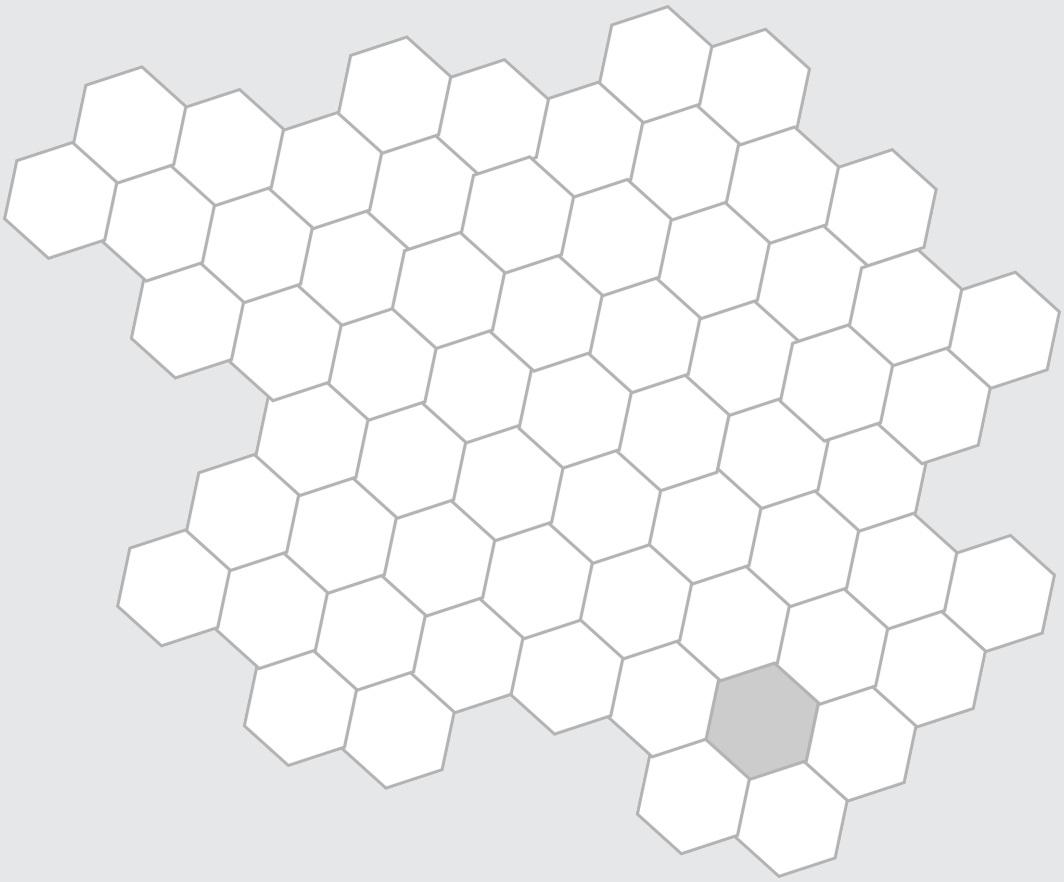
International Journal on

SOCIAL MEDIA

MMM: Monitoring, Measurement, and Mining



I, 2010, 1
ISSN 1804-5251

...OFFPRINT...

International Journal on SOCIAL MEDIA

MMM: Monitoring, Measurement, and Mining

I, 2010, 1

Editor-in-Chief: Jan Žižka

Publisher's website: www.konvoj.cz
E-mail: konvoj@konvoj.cz, SoNet.RC@gmail.com
ISSN 1804-5251

The Study of Sentiment Word Granularity for Opinion Analysis (A Comparison with Maite Taboada works)

Olga Kaurova
Department of French and Romance Philology, Faculty of Philosophy and Arts, Autonomous University of Barcelona, 08193 Bellaterra, Spain
e-mail: kaurovskiy@gmail.com

Mikhail Alexandrov
Department of French and Romance Philology, Faculty of Philosophy and Arts, Autonomous University of Barcelona, 08193 Bellaterra, Spain
e-mail: malexandrov@mail.ru

Natalia Ponomareva
Statistical Cybermetrics Research Group, School of Computing and IT, University of Wolverhampton, Stafford Str, WV1 1SB Wolverhampton, UK
e-mail: nata.ponomareva@wlv.ac.uk

Abstract

Sentiment Analysis (SA) is an area of NLP related to automatic evaluation of people's opinions and their attitudes to various objects and events. Nowadays OA has become an important part of Social Network Analysis, and researchers suggest different tools for solution of this problem. The semantic orientation calculator (SO-CAL) developed in Maite Taboada's group is one such effective tool, which uses dictionaries of sentiment words over a detailed sentiment scale (5 positive and 5 negative levels). In the paper we study the influence of granularity levels of sentiment words on the accuracy of sentiment classification in order to verify the possibility of using lesser granularity without a substantial decrease in performance. We exploit one- and two-parameter linear regression models as a classification method and product reviews of different categories (books, cars, movies, etc.) as a corpus. The results show that there is no significant difference between one-/ and two-parameter models; neither is there a need for a fine-grained granularity of sentiment.

