiti.limsi.fr

Test	Model	Exact match			Partial match		
		P	R	F	P	R	F
Corpus used	Config 1	**0.40**	**0.27**	**0.32**	0.76	**0.52**	**0.62**
	Config 2	0.39	0.26	0.31	0.76	**0.52**	**0.62**
	Config 3	0.40	0.25	0.30	0.77	0.48	0.59
	Config 4	0.40	0.24	0.30	**0.78**	0.47	0.59

Table 3: Evaluation of *source* and *target* extraction in French tweet corpus.

Class	Exact match			Partial match		
	P	R	F	P	R	F
SOURCE	0.64	0.38	0.48	0.76	0.45	0.57
TARGET	0.34	0.20	0.25	0.79	0.47	0.59
All	0.40	0.24	0.30	0.78	0.47	0.59

Table 4: Results of best model on the test corpus

of *source* and *target* correctly extracted over the total number of *source* and *target* marked the corpus used) and F-measure (the harmonic average of precision and recall). We show two types of results. the first is the results achieved by *Exact match* and the second by *Partial match*. We consider that there is an ***Exact match*** when the tokens obtained with our model match exactly those of the standard test annotation and we consider a ***Partial match*** when the obtained token are included. For example, *governor* partially matches *The governor*. Overall, with *Exact match*, configuration 1 is the best performing configuration. Results show that, we performed a good results with *Partial match*. Compared to other configurations, configuration 4 gives the best results on precision (Precision 0.78), and configuration 1 and 2 give the same results and the best results on recall and F-measure (recall 0.52 and F-measure 0.62). Contrary, with the sentiment lexicon as feature, we increase precision, decrease recall and f-measure. Brown Clustering is good feature if we want to have a good precision.

If CRF is relevant for extracting target and source candidates, how can we link them to the sentiments also expressed in the sentences? In figure 1, a sentence is annotated after the exploration of the lexical network. Sentiment tokens are represented by red points. Target and Source obtained with CRF are colored (in blue and yellow). Arrows correspond to the paths identified in the network. The more the arrow is thick the more the path is valuated.

5 Conclusions and Future Work

A combination of CRF and huge lexical network exploration seems promising for explaining sentiments in social networks. By experimenting with the CRF model, we found that the results varied depending on the features. The best results are obtained with the features: *lemmatization, cluster ID and Part Of Speech tagging*.

The first advantage of this method is that we can detect multiple tokens (e.g. *parc eolien terrestre*, *La France*). Another advantage of this method is that it is efficient even if the sentence contains misspelling. For example, the system identify *modèle de dévelopement durable* (sustainable development model) even if the word *dévelopement* is misspelled. Finally, the main advantage of this contribution is not to restrict sentiment, source and target identification to the case in which sentiment word is present. Indeed, in most cases people express sentiments implicitly without using these sentiment words. An emotion cannot be limited to something a person feels about a fact and not the sentiment that a person expresses about this fact. Thus, it could be common to explicitly express sentiments about things, but it is more common to feel emotions without expressing them explicitly. Our method take into account this fact and try to identify source and target beyond the explicit cases.

The principal limitation of our method is the length of the sentences in the considered corpora (size of the tweets). In many sentences, there is no source or no target. Results are significantly reduced. Moreover, a quantitative study has to be performed on step 3 and 4 to evaluate the quality of the computed relations between sentiment and targets/sources.

Prospects associated with this work are numerous. First, in this work we focus only on the targets/sources expressed in sentences and we now have to focus on inter-sentence relationships at paragraph level. In future work, we are going to use the best model we obtained on health forum messages with longer sentences. We will also compare our method to identify relations between sentiments and source/target with the methods of the state of the art. We will also adapt CRF to extract directly relations. Finally, we will present to users the part of the network used to identify relations in order to help their interpretation

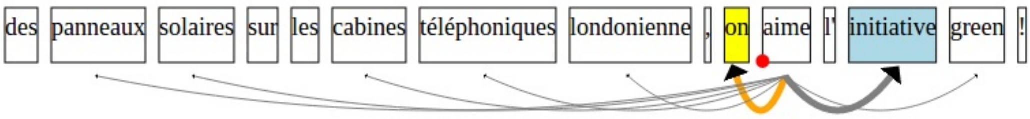

Figure 1: Relation example

References

Abdaoui Amine, Azé Jérôme, Sandra Bringay, and Pascal Poncelet. 2014. Feel : French extended emotional lexicon. volume ISLRN: 041-639-484-224-2. ELRA Catalogue of Language Resources.

Sandra Bringay, Eric Kergosien, Pierre Pompidor, and Pascal Poncelet. 2014. Identifying the targets of the emotions expressed in health forums. In *CICLing (2)*, pages 85–97.

Peter F Brown, Peter V Desouza, Robert L Mercer, Vincent J Della Pietra, and Jenifer C Lai. 1992. Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–479.

Xiaowen Ding and Bing Liu. 2007. The utility of linguistic rules in opinion mining. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, pages 811–812, New York, NY, USA. ACM.

Minqing Hu and Bing Liu. 2004. Mining opinion features in customer reviews. In *Proceedings of the 19th National Conference on Artifical Intelligence*, AAAI'04, pages 755–760. AAAI Press.

Mathieu Lafourcade and Alain Joubert. 2012. Increasing Long Tail in Weighted Lexical Networks. In *Cognitive Aspects of the Lexicon (CogAlex-III), COLING*, page 16, France, December.

Thomas Lavergne, Olivier Cappé, and Fran¸cois Yvon. 2010. Practical very large scale crfs. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 504–513, Uppsala, Sweden.

Saif M. Mohammad and Peter D. Turney. 2010a. Emotions Evoked by Common Words and Phrases : Using Mechanical Turk to Create an Emotion Lexicon. In *Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 26–34, Stroudsburg, PA, USA. ACL.

Saif M. Mohammad and Peter D. Turney. 2010b. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, CAAGET '10, pages 26–34, Stroudsburg, PA, USA. Association for Computational Linguistics.

Andrius Mudinas, Dell Zhang, and Mark Levene. 2012. Combining lexicon and learning based approaches for concept-level sentiment analysis. In *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*, WISDOM '12, pages 5:1–5:8, New York, NY, USA. ACM.

Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. 2011. Affect analysis model: Novel rule-based approach to affect sensing from text. volume 17, pages 95–135, New York, NY, USA, January. Cambridge University Press.

Martin Riedmiller and Heinrich Braun. 1992. Rprop-a fast adaptive learning algorithm. In *Proc. of ISCIS VII), Universitat*. Citeseer.

Carlo Strapparava and Rada Mihalcea. 2008. Learning to identify emotions in text. In *Proceedings of the 2008 ACM Symposium on Applied Computing*, SAC '08, pages 1556–1560, New York, NY, USA. ACM.

Mike Donald Tapi-Nzali, Aurélie Névéol, and Xavier Tannier. 2015. Analyse d'expressions temporelles dans les dossiers électroniques patients. In *Actes de la Conférence Traitement Automatique des Langues Naturelles (TALN 2015)*, Caen, France, June.

Yla R. Tausczik and James W. Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. volume 29, pages 24–54.

Yuanbin Wu, Qi Zhang, Xuanjing Huang, and Lide Wu. 2009. Phrase dependency parsing for opinion mining. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, pages 1533–1541, Stroudsburg, PA, USA. Association for Computational Linguistics.