Affect Sensing using Lexical Means: Comparison of a Corpus with Movie Reviews and a Corpus with Natural Language Dialogues

Alexander Osherenko
Multimedia Concepts and Applications, Faculty of Applied Informatics
University of Augsburg, Germany
osherenko@informatik.uni-augsburg.de

ABSTRACT
Affect sensing using pure lexical means is a problem that has been standing in the centre of attention of research community for a very long time and there exist promising approaches for solving this problem. The questions that have been investigated so far refer to the emotional meaning of a text is and its computation. Although the answers to these questions are certainly important, it is probably too premature to provide them without knowing differences between corpora used for lexical affect sensing. This contribution provides a comparison of two emotional textual corpora – a carefully built and worded movie review corpus and the spontaneous SAL corpus – and presents features for the classification of affective utterances.

Author Keywords
Lexical affect sensing, emotion research.

INTRODUCTION

Affect sensing is an important field of study. Different efforts have been already made in this research area studying textual, facial, physiological aspects etc. [3]. There is a great application potential of affect sensing, such as storytelling, interface agent applications, learning systems, telephony etc.

This paper compares affect sensing using pure textual means in two text corpora that can be seen in this context as two extremes of affect expressivity – a carefully built and worded movie review corpus and a spontaneous speech corpus. Thus, the question arises of whether the two corpora can be still analyzed using the same features or whether the feature set should vary according to some specific properties of the particular corpus.

This paper doesn't discuss advantages or drawbacks of particular groups of approaches to lexical affect sensing (keyword spotting, lexical affinity, statistical natural language processing, hand-crafted approach). For exact description of these topics see [4].

CORPORA

Two textual corpora are compared – a corpus with movie reviews [1] and the SAL corpus [6].

The movie review corpus contains 84 film reviews rated by a human and mapped onto classes from one to three stars. There are 28 reviews for each rating in the corpus (Fig. 1).

> [Rating: 3]
>
> Simply put, Sofia Coppola's Lost in Translation is an amazing motion picture. There may be some controversy over whether she truly wrote the screenplay on her own (there are sequences that argue that she at least had help from someone with a little more experience in life and marriage), but that doesn't impact the final analysis. …
>
> … As good as The Virgin Suicides is, Lost in Translation is superior in almost every way. When Top 10 lists are released at the end of the year, this title will feature prominently on a number of them (including mine).

Figure 1: An example of a movie review

The SAL corpus (Sensitive Artificial Listener) is a set of affective dialogues presenting four characters equipped with different responses in differing emotional states: optimistic and outgoing (Poppy), confrontational and argumentative (Spike), pragmatic and practical (Prudence), depressing and gloomy (Obadiah). These dialogues are annotated by four critics (cc, dr, em, jd) with FEELTRACE data [2]. The corpus consists of 27 dialogues (569 annotated utterances from critic cc). The FEELTRACE data for the evaluation affect score are mapped onto three categories ("negative", "neutral", "positive") to facilitate a comparison of the two corpora.

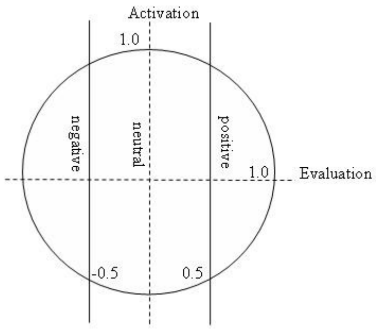

Figure 2: Affect segments representing evaluation

The FEELTRACE data are mapped onto affect segments as shown in Figure 2 (“positive”, “neutral”, and “negative” according to the value of the evaluation dimension). Radius 1 is predetermined by FEELTRACE, whereas the values -0.5 and 0.5 are chosen for the sake of symmetry. Note that there are also such utterances that change their affect meaning from positive to negative or vice versa. These inconsistent utterances occur very seldom and can be excluded from the analysis without significantly influencing the results.

