Expert Systems With Applications 168 (2021) 114231

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA

Baris Ozyurt [a,*], M. Ali Akcayol [b]

[a] Department of Computer Science, Graduate School of Informatics, Gazi University, Ankara, Turkey
[b] Department of Computer Engineering, Faculty of Engineering, Gazi University, Ankara, Turkey

ARTICLE INFO

Keywords:
Sentiment analysis
Opinion mining
Aspect based sentiment analysis
Aspect extraction

ABSTRACT

With the widespread use of social networks, blogs, forums and e-commerce web sites, the volume of user generated textual data is growing exponentially. User opinions in product reviews or in other textual data are crucial for manufacturers, retailers and providers of the products and services. Therefore, sentiment analysis and opinion mining have become important research areas. In user reviews mining, topic modeling based approaches and Latent Dirichlet Allocation (LDA) are significant methods that are used in extracting product aspects in aspect based sentiment analysis. However, LDA cannot be directly applied on user reviews and on other short texts because of data sparsity problem and lack of co-occurrence patterns. Several studies have been published for the adaptation of LDA for short texts. In this study, a novel method for aspect based sentiment analysis, Sentence Segment LDA (SS-LDA) is proposed. SS-LDA is a novel adaptation of LDA algorithm for product aspect extraction. The experimental results reveal that SS-LDA is quite competitive in extracting products aspects.

1. Introduction

Today, with the widespread use of internet applications, huge amount of short text data containing user opinions is present in social media messages and in user reviews about products and services. Therefore, mining user opinions and sentiment analysis have been hot research areas. Sentiment analysis is made at different granularity levels. For a rough analysis of user opinions, the sentiment polarity of the review is classified as positive or negative as a whole. However, sentiment analysis at this level is not satisfactory in most cases. For a more thorough analysis, Aspect Based Sentiment Analysis (ABSA) should be hired. In ABSA, besides sentiment polarity, product aspects on which users express opinions are also extracted.

Sentiment analysis is closely related with Natural Language Processing (NLP), so it has great dependency on the language. Most of the sentiment analysis works are done in English language; however, there is need for research in other languages. In this study, we worked on Turkish language because there are not enough number of ABSA works in this language.

Turkish is an agglutinative language. Turkish words are formed by adding many suffixes to the root words, which makes the language very productive and flexible. When Turkish words are used in the context of a sentence, they can get many inflectional and derivational suffixes (Dehkharghani et al., 2016a). It is quite ordinary that a sentence in English can be expressed by a single word in Turkish (Oflazer and Bozsahin, 1994). For example, the Turkish word "gelemeyecegim" is equivalent to English sentence "I will not be able to come". This agglutinative structure of the language makes NLP tasks very complex. However, besides the complexity, the suffixes at the end of the words also give some clues about the usage of the words and help in NLP tasks.

In this paper, SS-LDA, a novel method for ABSA and for extracting product aspects from user reviews is proposed. SS-LDA is based on LDA, which is the most known topic modeling algorithm. LDA was proposed by Blei et al. (2003) and since that time it has been used in discovering latent topics in documents in many implementations. The proposed SS-LDA method is completely unsupervised and requires no annotated training data. The dataset in this study consists of smartphone reviews taken from www.hepsiburada.com, which is a popular Turkish e-commerce web site.

The main contributions of the study in this paper can be summarized as follows:

- A novel aspect extraction method is proposed for ABSA.

* Corresponding author at: Odeon Software Technology, Guzeloba Mahallesi, 2171 Sokak, Muratpasa, Antalya, Turkey.
E-mail addresses: b.ozyurt@gmail.com (B. Ozyurt), akcayol@gazi.edu.tr (M.A. Akcayol).

https://doi.org/10.1016/j.eswa.2020.114231
Received 12 March 2020; Received in revised form 2 November 2020; Accepted 2 November 2020
Available online 16 November 2020
0957-4174/

- A novel adaptation of LDA for short texts is proposed in this study, which is named as SS-LDA. The proposed SS-LDA can also be used in other studies where the conditions suit the SS-LDA's requirements.
- There are only a few ABSA studies on the Turkish language and their success rates are low. This study has a better success rate for extracting aspects of short and long sentences.