Key words opinion analysis; sentiment classification

Biographical note

Olga Kaurova graduated from the Saint Petersburg State University as a specialist in theoretical and applied linguistics in 2009. She is currently a graduate student at the Autonomous University of Barcelona in Spain (International Master's Program in Natural Language Processing & Human Language Technology. Her area of research interests – Sentiment and Subjectivity Analysis, Language Acquisition.

Mikhail Alexandrov is a member of the fLexSem Research group at the Autonomous University of Barcelona in Spain. He is a professor of the Academy of National Economy under Russian Government. He is an applied mathematician and author of numerous publications related to mathematical modelling and natural language processing. His current topics of research are machine learning (inductive modelling, clustering) and internet technologies (social networking).

Natalia Ponomareva is a PhD student at the University of Wolverhampton. She received her master's degree from the Technical University of Valencia. Her research interests include Sentiment, Sentiment Transfer and Machine Learning for NLP. She is author of more than 10 scientific publications in international conferences and journals.

Introduction

We always have to deal with subjectivity in our everyday life, have to take into consideration other people's opinions, and now with the growth of the WWW we get a quick and easy access to a great quantity of subjective information – opinionated texts: users' reviews, forums, blogs, etc. Analysis of this opinionated web content is becoming increasingly important both for individual and for business aims: for example, consulting consumer reports when choosing a brand of washing machine to buy, or monitoring the company's efficiency and satisfaction of its customers. Many online shopping sites, e.g. Amazon and eBay, give customers the possibility to leave their comments and reviews of the products they purchased. Moreover, there are even special sites, devoted to user's opinions: epinions.com and others. Thus, easy access to subjective data, on the one hand, and their large quantities and low level of order, on the other hand, determine the rapid development and great importance of Sentiment Analysis, which nowadays occupies a significant place in Natural Language Processing.

Related work

Sentiment analysis is a broad area of NLP, which concerns the automatic determination of text subjectivity (whether a text is objective or subjective), polarity (positive or negative) and sentiment strength (strongly or weakly positive/negative). One of the main tasks of sentiment analysis is a binary sentiment classification which aims to assign to an opinionated document either an overall positive or an overall negative opinion (sentiment polarity classification or polarity classification). There are two different approaches to achieving this aim: a lexical (lexicon-based) approach (Turney, 2002) and a machine learning approach (Pang et al., 2002). The machine learning approach uses collections of labelled texts as training data in order to build automated classifiers. The lexical approach is based on semantic orientation (SO) lexicons (words with their semantic orientation) (Hatzivassiloglou and McKeown, 1997), and calculates overall sentiment by aggregating the values of those words presented in a text or a sentence. Besides polarity classification of documents, other sentiment classification tasks have been receiving lots of attention in the research community: sentiment classification of subjective expressions (Wilson et al., 2005; Kim and Hovy, 2004), subjective sentences (Pang and Lee, 2004) and topics (Yi et al., 2003; Nasukawa and Yi, 2003; Hiroshi et al., 2004). These tasks analyse sentiment at a fine-grained level and can be used to improve the effectiveness of sentiment classification, as shown in the study of Pang and Lee (2004).

Maite Taboada's SO-CAL, which the present study is based on, belongs to the lexical approach. It is an automated system which uses low-level semantic and syntactic information to calculate the overall polarity of texts. So-CAL uses SO-

Dictionaries, which are lists of manually-tagged sentiment words. The current version consists of four open-class dictionaries (nouns, adjectives, adverbs and verbs) and one closed class-dictionary of intensifiers. The integer SO value assigned to each word varies between -5 and 5. The calculation of the sentiment orientation of a text is accomplished, roughly speaking, by summing the values of all the words occurring, also taking into account negation, intensification and other language phenomena. The use of a 10-point scale (excluding zero) of SO seems to be a compromise between an attempt to capture clear differences in word meaning on the one hand, and the difficulty in assigning extremely fine-grained values to out-of-context words on the other hand. The numerical values were chosen to reflect both the prior polarity and strength of the word, averaged across likely interpretations (Brooke et al., 2009).

