Assigning Connotation Values to Events

Tommaso Caselli†, Irene Russo‡, Francesco Rubino‡

Dept. of Chinese and Bilingual Studies†   Dip. Linguistica "T. Bolelli"†   Istituto di Linguistica Computazionale‡

Hong Kong Polytechnic University   Università degli Studi di Pisa   CNR

Hum Hong, HKSAR   Via S.ta Maria, 36 56126 Pisa   Via G. Moruzzi, 1 56124 Pisa

E-mail: tommaso.caselli@connect.polyu.hk, irene.russo@ilc.cnr.it, francesco.rubino@ilc.cnr.it

Abstract

Sentiment Analysis (SA) and Opinion Mining (OM) have become a popular task in recent years in NLP with the development of language resources, corpora and annotation schemes. The possibility to discriminate between objective and subjective expressions contributes to the identification of a document's semantic orientation and to the detection of the opinions and sentiments expressed by the authors or attributed to other participants in the document. Subjectivity word sense disambiguation helps in this task, automatically determining which word senses in a corpus are being used subjectively and which are being used objectively. This paper reports on a methodology to assign in a semi-automatic way connotative values to eventive nouns usually labelled as neutral through syntagmatic patterns that express cause-effect relations between emotion cause events and emotion words. We have applied our method to nouns and we have been able reduce the number of OBJ polarity values associated to event noun.

Keywords: cause emotion events, event connotative values, sentiment polarity

1.   Introduction

Sentiment Analysis (SA) and Opinion Mining (OM) have become a popular task in recent years in NLP and a central topic for the management of Word-of-Mouth content in social media. Different language resources (LRs) such as lexica (WordNet-Affect (Valitutti and Strapparava, 2004), SentiWordNet (Esuli and Sebastiani, 2006; Baccianella et al., 2010)), corpora (MPQA corpus (Stoyanov et al., 2004; Stoyanov et al., 2005)) and annotation schemes (Wiebe et al; 2005; Boldrini et al., 2010) have been developed and made publicly available. One of the basic assumptions in this research area is the possibility to discriminate between objective and subjective expressions which contribute to the identification of a document's semantic orientation (polarity) and, thus, to the detection of the opinions and sentiments expressed by authors or attributed to other participants in the document. Two interesting recent sub-tasks which have been developed in this field are i.) the detection of the emotion cause event (ECE, Lee et al., 2010; Chen et al., 2010), which focuses on the identification of the phrase (if present) mentioning an event which causes a certain emotional state/response; and ii.) subjectivity word sense disambiguation (SWSD, Akkaya et al. 2009) which aims at automatically determining which word senses in a corpus are being used subjectively and which are being used objectively. The following examples are reported in order to clarify these tasks. Example 1. refers to the first subtask, ECE detection, where the ECE is reported in bold and the associated emotional state in italics; example 2. is taken from Akkaya et al. (2009). The words in bold refer to the word to be disambiguated and the labels S and O refer to their subjective or objective word sense.

- Non poteva mancare un accenno alla **strage di Bologna**, che costringe l' animo a *infinita vergogna*.

  [There was a mention of the **Bologna massacre**, that forces us to feel *ashamed*.]

- His **alarm** grew. S

  Will someone shut that darn **alarm** off? O

  The **alarm** went off. O

This work aims at unifying the techniques and methodologies developed for the identification of ECEs and extend the SWSD task in order to identify polarity values in context for a specific set of words, namely event denoting words. At this stage of development, we have concentrated on events realized by nouns. The outcome of our system will be a lexicon of event nouns with associated polarity values which express connotative information on events. Furthermore, the lexicon can be incrementally extended by means of parsed corpora of different size (and domain). This information could be used for the improvement of polarity detection and implicit opinion mining but also for the enrichment/advance of lexical resources with new connotative values.