The literature comparison was made in three ways in this study: First, besides our dataset, SS-LDA has also been tested on SemEval-2016 (Pontiki et al., 2016) Turkish restaurant reviews and has been compared with other test results in SemEval-2016 workshop on the same dataset. Second, SS-LDA has been compared with two other LDA adaptations for short texts. They are Sentence-LDA proposed by Jo and Oh (2011) and Biterm Topic Model proposed by Yan et al. (2013). It has also been compared with traditional LDA to make a comparison and to see how traditional LDA performs on aspect extraction without any modification for short texts. In recent years, deep learning and word embeddings have emerged as important approaches in sentiment analysis studies. Therefore, SS-LDA has also been compared with a study that makes use of deep learning and word embeddings approaches. All these tests and comparisons have proved that SS-LDA is quite competitive in extracting product aspects from user reviews.

2. Related work

As mining user opinions have been of great interest, many sentiment analysis researches have been published in literature. Especially within the last decade, research community has concentrated on sentiment analysis more.

Maharani et al. proposed an aspect extraction method using natural language processing and syntactical analysis of sentences (Maharani et al., 2015). They used a rule based approach according to the POS tags of words. Liu et al. proposed a rule based approach for ABSA, which is quite different from other rule based approaches (Liu et al., 2016). In traditional rule based methods, rules are defined and then they are applied. Liu et al. gathered all rules previously proposed and worked on determining the rule set, which will yield the best result. They proposed a heuristic algorithm depending on greedy search and simulated annealing to find the optimum set of rules.

Liao et al. proposed a method for ABSA on Chinese social media messages in which they used a graph data structure (Liao et al., 2016). The nodes in the graph hold the aspect terms and the edges carry a weight according to sentimental association between the nodes. They combined this graph data structure with word embeddings method. Recently, word embeddings method has been widely used in sentiment analysis works. Rezaeinia et al. have worked on enhancing the usage of word embeddings in sentiment analysis (Rezaeinia et al., 2019). They proposed considering POS tags, lexicon-based approaches, word position algorithm together with applying the word embeddings algorithm and showed greater level of success.

Mowlaei et al. proposed two sentiment lexicon generation methods for sentiment analysis (Mowlaei et al., 2020). One method depends on the frequency of words in positive and negative sentences in reviews and the other method is based on genetic algorithms. They verified their work by the tests made on product reviews from different domains and on the dataset of Hu and Liu (2004).

Sentiment analysis studies in two categories are in our special interest area: Sentiment analysis works in Turkish language and ABSA works based on topic modeling and LDA. These studies are summarized in the upcoming sections.

2.1. Sentiment analysis studies in Turkish language

There are not many sentiment analysis publications in Turkish language. Kaya et al. did binary (positive, negative) classification work on political news using a mixture of machine learning and lexicon based methods (Kaya et al., 2012). Cetin et al. worked on product reviews and used a supervised machine learning method (Cetin and Amasyali, 2013). They used TF, TF-IDF and Delta TF-IDF term weighting techniques. Meral et al. worked on classifying Twitter messages (Meral and Diri, 2014). In constructing the machine learning feature vector, instead of word n-grams, they used character n-grams to be able to handle the jargons and abbreviations used in Twitter. They used character 2-g and 3-g.

All above sentiment analysis studies are at document level. ABSA works in Turkish language are as follows: Akbas made an ABSA work on Twitter messages in her Ms. Thesis (Akbas, 2012). In her study, she did not make aspect extraction from tweets. What she did is to classify the tweets according to topics. Dehkharghani et al. made sentiment analysis study in which they made sentiment analysis at different granularity levels including ABSA (Dehkharghani et al., 2016b). However, in their study, they did not make automatic aspect extraction. Instead, they manually constructed a word list for product aspects and searched these words in the user reviews.

2.2. Aspect based sentiment analysis studies using LDA

LDA algorithm is effective on long documents and on large corpus. The co-occurrence pattern of a word in the document is important for LDA. Because of the lack of co-occurrence patterns and data sparsity problem, LDA is not effective on short texts without any modification or adaptation. Adaptation of LDA for short texts is a hot research area.

One method in literature to overcome the data sparsity problem is to aggregate short texts to create pseudo long documents. This method is often used on Twitter analysis by combining tweets. Weng et al. combined Twitter messages posted by the same user (Weng et al., 2010). Shared common words among tweets is another aggregation method implemented by Hong and Davison (2010). Nimala et al. did hashtag-based tweet aggregation strategy in their study (Nimala et al., 2018).

Besides, creating pseudo long documents by aggregating short texts, there are many other studies for the adaptation of LDA to use on short texts.