Problem settings

The present study is founded on the work of Maite Taboada's research group, namely on their Semantic Orientation Dictionaries and a corpus of reviews, which we were kindly provided with by them. Our main objectives are:

- To study the influence of the adopted sentiment granularity scale on the accuracy of sentiment classification within the framework of the regression model.
- To compare performances of one- and two-parameter models, where the former model takes into account a summed contribution of positive and negative words while the latter model considers positive and negative scores separately as independent variables.
- To analyse the accuracy of regression models for each of the object categories of the corpus.
- To construct an integral model (built on all categories) and test its applicability to individual categories.

Models for decision-making

Source data and vocabularies

The corpus that served as material for the present work was developed by Maite Taboada's research group (Taboada et al., 2006). It is a collection of 400 texts obtained from the website Epinions (`www.epinions.com`). The reviews are divided into 8 object categories: books, cars, computers, cookware, hotels, movies, music and phones. Each category contains a set of 50 reviews: 25 positive and 25 negative. The reviews vary greatly in length: from several phrases to several pages. All the texts are written in English. General characteristics of the corpus are given in the table below.

We use four manually-ranked SO Dictionaries (nouns, adjectives, adverbs and verbs), where the integer SO value assigned to each word varies between 5 and 5. The dictionary of intensifiers is left out for the fact that we do not take into account negation, intensification, modality, etc. in this study. In Table 2 we present the size of the dictionaries.

Table 1: Characteristics of the experimental corpus

Characteristics	Value
Total number of reviews	400
Number of categories	8
Number of reviews in a category	50
Review's sentiment value	+1 / −1

Table 2: The size of SO-CAL dictionaries

Dictionary	No. of Entries
Adjectives	2257
Adverbs	745
Nouns	1142
Verbs	903

Document parameterisation

To examine the influence of SO granularity we introduced 5 different models of roughening the granularity scale (Table 3). According to these models we modified SO values in SO Dictionaries and performed parameterisation of each review of the experimental corpus. Table 4 presents an example of an output file after applying our document parameterisation procedure implemented in Python (fragment from the category BOOKS).

Table 3: Models of sentiment contribution

Model	SO-scale	Modified SO-scale
Model 1	−5, −4, −3, −2, −1, 1, 2, 3, 4, 5	−5, −4, −3, −2, −1, 1, 2, 3, 4, 5
Model 2	[−5, −4], −3, [−2, −1], [1, 2], 3, [4, 5]	−3, −2, −1, 1, 2, 3
Model 3	[−5, −4, −3], [−2, −1], [1, 2], [3, 4, 5]	−2, −1, 1, 2
Model 4	[−5, −4], [−3, −2, −1], [1, 2, 3], [4, 5]	−2, −1, 1, 2
Model 5	[−5, −4, −3, −2, −1], [1, 2, 3, 4, 5]	−1, 1

Regression models

We chose linear regression as the principal method of our analysis for several reasons. First of all, when considering one-parameter regression model with joint contribution of positive and negative words, our method becomes similar to the approach implemented in SO CAL. Second, this model can easily be adapted to multi-scaled sentiment classification although in this work only binary classification was carried out. Finally, a small amount of training data does not allow the exploitation of more sophisticated machine learning algorithms.

One of the objectives of this study is to check whether a two-parameter regression model (or Separate model), where contributions of positive and negative words are considered separately as independent parameters, has any advantage

Table 4: Fragment of an output of the Python program for SO calculation

file	name	npw	nnw	PS 1	NS 1	PS 2	NS 2	PS 3	NS 3	PS 4	NS 4	PS 5	NS 5
No 01	txt	13	9	19	−14	15	−11	15	−11	13	−9	13	−9
No 02	txt	8	8	13	−22	10	−15	10	−13	8	−10	8	−8
No 03	txt	44	23	73	−44	56	−30	55	−28	45	−25	44	−23
No 04	txt	10	6	15	−8	10	−7	10	−7	10	−6	10	−6
No 05	txt	40	18	65	−27	48	−21	47	−21	41	−18	40	−18
...	...	...	...	...	...	...	...	...	...	...	...	...	...

Legend: npw = number of positive words; nnw = number of negative words;
NS = negative score; PS = positive score

over a one-parameter model (or Joint model), where contributions of sentiment words are summed up.

Formula (1) presents Joint and Separate models we are going to construct and compare.

$$\text{Joint Model: } F_j = A_0 + A_1(\text{PosScore+NegScore}) \tag{1}$$

$$\text{Separate Model} F_s = A_0 + A_1\text{PosScore} + A_2\text{NegScore},$$

where A_0, A_1, A_2 are unknown coefficients.

