Lexicon Based Sentiment Analysis of Urdu Text Using SentiUnits

Conference Paper · November 2010

DOI: 10.1007/978-3-642-16761-4_4 · Source: DBLP

3 authors:

Afraz Zahra Syed
Vietnam National University, Hanoi

Aslam Muhammad
University of Engineering and Technology, Lahore

Ana Maria Martinez-Enriquez
Center for Research and Advanced Studies of the National Polytechnic Institute

Lexicon Based Sentiment Analysis of Urdu Text Using SentiUnits

Afraz Z. Syed[1], Muhammad Aslam[2], and Ana Maria Martinez-Enriquez[3]

[1,2] Department of CS & E, U.E.T., Lahore, Pakistan
afrazsyed@uet.edu.pk, maslam@uet.edu.pk
[3] Department of CS, CINVESTAV-IPN, D.F. Mexico
ammartin@cinvestav.mx

Abstract. Like other languages, Urdu websites are becoming more popular, because the people prefer to share opinions and express sentiments in their own language. Sentiment analyzers developed for other well-studied languages, like English, are not workable for Urdu, due to their scriptic, morphological, and grammatical differences. As a result, this language should be studied as an independent problem domain. Our approach towards sentiment analysis is based on the identification and extraction of SentiUnits from the given text, using shallow parsing. SentiUnits are the expressions, which contain the sentiment information in a sentence. We use sentiment-annotated lexicon based approach. Unluckily, for Urdu language no such lexicon exists. So, a major part of this research consists in developing such a lexicon. Hence, this paper is presented as a base line for this colossal and complex task. Our goal is to highlight the linguistic (grammar and morphology) as well as technical aspects of this multidimensional research problem. The performance of the system is evaluated on multiple texts and the achieved results are quite satisfactory.

Keywords: Natural language processing, computational linguistics, sentiment analysis, opinion mining, shallow parsing, Urdu text processing.

1 Introduction

Natural Language Processing (NLP) or Computational Linguistics (CL) is a challenging field in Artificial Intelligence. The philosophical, psychological and conceptual nature of natural language entails complexity to process it. NLP applications can be seen as bi-dimensional problems. For some applications, computational aspect is more important. For instance, spell checkers, machine translators, grammar checkers, and human computer interaction based applications. These solutions conceal original issues in modeling aspects of the language processing and therefore, these are of no real conceptual interest. But on the other hand, there are some applications for which linguistics is a major concern. In this case, the goal is not only to simulate human language processing, but also to understand and manipulate the conceptual and psychological knowledge. Poetry generation, story generation, intelligent information retrieval, and sentiment analysis lay in this category.

G. Sidorov et al. (Eds.): MICAI 2010, Part I, LNAI 6437, pp. 32–43, 2010.

For both types of computational linguistic applications mentioned above, English is a very well studied language. Among other areas of natural language processing, an emerging field is sentiment analysis (SA). Since the last decade, this area is the center of focus for NLP researchers. The need for sentiment analysis is the outcome of sudden increase in the opinionated or sentimental text, which is in the form of blogs, reviews, and discussions [1]. The main feature of such text is its variability in structure, style, and vocabulary. Researchers have made great attempts to cope with this variability issue. Some of the contributions are quite successful like [2] [3] [4]. In theses contributions SA is presented as a classification problem.

A text can be categorized as objective or subjective. The objective text is fact based neutral text, but opinions, reviews and discussions are all in the category of subjective text, which exhibit some feeling or sentiment. The main goal of a sentiment analyzer is to classify the subjectivity orientation towards positive or negative. That is why sentiment analysis is sometimes referred to as subjectivity analysis.

The approaches mentioned above are mostly addressed for English text, but these are not able to handle Arabic and Arabic-orthography-based language like Urdu, Persian, and some other local dialects of south Asia. These languages have altogether different script, morphology, and grammatical rules [5]. With increasing popularity of Internet, like other languages, Urdu websites are also becoming more popular as people prefer to share opinions and express sentiments in their own language.