Lin et al. proposed the Joint Sentiment/Topic Model (JST) for aspect and sentiment extraction from user reviews (Lin and He, 2009). In order to model document sentiments, they added an additional sentiment layer to LDA between the document and the topic layer. In JST, sentiments and topics are simultaneously extracted from the text. The dataset in this study is a collection of movie reviews.

Jo et al. proposed Sentence-LDA and Aspect Sentiment Unification Model (ASUM), which is a similar method to JST (Jo and Oh, 2011). The difference is that in ASUM, each sentence in a user review is assumed to be about single product aspect which is not so in JST. They used electronic devices and restaurant reviews datasets to evaluate Sentence-LDA and ASUM.

Xianghua et al. proposed an adaptation of LDA for short text that they name as MG-LDA (Xianghua et al., 2013). They state that when LDA algorithm is applied on user reviews, not only it discovers the product aspects as topics, but also it discovers the topics other than product aspects. In MG-LDA, topics about product aspects are discovered by local topics and other topics are discovered by global topics. To discover the global topics, LDA algorithm is applied by involving all reviews. To discover the local topics (product aspects), LDA is applied by using sliding window method. They tested MG-LDA on Chinese online social reviews.

Xiong et al. proposed WSTM (Word-pair Sentiment-Topic Model) for the short text reviews (Xiong et al., 2018). In WSTM, they modeled the generative process of sentiment and aspect terms simultaneously. They used a sliding window and in each step of the sliding window, they generated word-pairs and then used these word pairs in their topic model.

García-Pablos et al. proposed W2VLDA method for aspect extraction and sentiment polarity detection on user reviews (García-Pablos et al., 2018). They used LDA algorithm in combination with continuous word embeddings, word2vec and maximum entropy classifier. They tested their system on restaurant and electronic devices (laptop, digital-slr) reviews and on SemEval-2016 task 5 datasets (Pontiki et al., 2016).

Tang et al. proposed joint aspect based sentiment topic (JABST) model that jointly extracts multi-grained aspects and opinions through modeling aspects, opinions, sentiment polarities and granularities simultaneously (Tang et al., 2019). They compared JABST with JST and ASUM on various datasets and showed that JABST outperforms them.

3. Proposed method

In designing the system and designing the aspect extraction method, we inspected all issues that should be considered in ABSA. First issue is that the same product aspect can be expressed with very different words and very different phrases. The system should not extract them as different aspects; it should be able to group these different words and phrases pointing to the same aspect under the same category. Second issue is that a product aspect is not always expressed with a single word, sometimes it is expressed by phrases having some number of words. Third issue is that a sentence in a product review may be about a single aspect or it may be about several aspects of the product. Fourth issue is that, when a user expresses an opinion about a product, he does not always talk about an aspect of the product; he may express a general liking about it without pointing to any aspect.

To clarify, these issues are illustrated on some sample product review sentences below. We worked on Turkish reviews, but here for simplicity, all sample sentences are given in English.

Sentence 1: Photo shoot of the phone at night is *great*.

→ Talks about the "camera" aspect of the phone. The "camera" aspect is expressed with two different words (underlined) which are all related with "camera". Also "camera" aspect is referred without the word "camera".

Sentence 2: The zoom time of the camera at low light is *not satisfactory*.

→ Talks about the "camera" aspect again and the underlined words all about the "camera" aspect. These words are quite different from the words in the previous sentence, which is also about the "camera" aspect.

→ The user expresses an opinion about the smartphone but does not talk about any product aspect. He expresses a general liking about it.

Under the consideration of all these issues, the developed system should be able to handle the following situations correctly:

1. When a product aspect is expressed with more than one word, the system should not extract each word as a different product aspect. For example in Sentence 3, the words "touch", "screen" and "sensitivity" should not be extracted as individual aspects. Instead, the whole phrase "Touch screen sensitivity" should be extracted as a single aspect.
2. When a sentence covers more than one product aspect as in Sentences 3 and 4, the system should be able to extract each of the aspects in the sentence. It should not assume that a sentence is about a single aspect.
3. When the same product aspect is expressed with very different words and phrases, the system should not extract them as different aspects; rather it should group them under the same aspect. For example, the phrases in Sentence 1 and Sentence 2 should grouped under the "camera" aspect.
4. When a user expresses an opinion without talking about any aspect as in Sentence 5, the system should determine this and not try to extract aspect from such sentences.