The remaining of the paper is organized as follows: Section 2 will report on previous works and on the analysis of the Italian version of SentiWordNet (It-SWN) with respect to the polarity values associated to eventive synsets. Section 3 will describe the methodology we adopted in order to identify text portions expressing a relation of cause effect between an (eliciting) event and an (associated) emotion. Section 4 will present an analysis of the results obtained and compare them with the values provided by It-SWN. Finally Section 5 reports on the conclusion and future research directions.

2. Connotative values of event nouns in SentiWordNet

Emotional states are often triggered by the perception of external events (pre-events) (Wierzbicka, 1999), but, on the contrary, they can also be the cause of events (post-events; Huang, 2010). This suggests to consider the relations between emotional states and related events as a tri-tuple of two pairs (Huang, 2010):

- <<pre-events, emotional state> <emotional state, post-event>>

As already stated, in this study we concentrate on the first pair of the tri-tuple, namely pre-events (or ECE), and emotional states. Previous works on this task have been carried out for Chinese and English (Chen et al., 2010, Lee et al., 2010). ECE can be explicitly expressed as arguments, propositions, nominalizations and nouns. However, the happening of an event in the world is considered as bearing a neutral/objective polarity. Through the exploitation of WordNet's Inter-Lingual Index (ILI), we have associated the polarity values of SentiWordNet 3.0 (Baccianella et al., 2010) to the Italian entries in MultiWordNet (Pianta et al., 2002), thus obtaining direct access to the polarity values and providing Italian with a preliminary version of a useful language resource for Sentiment Analysis and Opinion Mining. Henceforth, we will refer to this lexicon as It-SWN. By means of simple database queries we have extracted all eventive synsets from It-SWN. Their figures are reported in Table 1. In particular, we focused on verbs, as they are the prototypical PoS which give rise to events, and on nominal events, i.e those noun synsets whose supertypes correspond to "*event*", "*phenomenon*", "*state*" and "*act*".

eventive synsets PoS	# synsets	# synsets w/o polarity	# synsets w/ polarity
nouns	4,925	3,183 (64.6%)	1,742 (35.4%)
verbs	4,985	3,586 (71.9%)	1,339 (28.1%)

Table 1 - Figures of the eventive synsets in It- SWN

As the data show, very few events display a polarity value (either negative or positive). On the other hand, as competent speakers we experience everyday that words have connotative meanings, encoding the speaker/writer attitude to the denoted concept or entity. Adopting a wider perspective, we can state that connotative meaning expresses opinionated content related to commonsense knowledge about a certain concept or entity. Under this perspective, it is a matter of opinions and attitudes about events. For instance, consider the following examples, "nascita" [birth] and "incidente" [incident]. Both words denote or express two different events. As things which happen or take place in the world, they should not be considered as bearer of polarity values. Nevertheless, we intuitively associate a positive value with respect to the

former event, while, with respect to the latter, "incidente" [incident], we tend to associate negative feelings. If we look in It-SWN for this two lemmas, we obtain the following information. Notice that we have reported the synset ID for each lemma plus the gloss (in italics) and the polarity values where "pos" stands for positive and "neg" stands for negative:

- [nascita#0732032] - *the event of being born*; pos=0 neg=0

- [nascita#0732032] - *the event consisting of the start of something*; pos=0 neg=0

- [incidente#07301336] - *an unfortunate mishap; especially one causing damage or injury*; pos=0 neg=0.75

- [incidente#07301950] - *a serious accident (usually involving one or more vehicles)*; pos=0 neg=0

- [incidente#13978033] - *a public disturbance*; pos=0 neg=0.25

From a theoretical point of view, we can argue that it is quite difficult to fix sentiment values for words senses out of context because it is well known that specific patterns can increase, decrease or reverse the previous established polarity. Facing this problem two heuristics are possible: weighting the polarity of words on the basis of frequency of occurrences in specific context or augmenting lexical resources with patterns that in a semi-compositional way operationalize changes in polarity. In this work we implement the first type of heuristics, proposing the identification of connotative/polarity values for event nouns based on their co-occurrences in a syntagmatic relation that codify their role as emotion cause events, i.e. the tuple <pre-events, emotional state>. However, the second approach is equally promising and it will be tested/implemented when more rules will be identified.