Keeping this fact in mind, for our research, in this paper, we propose a sentiment analyzer for Urdu language, which is spoken and understood in major part of Asia. As already mentioned Urdu is a quite different language from English in both script and morphology. Despite, script similarities with Arabic (spoken in Saudi Arabia, United Arab Emirates, and many other Arab and African countries) and Persian (spoken in Iran, Afghanistan, and many states of former Soviet Union), and morphological similarities with Hindi (spoken in India), Urdu has its own requirements, as far as CL is concerned. Literature survey shows that [5] and [6] approaches used for other languages are not applicable for the proper handling of Urdu text. This fact entails the need of an updated or even altogether new sentiment analysis model.

The rest of the paper proceeds as follows: Section 2 depicts SA in general with some related works. Section 3 gives a brief overview of Urdu language; our focus in this section is on highlighting the major concerns (linguistic as well as technical) in sentiment analysis of Urdu text. We also present a comparison of Urdu with English, Arabic and Persian. Section 4 describes the SentiUnits (words and phrases carrying sentiment information of a sentence) and their attributes in details. Section 5 describes our methodology through a system model and its validation process. Finally, in Section 6 we conclude our effort and discuss some future directions.

2 Sentiment Analysis

John Locke (1632-1704) said, "Man is by nature a social animal". So, man always seeks for suggestions, opinions, and views from other people in society for his survival and proper decisions making. In this modern era of computer and technology we are living in virtual communities and societies. Now, Internet forums, blogs, consumer reports, product reviews, and other type of discussion groups have opened new horizons for human mind. That is why, from casting a vote to buying a latest gadget we search for opinions and reviews from other people on the Internet. This explosion of opinionated text has fashioned an exciting area in text analysis, which is referred by many names like sentiment analysis, opinion mining, subjectivity analysis, and appraisal extraction [1]. For this paper we use the term Sentiment analysis.

2.1 Sentiment Analysis Related Work

Sentiment analysis is considered as a classification problem [7]. The opinionated text is classified as positive or negative e.g., thumbs up or thumbs down. Some classifiers use a multi-point-range e.g., five star scale for movie reviews, etc. The classification usually starts from term or phrase level and moves towards sentence and then document level. Usually, the output of one level becomes input to the next [1]. There are, a number of features on the basis of which, the text is classified. For example, frequency and position of a term or information of a part of speech like adjectives and adverbs [3] [4] [8]. However, an important aspect is to define semantic orientation of words and phrases. In this regard, the following approaches are used [9]:

a) Lexicon based approaches. These approaches are based on lexicons of words/phrases or expressions, which are pre-tagged with sentiment, subjectivity or polarity information about each entry. This lexicon can be manually created as well as automatically generated, using machine learning methods for example [7] and [10]. Some researchers have used WordNet for sentiment mining like [11].

Each word in the text (to be analyzed) is compared with the lexicon entry. As a result a positive or negative orientation score is attached to it. The total score of a sentence is then computed as a sum of word scores. Consider a sentence, "This is a high quality camera" in which, the phrase "high quality" is the only semantic unit containing sentiment information (positive). All other words are neutral and hence the sentence is classified as a positive comment.

b) Machine learning approaches. The main focus of these approaches is on feature vectors, which are selected according to domain. A classifier is trained by a tagged corpus. Feature selection is a crucial issue, which can highly affect the results [9].

Domain-Specific Contributions. It is observed that sentimental analysis of the text is highly domain specific in nature [1]. Thus, development of a generalized solution for analyzing all domains is still an open research direction. In literature this is referred to as Domain Adaptation [12] [13]. Consequently, most of the contributions try to spot a particular domain. For example, analysis of reviews related to products and movies. Such texts are relatively easier to handle due to specified targets and their attributes [9]. On the contrary, political speeches and discussions are perhaps the most complex to handle. In [14] is pinpointed an issue and evaluated whether the speech was in favor or opposition. Other challenging domains of research are News texts in which organizations, personalities, and events etc are center of focus [2].

