10.12871/CLICIT2014173

A context based model for Sentiment Analysis in Twitter for the Italian Language

Andrea Vanzo(†), **Giuseppe Castellucci**(‡), **Danilo Croce**(†) **and Roberto Basili**(†)
(†)Department of Enterprise Engineering
(‡)Department of Electronic Engineering
University of Roma, Tor Vergata
Via del Politecnico 1, 00133 Roma, Italy
{vanzo,croce,basili}@info.uniroma2.it, castellucci@ing.uniroma2.it

Abstract

English. Recent works on Sentiment Analysis over Twitter leverage the idea that the sentiment depends on a single incoming tweet. However, tweets are plunged into streams of posts, thus making available a wider context. The contribution of this information has been recently investigated for the English language by modeling the polarity detection as a sequential classification task over streams of tweets (Vanzo et al., 2014). Here, we want to verify the applicability of this method even for a morphological richer language, i.e. Italian.

Italiano. *Studi recenti per la Sentiment Analysis in Twitter hanno tentato di creare modelli per caratterizzare la polaritá di un tweet osservando ciascun messaggio in isolamento. In realtà, i tweet fanno parte di conversazioni, la cui natura può essere sfruttata per migliorare la qualità dell'analisi da parte di sistemi automatici. In (Vanzo et al., 2014) è stato proposto un modello basato sulla classificazione di sequenze per la caratterizzazione della polarità dei tweet, che sfrutta il contesto in cui il messaggio è immerso. In questo lavoro, si vuole verificare l'applicabilità di tale metodologia anche per la lingua Italiana.*

1 Introduction

Web 2.0 and Social Networks allow users to write about their life and personal experiences. This huge amount of data is crucial in the study of the interactions and dynamics of subjectivity on the Web. Sentiment Analysis (SA) is the computational study and automatic recognition of opinions and sentiments. Twitter is a microblogging service that counts about a billion of active users. In Twitter, SA is traditionally treated as any other text classification task, as proved by most systems participating to the *Sentiment Analysis in Twitter* task in SemEval-2013 (Nakov et al., 2013). A Machine Learning (ML) setting allows to induce detection functions from real world labeled examples. However, the shortness of the message and the resulting semantic ambiguity represent a critical limitation, thus making the task very challenging. Let us consider the following message between two users:

`Benji`: *@Holly sono completamente d'accordo con te*

The tweet sounds like to be a reply to the previous one. Notice how no lexical or syntactic property allows to determine the polarity. Let's look now at the entire conversation:

`Benji` : *@Holly con un #RigoreAl90 vinci facile!!*
`Holly` : *@Benji Lui vince sempre però :) accanto a chiunque.. Nessuno regge il confronto!*
`Benji` : *@Holly sono completamente d'accordo con te*

The first is clearly a positive tweet, followed by a positive one that makes the third positive as well. Thus, through the conversation we can disambiguate even a very short message. We want to leverage on this to define a context-sensitive SA model for the Italian language, in line with (Vanzo et al., 2014). The polarity detection of a tweet is modeled as a sequential classification task through the SVM^{hmm} learning algorithm (Altun et al., 2003), as it allows to classify an instance (i.e. a tweet) within an entire sequence. First experimental evaluations confirm the effectiveness of the proposed sequential tagging approach combined with the adopted contextual information even in the Italian language.

A survey of the existing approaches is presented in Section 2. Then, Section 3 provides an account of the context-based model. The experimental evaluation is presented in Section 4.

2 Related Work

The spread of microblog services, where users post real-time opinions about "everything", poses different challenges in Sentiment Analysis. Classical approaches (Pang et al., 2002; Pang and Lee, 2008) are not directly applicable to tweets: they focus on relatively large texts, e.g. movie or product reviews, while tweets are short and informal and a finer analysis is required. Recent works tried to model the sentiment in tweets (Go et al., 2009; Davidov et al., 2010; Bifet and Frank, 2010; Zanzotto et al., 2011; Croce and Basili, 2012; Si et al., 2013). Specific approaches, e.g. probabilistic paradigms (Pak and Paroubek, 2010) or Kernel based (Barbosa and Feng, 2010; Agarwal et al., 2011; Castellucci et al., 2013), and features, e.g. *n*-grams, POS tags, polarity lexicons, have been adopted in the tweet polarity recognition task.