3.   Identifying syntagmatic patterns

In order to discover reliable syntagmatic patterns for the identification of cause-event emotions, we have exploited data elicited from "*Il Corriere della Sera*" website visitors during December 2010. Visitors were asked to to describe the year 2010 with 10 words. 2,378 people participated in the data collection for a total of 22,469 words. The working hypothesis we have applied to these data after a manual exploration of a sample is that the way people have "answered" to the "elicitation" of the description for the year 2010 is very similar to the semantic association responses used in much research programs in cognitive science and psycholinguistics (Miller, 1969; Spence and Owens, 1990), according to which the semantic association is triggered by the textual co-occurrence of the stimulus-response pairs. From these data we find out associations between events and emotional states in order to identify preliminary couples of emotions and emotion cause events, for a total of 18,240 nominal couples. By means of an Italian adapted version of WN-Affect (It-WordNet-Affect), we have extracted all nominal couples where at least one item has an associated sense corresponding to the "*emotion*" category and the other item corresponded to the category of "*emotion eliciting situation*". We have filtered the set of emotion eliciting situation items by selecting only those nouns whose top nodes uniquely belong to the eventive ontological classes, as stated above. In this way, we have obtained two lists of keywords, one of emotion words, composed by 133 lemmas, and another of possible emotion event causes (ECEs) of 161 lemmas. This data set has been exploited to identify relevant syntagmatic patterns for the detection of nominal emotion causes. The pattern extraction phase has been performed on a parsed version of a large corpus of Italian, the La Repubblica Corpus (Baroni et al., 2004). We have developed a specific pattern extractor which takes as input the couples of the seed words we have obtained (133 emotion keywords and 161 ECEs) and extracts all combinations of emotion keywords and its/their associated ECEs occurring in the same sentence, with a distance ranging from 1 to 8 possible intervening parts-of-speech. We have thus obtained 1,339 possible patterns. This set has been cleaned both on the basis of pattern frequencies and through manual exploration. Finally, 47 patterns were selected and considered as good

co-occurrence contexts. In addition to this, we have extended this set with a list of 6 intrasentential relevant collocational patterns manually identified on the basis of the authors' judgments which express cause-effect relationships between a nominal cause emotion event and an associated emotion word. In Table 2 we report some the parts-of-speech patterns we have identified. The parts-of-speech abbreviations have been obtained from the Tanl tagset[1] , based on the ILC-PAROLE tagset and conformant to the EAGLES international standard. The elements in brackets are optional, i.e. they do not necessarily has to be in the pattern. In addition to this, notice that to avoid biases on the cause-effect relationship, each pattern specifies the position of the emotion word (S_E) and that of the nominal cause emotion event (s_{ECE}).

Extracted patterns from La Repubblica	S_E E S_{ECE} | S_E E RI S_{ECE} | S_{ECE} E R S | S_{ECE} V RI A S_E …
Manually identified patterns	S_E per RD/RI (A) S_{ECE} | RD/RI S_{ECE} E/EA S suscitare RI/RD (A) S_E …

Table 2 – Instances of the extracted syntagmatic patterns

To clarify, we report below some instances of patterns extracted from the La Repubblica corpus:

- *amore$_S$ [per]$_{PER}$ la$_{RD}$ musica$_S$* [love for music]
    Pattern: S_E PER RD S_{ECE}
- *paura$_S$ del$_{EA}$ terremoto$_S$* [fear of the earthquake]
    Pattern: S_E EA S_{ECE}
- *la$_{RD}$ morte$_S$ di$_E$ Lazzati$_S$ [ha scuscitato]$_{SUSCITARE}$ un$_{RI}$ grande$_A$ dolore$_S$* [the death of Lazzati caused a big grief]
    Pattern: RD S_{ECE} E S SUSCITARE RI A S_E

4. Amending SentiWordNet polarity values

We applied the 53 validated pattens on two Italian corpora La Repubblica and ItaWaC[2] (Baroni et al.; 2009) in order to extract contextual data on connotative meaning of events. In this way, we have been able to identify 143 nominal lemmas which represent elegible cause-events of emotions. Since we have used WN-Affect as a starting point, where emotion words are "classified" as being positive, negative, ambiguous or neutral with no polarity measure, has posed some issues on the following aspects:

- dealing with ambiguous emotions: in this case, we will consider the emotion cause event as perfectly balanced between positive and negative polarity;

- dealing with neutral emotions: in this case, we will not assign any polarity value and consider the emotion cause event as occurring in a objective, unemotional context.

