УДК 81'33

Comparable Domain Dependency in Sentiment Analysis

Jeremy Reffin[a],
Taras E. Zagibalov[a] and Ekaterina O. Belyatskaya[b]
[a] University of Sussex
Sussex House Brighton BN1 9RH United Kingdom
[b] Siberain Federal University
79 Svobodny, Krasnoyarsk, 660041 Russia [1]

Sentiment analysis (or opinion mining) is concerned not with the topic of a document, or its factual content, but rather with the opinion expressed in a document. In this paper we present a number of experiments on a word-based sentiment analysis on two corpora representing two related domains: film reviews and book reviews. We find that even close domains are very difficult to process without utilising in-domain data. We also indicate certain characteristics of features that affect cross-domain performance of sentiment classifiers.

Keywords:sentiment analysis; domain-dependency; opinion mining

Introduction

Opinion orientation is usually a three-way classification of positive, negative, or neutral, and can be applied to different levels of the text: phrases, sentences, documents or collections of documents. Ways in which opinions are expressed can vary not only between languages, but also within languages (so-called "domain-dependency"). In this paper we investigate the cross-domain portability of different kinds of word-based features. The key issue to be investigated here is whether lexical features developed to discern sentiment in one domain can perform similarly in a different but closely related domain.

The paper is structured as follows: in Related Work we briefly cover studies in the field of sentiment analyses, more specifically those dealing with supervised word-based approaches in the context of cross-domain studies. In Corpus we present the data we use for our experiments. The section Classifier Development presents the approaches we used to develop classifiers for sentiment in a particular domain (film reviews) and the classifiers' performance in that domain. In Cross-Domain Comparisons we apply these classifiers to a different but related domain (book reviews) and compare their performance.

Related Work

Most work on sentiment classification has used approaches based on supervised machine learning. For example, (Pang, Lee, & Vaithyanathan, 2002) collected movie

reviews that had been annotated with respect to sentiment by the authors of the reviews, and used this data to train supervised classifiers. A number of studies have investigated the impact on classification accuracy of different factors, including choice of feature set, machine learning algorithm, and pre-selection of the segments of text to be classified.

While supervised systems generally achieve reasonably high accuracy, they do so only on test data that is similar to the training data. To move to another domain one would have to collect annotated data in the new domain and retrain the classifier. (Engström, 2004) reports decreased accuracy in cross-domain classification since sentiment in different domains is often expressed in different ways. (Read, 2005)[1] also observed significant differences between the accuracy of classification of reviews in the same domain but published in different time periods.

Corpus

We used two corpora representing two related domains: film reviews and book reviews. The former corpus was made by (Pang & Lee, 2004) and is frequently used for sentiment analyses experiments.

The corpus of film reviews contains 1000 positive and 1000 negative reviews all written before 2002, with a cap of 20 reviews per author (312 authors total) per category. This corpus is widely used for sentiment classification experiments and researchers report different results, ranging from 70 % of accuracy in weakly supervised experiments by (Read & Carroll, 2009) to more than 86 % in supervised classification by Pang and Lee.

The domain of film reviews is reported to be difficult for automatic sentiment analysis

(Turney, 2002). Indeed, the collection of film reviews consists of mostly long and very well-written reviews featuring rich vocabulary and a professional writing style. The average length of a positive review is 788 words, that of a negative review is slightly shorter: 707 words. Positive and negative reviews have vocabulary which is very similar in size, consisting of 36806 and 34542 words respectively, with 50920 unique words in the whole corpus. The large size of the vocabulary can be attributed not only to professional writing but also to a high number of proper names (film titles, names of actors, characters, film directors, different locations where an action takes place and so on). The wide variety of the words used in the reviews means that their frequency might be low and this may adversely affect performance of a classifier that uses frequency-based methods.

The content of the reviews is also difficult to analyse automatically. The main reason for this is a very complex and ambiguous structure of the reviews, which usually touch upon different aspects of a film, including its plot, performance of actors, camera work, historical background etc. All of these aspects may receive different sentiments which can contradict to the overall opinion. For example, consider the following example of a positive review of a film:

on a return trip from new york where he was trying to get a job, dunne is in a horrible train accident that he is the only survivor of.