We build linear regression models for different levels of sentiment granularity in order to find out whether finer-grained sentiment scales can significantly improve the results of classification. Besides individual models for each object category (categorial models), an integral (or multicategorial) model based on the whole dataset is constructed. It is used to verify the possibility of applying the same model to all categories without a substantial decrease in performance.

Prior to model testing we apply some modification to the variables. First of all, in order to avoid the dependency of sentiment scores on text length we normalise them on the total number of sentiment-words in a review. In order to be able to compare models of different levels of granularity, we adjust their variables to the same scale, namely, (-1, 1), by introducing a scale factor. The resultant forms of variables for different models are presented in Table 5 (Np stands for the number of positive sentiment-words in the text, and Nn – that of negative words). Sentiment scores before and after application of normalisation and scaling are shown in Figures 1 and 2.

Table 5: Normalisation and scaling of coefficients for model testing

Model	Joint Sentiment Score	Separate Sentiment Scores	
1	(PS 1 + NS 1) / (N_p+N_n) / 5	PS 1 / (N_p+N_n) / 5	NS 1 / (N_p+N_n) / 5
2	(PS 2 + NS 2) / (N_p+N_n) / 3	PS 2 / (N_p+N_n) / 3	NS 2 / (N_p+N_n) / 3
3	(PS 3 + NS 3) / (N_p+N_n) / 2	PS 3 / (N_p+N_n) / 2	NS 3 / (N_p+N_n) / 2
4	(PS 4 + NS 4) / (N_p+N_n) / 2	PS 4 / (N_p+N_n) / 2	NS 4 / (N_p+N_n) / 2
5	(PS 5 + NS 5) / (N_p+N_n)	PS 5 / (N_p+N_n)	NS 5 / (N_p+N_n)

Legend: PS = positive score; NS = negative score

All constructed regression models are checked for their statistical significance by a global test (F-test) and tests on individual variables (t-test).

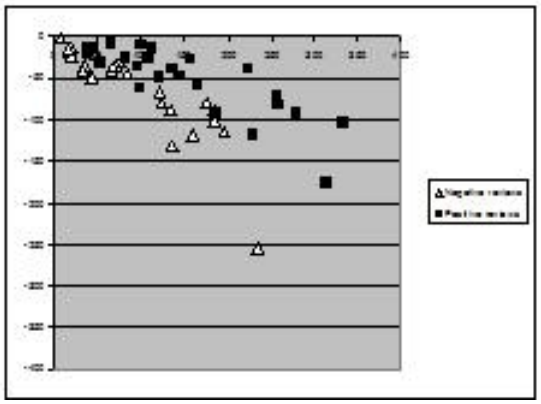


Figure 1:
Distribution of scores before normalisation

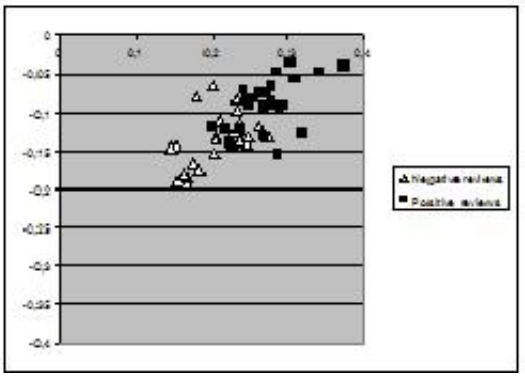


Figure 2:
Distribution of scores after normalisation

Model accuracy and model comparison

In as far as the scale of sentiment of reviews under consideration is binary: +/−1, the coefficients of determination R2 and the values of standard errors are not representative for the evaluation and comparison of regression models. Therefore we apply a cross-validation technique to estimate model accuracies (accuracy in our case means the probability of correct classification of a review as negative or positive). Taking into account the fact that the experimental data are not very large and the level of noise is high, the leave-one-out cross-validation method is used.

In order to compare model accuracies the z-test is used (it is justified because the number of observations in all experiments exceeds 30). We apply this statistical method to check the null-hypothesis that there is no statistically significant difference in the accuracies of models under consideration. Z is calculated according to the following formula (2):

$$Z = \frac{p_I - p_I I}{\sqrt{S_I^2 + S_I^2 I}} \qquad S = \sqrt{\frac{pq}{n}} \tag{2}$$

where p is the probability of correct classification of a review (i.e. accuracy); q – the probability of incorrect classification; n – number of observations; S – standard deviation; numbers I and II are the 2 models that are compared.

For a confidence level equal to 0.95 (α = 0.05) the null-hypothesis is confirmed if Z < 1.96.