In order to use It-SWN as a benchmark test suite and, at the same time, try to contribute to its extensions, we have developed the following heuristics:

- we compute the global event polarity by summing the different scores of eventive synsets and dividing it by the number of synsets, as already proposed in Denecke (2008);

- we will assign the obtained polarity value only to those synsets whose supersense(s) corresponds to one of the top eventive four nodes "*event*", "*act*", "*phenomenon*" and "*state*";

- a minimum threshold of 10 occurrences of the emotion cause event noun has been set in order to avoid biases due to hapax and idiosyncratic expressions;

- we rescale the polarity values to the measures proposed in SentiWordNet on the basis of the frequency of co-occurrences the emotion cause event with the emotion lemma, following a methods similar to that described in Braasch and Pedersen (2010).

[1]http://medialab.di.unipi.it/wiki/Tanl_POS_Tagset
[2]http://wacky.sslmit.unibo.it

4.1 Evaluation

The corpus exploration has shown that 3 lemmas of the extracted elegible cause-events are not present in our version of It-SWN, thus reducing the number of data available for comparison to 140 entries. Table 3 summarizes the results of the computation of the global polarity in It-SWN for the 140 remaining noun lemmas. Each row represents a polarity value or a polarity pattern, while the numeric values represent the number of lemmas which are affected by that polarity pattern. As a convention, we use the symbol "<" to refer to the order of the polarity value from the highest to the lowest. Similarly, we use the symbol "|" to refer to equal polarity values. To clarify, a polarity pattern like *OBJ < POS < NEG* is to be interpreted as follows: the highest polarity values is the objective (or null) polarity, followed by the negative polarity and, finally, by the positive polarity.

Polarity patterns	# of occurrences per lemma
OBJ	58
OBJ | NEG	2
OBJ < NEG	22
NEG < OBJ	1
OBJ < POS	17
OBJ < NEG | POS	6
OBJ | NEG | POS	1
OBJ < NEG < POS	26
NEG < OBJ < POS	2
OBJ < POS < NEG	5

Table 3 – It-SWN polarity patterns for the 140 elegible cause emotion event nouns

As Table 3 shows, we have identified 10 different polarity patterns. Not surprisingly, the vast majority of these patterns have the objective or null value (OBJ) either as the first member or as the only member of the pattern. In addition to this, it is also interesting to notice that the negative value is the second highest ranking polarity value. These polarity patterns will be used in our evaluation in order to identify discrepancies between the It-SWN polarity values and those that we will propose for the cause event nouns. Finally, the analysis of the 140 lemmas impact on 287 different eventive nominal synsets.

As for the identification of the polarity values through the syntagmatic patterns, we have first filtered the 140 nouns in It-SWN with the 10 occurrences threshold. This has left us with 120 elegible cause events nouns. In Table 4 we report the polarity patterns we found.

Polarity patterns	# of occurrences per lemma
NEG	9
POS	10
NEG < POS	24
NEG | POS	5
POS < NEG	9
NEG < OBJ | POS	17
NEG | POS < OBJ	1
NEG < POS < OBJ	36
POS < NEG < OBJ	8
POS < OBJ | NEG	1

Table 4 – Polarity patterns for the 120 elegible cause emotion event nouns identified through the syntagmatic patterns

For each valid elegible cause emotion event, we have computed the polarity values (i.e. positive, negative and objective meaning) and rescaled that to the global value of 1. This operation has been accomplished by taking into account the relative frequency of each emotion co-occurring with the emotion cause event. The percentages thus obtained have been mapped to the 8 polarity values (ranging from 0.125 to 1) of It-SWN by means of a heuristic which differentiate the mapping of the percentages to the 8 values according to the number of polarity values. In Table 5 we report the mapping when all the three polarity values (OBJ, NEG and POS) appear.