The distribution of affect segments as annotated by critic cc is the following: 'positive': 62, 'neutral': 447, 'negative': 59, 'undefined': 1 where the pairs show an affect segment and the number of utterances of this segment in dialogues. Affect segment “undefined” denotes inconsistent affect segments. Sample utterances from the SAL corpus are shown in Fig. 3.

[Affect segment: positive]

– Oh I'm pretty good I guess. It's nice to hear a cheery voice though.

[Affect segment: neutral]

– Well, there are some real idiots on the road aren't there.

[Affect segment: negative]

– Well that's true too, but if you dwell on that your not gonna get by life in a very (laugh) positive frame of mind.

Figure 3: Examples of SAL utterances

Note that the examples above are provided without a context in order to stress its importance in emotional comprehension – in some cases it is clear what meaning an utterance has, in some cases yet not clear.

Differences between the movie review corpus and the SAL corpus can be expressed as follows:

1. The movie reviews corpus contains affect words that can convey emotional meaning whereas the SAL corpus doesn’t normally have such obvious signs of emotional meaning.
2. The length of emotional texts in the movie reviews corpus is rather long (about 1000 signs) whereas the utterances of the SAL corpus can contain only one word.
3. The movie reviews are grammatically correct sentences whereas the utterances in the SAL corpus can be grammatically incorrect, contain repairs, repetitions and inexact wordings.
4. The movie reviews texts are usually consistent regarding the expression of reviewer’s opinion about the movie whereas some utterances in the SAL corpus can be inconsistent (can convey contradicting emotional meaning).
5. There is no connection between particular reviews in the movie review corpus whereas affect segment values for utterances in a dialogue can be seen as a continuous stream of information between utterances that influences the set of possible features e.g. the feature set can contain a feature for the previous average value.
6. A movie review is composed and rated by the same person whereas test persons and critics in the SAL corpus are different people. This raises a lot of questions e.g. if a critic has the same or at least a comparable impression of verbal entities regarding their affect meaning as a person he/she rates.

Due to these differences, we also expected different results for automated methods to affect sensing.

FEATURE EXTRACTION

For each corpus, we computed the following features:

1. Affect word features. Affect words from the Whissell’s dictionary of affect as in the lexical affinity approach. Chosen affect words can belong to every part-of-speech (POS) group out of 40 (see below), be it an adverb, a pronoun whatsoever. The set of affect words doesn’t contain every affect word that can appear in an affect utterance in every grammatical form. Section 4 shows that it is fairly sufficient to have only particular words in the learning model (cf. Table 2).
2. Average value features. Average values of affect scores of affect words found in a particular utterance.
3. Previous average (only in the SAL corpus). Average values of affect scores from previous utterance in a dialogue.
4. Part-of-speech (POS) tags and their combinations [5]. The features describe the number of corresponding POS groups in an utterance.

CLASSIFICATION

The extracted features are evaluated using the WEKA machine learning toolkit using the SMO classifier – a WEKA analogue of SVM [8].

Feature calculation

1. Affect word features. We performed tests with a different number of affect word features. In particular, we used all affect word features (8574) or the affect word features with the highest evaluation scores (positive or negative) according to four thresholds resulting into sets of 2451, 1307, 950 or 765 features.
2. Average value features. Average values for evaluation,

activation, and imagery are calculated from the emotional scores of affect words in the current affect text.

3. Previous average – the average from the previous affect text (only in the SAL corpus).
4. Part-of-speech (POS) tags and their combinations. The number of the features in this group is 36 conventional + 4 derived combinations (“NN or NNS or NNP or NNPS”; “VB or VBD or VBG or VBN or VBP or VBZ”; “JJ or JJR or JJS”; “RB or RBR or RBS”) and 1600 possible combinations. The values of these features are the numbers of POS or combination occurrences in an affect text (movie review or an utterance).