Handling Negations. The words like, "no, not, do not, don't, can't" are called negations. These words can altogether alter the sense of a sentence, so are very important to tackle. Different approaches are used to handle negations, e.g., [15] process negations as a part of post processing. They associate negating words with the subjective components of the sentence using co-location. This technique works in sentences like "I don't like", "This is not good" but is not effective in "No doubt it is amazing" [1]. In [16] is considered negations as part of appraisal expressions annotated with attitudes. In this way, negation becomes independent of location. Another approach is the use of POS tagged corpus [17].

3 Urdu Language

About 60.5 million speakers mainly in Indian subcontinent [19], Urdu is a widely spoken Indo-Aryan language. A large number of speakers exist in Pakistan, India, Afghanistan, Iran, and Bangladesh. Moreover, Urdu is the Pakistan official language and scheduled of India. Urdu orthography resembles Arabic, Persian, and Turkish. Cursive Arabic script and Nastalique writing style is used [20].

3.1 Urdu: In Comparison with Other Languages

Urdu vs. English. For English language sentiment analysis is well explored. In [1] and [9] very extensive and comprehensive surveys are presented. Although the core approaches used to handle English text (lexicon based and machine learning, described in detail in related work) can be used for Urdu text but modifications and adaptations are compulsory due to vast orthographic, morphological, and grammatical differences between both languages as described in preceding section.

Urdu vs. Arabic. A major language, which is comparable with Urdu, is Arabic. In computational linguistic realm Arabic is much mature than Urdu, and a number of approaches are proposed for Arabic text processing. Orthographically and morphologically both languages are very similar, but Urdu grammar is more inclined towards Sanskrit and Persian [20]. So, for appropriate processing of Urdu text we need to revise these approaches or require developing entirely new solutions.

Urdu vs. Hindi. Hindi is a major dialect of Urdu and there are minimal differences in the grammar of both languages. But their orthography and vocabulary are dissimilar. Urdu uses right to left Nastalique calligraphy style of Persio-Arabic script and draws vocabulary from Arabic and Persian. Whereas, Hindi uses left to right Devanagari (देवनागर) script and draws vocabulary from Sanskrit.

3.2 Major Concerns in Urdu Language Processing

As already mentioned Urdu NLP is not a very established area. There are a number of hurdles with which we have to cope before applying sentiment analysis. Here, we highlight some major concerns in Urdu processing for accomplishing this complex task. We present both linguistic and technical difficulties with literature review.

3.2.1 Technical Aspects

a) *Corpus.* Urdu websites are becoming popular day by day, even though these websites cannot be used for corpus construction, because such a task needs large amount of electronic text. This is an unfortunate fact that most of the Urdu websites use graphic formats i.e. gif or other image formats, to display Urdu text [21].
b) *Lexicon.* For lexicon based sentiment analysis we need a sentiment-annotated lexicon of Urdu words. Unluckily there is no such lexicon available or even developed to date. So, from conception to modeling and then implementation we have to cope with this challenging task.
c) *Word Segmentation.* Urdu orthography is context sensitive. The "حروف" (haroof, alphabets)[1] have multiple glyphs and shapes and are categorized as joiners and non-joiners. Moreover, word boundaries are not indicated by space. A single word can have a space in it, e.g., "خوب صورت" (khoob surat, beautiful). On the contrary, two different words can be written without space, e.g., "دستگیر" (dastgeer, benefactor).

This segmentation issue is divided into two sub problems [19]: a) Space-insertion, b) Space-deletion. This work emphasizes on the orthographic word "OW" instead of "word", as an example consider again the word "خوب صورت" (khoob surat, beautiful), orthographically it is a single word based on two lexical words.