In (Mukherjee and Bhattacharyya, 2012) contextual information, in terms of discourse relations is adopted, e.g. the presence of conditionals and semantic operators like *modals* and *negations*. However, these features are derived by considering a tweet in isolation. The approach in (Vanzo et al., 2014) considers a tweet within its context, i.e. the stream of related posts. In order to exploit this information, a Markovian extension of a Kernel-based categorization approach is there proposed and it is briefly described in the next section.

3 A Context Based Model for SA

As discussed in (Vanzo et al., 2014), contextual information about one tweet stems from various aspects: an explicit conversation, the user attitude or the overall set of recent tweets about a topic (for example a hashtag like *#RigoreAl90*). In this work, we concentrate our analysis only on the explicit conversation a tweet belongs to. In line with (Vanzo et al., 2014), a conversation is a sequence of tweets, each represented as vectors of features characterizing different semantic properties. The Sentiment Analysis task is thus modeled as a sequential classification function that associates tweets, i.e. vectors, to polarity classes.

3.1 Representing Tweets

The proposed representation makes use of different representations that allow to model different aspects within a Kernel-based paradigm.

Bag of Word (BoWK). The simplest Kernel function describes the lexical overlap between tweets, thus represented as a vector, whose dimensions correspond to the presence or not of a word. Even if very simple, the BoW model is one of the most informative representation in Sentiment Analysis, as emphasized since (Pang et al., 2002).

Lexical Semantic Kernel (LSK). In order to generalize the BoW model, we provide a further representation. A vector for each word is obtained from a co-occurrence Word Space built according to the Distributional Analysis technique (Sahlgren, 2006). A word-by-context matrix M is built through large scale corpus analysis and then processed through *Latent Semantic Analysis* (Landauer and Dumais, 1997). Dimensionality reduction is applied to M through Singular Value Decomposition (Golub and Kahan, 1965): the original statistical information about M is captured by the new k-dimensional space, which preserves the global structure while removing low-variance dimensions, i.e. distribution noise. A word can be projected in the reduced Word Space: the distance between vectors surrogates the notion of paradigmatic similarity between represented words, e.g. the most similar words of *vincere* are *perdere* and *partecipare*. A vector for each tweet is represented through the linear combination of its word vectors.

Whenever the different representations are available, we can combine the contribution of both vector simply through a juxtaposition, in order to exploit both lexical and semantic properties.

3.2 SA as a Sequential Tagging Problem

Contextual information is embodied by the stream of tweets in which a message t_i is immersed. A stream gives rise to a sequence on which sequence labeling can be applied: the target tweet is here labeled within the entire sequence, where contextual constraints are provided by the preceding tweets. Let formally define a conversational context.

Conversational context. For every tweet $t_i \in \mathcal{T}$, let $r(t_i) : \mathcal{T} \rightarrow \mathcal{T}$ be a function that returns either the tweet to which t_i is a reply to, or $null$ if t_i is not a reply. Then, the *conversational context* $\Lambda_i^{C,l}$ of tweet t_i (i.e., the *target tweet*) is the sequence of tweet iteratively built by applying $r(\cdot)$, until l tweets have been selected or $r(\cdot) = null$.