The word *horrible* bears negative sentiment but in this review it is used to describe a plot, not the film. In general, most horror films may contain a lot of negative words in their description regardless of their overall quality. The opposite is true of romantic love stories, reviews of

---

[1]   Available   at   www.cs.cornell.edu/people/pabo/movie-review-data/ (review corpus version 2.0)

which may have an excessive amount of positive vocabulary regardless of overall quality.

if there are any positive things to say about " message in a bottle, " it is that the performances by robin wright penn and paul newman, as garrett's stubborn, but loving father, are far above par to be in such a wasteful, " shaggy dog " love story, and that the cinematography by caleb deschanel takes great advantage of the beautiful eastern coast, and paints chicago as an equally alluring city.

To compare the above-described corpus we developed a **book reviews corpus** which intuitively should share a lot of features with film review corpus. The English book review corpus comprises reviews of books such as: S. Erickson (Guardians of the Moon, Memories of Ice), S. King (Christine, Duma Key, Gerald's Game, Different Season and others), S. Lem (Solaris, Star Diaris of Iyon Tichy, The Cybriad), A. Rise (Interview with the Vampire, The Tale of the Body Thief and others), J.K. Rowling (Harry Potter), J.R.R. Tolkien (The Hobbit, The Lord of the Rings, The Silmarillion), S. Lukyanenko (The Night Watch, The Day Watch, The Twilight Watch, The Last Watch), and some others. The reviews were published on the website www.amazon.co.uk.

We manually annotated each review as `POS' if positive sentiment prevails or `NEG' if the review is mostly negative. The corpus consists of 1500 reviews, half of which are positive and half negative. The annotation is simple and encodes only the overall sentiment of a review, for example:

[TEXT = POS]
Hope you love this book as much as I did. I thought it was wonderful!

[/TEXT]

The reviews contain a mean of 58 words (the mean length for positive and negative reviews being almost the same) and feature a more or less equal number of documents of different lengths (mostly in the range 15 to 75). This suggests that the book reviews are stylistically different from the film reviews.

Ways of Expressing Sentiments

Sentiment can be expressed at different levels in a language, from lexical and phonetic levels up to the discourse level. Judging from the corpus, English makes heavy use of **adjectives** to express sentiment: there are 1360 reviews in the book review corpus that use adjectives to express sentiment. Apart from adjectives, which are recognised as the main tool for expressing evaluation, other parts of speech are also often used in this function, most notably verbs and nouns. The reviews also feature adverbials and interjections.

As observed by some researchers, opinions delivered by **verbs** are more expressive compared to opinions expressed in other ways. This is explained by the fact that a verb's denotation is a situation and the semantic structure of the verb reflects linguistically relevant elements of the situation described by the verb. Appraisal verbs not only name an action, but also express a subject's attitude to an event or fact.

Consider the following examples:

I truly loved this book, and I KNOW you will, too!

The English verb loved describes a whole situation which is completed by the time of reporting it.

Table 1. Ways of expressing sentiment in the English Book Review Corpus (number of documents)

	Syntactic	Lexical				Phonetic
		Verbs	Adjectives	Nouns	Other	
Positive	432	312	708	225	325	12
Negative	367	389	652	238	407	16
Total	799	701	1360	463	732	28

This means that a subsequent shift in sentiment polarity is all but impossible:

* I truly loved this book, but it turned out to be boring.

Nouns can both identify an object and provide some evaluation of it. But nouns are less frequently used for expressing opinion compared to verbs: only 463 English reviews made use of a noun to describe opinion.

Although the corpora consist of written text and do not have any speech-related mark-up, some of the review authors used **speech-related** methods to express sentiment, for example:

A BIG FAT ZEEROOOOOOOOOOOOO for M.A

There are 799 instances of **Sentence-level** means of expressing sentiment (mostly exclamatory clauses, imperatives or rhetorical questions) and they are more frequent in positive reviews.

One particularly common sentiment-relevant sentence-level phenomenon is imperative, the review author is telling their audience `what to do', which is often to read a book or to avoid doing so.

Run away! Run away!

Pick up any Pratchett novel with Rincewind and re-read it rather than buying this one

Another way of expressing sentiment by means of syntactic structure is exclamatory clauses, which are by their very nature affective. This type of sentence is widely represented in both corpora.

It certainly leaves you hungering for more!

Buy at your peril. Mine's in the bin!