3.2.2 Linguistic Aspects

The following aspects are given, particularly, in comparison with English language:

a) *Variability in Morphology.* In English there are mostly hard and fast inflectional and derivational rules applied on morphemes. For example, "s" or "es" suffixes are mostly used to make plurals, like in "chair + s" and "dish+es". Exceptions are there but are very rare and can easily be handled. On the other hand, Urdu morphology is very complex. Inflection, derivation, compounding, and duplication are very common phenomenon. The plural can be indicated by a number of ways. For example, in the sentence: "بہت سارے پھول-" (bohat sare phool, A lot of flowers.) no plural suffix is used. And in, "پھولوں کے رنگ-" (pholoon kay rang, Colors of Flowers.) plural suffix "وں" (on) is used without any replacement. But in the sentence: "پودے سبزہیں-" (poday sabz hain, Plants are green.) plural suffix is "ے" (ay) and is replacing "ا"(aa).
b) *Flexibility in vocabulary.* Urdu has abridged several languages. It has words from languages like Arabic, Persian, Hindi, English Turkish and even more. The absorption power of Urdu is very exceptional and it enhances the beauty of the language. But, unfortunately this makes our work more challenging. Code switching, using multiple languages concurrently is also very common in Urdu writings,. For example, "کر دو mobile off" (Mobile off kar do, Turn off the mobile) means, "switch off the mobile".
c) *Case markers.* [21] Identifies eleven categories of POS tags in Urdu language: noun, verb, adjective, adverb, numeral, postposition, conjunction, pronoun, auxiliaries, case markers, and "حرف" (harf). Among all of them, the case markers are very dissimilar in nature for Urdu in comparison with other languages because they are

[1] We first write the Urdu word/sentence in Persio-Arabic script based Nastalique writing style. Pronunciation is enclosed in parenthesis, followed by English translation.

written with space. Therefore, they are considered as a distinct POS tagged word. This distinction adds to the ambiguity of the words with which they are semantically associated. There are four case markers ergative "نے" (ne), instrumentive "سے" (se), genitive "کا" (ka) and dative/accusative "کو" (ko).

4 SentiUnits

In an opinionated sentence, all terms are not subjective. Indeed, the sentimentality of a sentence depends, only on some specific words or phrases. Consider, the examples "Fatima is an adorable child." and "Irtaza is such a nice boy." underlined words are the expressions made of one or more words, which carry the sentiment information of the whole sentence. We recognize them as SentiUnits. We can judge, only these units, as the representatives of the whole sentence's sentiment. These are identified by shallow parsing based chunking. We consider two types of SentiUnits:

- *Single Adjective Phrases* are made of adjective head and possible modifiers, e.g. "بہت خوش" (bohat khush, very happy), "زیادہ بہادر" (zyada bhadur, more brave)
- *Multiple Adjective Phrases* comprise of more than one adjective with a delimiter or a conjunction in between, e.g. "بہت چالاک اور طاقتور" (bohat chalak aur taqatwar, very clever and strong).

4.1 Attributes of SentiUnits

A SentiUnit can be described by following attributes:

A) *Adjectives*(as head words). Conceptually, adjectives in Urdu can be divided into two types. First type describes quantity and quality, e.g. "کم" (kam, less), "بدترین" (budtareen, worst), "زیادہ" (ziyada, more). And the second distinguishes one person from other, e.g. "حسین" (haseen, pretty), "فطین" (fateen, intelligent).

Further, adjectives are categorized as marked, which, can be inflected for number and gender and unmarked which are usually Persian loan words. Also the adjectives inflected from nouns remain unmarked [22]. For examples, see Table 1.

Table 1. Types of Urdu adjectives as Marked and unmarked

Marked			Unmarked		
Male	**Female**	**Number**	**Persian loan**	**Inflected from Noun**	
اچھا کام	اچھا قلم	اچھے آم	تازہ	دفتری	دفتر
acha kaam	acha qalam	achay aam	tazah	daftary	daftar
good work	good pen	good mangoes	fresh	official	office

Attributive adjectives precede the noun they qualify. Arabic and Persian loan adjectives are used predicatively and appear in the form of phrases. See Table 2.