Experiments

Testing one-parameter and two-parameter models

First, we aim to find out whether the Separate model outperforms the Joint model (1). In order to compare them we construct integral models over all the categories (Table 7) and the 'best' category models (Table 6), i.e. CARS (during the experiments it was noticed that the category CARS gives consistently better results than the rest of the categories).

Table 6: Accuracies of one-/ and two-parameter models built on CARS

	Model 1	Model 2	Model 3	Model 4	Model 5
1 variable	0.8	0.76	0.78	0.78	0.8
2 variables	0.8	0.68	0.72	0.8	0.78

Table 7: One-/ and two-parameter multicategorial models

	Model 1	Model 2	Model 3	Model 4	Model 5
1 var					
M:	4.71*score−0.69	4.18*score−0.73	3.08*score−0.72	3.55*score−0.73	1.88*score−0.68
A:	0,780	0,774	0,760	0,746	0,714
2 var					
M:	5.52*pos+3.73*neg −1.03	5.08*pos+3.08*neg −1.16	2.36*pos+3.9*neg−0.23	6.71*neg+1.15	3.76*pos−2.56
A:	0,783	0,774	0,746	0,703	0,706

Legend: M = Model; A = Accuracy

It is interesting to draw attention to the fact that when building regressions for two-parameter Model 5 the equation always transforms into a one-parameter one. This is due to the fact that for this model the overall sentiment contribution is equal to the number of sentiment-words. There is a functional dependency between the parameters pos and neg (neg = pos − 1), which makes regression impossible.

To compare the models we firstly apply the z-test (formula (2); n=400) to multicategorial models (Table 7), namely models 1, 2 and 3 as the other two have been transformed into single-parameter. The result is that for every pair Z<1.96; the null-hypothesis is confirmed, therefore, there is no statistically significant difference between the models. We then carry out the same test (formula (2); n=50) for the category CARS (Table 6) with the same result.

We conclude that in as far as there is no statistically significant difference, a one-parameter model is preferable.

Testing model granularity

Comparison of regression models for different granularity levels of sentiment words is accomplished using the one-parameter model as it proved to be preferable in the previous section. The constructed models and their accuracies for all individual categories are presented in Table 8 (A stands for model accuracy). We should note that regression could not be built on the category BOOKS due to a high level of inconsistency of sentiment-words contributions.

Table 8: One-parameter categorial models of different granularity scale

M:	books	cars	computers	cookware	hotels	movies	music	phones
1	–	8.37*X−1.07 A=0.8	5.77*X−0.85 A=0.8	4.4*X−0.71 A=0.74	5.54*X−1.1 A=0.8	4.48*X−0.49 A=0.76	4.54*X−0.58 A=0.72	3.79*X−0.61 A=0.7
2	–	6.55*X−1.02 A=0.76	5.06*X−0.87 A=0.82	4.18*X−0.78 A=0.74	5.08*X−1.16 A=0.82	4.03*X−0.54 A=0.74	3.9*X−0.6 A=0.72	3.31*X−0.62 A=0.7
3	–	4.61*X−0.99 A=0.78	3.85*X−0.88 A=0.76	3.44*X−0.85 A=0.76	3.95*X−1.18 A=0.78	2.98*X−0.53 A=0.76	2.7*X−0.56 A=0.7	2.25*X−0.6 A=0.7
4	–	5.58*X−1.1 A=0.78	4.13*X−0.89 A=0.72	3.64*X−0.79 A=0.7	4.46*X−1.1 A=0.78	3.49*X−0.55 A=0.74	3.27*X−0.58 A=0.7	2.69*X−0.59 A=0.68
5	–	2.91*X−1.04 A=0.8	2.29*X−0.86 A=0.7	2.27*X−0.83 A=0.74	2.48*X−1.08 A=0.74	1.97*X−0.53 A=0.72	1.51*X−0.47 A=0.62	1.19*X−0.49 A=0.6

Legend: M = Model

The models are compared using the z-test (formula (2)) where n = 50 for categorial models and n=400 for the integral model. The comparison is carried out in pairs: Model 1 is compared to all others. In Table 9 we present the results of the comparison of Model 1 to the one with the lowest accuracy within a category. Table 10 shows the comparison of multicategorial Model 1 to other multicategorial models. In as far as the maximum value of z-statistics for all categories (except for multicategorial Model 5) is less than Z-criterial, we infer that there is no significant difference between any compared models. Figure 3 presents a comparison of accuracy of models of different granularity within one-parameter regression for the best and the worst category.