The example below also features an imperative sentence used to express negative sentiment. This review also lacks any explicit sentiment markers. The negative appraisal is expressed by the verbs `stab' and `burn' that only in this context show negative attitude.

Stab the book and burn it!

The reviews often use different means of expressing sentiment, many of which are difficult (if at all possible) to process automatically. Often opinions are described through adjectives (86 % of reviews contain adjectives). The second most frequent way of expressing sentiment is through verbs (59 % of reviews have sentiment-bearing verbs). Less frequent is the noun, in 39 % of reviews. Sentence-level and discourse-level sentiment phenomena are found in 56 % of reviews. 3 % of reviews contain phonetic phenomena.

Experiments

The purpose of the experiments presented in this section is to find if features extracted from

one corpus can be effectively used for sentiment classification of reviews from another corpus representing a different domain. We will use two types of classifiers widely used for sentiment classification: score-based and supervised machine learning. We will use the movie review corpus for extracting features as it is larger and contains a richer vocabulary.

Word Based Classification

The 'bag of words' approach searches within a set of documents for correlations between words employed and the sentiment being expressed. The assumption is that words thus correlated will be an indicator of sentiment; the words used are 'causing' the sentiment to be expressed. Table 2 presents lists of 'negative' and Table 3 – 'positive' words developed using different types of this approach from the movie reviews corpus.

Raw Data Analysis. This methodology counts the number of mentions of a word in the corpus and calculates the relative weighting of these mentions in positive and negative reviews. The 'negative words' are those with the highest relative weighting of mentions in negative reviews versus positive reviews. Conversely, the 'positive words' are those with the highest relative weighting of mentions in positive reviews versus negative reviews. Note that this approach sums up multiple references to a word within a particular review – thus twenty mentions of the film title in a single positive review of Shrek would give 'shrek' a high score across the entire database. The first columns in both Tables show the results achieved applying the proposed ranking methodology directly to the (entire) corpus. The lists are dominated by specific references to individual terrible or popular films or actors. The words 'seagal', 'bronson', 'dalmatians', 'silverstone' and 'avengers' are linked to negative reviews, while 'shrek', 'donkey', 'farquaad', 'gattaca' and

'niccol' are associated with positive reviews. This suspect methodology also exacerbates the general phenomenon that 'correlation does not equal causation'; many mentions of the word 'donkey' in a review of Shrek suggested to the algorithm that the word 'donkey' would elsewhere be an indicator of positive sentiment.

Limiting to a Single Count per Review. Column 2 shows the results if the methodology is altered so that multiple mentions of a given word in any single review counts only once for the overall count. Most individuals and films are eliminated while the 'negative' list now contains words that seem intuitively appropriate: 'incoherent', 'insulting', 'excrutiatingly', 'illogical', 'sucks', 'ludicrous' and so forth. The positive list remains more surprising, however; 'lovingly' and 'masterfully' sit alongside 'en'(??), 'soviet' and 'online'. A review of the frequency of occurrence points to the main problem here, however. The words in the positive list all occur fewer than 15 times in the entire corpus of 2,000 reviews – and have all only been seen in positive reviews. So sparsity of data is giving undue prominence to marginal words.

Dealing with Sparse Data. In column 3 we exclude all words that occur in fewer than 20 reviews (this is 1 % of the total), thereby eliminating all words for which there is insufficient data to estimate the correlation with sentiment accurately. Excluding very rare words also makes sense if we are restricted to a limited number of feature words for the sentiment analysis. (At this stage we also excluded all tokens of length less than three to remove punctuation marks or short words. This change had no impact on these lists). Most of the found words in this list seem to be sensible in the context of sentiment analysis. Only 'seagal' (making a reappearance) and 'freddie' stand out from the negative list – and it might be argued that the allegedly poor quality of Steven Seagal and Freddie Kruger films is such that

Table 2. Negative words.

