Valentino: A Tool for Valence Shifting of Natural Language Texts

Marco Guerini, Carlo Strapparava, Oliviero Stock

FBK-irst
38050 Povo, Trento, Italy
{guerini, strappa, stock}@fbk.eu

Abstract

In this paper a first implementation of a tool for valence shifting of natural language texts, named *Valentino* (VALENced Text INOculator), is presented. Valentino can modify existing textual expressions towards more positively or negatively valenced versions. To this end we built specific resources gathering various valenced terms that are semantically or contextually connected, and implemented strategies that uses these resources for substituting input terms.

1. Introduction

Accurate wording is essential in verbal communication. We can present or view the same information from a particular angle, in a biased or even unfair way, through an accurate choice of words and images. We may load description of a specific situation with vivid, connotative words and figures of speech, without changing the basic content. These words have the capability to provide an affective connotation to the text and reveal the affective disposition of the speaker or induce a similar disposition on the recipient. They have an important role in persuasion and for this reason they are very used in political speeches and/or advertisement.

While there is the active NLP field of opinion mining and sentiment analysis (Turney, 2002; Lin et al., 2006) on the other side, given the large amount of available texts, it would be conceivable to exploit NLP techniques to *slant* original writings toward specific biased orientation, keeping as much as possible the same meaning, see (Hirst and Budanitsky, 2005) for some initial work along this direction.

In this paper, we present a tool for modifying existing textual expressions towards more positively or negatively valenced versions as an element of a persuasive system. For instance a strategic planner may decide to intervene on a draft text with the goal of "coloring" it emotionally. When applied to a text, the changes invoked by a strategic level may be uniformly negative or positive; they can smooth all affective peaks; or they can be introduced in combination with deeper rhetorical structure analysis, resulting in different types of changes for key parts of the texts.

Valentino is meant to be an easily pluggable component. The only information it requires in input is a coefficient (included between 1 and -1) that represents the designed valence for the final expression.

2. Resources

For affective persuasion and the task of positive (or negative) slanting of texts, we drove a preliminary qualitative study with 5 human subjects to understand how people modify the valence of existing texts. The subject were given 4 pieces of text from news and asked to modify their valence (neutral to positive, neutral to negative, negative to neutral and positive to neutral). The insight gained from the study showed that (a) people usually modify single words, (b) sometimes use paraphrases (c) sometimes add or subtract words that play the role of downtoners or intensifiers.

Point (a): We found that there are different classes of valenced terms that are addressed, like adjectives, adverbs, quantifiers, terms indicating strength of belief, etc. We built a resource that gathers these terms in vectors (OVVTs). At present there are 3700 OVVTs in the resource. We used the *WordNet* `antonymy` relation as an indicator of terms that can be "graded". We built four groups of terms that can be potentially used (one group for each POS). Moreover, we populated the vectors using other specific *WordNet* relations (`similar_to` relation for adjectives, `hyponym` relation for verbs and nouns). Finally the valence of *WordNet* synsets, taken from *SentiWordNet* scores - (Esuli and Sebastiani, 2006), was added to the corresponding lemmata. *SentiWordNet* is a lexical resource in which each *WordNet* synset is associated to three numerical scores: `Obj(s)`, `Pos(s)` and `Neg(s)`. These scores represent the objective, positive and negative valence of the synset. An example of *SentiWordNet* items is given in Table 1.

Thus, an OVVT is composed of several "terms" (synsets prima facie, by assuming that all the terms in the synset have the same valence) with similar semantic reference (e.g. beauty) but different valence (see Figure 1, each entry in the OVVTs takes the form `lemma#pos#sense-number`[1]).

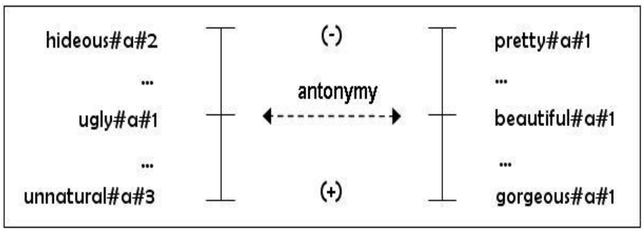

Figure 1: An example of OVVT

Point (b): For mimicking the use of paraphrases, we collected from *WordNet* words gloss, starting from the as-

[1] `Sense-number` is the number of sense in *WordNet*.