It-SWN values	Correponding percentages
0.125	0.1% - 25%
0.25	26% - 38%
0.375	39% - 51%
0.5	52% - 63%
0.625	64% 76%
0.75	77% - 89%
0.825	90% - 99%

Table 5 – Mapping between the It-SWN polarity values and the frequency percentage between emotion cause event and emotion noun

By analyzing the polarity patterns and comparing them with their corresponding patterns forms in It-SWN,we have identified the following data:

- no pattern repeats in either of the two resources. Clearly, such a result could be influenced by the fact that we were extracting data from the corpora by exploiting patterns which specifically identified relations of cause-effect between an eventive nouns and an emotion keyword. Nevertheless, the OBJ value is still present, thus suggesting that the approach we have adopted, even if not perfect, is correct. As a matter of fact our aim was to improve the polarity analysis of events, so that they could be applied in open domain texts in order to facilitate implicit SA and O.

- 50 new polarity patterns have been identified; these patterns are instances of shifts from pure OBJ pattern in It-SWN to new polarity values which either completely exclude the old OBJ value (e.g. from OBJ to POS) or spread the polarity values along the three classes (e.g. from OBJ to NEG < POS < OBJ);

- 37 polarity patterns have different polarity configurations with respect to the polarity patterns of It-SWN and, at the same time, they have a different hierarchy of the polarity values in the pattern. For instance, the pattern OBJ < POS for the cause emotion noun "*sciopero*" is changed to NEG < OBJ | POS.

- 32 polarity patterns have different polarity configurations with respect to the polarity patterns of It-SWN, but at the same time, they reinforce the polarity hierarchy of the NEG and OBJ values. For instance, a pattern like OBJ < NEG < POS is changed to NEG < POS < OBJ.

- 1 pattern introduces new polarity values which were not present in It-SWN.

The data for the polarity patterns obtained from the corpus data can be grouped into three groups, namely: a) those which shift the polarity values from OBJ to POS and NEG, thus assigning a strong polarity value to events; b) those which downgrade the OBJ value and at the same time suggest a different configuration, or hierarchy, of the POS and NEG values with respect to the one proposed in It-SWN; and c) those which downgrade the OBJ value and at the same time reinforce the configuration, or hierarchy, of the POS and NEG values. With respect to the data encoded in It-SWN, our resource clearly tends to maximize the polarity values of NEG and POS, thus providing more "opinionated" data. Thus, in the perspective of using It-SWN as test set for the validation of our approach and collected data, we need to keep in mind this aspect. So far, the best conclusion we can draw, is that events do not always have a null polarity value (i.e. OBJ) but are perceived as bearer of polarity, i.e. they have a connotative value. The newly proposed polarity values go in this direction. In addition to this, it is also interesting to point out that, if we exclude the OBJ value, in 32 cases (26.6%) the corpus data are in-line with respect to the values in It-SWN and that in 37 cases (30.8%) there is a shift in the hierarchy of the POS and NEG values, thus suggesting modification to the "weight" of POS and NEG in It-SWN.

Going through the corpus frequency and co-occurrences, it is interesting to observe that on a total of 91,350 occurrences of the syntagmatic patterns for emotion cause events - emotions, the vast majority of them is composed by negative emotions (53,927), followed by positive emotions (31,009), ambiguous emotions (4,958), and, finally, by neutral emotions (1,256).

5. Conclusion and future works

This work has described a methodology to assign in a semi-automatic way connotative values to events through syntagmatic patterns that express cause-effect relations between events (emotion cause events. ECEs) and emotion words. At this level of development, we focused on nominal events.

The results we obtained are encouraging as we identified a reduction of the OBJ polarity value and an increase of

the values NEG and POS. In particular, we have also discovered higher occurrence of negative emotions with respect to positive ones, a trend that should be tested on a wider dataset.