raw score	1 per review	>20 reviews	in dictionary	Contribution
seagal	3000	insulting	insulting	bad
jawbreaker	hudson	sucks	ludicrous	worst
webb	incoherent	ludicrous	stupidity	plot
magoo	insulting	stupidity	idiotic	script
hudson	excrutiatingly	seagal	turkey	boring
jakob	fairness	idiotic	unintentional	nothing
lambert	illogical	turkey	freddie	stupid
stigmata	sans	unintentional	laughably	why
heckerling	sucks	uninvolving	inept	supposed
bats	ludicrous	freddie	forgot	least
sammy	feeble	laughably	sloppy	unfortunately
bronson	predator	inept	lame	looks
dalmatians	unimaginative	forgot	wasted	ridiculous
farley	wasting	sloppy	lousy	waste
silverstone	furniture	lame	awful	mess
3000	mediocrity	wasted	chuckle	reason
avengers	silverstone	lousy	miscast	have
schumacher	wisecracking	awful	poorly	should
incoherent	stupidity	chuckle	garbage	awful
modine	amateur	miscast	ridiculous	maybe

their mention in a review would be an accurate reflection of sentiment.

Eliminating Words not in a Dictionary. In column 4, a final change is made, aimed at eliminating any non-words or words not seen in a general directory. The entries are looked up in an English dictionary and any words that do not appear are eliminated. We used the Unix standard dictionary which removed a number of perfectly reasonable words ranging from 'sucks' (a recent word-usage but an old word) to 'captures' (dictionary only has stems). A more comprehensive dictionary might be expected to improve matters further.

Column 4 therefore stands as a 'final' proposed methodology for selecting lexical features by hand: rank all words (counted singly per review, minimum length of two letters) for correlation to sentiment, excluding any non-

words, and excluding any words for which there were 20 or fewer occurrences in the dataset. Take the best discriminating 40 candidates (20 each side). This seems to be a justifiable methodology and could be applied generally to a wide range of tasks and genres.

Approach Based on a Contribution 'Score'. Column 5 shows an entirely different approach. If we are going to use a limited list of features then we want those that on average will have the biggest impact on the decision. This suggests we should take into account both the discriminating ability and the frequency of occurrence :

Word score = (2 * (p-0.5)) * (n / T)

where   p = probability that word is found in a positive review

n = number of reviews in which word occurs

Table 3. Positive words. Note: columns 1 and 2 all at 100 % positive but <15 occurrences always.

raw score	1 per review	>20 reviews	in dictionary	Contribution
shrek	lovingly	outstanding	outstanding	life
ordell	en	finest	marvelous	also
gattaca	melancholy	marvelous	magnificent	both
argento	missteps	magnificent	wonderfully	great
guido	ideals	captures	chilling	best
leila	masterfully	wonderfully	anger	world
sweetback	gattaca	chilling	damon	many
lambeau	tobey	breathtaking	uplifting	performances
mallory	meryl	anger	debate	perfect
taran	ideology	damon	offbeat	performance
maximus	criticized	uplifting	gripping	most
fei	comforts	maintains	effortlessly	true
apostle	burbank	depicted	poignant	very
donkey	uncut	debate	beautifully	especially
sethe	sullivan	offbeat	religion	american
farquaad	soviet	gripping	vulnerable	different
camille	online	effortlessly	flawless	family
rounders	notoriety	poignant	everyday	well
niccol	niccol	beautifully	commanding	quite
lumumbra	methodical	religion	sincere	others

T = total number of reviews

This is the average score of the feature in any given review. (For ranking purposes we can dispense with the factor 2/T). Column 5 shows for this score the most extreme ranking words (counted singly per review) negatively and positively. The list contains much more common 'judgemental' words ('bad', 'worst', 'mess' 'boring', 'waste', 'great', 'best', 'perfect'). There are also some words hinting at stylistic differences between a positive and a negative review – 'why', 'supposed', 'should', 'have', 'maybe' for negative reviews, 'both', 'many', 'most', 'very', 'true' for positive reviews.

The following figure shows the accuracy of each of the feature sets in columns 1-5 of tables 2 and 3 when tested against the corpus. We recall that 50 % performance is at chance level. As expected, performance improves as we sort out some of the problems with our original methodology (columns 1-3). Performance peaks at ~72 %, roughly 40 % of the way between chance and perfect performance.

The "Contribution score" methodology (column 5) – which tries to include those words that will have the greatest impact – improves the score further to 75 %, halfway between chance and perfect performance, even though there is no overlap with the words in the >20 reviews feature set.