Despite the fact that a change of the granularity scale does not give a statistically significant difference in the performance, it cannot escape our attention that the accuracy monotonically decreases with the roughening of the granularity scale (when an integral model is considered). Therefore, we do not exclude the possibility that a greater amount of data will reveal the higher impact of fine-grained sentiment scales.

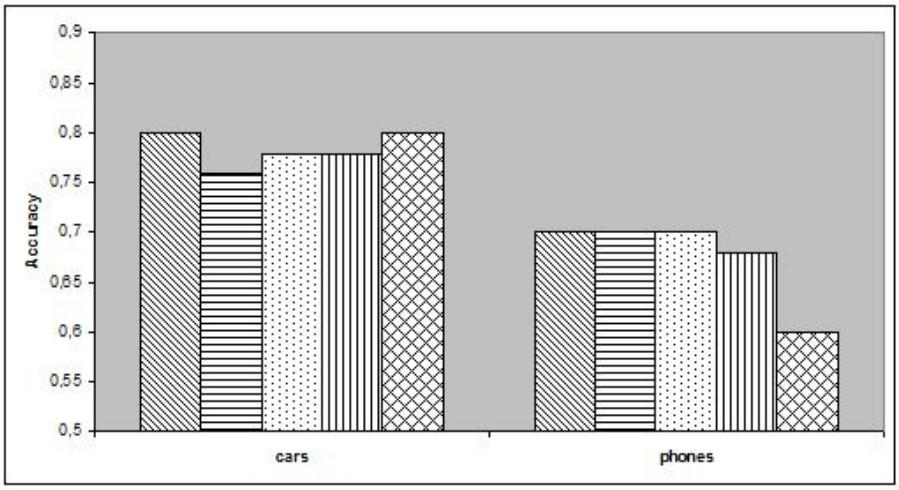
Table 9: Max z-statistics for each category

	cars	computers	cookware	hotels	movies	music	phones
Z-statistics	0.48	1.16	0.45	0.71	0.46	1.07	1.05
Z-criterial (5%)	1.96	1.96	1.96	1.96	1.96	1.96	1.96

Table 10: Z-statistics for multicategorial model

	Model 2	Model 3	Model 4	Model 5
Z-statistics	0.204	0.672	1.132	2.153
Z-criterial (5%)	1.960	1.960	1.960	1.960

Figure 3: Comparison in accuracy of five one-parameter models for CARS and PHONES



Testing the multicategorial model

For the purpose of studying the possibility of domain transfer and application of regression models to a corpus of unknown subject we apply one-parameter multicategorial models (Table 7) to each object category comparing the accuracy to that of corresponding categorial models. In Table 11 there are accuracies of local categorial models and of applied multicategorial models. Figures 4 and 5 show the comparison of accuracy for multicategorial and local models of type 1 and 5 for all categories.

To examine the proximity of the multicategorial model to categorial ones confidence intervals for accuracy in each category are calculated:

$$P = p \pm Z_\alpha \sigma \tag{3}$$

where P is a range of possible accuracy values; $Z_\alpha/2$ - quartile measure (α=5%, Z=1.96); σ – mean square value of accuracy. The results for Model 1 are presented in Table 12. The accuracy of every multicategorial model lies within the confidence interval of accuracy of categorial models; the multicategorial model is therefore representative for every object category.

Table 11: Accuracy of multicategorial models for each category

	Model 1	Model 2	Model 3	Model 4	Model 5
books_all	0.56	0.60	0.56	0.56	0.56
books_local	–	–	–	–	–
cars_all	0.86	0.80	0.80	0.82	0.80
cars_local	0.80	0.76	0.78	0.78	0.80
comp_all	0.84	0.82	0.80	0.74	0.70
comp_local	0.80	0.82	0.78	0.72	0.70
cook_all	0.78	0.76	0.78	0.78	0.76
cook_local	0.74	0.74	0.76	0.70	0.74
hotels_all	0.78	0.78	0.76	0.78	0.72
hotels_local	0.80	0.82	0.78	0.78	0.74
movies_all	0.78	0.80	0.82	0.76	0.76
movies_local	0.76	0.74	0.76	0.74	0.72
music_all	0.68	0.72	0.68	0.68	0.64
music_local	0.72	0.72	0.70	0.70	0.62
phones_all	0.74	0.72	0.70	0.68	0.68
phones_local	0.70	0.70	0.70	0.68	0.60

Table 12: Confidence intervals for model accuracy (Model 1)

	P	Multicat. Accuracy
cars	0.80±0.162	0.86
comp	0.80±0.167	0.84
cook	0.74±0.174	0.78
hotels	0.80±0.162	0.78
movies	0.76±0.172	0.78
music	0.72±0.184	0.68
phones	0.70±0.187	0.74