Table 2. Types of Urdu adjectives as attributive and predicative

Attributive (precede the noun they qualify)		**Predicative**
Adjective	**Noun**	**Persian and Arabic based**
مزیدار	مزہ	معلوم ہونا
(mazedar, tasty)	(maza, taste)	(maloom hona, to be known)

The postpositions "سے" (say), "سی" (si), "سا" (sa) and "والا" (wala), "والی" (wali), "والے" (walay) are very frequently used with noun to make adjectives. Examples are listed in Table 3. Whereas Table 4 shows the derivation of adjectives from nouns:

Table 3. Use of postpositions with adjectives

Noun	**With postposition *si, sa, say***	**With postposition *wala, wali, walay***
پھول	پھول سی	اوپر والی
(phool, flower)	(phool si, like flower)	(oopar wali, the upper one)
چاند	چاند سا	اچھے والے
(chand, moon)	(chand sa, like moon)	(achay walay, the good ones)

Table 4. Use of postpositions with adjectives

Noun	**Adjectives**
برف (barf, ice)	برفیلا (barf-eela, icy)
درد (dard, pain)	دردناک (dard-nak, painful)
بھوک (bhook, hunger)	بھوکا (bhooka, hungry)

B) *Modifiers.* These are classified as absolute, comparative and superlative:

a) Absolute. Simple adjectives without modifiers make absolute expressions, e.g. "یہ لباس مہنگا ہے۔" (Yeh libaas mehnga hay, This dress is expensive.)
b) Comparative. Two comparative modifiers are there: "سے" (say) or "سے ذیادہ" (say zyadha), e.g. "یہ لباس اس سے مہنگا ہے۔" (Yeh libaas us say mehnga hay, This dress is more expensive than that). Or "یہ لباس اس سے زیادہ مہنگا ہے۔" (Yeh libaas us say zyadah mehnga hay, This dress is more expensive than that.)
c) Superlative. For superlatives "سب سے" (sab say) and "سب میں" (sab main) are used, e.g."یہ لباس سب سے مہنگا ہے۔" (Yeh libaas sab say mehnga hay, This dress is the most expensive), and "یہ لباس سب میں مہنگا ہے۔" (Yeh libaas sab main mehnga hay, This dress is the most expensive).

C) *Orientation.* Orientation describes the positivity or negativity of an expression, e.g. "اچھا" (acha, good) have positive orientation.
D) *Intensity.* This is the intensity of orientation, e.g. "بہتر" (behtar, better).
E) *Polarity.* A polarity mark is attached to each lexicon entry to show its orientation.
F) *Negations.* Negations are the polarity shifters, e.g. "ارتضی اچھا ہے۔" (Irtaza acha hay, Irtaza is nice.) is a positive sentence but with the use of negation its polarity shifts to negative, i.e. "ارتضی اچھا نہیں ہے۔" (Irtaza acha naheen hay, Irtaza is not nice.)

5 Methodology

The research work is divided into two main tasks:

- To create sentiment-annotated lexicon for the inclusion of information about the subjectivity of a word/phrase in addition to its orthographic, phonological, syntactic, and morphological aspects.
- To build an appropriate classification model for the processing and classification text in accordance with the inherent sentiments.

From the approaches discussed in Section 2, we use a lexicon-based approach. Fig. 1 shows the context model, in which the classification system represents the process. Whereas, sentiment annotated lexicon holds the sentiment orientation of each entry.