From an engineering perspective, it is usually worth combining two methods with apparently little overlap both of which have moderate performance. Combining the ">20 reviews" feature set with the "Contribution score" feature set here gives a slightly improved score of 77.0 %, at the cost of a feature set of twice the size.
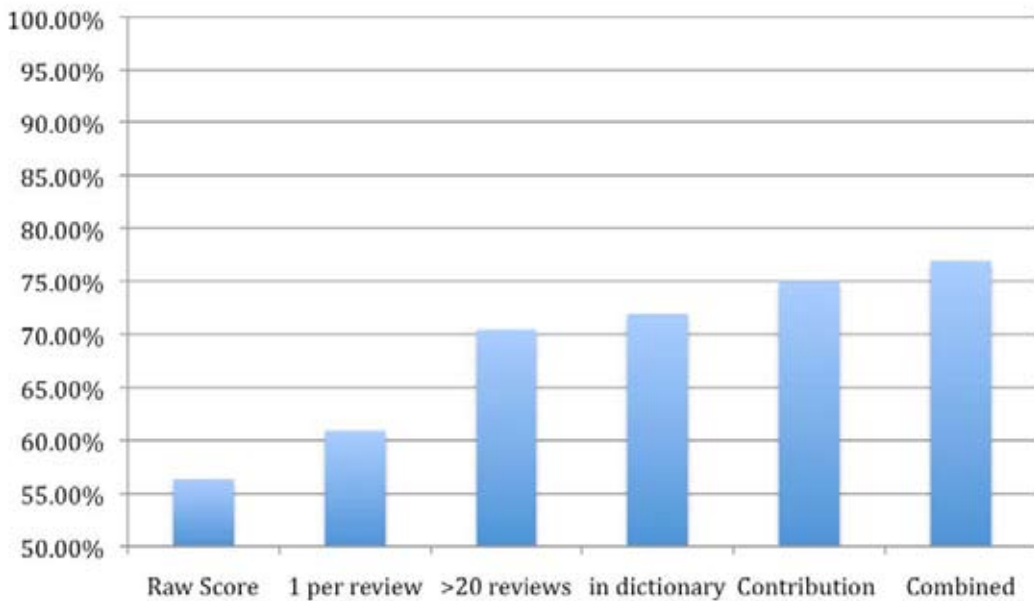
Fig. 1. Accuracy of sentiment analysis for hand-crafted categorizer using feature sets as per tables 2 and 3

Machine Learning Approach

Three different Naïve Bayes classifiers were trained, differing only in the form of the feature set that was offered to the trainer. Diverse behaviour was observed across these three classifiers, the implications of which are discussed.

For each classifier, we divided the corpus into 5 segments (each of 20 %) and conducted 5-fold cross validation to produce 5 classifiers, each of which was trained on 80 % of the data and tested on the remaining 20 %. The result was then averaged to provide an overall performance score.

Bayes I. In this version, each review's feature set contained the word count for every word mentioned in that review. Words absent in that review (but found elsewhere) were not included. Table 2 summarises the results of this analysis. The calculated accuracy was 73.4 % +/- 2.3 % (2 standard errors) a comparable level to the 'best' hand-crafted classifier.

Bayes II. Each review's feature set indicated solely the presence of a particular word (not its frequency). Words absent in that review (but found elsewhere) were not included. In an influential early study, (Pang, Lee, & Vaithyanathan, 2002) had found that testing for presence as opposed to frequency had improved performance of a Naïve Bayes classifier in a very similar movie review sentiment analysis task. However, in this instance, a 5-fold cross validation analysis found no significant difference in performance between the previous Bayes classifier and this version (Bayes II), whose accuracy was measured at 72.1 % +/- 2.8 % (2 standard errors)).

Bayes III. (Bird, Klein, & Loper, 2009) utline a somewhat different approach to creating a feature set for the movie review task. They propose identifying the 2,000 most frequent words within the entire corpus; each review's feature set then contains an indication of whether each of these 2,000 words is present or absent

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average	Std Err
Accuracy	75.0 %	76.3 %	71.0 %	70.3 %	74.3 %	73.4 %	1.16 %

in that review. Data for these 2,000 keys is thus complete across all reviews. So there is a clear contrast to Bayes I/II – in which the presence of every word was retained (about 36,000 distinct words) but absent words in any single review were not highlighted as being absent.