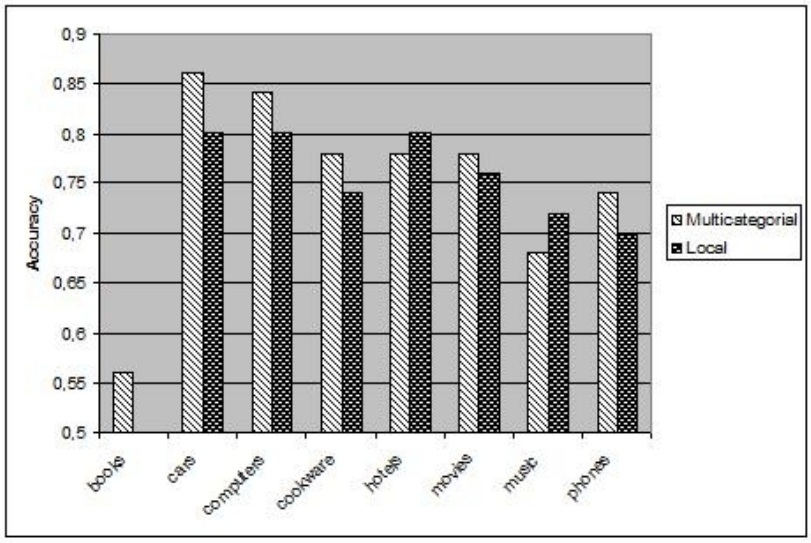
Conclusions

Discussion

In this study we have compared different regression models in the framework of binary sentiment classification (positive/negative). Applying the z-test to the results of leave-one-out cross-validation we firstly compared one-/ and two-parameter models, both local and multicategorial. The null-hypothesis was confirmed, which implies that there is no statistically significant difference between the given types of models. We therefore infer that in as far as there is no difference there is no reason to use a more complex model, namely the two-parameter one.

Comparison of five models of different granularity levels has shown the same result: there is no statistically significant difference between different types of categorial model (cf. Table 9). As for multicategorial models – the difference only appears at the level of model 5. The summary comparison of accuracy for the best and the worst category, Figure 3, showed that a model with a rougher granularity

Figure 4: Comparison of multicategorial and local models (for model 1)



scale can be used with no loss in performance.

The construction of the multicategorial model on the whole corpus produces a level of accuracy which correlates to the results obtained by Maite Taboada's research group (0,78 in our case vs. 0,83 – Brook, 2009). Note, that in contrast to Maite we did not take negation and intensification into account.

Applying a multicategorial model to local categorial ones (Table 14, Figures 4 and 5) showed that the multicategorial model is representative for every object category and can be successfully applied even to 'bad' categories.

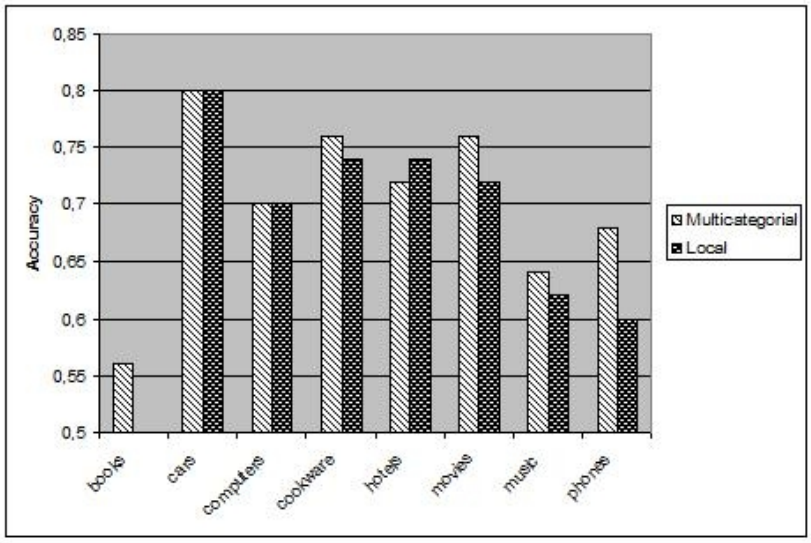
To sum up, we conclude that our results are not inferior to Maite Taboada's: simpler models can be used for the purpose of sentiment classification with no loss in performance as reducing the sentiment scale would in turn reduce the subjective influence of the raters (note that SO Dictionaries used in SO-CAL are manually rated sentiment-words and the finer the scale, the more difficult it is to assign correctly the value of the intensity of the emotion).