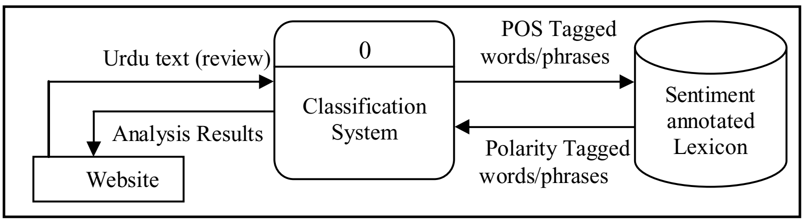

Fig. 1. Context model of the system

5.1 Sentiment-Annotated Lexicon

The fundamental part of our research is the construction of sentiment-annotated lexicon. A SentiUnit is classified on the basis of orientation and intensity. Orientation is predicted by marked polarity and intensity is calculated by analyzing the modifiers. For example the intensity of "بہت ذیادہ" (bohat zyadah, much more) is more than "ذیادہ" (zyadah, more). The lexical construction tasks are divided as follows:

- Identify the sentiment-oriented words/phrases in Urdu language.
- Identify morphological rules, e.g. inflection or derivation.
- Identify grammatical rules, e.g. use of modifiers.
- Identify semantics between different entries, e.g. synonyms, antonyms, and cross-references.
- Identify and annotate polarities to the entries.
- Identify modifiers and annotate intensities.
- Differentiate between multiple POS tags for same entries.
- Construct lexicon

5.2 Sentiment Classification

A high level model of the classification system is presented in the Fig. 2.

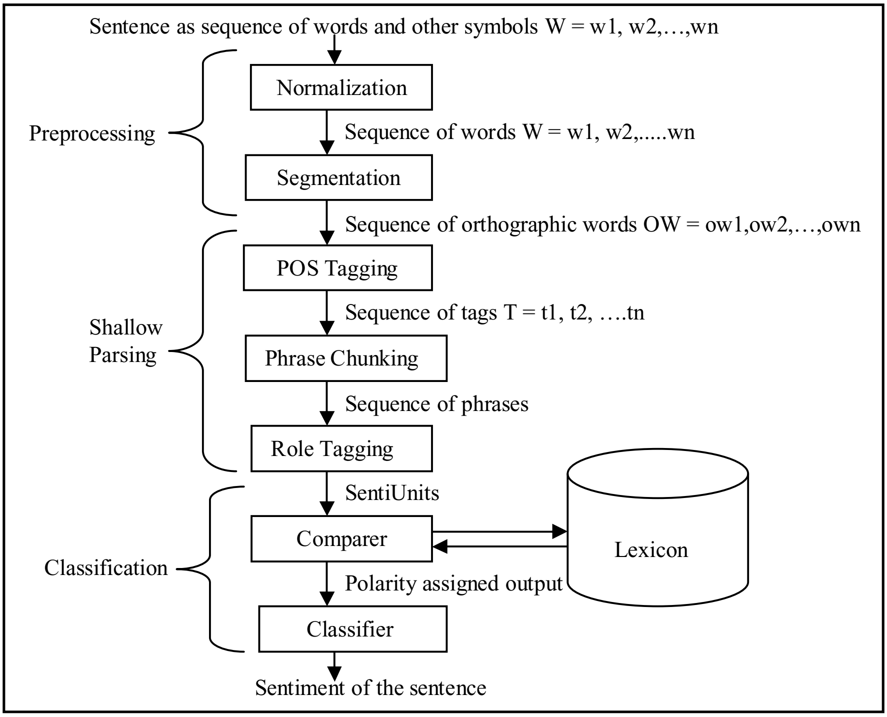

Fig. 2. Classification process of the system

The process of sentiment analysis is composed of three phases (see Fig.2):

Preprocessing. This phase prepares text for sentiment analysis. It is usually, based on removal of punctuations and striping of HTML tags [9]. But, due to Urdu's orthographic characteristics, (i.e., optional use of diacritics and ambiguity in word boundaries) [19], we add two more tasks:

Diacritic omission. In Urdu diacritics are optional and their use is highly author dependent. So as a regular practice these are removed during text normalization [19].
Word segmentation. As mentioned in section 3.2.1, Urdu orthography is context sensitive and word boundaries are not always identified by space like in English. So the outputs of the preprocessing phase are the orthographic words which can/cannot have space within.