For our data, this "Bayes III" design produced performance comparable with that found by (Pang, Lee, & Vaithyanathan, 2002) on a very similar task with an apparently comparable movie-review dataset (2,053 reviews). (Pang, Lee, & Vaithyanathan, 2002) recorded a performance of 81.0 % accuracy with a Naïve Bayes classifier (using word presence/absence as the feature) and a feature set of 16,165 features. Using the 2,000 most frequent features, the Bayes III approach here recorded an accuracy of 80.7 % +/- 1.8 % (2 standard errors). It is possible that the performance could have been tweaked higher with a larger number of features.

So the results of this analysis seemed clear – Bayes I, II, III : 73.4 %, 72.1 %, 80.7 %; the Naïve Bayes classifier is more effective if provided on training (and testing) with a smaller but consistent set of the most frequent features in each review. However, the performance in the book review, analysed in the next section, provides a significant twist to this analysis.

Cross-Domain Comparisons

We have created six hand-crafted word-based classifiers and three forms of Naïve Bayes classifier based on differing feature sets (Bayes I, II and III). Figure 2 compares movie review (blue bars, left) and book review (red bars, right) performance for all of these approaches.

Original Methodology. The first four sets of results summarise the original hand-crafting methodology with various improvements in place. Performance against the book review corpus is very poor, rising from chance for the original raw score to only 10 % of the way between chance

and perfect performance for the 'best' version ('in dictionary').

Contribution Score and Combined. Although apparently only marginally better on the movie review task, the 'Contribution Score' approach clearly out-performs the original methodology on the book reviews, improving to a performance 30 % between chance and perfect performance (64.7 % accuracy). This is surprisingly good given that classifiers of this kind generally show poor discrimination performance across different genres. The result suggests that perhaps movie reviews and book reviews are sufficiently close to one another to bear comparison using statistical approaches of this kind.

The Contribution score list is biased towards higher frequency and simpler judgemental words than the original methodology based on summing up word counts. This appears to be closer to the vocabulary employed in these particular book reviews, which are shorter and written by amateur authors as opposed to the longer, more nuanced and 'professional' movie reviews. A combined approach (original methodology plus Contribution score) marginally improves performance.

Bayes approach. On movie reviews, the performance of Bayes I and II was roughly comparable to or slightly below that of the best hand-crafted classifiers. Typically one would expect a Naïve Bayes classification methodology to out-perform hand-classifiers. However, for the book reviews, Bayes I and II appear to out-perform the best 'simple' approach, Bayes I and Bayes II achieving accuracy scores of 67.3 % +/- 0.8 % and 72.3 % +/- 0.9 % (two standard errors), respectively.

We note that Bayes II significantly out-performed Bayes I on this "off-topic" corpus of book reviews. This may be attributed to fact that building a classifier based only on the presence or absence of features we are, in effect, 'normalising'
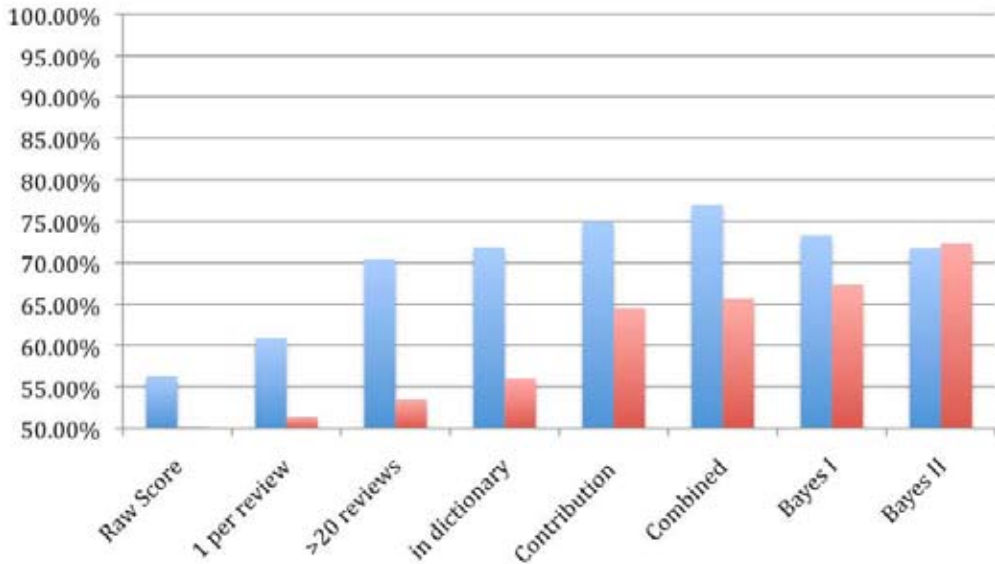
Fig. 2. Accuracy of sentiment analysis (movie reviews – left bars; book reviews – right bars).