As future works, we are planning to assign new polarity values to all eventive synsests in It-SWN (including verbs) and run an evaluation experiment on polarity classification by exploiting this newly created resource.

Acknowledgments

References

Akkaya C., J. Wiebe, and R. Mihalcea**.** (2009). Subjectivity Word Sense Disambiguation. *Proceedings of EMNLP 200.*

Baccianella S., A. Esuli and F. Sebastiani. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *Proceedings of the 7th conference on International Language Resources and Evaluation (LREC 2010)*, Malta, May 2010.

Baroni M., S. Bernardini, F. Comastri, L. Piccioni, A. Volpi, G. Aston and M. Mazzoleni. (2004). Introducing the la Repubblica corpus: A large, annotated, TEI(XML)-compliant corpus of newspaper Italian. In *Proceedings of the 4th conference on International Language Resources and Evaluation (LREC 2004)*, Lisboa, May 2004.

Baroni M., S. Bernardini, A. Ferraresi and E. Zanchetta. (2009). The WaCky Wide Web: A collection of very large linguistically processed Web-crawled corpora. In *Journal of Language Resources and Evaluation* 43 (3), pp. 209-226.

Boldrini E, A. Balahur, P. Martinez-Barco and A. Montoyo. (2010). EmotiBlog: a finer-grained and more precise learning of subjectivity expression models. *Proceedings of the Fourth Linguistic Annotation Workshop (LAW IV '10).*

Braasch, A. & B.S. Pedersen. (2010). Encoding Attitude and Connotation in Wordnets. *The 14th EURALEX International Congress*, Leeuwarden , The Netherlands.

Chen Y., S.Y.M. Lee, S. Li, and C. Huang. (2010) Emotion Cause Detection with Linguistic Constructions. *Proceeding of the 23rd International Conference on Computational Linguistics (COLING 2010)*.

Denecke, K. (2008). Using SentiWordNet for Multilingual Sentiment Analysis. *IEEE 24th International Conference on Data Engineering - ICDEW 2008.*

Esuli, A. and F. Sebastiani. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining**.** In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006)*, Genova, May 2006.

Huang, C. (2010). Emotions as Events (and Cause as Pre-Events). Communication at the Chinese Temporal/discourse annotation workshop, Los Angeles, June 2010.

Miller, G. (1969). The organization of lexical memory: Are word associations sufficient? In G.A. Talland & N.C. Waugh (eds.), *The pathology of memory*. New York: Academic Press, pp. 223-237.

Lee S.Y.M., Y. Chen, C. Huang. (2010). A Text-driven Rule-based System for Emotion Cause Detection. *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*.

Pianta E., L. Bentivogli and C. Girardi. (2002) MultiWordNet: Developing and Aligned Multilingual Database. In *Proceedings of the First International Conference on Global WordNet*, Mysore, India, January 21-25, 2002, pp. 293-30.

Schulte Im Walde, S. (2008). Human associations and the choice of features for semantic verb classification. *Research on Language and Computation*, 6(1), pp. 79-111.

Spencer, D. P. and K. C. Owens. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, 19, pp. 317–330.

Stoyanov, V., C. Cardie, D. Litman, and J. Wiebe. (2004). Evaluating an Opinion Annotation Scheme Using a New Multi-Perspective Question and Answer Corpus. *Working Notes of the 2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*.

Stoyanov, V., C. Cardie, and J. Wiebe. (2005). Multi-Perspective Question Answering Using the OpQA Corpus.*Proceedings of HLT/EMNLP 2005*, Vancouver, Canada.

Strapparava C. and A. Valitutti. (2004) WordNet-Affect: an affective extension ofWordNet". *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, May 2004.

Wiebe, J., T. Wilson, and C. Cardie. (2005). Annotating expressions of opinions and emotions in language. Language Resources and Evaluation 39 (2-3)*, pp.165-210.

Wierzbicka, A. (1999). *Emotion Across Languages and Cultures: Diversity and Universals*. Cambridge, CUP.