SentiUnit extraction. After preprocessing, a shallow parsing is applied to identify SentiUnits. At the same time, negation is considered because; negation can altogether change the polarity of the sentence.

Classification. The extracted SentiUnits are compared with lexicon and their polarities are calculated for classification as positive or negative. After calculating the sentence polarity "s" a total post polarity "p" is calculated by adding all sentence polarities, i.e. p = s1 + s2 + s3 …..sn.

Fig. 3 shows execution of a single sentence. “گاڑی کا یہ ماڈل خوب صورت نہیں-” (Gari ka yeh model khoobsurat naheen hay, This model of the car is not beautiful).

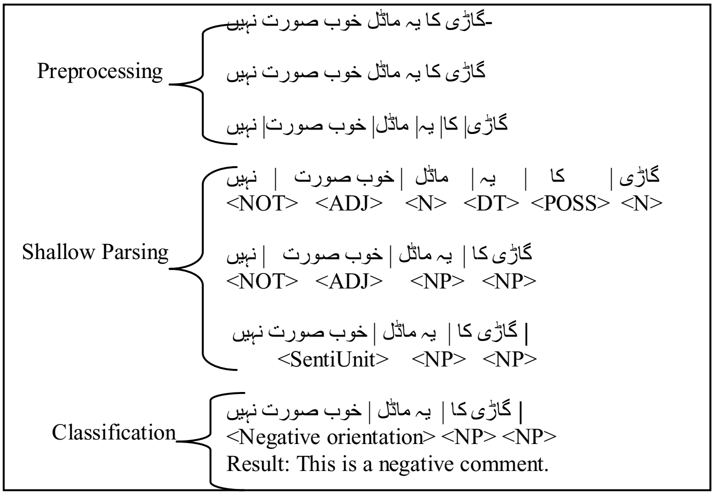

Fig. 3. Example execution of the analyzer

5.3 Validation

Due to deficiency of publicly accessible corpus of Urdu text for sentiment analyze, a sentiment corpus was collected and constructed from domains like movies and products (electronic appliances from main brands in Pakistan) reviews. We processed 753 reviews from which 435 are movies and 318 product reviews. Positive (361) and negative (392) documents are included in both categories. The experimented results performed are shown in Table 5. The identification and extraction of SentiUnit is quite practical. Furthermore, sentiment annotated lexicon can extend the same model.

Despite morphological complexity of adjectives and propositions used, we are hopeful to proceed on the same line, by including other speech parts.

Further, we have considered certain observations during the evaluation:

- The classification accuracy for SentiUnits with unmarked adjectives is about 75%, and for marked adjectives is 71%.
- The SentiUnits with adjectives made by postpositions combined with nouns, cause errors and hence, an improved algorithm is required.
- On the other hand, adjectives made by inflected nouns, entail the best results, with an accuracy of 80-85%.
- The most frequent modifiers are “ذیادہ” (zyadah, more) and “کم” (kam, less).
- Negations are less problematic in Urdu sentiment analysis, because these are present only in specific patterns.

Table 5. Experiment results from sentiment corpora

Domains	Total Number	Orientation	Number	Over all Accuracy (%)
Movies C1	435	Positive	215	72%
		Negative	220	
Product C2	318	Positive	146	78%
		Negative	172	

6 Conclusions and Future Work

Despite, the developments in sentiment analysis of English text, it is a fact that for Urdu language this domain is still an open challenge. In this paper, we effectively identify the major concerns and explore the possible solutions for Urdu language. We also present a comprehensive overview of adjectives and their modifiers, with respect to the task of SentiUnits extraction. In consequence, our approach can serve as a baseline for this issue.

Among a number of possible future works, the most important is the extension of the lexicon. Adding more adjectives and modifiers as well as other parts of speech can trigger this task. Moreover, for achieving better results the SentiUnits should be annotated with their respective targets.