A markovian approach. The sentiment prediction of a target tweet can be seen as a sequential classification task over a context, and the *SVM*^{hmm} algorithm can be applied. Given an input sequence $\mathbf{x} = (x_1 \dots x_l) \subseteq \mathcal{X}$, where $\mathbf{x}$ is a tweet context, i.e. the *conversational context* previously defined, and x_i is a feature vector representing a tweet, the model predicts a tag sequence $\mathbf{y} = (y_1 \dots y_l) \in \mathcal{Y}^+$ after learning a linear discriminant function $F : \mathcal{P}(\mathcal{X}) \times \mathcal{Y}^+ \rightarrow \mathbb{R}$ over input/output pairs. The labeling $f(\mathbf{x})$ is defined as: $f(\mathbf{x}) = \arg\max_{\mathbf{y}\in\mathcal{Y}^+} F(\mathbf{x}, \mathbf{y}; \mathbf{w})$. In these models, F is linear in some combined feature representation of inputs and outputs $\Phi(\mathbf{x}, \mathbf{y})$, i.e. $F(\mathbf{x}, \mathbf{y}; \mathbf{w}) = \langle \mathbf{w}, \Phi(\mathbf{x}, \mathbf{y}) \rangle$. As Φ extracts meaningful properties from an observation/label sequence pair $(\mathbf{x}, \mathbf{y})$, in *SVM*^{hmm} it is modeled through two types of features: interactions between attributes of the observation vectors x_i and a specific label y_i (i.e. **emissions** of a tweet w.r.t. a polarity class) as well as interactions between neighboring labels y_i along the chain (i.e. **transitions** of polarity labels in a conversation context.). Thus, through *SVM*^{hmm} the label for a target tweet is made dependent on its context history. The markovian setting acquires patterns across tweet sequences to recognize sentiment even for truly ambiguous tweets. Further details about the modeling and the *SVM*^{hmm} application to tweet labeling can be found in (Vanzo et al., 2014).

4 Experimental Evaluation

The aim of the experiments is to verify the applicability of the model proposed in (Vanzo et al., 2014) in a different language, i.e. Italian. In order to evaluate the models discussed above in an Italian setting, an appropriate dataset has been built by gathering[1] tweets from Twitter servers. By means of Twitter APIs[2], we retrieved the whole corpus by querying several Italian hot topics, i.e. `expo`, `mose`, `renzi`, `prandelli`, `mondiali`, `balotelli` and commonly used emoticons, i.e. `:)` and `:(` smiles. Each tweet t_i and its corresponding conversation $\Lambda_i^{C,l}$ have been included into the dataset if and only if the conversation itself was available (i.e. $|\Lambda_i^{C,l}| > 1$). Then, three annotators labeled each tweet with a sentiment polarity label among `positive`, `negative`, `neutral` and `conflicting`[3], obtaining a inter-annotator agreement of 0.829, measured as the mean accuracy computed between annotators pairs.

[1]The process has been run during June-July 2014

[2]`http://twitter4j.org/`

[3]A tweet is said to be conflicting when it expresses both a positive and negative polarity

As about 1,436 tweets, including conversations, were gathered from Twitter, a static split of 64%/16%/20% in *Training/Held-out/Test* respectively, has been carried out as reported in Table 1.

	train	**dev**	**test**
Positive	212	61	69
Negative	211	42	92
Neutral	387	72	87
Conflicting	129	26	48
	939	201	296

Table 1: Dataset composition

Tweets have been analyzed through the *Chaos* natural language parser (Basili et al., 1998). A normalization step is previously applied to each message: fully capitalized words are converted in lowercase; reply marks, hyperlinks and hashtags are replaced with the pseudo-tokens, and emoticons have been classified with respect to 13 different classes. LSK vectors are obtained from a Word Space derived from a corpus of about 3 million tweets, downloaded during July and September 2013. The methodology described in (Sahlgren, 2006) with the setting discussed in (Croce and Previtali, 2010) has been applied.

Performance scores are reported in terms of Precision, Recall and F-Measure. We also report both the F_1^{pnn} score as the arithmetic mean between the F_1s of *positive*, *negative* and *neutral* classes, and the F_1^{pnnc} considering even the *conflicting* class. It is worth noticing that a slightly different setting w.r.t. (Vanzo et al., 2014) has been used. In this work we manually labeled every tweet in each conversation and performance measures considers all the tweets. On the contrary in (Vanzo et al., 2014) only the last tweet of the conversation is manually labeled and considered in the evaluation.