Results

Table 1 shows results of affect sensing in the movie review corpus and the SAL corpus. The values are averaged over affect classes.

s#	Movie Features	R (%)	P (%)	SAL Features	R (%)	P (%)
1	#W: 8574	54.76	54.54	#W: 8574	43.74	46.43
2	#W: 2451	54.76	56.66	#W: 2451	40.30	51.01
3	#W: 1307	59.52	60.95	#W: 1307	36.52	42.87
4	#W: 950	59.52	61.83	#W: 950	39.33	53.29
5	#W: 765	54.76	55.95	#W: 765	38.61	54.59
6	#W: 8574 #POS: 40	54.76	54.54	#W: 8574 #POS: 40	42.73	44.42
7	#W: 2451 #POS: 40	55.95	56.92	#W: 2451 #POS: 40	39.98	47.90
8	#W: 1307 #POS: 40	63.09	62.96	#W: 1307 #POS: 40	37.62	47.58
9	#W: 950 #POS: 40	59.52	61.82	#W: 950 #POS: 40	38.31	49.70
10	#W: 765 #POS: 40	55.95	56.38	#W: 765 #POS: 40	40.25	58.56
11	#W: 8574 #POS: 1640	59.52	57.91	#W: 8574 #POS: 1640	41.48	45.36
12	#W: 2451 #POS: 1640	63.09	62.82	#W: 2451 #POS:1640	40.57	46.82
13	#W: 1307 #POS: 1640	64.28	64.45	#W: 1307 #POS:1640	40.21	45.55
14	#W: 950 #POS: 1640	63.09	63.42	#W: 950 #POS:1640	38.70	43.76
15	#W: 765 #POS: 1640	64.28	65.0	#W: 765 #POS: 1640	36.12	38.63
16	#W: 8574 A: 3	53.57	53.61	#W: 8574 A: 3	42.57	44.76
17	#W: 2451 A: 3	53.57	55.14	#W: 2451 A: 3	39.89	50.07
18	#W: 1307 A: 3	57.14	58.89	#W: 1307 A: 3	36.67	44.23
19	#W: 950 A: 3	60.71	63.51	#W: 950 A: 3	38.84	53.39
20	#W: 765 A: 3	50.0	52.36	#W: 765 A: 3	38.69	55.88
21				#W: 8574 A: 3 Pr: 3	42.30	44.86
22				#W: 2451 A: 3 Pr: 3	40.38	51.41
23				#W: 1307 A: 3 Pr: 3	36.67	43.92
24				#W: 950 A: 3 Pr: 3	38.72	50.20
25				#W: 765 A: 3 Pr: 3	38.76	56.90

Table 1: Results of affect sensing

The Features column designates the features used where “#” column shows the row number, #W is the number of affect words from the Whissell’s dictionary of affect, A represents the average score feature and Pr the previous average score (only in the SAL corpus). The R, P columns are the conventional recall and precision measures. The number after the colon represents the number of values for this feature.

Feature comparison

It is impossible to give a clear answer to the question of whether affect words influence affect sensing in an obvious manner (cf. Table 1). Even with a smaller intersection of words from the Whissell’s dictionary of affect it is possible to achieve better results for affect sensing (cf. rows 2 and 3; rows 3 and 4; rows 7 and 8 in the Movie corpus or rows 9 and 10 in the SAL corpus). Note that the result values don’t introduce monotonicity regarding the number of affect words (higher words number means higher precision and recall values) e.g. in the SAL corpus row 1 contains the highest value of the recall measure whereas row 5 contains the highest value of the precision measure.

Table 2 presents numerical characteristics of occurrences of affect words in the Whissell’s dictionary of affect. Note the lowest intersection of the set of the words from the Whissell’s dictionary of affect and the set of words from the movie review corpus is 5.08% whereas the intersection of the set of the words from the Whissell's dictionary of affect and the set of words from the SAL utterances is not greater than 2.04%.