Future work

Possible developments of the present work might include construction of models for a more detailed scale of sentiment, that is, when reviews are classified not only within the binary polarity of positive vs. negative, but using a triple classification: positive, negative, neutral. The models for triple classification can be further analysed for the possibility of constructing models of 5 levels of sentiment classification: very positive, positive, neutral, negative, very negative.

The other option is testing the usage of Bayes classifiers for obtaining sentiment assessments and comparing the results with those when using regression

Figure 5: Comparison of multicategorial and local models (for model 5)



models.

In this study we tested our regression models on a corpus containing reviews obtained from the Internet, these reviews being written by ordinary users on product categories such as books, hotels, computers, etc. It would be useful to apply regression and Bayes models to specialised subject areas instead (economics, physics, etc), for example, with the aim of facilitating critical reviews of articles on a given area of knowledge.

References

Brooke, J., Tofiloski, M., Taboada, M. (2009): Cross-Linguistic Sentiment Analysis: From English to Spanish. In: *Proceedings of the 7th International Conference on Recent Advances in Natural Language Processing*. Borovets (Bulgaria), pp. 50–54.

Cramer, H. (1946): *Mathematical methods of statistics*. Cambridge, .

Hatzivassiloglou, V., McKeown, K. (1997): Predicting the semantic orientation of adjectives. In: *Proceedings of 35th Meeting of the Association for Computational Linguistics*. Madrid (Spain), pp. 174–181.

Hiroshi, K., Tetsuya, N., Hideo, W. (2004): Deeper sentiment analysis using machine translation technology. In: *Proceedings of the 20th international conference on computational linguistics (COLING)*. Geneva (Switzerland), pp. 494–500.

Kim, S.-M., Hovy, E. (2004): Determining the sentiment of opinions. In: *Proceedings of the 20th international conference on computational linguistics (COLING)*. Geneva (Switzerland), pp. 1367–1373.

Nasukawa, T., Yi, J. (2003): Sentiment analysis: capturing favorability using natural language processing. In: *Proceedings of the 2nd international conference on knowledge capture*. Florida (USA), pp. 70–77.

Pang, B., Lee, L. (2004): A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: *Proceedings of the 42nd annual meeting of the Association for Computational Linguistics (ACL)*. Barcelona (Spain), pp. 271–278.

Pang, B., Lee, L. (2008): Opinion mining and sentiment analysis. *Foundation and Trends in Information Retrieval*, Vol. 2, No. 1–2, pp. 1–135.

Pang, B., Lee, L., Vaithyanathan, S. (2002): Thumbs up? Sentiment classification using Machine

Learning techniques. In: *Proceedings of Conference on Empirical Methods in NLP* , pp. 79–86.
Taboada, M., Anthony, C., Voll K. (2006): Creating semantic orientation dictionaries. In: *Proceedings of 5th International Conference on Language Resources and Evaluation (LREC)*. Genoa (Italy), pp. 427–432.
Taboada, M., Grieve, J. (2004): Analyzing appraisal automatically. In: Qu, Y., Shanahan, J. G., Wiebe, J. (eds.) *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS-04-07)*. Stanford University, CA : AAAI Press, pp. 158–161.
Turney, P. (2002): Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: *Proceedings of 40th Meeting of the Association for Computational Linguistics* , pp. 417–424.

Yi, J., Nasukawa, T., Niblack,W., Bunescu, R. (2003): Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques. In: *Proceedings of the 3rd IEEE international conference on data mining (ICDM)*. Florida (USA), pp. 427–434.
Wilson, T., Wiebe, J., Hoffmann, P. (2005): Recognizing contextual polarity in phrase-level sentiment analysis. In: *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*. Vancouver (B.C., Canada), pp. 347–354.

We would like to thank Maite Taboada for kindly providing SO-Dictionaries and a review corpus for the present study.

Call for Papers

Papers to be included in the next issue should be preferably focused on topics related to social-networks in one or more of the following subjects (the list is indicative rather than exhaustive):

Sentiment/Opinion Analysis in Natural-Language Text Documents

Algorithms, Methods, and Technologies for Building and Analysing Social Networks

Applications in the Area of Social Activities

Knowledge Mining and Discovery in Natural Languages Used in Social Networks

Medical, Economic, and Environmental Applications in Social Networks

Submitted papers should not have been previously published nor be currently under consideration for publication elsewhere. Each of the submitted research papers should not exceed 26 pages. All papers are refereed through a peer review process.

Submissions should be send in the PDF form via email to the following address: `SoNet.RC@gmail.com`

Accepted papers are to be prepared according to the instructions available at `http://www.konvoj.cz/journals/mmm/`.