4.1 Experimental Results

Experiments are meant to verify the ability of a context-based model in the Italian setting. As a baseline we considered a multi-class classifier within the *SVM*^{multiclass} framework (Tsochantaridis et al., 2004). Each tweet in a conversation is classified considering it in isolation, i.e. without using contextual information. In Table 2, performances of the Italian dataset are reported, while Table 3 shows the outcomes of experiments over the English dataset (Vanzo et al., 2014). Here, *w/o conv* results refer to a baseline computed with the *SVM*^{multiclass} algorithm, while *w/ conv* results refer to the application of the model described in the

	Precision				Recall				$\mathbf{F_1}$				$\mathbf{F_1^{pnn}}$	$\mathbf{F_1^{pnnc}}$
	pos	*neg*	*neu*	*conf*	*pos*	*neg*	*neu*	*conf*	*pos*	*neg*	*neu*	*conf*		
BoWK														
w/o conv	.705	.417	.462	.214	.449	.109	.690	.438	.549	.172	.553	.288	.425	.390
w conv	.603	.580	.379	.375	.507	.435	.701	.063	.551	.497	.492	.107	.513	.412
BoWK+LSK														
w/o conv	.507	.638	.416	.000	.493	.402	.793	.000	.500	.493	.545	.000	.513	.385
w conv	.593	.560	.432	.368	.464	.457	.736	.146	.520	.503	.545	.209	.523	.444

Table 2: Evaluation results of the Italian setting.

	Precision			Recall			$\mathbf{F_1}$			$\mathbf{F_1^{pnn}}$
	pos	*neg*	*neu*	*pos*	*neg*	*neu*	*pos*	*neg*	*neu*	
BoWK										
w/o conv	.713	.496	.680	.649	.401	.770	.679	.444	.723	.615
w/ conv	.723	.511	.722	.695	.472	.762	.709	.491	.741	.647
BoWK+LSK										
w/o conv	.754	.595	.704	.674	.486	.804	.712	.535	.751	.666
w/ conv	.774	.554	.717	.682	.542	.791	.725	.548	.752	.675

Table 3: Evaluation results on the English language from (Vanzo et al., 2014)

previous sections with the *SVM*^{hmm} algorithm. In the last setting, the whole *conversational context* of each tweet is considered.

Firstly, all *w/o conv* models beneficiate by the lexical generalization provided by the Word Space in the LSA model. In fact, the information derived from the Word Space seems beneficial in its relative improvement with respect to the simple BoW Kernel accuracy, up to an improvement of 20.71% of $\mathbf{F_1^{pnn}}$, from .425 to .513. However, it is not always true, in particular w.r.t. the conflicting class where the smoothing provided by the generalization negatively impact on the classifiers, that are not able to discriminate the contemporary presence of positive and negative polarity.

Most importantly, the contribution of conversations is confirmed in all context-driven models, i.e. *w/conv* improves w.r.t. their *w/o conv* counterpart. Every polarity category benefits from the introduction of contexts, although many tweets annotated with the conflicting (*conf*) class are not correctly recognized: contextual information unbalances the output of a borderline tweet with the polarity of the conversations. The impact of conversational information contribute to a statistically significant improvement of 20.71% in the BoWK setting, and of 1.95% in the BoWK+LSK setting.

In (Vanzo et al., 2014) a larger dataset (10,045 examples) has been used for the evaluation of contextual models in an English setting. The dataset is provided by *ACL SemEval-2013* (Nakov et al., 2013). Results are thus not directly comparable, as in this latter dataset, where even tweets without a conversational contexts are included, only the target tweet is manually labeled and the labels of remaining tweets have been automatically predicted in a semi supervised fashion, as discussed in (Vanzo et al., 2014). Additionally, the conflicting class, where a lexical overlap is observed with both positive and negative classes, is not considered. However, results in Table 3 show that the BoWK setting benefits by the introduction of the lexical generalization, given by the LSK, with a performance improvement of 8.29%. When the focus is held within the same Kernel setting, in both BoWK and BoWK+LSK, the conversational information seems to be beneficial as increases of 5.20% and 1.35%, respectively, are observed.

5 Conclusions

In this work, the role of contextual information in supervised Sentiment Analysis over Twitter is investigated for the Italian language. Experimental results confirm the empirical findings presented in (Vanzo et al., 2014) for the English language. Although the size of the involved dataset is still limited, i.e. about 1,400 tweets, the importance of contextual information is emphasized within the considered markovian approach: it is able to take advantage of the dependencies that exist between different tweets in a conversation. The approach is also largely applicable as all experiments have been carried out without the use of any manual coded resource, but mainly exploiting unannotated material within the distributional method. A larger experiment, eventually on an oversized dataset, such as *SentiTUT*[4], will be carried out.