#W	#M.I.	#M.A.	M./W. (%)	#S.I.	#S.A.	S./W. (%)
8574	4815	6647	56.15	175	25	2.04
2451	1395	10067	16.27	41	159	0.47
1307	750	10712	8.74	21	179	0.24
950	541	10921	6.30	17	183	0.19
765	436	11026	5.08	10	190	0.11

Table 2: Intersections of affect words

where #W is the number of affect words from the Whissell's dictionary of affect, #M.I. – the intersection of the set of words from the movie review corpus and the set from the Whissell's dictionary of affect, #M.A. – the number of words from the movie review corpus that are absent in the set of affect words from the Whissell's dictionary of affect, M./W. – the ratio of the set of words from the movie review corpus and the dictionary of affect, #S.I. – the intersection of the set of words from the SAL scenario and the set from the Whissell's dictionary of affect, #S.A. is the number of words from the SAL that are absent in the set of affect words from the Whissell's dictionary of affect, S./W. is the ratio of the set of words from the SAL corpus and the dictionary of affect.

Our classification algorithm provided better results for the movie review corpus (the precision measure for a three class problem – "negative", "neutral", "positive" – is on average 58.98%) than for the SAL corpus (the precision measure for a three class problem – "*", "**", "***"– is on average 48.20%) independently of the number of affect word features. Another interesting observation is that a reduction of affect word features for the movie corpus yielded better results. Only when reducing the set to 765 affect words the results were similar again to the results obtained for 8574 words. In contrast, the best results for the SAL corpus were obtained when using the full set of 8574 features.

The POS feature improves affect sensing in the movie review corpus and at least influences the processing in the SAL corpus (cf. rows 1-5 with rows 11-15) although the influence of POS groups on affect sensing in the SAL corpus is less obvious. A higher number of POS groups (1640 vs. 40) always led to higher recall and precision values.

The average affect score features had no essential influence on the recognition rate. In the SAL corpus, for example, using 1307 affect word features the recall measure was 36.52% and precision measure 42.87%, whereas using 1307 affect word features and the average score features – 36.67% and 44.23% resp. (cf. rows 1-5 with rows 16-20 in both corpora). Note the results in row 19 (38.84%-53.39%).

The previous average feature influences affect sensing. Worse results are achieved in rows 18 and 23, better results in rows 20 and 25.

In sum, all the features above can positively influence classification results although it is sometimes intuitively difficult to find a correct balance between feature values.

CONCLUSION

This contribution conducted a comparison of two emotional textual corpora. It also provided features for classifying affect utterances in these corpora. The best classification results for three affect segments are – for the recall measure 64.28% and 43.74%, for the precision value 65% and 55.88% in the movie review corpus and SAL corpus respectively.

In our future research, the following areas will be investigated more thoroughly taking into consideration the differences between emotional text corpora:

1. Using stemming algorithms to recognize affect word features
2. Testing of different algorithms for feature selection
3. Considering a longer dialogue history both for affects score features as well as POS features.

REFERENCES

1. Berardinelli, J. (2006). URL: http://movie-reviews.colossus.net.
2. Cowie, R. et al. (2000). 'FEELTRACE': An instrument for recording perceived emotion in real time. In: Proceedings of the ISCA Workshop on Speech and Emotion, Northern Ireland. pp. 19–24.
3. HUMAINE (2006). Research on Emotions and Human-Machine Interaction. URL: emotion-research.net.
4. Liu, H. & Lieberman, H. & Selker, T. A (2003). Model of Textual Affect Sensing using Real-World Knowledge. Proceedings of IUI 2003. Miami, Florida.
5. Mitchell P. M. & Santorini B. & Marcinkiewicz M.A.. (1993). Building a large annotated corpus of english: The penn treebank. Computational Linguistics, 19:313-330.
6. Semantic Affect Listener – SAL (2006). URL: http://emotion-research.net/ws/wp5/ellsal.ppt.
7. Whissell C. (1989). The dictionary of Affect in Language. In: R Plutchik & H Kellerman ed. Emotion: Theory, research and experience vol. 4 Academic Press: New York.
8. Witten I. H. & Frank E. (2005). Data Mining: Practical Machine Learning Tools and Techniques, Second Edition.