POS	Offset	PosScore	NegScore	SynsetTerms
a	602378	0.0	0.875	wrong#a#1 incorrect#a#1
r	60640	0.75	0.0	better#r#1
n	7017251	0.0	0.0	victory#n#1 triumph#n#1

Table 1: examples of *SentiWordNet* entries

sumption that the definition of a word is a paraphrase used to describe that word.

Point (c): For insertion or deletion of words that play the role of downtoners or intensifiers we created specific OVVTs (that we call Modifiers-OVVTs). In this case the words were gathered according to a criterion of *contextual* connection rather than *semantic* connection. That is to say: instead of using *WordNet* semantic relations as for point (a) we used information extraction techniques on the BNC corpus to find contextual connections between words and respective modifiers. In particular we started extracting from `FrameNet` (Baker et al., 1998) the verbs in the frames *Adducing*, *Discussion*, *Statement*, *Awareness*, *Expectation*, *Telling*, *Chatting*, *Opinion*[2]. Then we looked up in the BNC corpus to find the adverbs associated to these verbs. We considered a window of one token preceding and following the verb. An example of a Modifiers-OVVT, for the verb "assert" is given in figure 2.

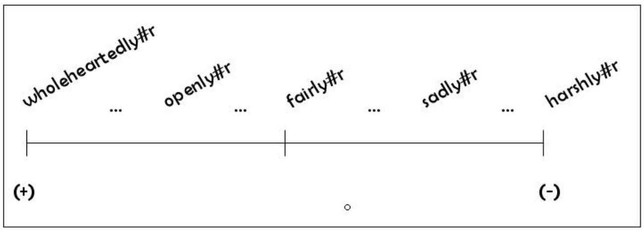

Figure 2: An example of Modifiers-OVVT associated to the verb "assert"

3. Strategies

At the current stage of implementation only a simple POS analysis (together with named entity recognition and morphological analysis) without contextual information is performed. For this task we used the TextPro package, see (Pianta and Zanoli, 2007) and (Zanoli and Pianta, 2007).

Various strategies have been implemented, mimicking those performed by humans:

Paraphrase: if a lemma has only one sense, then the gloss of the word is inserted in the text. The gloss is then valenced, but no more paraphrases are allowed. This augments (a) variety in the output text and (b) the possibility of further valencing the original text (see Table 2 for an example).

Use of OVVTs considering only the most frequent senses: for every lemma the candidate substitutes are chosen by searching in the OVVTs up to the third sense of that lemma (e.g. given `big#a` it is first searched `big#a#1`, in case of failure `big#a#2` and eventually `big#a#3`).

Candidate lemmas selection: After these two steps there is the necessity to choose among the candidates lemmas. This choice is performed by using lists of persuasive words that we collected from a CORpus of tagged Political Speeches (CORPS), see (Guerini et al., 2008). If the valence coefficient for the final expression is negative then the "negative-focus" words list is accessed, if it is positive then the "negative-focus" words list is accessed. Next, the candidate with highest ranking is selected.

Strengthening/weakening by modifying adjectives grade: if the chosen lemma is "too weak" (e.g. the output valence should be -1 but the most valenced candidate for substitution is -0.125), the superlative form is used. Also the opposite situation is considered: if the chosen lemma should be in the superlative form (according to the morphology of the substituted term), but the output valence is already met, then the superlative form is discarded.

Insertion or deletion of downtoners and intensifiers: this strategy behaves similarly to the previous strategy. If the chosen lemma (a verb of assertion, opinion, etc. as described in the previous section) is "too weak", an adverb is chosen from the corresponding Modifiers-OVVT, to strengthen the verb. Also the opposite situation is considered: if the original lemma has a modifier, but the output valence is already met, then the modifier is discarded.

Morphology synthesis: As a final step the chosen lemma is synthesized according to the chosen morphology (either the morphology of the original lemma, or the modified morphology as defined in the aforementioned strategy that works on adjectives grade).

Named entity blocking: to prevent cases like "Super Bowl" shifting to "*Giant* Ball". Named entities are left as they are in text.

In Table 3 various examples of valence shifting of the sentence "Bob admitted that John is absolutely the best guy" are given. On the left the coefficient of shifting is indicated. On the right the corresponding output with: lemmata chosen from OVVTs in italic, words that further underwent grade modification between parentheses and added modifiers between square bracket.

4. Advantages and limits

SentiWordNet scores: even though there are some drawbacks in *SentiWordNet* scores (e.g. words that should be clearly valenced that are not, words that are too much valenced) *Valentino* performs reasonably well.