[4] http://www.di.unito.it/~tutreeb/sentiTUT.html

References

Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media*, LSM '11, pages 30–38, Stroudsburg, PA, USA. Association for Computational Linguistics.

Y. Altun, I. Tsochantaridis, and T. Hofmann. 2003. Hidden Markov support vector machines. In *Proceedings of the International Conference on Machine Learning*.

Luciano Barbosa and Junlan Feng. 2010. Robust sentiment detection on twitter from biased and noisy data. In Chu-Ren Huang and Dan Jurafsky, editors, *COLING (Posters)*, pages 36–44. Chinese Information Processing Society of China.

Roberto Basili, Maria Teresa Pazienza, and Fabio Massimo Zanzotto. 1998. Efficient parsing for information extraction. In *Proc. of the European Conference on Artificial Intelligence*, pages 135–139.

Albert Bifet and Eibe Frank. 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.

Giuseppe Castellucci, Simone Filice, Danilo Croce, and Roberto Basili. 2013. Unitor: Combining syntactic and semantic kernels for twitter sentiment analysis. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013)*, pages 369–374, Atlanta, Georgia, USA, June. Association for Computational Linguistics.

Danilo Croce and Roberto Basili. 2012. Grammatical feature engineering for fine-grained ir tasks. In Giambattista Amati, Claudio Carpineto, and Giovanni Semeraro, editors, *IIR*, volume 835 of *CEUR Workshop Proceedings*, pages 133–143. CEUR-WS.org.

Danilo Croce and Daniele Previtali. 2010. Manifold learning for the semi-supervised induction of framenet predicates: An empirical investigation. In *Proceedings of the 2010 Workshop on GEometrical Models of Natural Language Semantics*, GEMS '10, pages 7–16, Stroudsburg, PA, USA. Association for Computational Linguistics.

Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Enhanced sentiment learning using twitter hashtags and smileys. In Chu-Ren Huang and Dan Jurafsky, editors, *COLING (Posters)*, pages 241–249. Chinese Information Processing Society of China.

Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *Processing*, pages 1–6.

G. Golub and W. Kahan. 1965. Calculating the singular values and pseudo-inverse of a matrix. *Journal of the Society for Industrial and Applied Mathematics*, 2(2).

T. Landauer and S. Dumais. 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104(2):211–240.

Subhabrata Mukherjee and Pushpak Bhattacharyya. 2012. Sentiment analysis in twitter with lightweight discourse analysis. In *Proceedings of COLING*, pages 1847–1864.

Preslav Nakov, Sara Rosenthal, Zornitsa Kozareva, Veselin Stoyanov, Alan Ritter, and Theresa Wilson. 2013. Semeval-2013 task 2: Sentiment analysis in twitter. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013)*, pages 312–320, Atlanta, Georgia, USA, June. Association for Computational Linguistics.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May. European Language Resources Association (ELRA).

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of EMNLP*, pages 79–86.

Magnus Sahlgren. 2006. *The Word-Space Model*. Ph.D. thesis, Stockholm University.

Jianfeng Si, Arjun Mukherjee, Bing Liu, Qing Li, Huayi Li, and Xiaotie Deng. 2013. Exploiting topic based twitter sentiment for stock prediction. In *ACL (2)*, pages 24–29.

Ioannis Tsochantaridis, Thomas Hofmann, Thorsten Joachims, and Yasemin Altun. 2004. Support vector machine learning for interdependent and structured output spaces. In *Proceedings of the twenty-first international conference on Machine learning*, ICML '04, pages 104–, New York, NY, USA. ACM.

Andrea Vanzo, Danilo Croce, and Roberto Basili. 2014. A context-based model for sentiment analysis in twitter. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics*, pages 2345–2354, Dublin, Ireland, August. Dublin City University and Association for Computational Linguistics.

Fabio M. Zanzotto, Marco Pennacchiotti, and Kostas Tsioutsiouliklis. 2011. Linguistic Redundancy in Twitter. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 659–669, Edinburgh, Scotland, UK., July. Association for Computational Linguistics.