for differences in length. As noted earlier, the book reviews are much shorter than the movie reviews and therefore have much less opportunity to repeat salient features on multiple occasions. Bayes I is sensitive to the number of features in a review, Bayes II is not. Bayes I will therefore be at a relative disadvantage when presented with the shorter book reviews – it is in effect expecting 'more' positive features per review to classify something as a positive review.

The most unexpected results we receive from evaluation of Bayes III on the 'off-topic' book reviews: Bayes III fails to outperform a chance baseline of 50 %. Thus the improved performance in movie reviews achieved by altering the structure of feature sets has been at the cost of a collapse in the performance "off topic". So it seems that the Bayes III feature set locked the Bayes III classifier on to key features of the data set that were highly discriminatory in the case of movies and useless in the case of book reviews. Bird et al. themselves point out that two of the most "informative" features in the Bayes III classification are genre specific – "seagal" and "damon".

Conclusion

Experiments described above show that simple lexical features based on counts do not perform well either for in-domain classification or for cross-domain classification. Obviously, some domain-specific words (e.g. "shreck") are useless for the book-review classification task. A less expected result was shown by the word-list containing only items that were found in a dictionary in the cross-domain experiment. This list should not contain too many domain-specific items but still it performs poorly on the book-reviews. This means that even in-dictionary words are often domain-specific. The Contribution lexicon is relatively domain-independent because the approach calculates sentiment-related score for words. This score reflects a word's contribution to overall sentiment thus these words are more sentiment-related than domain-related (bearing in mind that sentiments are more universal than topics it explains why such words perform better on other domains).

Among the supervised classifiers Bayes II is better as it is independent of the size of

the corpora. As for the Bayes III, the top 2000 words tend to be very domain-dependent and not necessarily sentiment-related and the total number of words being too restrictive prevents a supervised classifier from getting good results.

Finally we can conclude is that 1) even for close domains the problem of domain-dependency remains considerable; 2) in cross-domain sentiment analysis the best results can be achieved with features that a) are less domain-specific (e.g. do not have anything to do with in-domain word frequency as in Bayes I and Bayes III) and b) are more sentiment-related (the Contribution scores rather than frequencies).

References

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O'Reilly Media.

Engström, C. (2004). Topic Dependence in Sentiment Classification.

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on. In *the 42nd Annual Meeting on Association of Computational Linguistics*. Barcelona, Spain..

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In *Conference on Empirical Methods in Natural Language Processing* (pp. 79-86).

Read, J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Student Research Workshop* (p. 43). Ann Arbor, Michigan.

Read, J., & Carroll, J. (2009). Weakly supervised techniques for domain-independent sentiment classification. In *Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion* (p. 45–52). ACM.

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Annual Meeting of Assosiation of Computational Linguistics* (pp. 417-424). Philadelphia, Pennsylvania.

Топиковая зависимость в тематически близких текстах в контексте автоматического анализа оценочной составляющей

Д. Реффин[a],
Т.Е. Загибалов[a], Е.О. Беляцкая[б]
[a] Университет Сассекса,
Великобритания,
[б] Сибирский федеральный университет,
Россия 660041, Красноярск, пр. Свободный, 79

Анализ оценочной составляющей направлен не на анализ тематического или содержательного контента, а на анализ содержащихся в тексте оценок и субъективных высказываний. В настоящей публикации мы представляем результаты экспериментов по автоматическому анализу оценочной составляющей при помощи лексикона на материале двух корпусов жанрово и тематически близких текстов: ревью фильмов и ревью книг. Мы обнаружили, что даже для тематически близких текстов эффективная классификация оценки затруднительна без использования информации из обрабатываемого корпуса. Мы также выявили определённые характеристики лексикона, которые оказывают влияние на классификацию оценки в тексте.

Ключевые слова: анализ оценочной составляющей.