[2]We started from these classes because in the preliminary study we found that these kind of verbs were the most affected by this strategy

Original expression:	"He would likely go"
Selected gloss:	"likely = with considerable certainty"
Shifted Output:	"He would (with *wide certitude*) go"

Table 2: An example of paraphrase

CF. 1.0	Bob *[wholeheartedly] admitted* that John is *absolutely (a superb) hunk*
CF. 0.5	Bob *[openly] admitted* that John is *highly* the *redeemingest signor*
CF. 0.0	Bob *admitted* that John is *highly (a well-behaved) sir*
CF. -0.5	Bob *[sadly] confessed* that John is *nearly (a well-behaved) beau*
CF. -1.0	Bob *[harshly] confessed* that John is *pretty (an acceptable) eunuch*

Table 3: An example of *Valentino* shifting capabilities

Advantages of using only the most frequent senses of words: an example starting from the sentence: "He was a great singer"

- without taking into account the senses frequencies order: "he was a *pregnant*[3] singer"
- by searching among most frequent senses (1_{st} to 3_{rd}): "he was a *giant* singer"

Advantages of using the list of persuasive words: the word "giant" has been chosen from the following bunch of candidate lemmata (score 0.375): `elephantine#a#1` - `gargantuan#a#1` - `giant#a#1` - `jumbo#a#1`

5. Applications Scenario

There are many applied scenarios: edutainment systems that should adapt the output to the audience, news agencies wishing to deliver valenced information, conflict management systems that adapt the messages according to the stage of the conflict (fostering escalation or de-escalation) and so on.

An interesting technological scenario is for Embodied Conversational Agents' applications. Often these applications rely on canned, pre-compiled text. Different emotion intensity realizations of the same message are obtained only via facial expression, see for example (Guerini et al., 2007). With *Valentino* the pre-compiled text can be automatically valenced according to emotion intensity, augmenting the output effect.

6. Conclusions and Future Work

In this paper we presented the first implementation of *Valentino*, a tool for modifying existing textual expressions towards more positively or negatively valenced versions. We presented the main resources we built and strategies we implemented for the system together with some application scenarios.

We plan to collect other Modifiers-OVVTs and to implement various strategies based on LSA similarity techniques to further improve the performances of our system, e.g.:

- At present "newspaper article" is (negatively) shifted to "newspaper *lemon*" because "article" is taken in the primary sense of "artifact", and "lemon" in the secondary sense of "an artifact that is defective or unsatisfactory". By using LSA techniques we can prevent such cases.
- A filter to rule out cases of incongruence between adjacent words once chosen. For example "toughest eunuch" is a correct but incongruent realization (with coefficient -1) of "tough guy".

We also want to explore higher reasoning strategies (at present a "word by word" approach is used, every token is considered and modified in isolation, without considering the context). The first steps will be to reason on whole constituents valence modification and to address the problem of negations (like in "not bad").

[3] Here "pregnant" is in his secondary sense of "significant" which is correct but sounds odd.

7. References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The berkeley framenet project. In Christian Boitet and Pete Whitelock, editors, *Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics*, pages 86–90. Morgan Kaufmann Publishers.

A. Esuli and F. Sebastiani. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation*, pages 417–422, Genova, IT.

M. Guerini, O. Stock, and M. Zancanaro. 2007. A taxonomy of strategies for multimodal persuasive message generation. *Applied Artificial Intelligence Journal*, 21(2):99–136.

M. Guerini, C. Strapparava, and O. Stock. 2008. Resources for persuasion. In *Proceedings of LREC 2008*, Marrakech, Morocco.

Graeme Hirst and Alexander Budanitsky. 2005. Correcting real-word spelling errors by restoring lexical cohesion. *Natural Language Engineering*, 11(1):87–111, March.

W.H. Lin, T. Wilson, J. Wiebe, and A. Hauptmann. 2006. Which side are you on? identifying perspectives at the document and sentence levels. In *Proceedings of the*

Tenth Conference on Computational Natural Language Learning (CoNLL-X).

E. Pianta and R. Zanoli. 2007. Tagpro: a system for italian pos tagging based on svm. *Intelligenza Artificiale, Numero Speciale Strumenti di Elaborazione del Linguaggio Naturale per l'Italiano*, 4(2):8–9, June.

P.D. Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*, pages 417–424, Philadelphia, Pennsylvania.

R. Zanoli and E. Pianta. 2007. Entitypro: exploiting svm for italian named entity recognition. *Intelligenza Artificiale, Numero Speciale Strumenti di Elaborazione del Linguaggio Naturale per l'Italiano*, 4(2):69–70, June.