References

1. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundation and Trends in Information Retrieval 2(1-2), 1–135 (2008)
2. Bautin, M., Vijayarenu, L., Skiena, S.: International sentiment analysis for news and blogs. In: International Conference on Weblogs and Social Media, ICWSM (2008)
3. Hatzivassiloglou, V., Wiebe, J.: Effects of Adjective Orientation and Gradability on Sentence Subjectivity. In: 18th International Conference on Computational Linguistics, New Brunswick, NJ (2000).
4. Turney, P.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: ACL, Ph, PA, pp. 417–424 (July 2002)
5. Riaz, K.: Challenges in Urdu Stemming. Future Directions in Information Access, Glasgow (August 2007)
6. Akram, Q., Naseer, A., Hussain, S.: Assas-band, an Affix-Exception-List Based Urdu Stemmer. In: 7th Workshop on Asian Language Resources, IJCNLP 2009, Singapore (2009)
7. Melville, P., Gryc, W., Lawrence, R.D.: Sentiment analysis of blogs by combining lexical knowledge with text classification. In: Conference on Knowledge Discovery and Data Mining (2009)
8. Bloom, K., Argamon, S.: Unsupervised Extraction of Appraisal Expressions. In: Farzindar, A., Kešelj, V. (eds.) Canadian AI 2010. LNCS (LNAI), vol. 6085, pp. 290–294. Springer, Heidelberg (2010)
9. Annet, M., Kondark, G.: A comparison of sentiment analysis techniques: Polarizing movie blogs. In: Bergler, S. (ed.) Canadian AI 2008. LNCS (LNAI), vol. 5032, pp. 25–35. Springer, Heidelberg (2008)

10. Bloom, K., Argamon, S.: Automated learning of appraisal extraction patterns. In: Gries, S.T., Wulff, S., Davies, M. (eds.) Corpus Linguistic Applications: Current Studies, New Directions. Rodopi, Amsterdam (2009)
11. Andreevskaia, A., Bergler, S.: Mining WordNet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses. In: EACL 2006, Trent, Italy (2006)
12. Mansour, Y., Mohri, M., Rostamizadeh, A.: Multiple source adaptation and the Renyi divergence. In: Uncertainty in Artificial Intelligence, UAI (2009)
13. Tan, S., Cheng, Z., Wang, Y., Xu, H.: Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis. In: Advances in Information Retrieval, vol. 5478, pp. 337–349 (2009)
14. Bansal, M., Cardi, C., Lee, L.: The power of negative thinking: Exploring label disagreement in the min cut classification framework. In: International Conference in Computational Linguistics, COLING (2008)
15. Hu, M., Lui, B.: Mining and summarizing customer reviews. In: Conference on Human Language Technology and Empirical Methods in Natural Language Processing (2005)
16. Whitelaw, C., Garg, N., Argamon, S.: Using appraisal taxonomies for sentiment analysis. In: SIGIR (2005)
17. Na, J.-C., Sui, H., Khoo, C., Chan, S., Zhou, Y.: Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews. In: Conference of the International Society of Knowledge Organization (ISKO), pp. 49–54 (2004)
18. Muaz, A., Khan, A.: The morphosyntactic behavior of 'Wala' in Urdu Language. In: 28th Annual Meeting of the South Asian Language Analysis Roundtable, SALA 2009, University of North Texas, US (2009)
19. Durrani, N., Hussain, S.: Urdu Word Segmentation. In: 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL HLT 2010), Los Angeles, US (2010)
20. Riaz, K.: Stop Word Identification in Urdu. In: Conference of Language and Technology, Bara Gali, Pakistan (August 2007)
21. Ijaz, M., Hussain, S.: Corpus based Urdu Lexicon Development. In: Conference on Language Technology (CLT 2007), University of Peshawar, Pakistan (2007)
22. Schmidt, R.: Urdu: An Essential Grammar. Routlege Publishing, New York (2000)