Considering all these issues, the proposed method consists of two phases: First is the sentence segmentation phase in which sentences are split into segments such that each segment covers a single product aspect. For example, the above sample sentences should be segmented as follows:

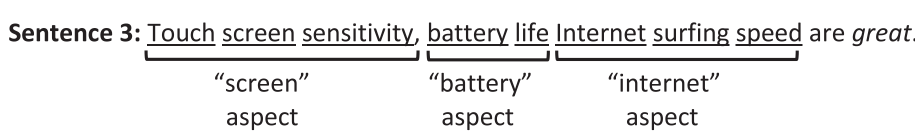
Sentence 1

- Segment-1: "Photo shoot of the phone at night is *great*"

Sentence 2

- Segment-1: "The zoom time of the camera at low light is *not satisfactory*"

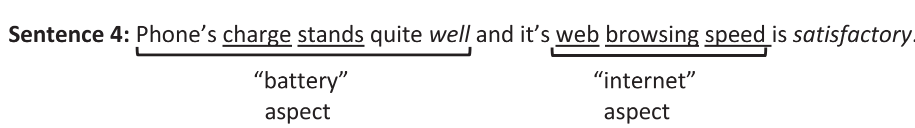
Sentence 3



→ Talks about three different product aspects in one sentence.

- Segment-1: "Touch screen sensitivity"
- Segment-2: "battery life"
- Segment-3: "Internet surfing speed are *great*"

Sentence 4



→ Talks about two different product aspects in one sentence. These two aspects are covered in also Sentence 3 but the words used to express them are not very similar.

Sentence 5: I like it.

- Segment-1: "Phone's charge stands quite *well* and",
- Segment-2: "it's web browsing speed is *satisfactory*."

Sentence 5

Table 1
Grouping of sentence segments under aspect categories.

Aspect category	Sentence segments grouped under aspect category
Camera	- Photo shoot of the phone at night is *great* (Sentence 1, Segment-1)
	- The zoom time of the camera at low light is *not satisfactory* (Sentence 2, Segment-1)
Screen	- Touch screen sensitivity (Sentence 3, Segment-1)
Battery	- battery life (Sentence 3, Segment-2)
	- Phone's charge stands quite *well* and (Sentence 4, Segment-1)
Internet	- Internet surfing speed are *great* (Sentence 3, Segment-3)
	- it's web browsing speed is *satisfactory* (Sentence 4, Segment-2)
No aspect	- I like it (Sentence 5, Segment-1)

- Segment-1: "I like it"

At the end of this phase, we obtain a collection of sentence segments each covering a single product aspect. Second phase is the grouping phase. The sentence segments pointing to the same product aspect with different words and phrases should be grouped under the same aspect category. The segments in the above samples sentences should be grouped as in Table 1.

Sentence segmentation and grouping of sentence segments are the two main phases of the proposed method. Additionally, there are also some intermediate tasks in the whole system. The overall outline of the system is illustrated in Fig. 1. The implementations of sentence segmentation, grouping and the other details of the system are explained in upcoming sections.

3.1. Text preprocessing

Text preprocessing consists of three tasks: 1. Correcting mistypes. 2. Elimination of stop words, product and company names. 3. Morphological analysis (POS tagging, lemmatization, determination of root and suffixes of words). Correcting mistypes and morphological analysis tasks are done by using Zemberek (Zemberek NLP), which is a Turkish morphological analysis tool. For the elimination of product and company names, a list containing product and company names has been manually constructed by us.

3.2. Sentence polarity detection

For the sentiment polarity detection task, a Turkish sentiment dictionary has been manually constructed by us. In constructing the sentiment dictionary, besides compiling the Turkish sentiment terms, some sentiment terms were taken from English sentiment dictionaries by translating them to Turkish. The sentiment dictionary does not only contain single terms, but also it contains sentiment phrases consisting of two and three words.

Sentiment polarity detection is a complex task in sentiment analysis. There are many issues to be considered like irony detection, sarcasm, implicit sentiment polarity detection and so on. In this study, the main focus is about product aspect extraction, so basic level sentiment polarity detection has been implemented. In detecting sentences' sentiment polarity, we first look for the existence of sentiment terms by using the sentiment dictionary. If a sentiment term exists, then the sentence is checked for the existence of a negation word or a negation suffix. In Turkish, negation is done by both negation words and negation suffixes. Lastly, the sentiment polarity is detected by considering antonyms like "ama (*but*)", "fakat (*however*)". If two statements are in association with each other with an antonym and the polarity of one of the statements is known, then the sentiment polarity of the other statement is set oppositely. If a sentence does not contain any sentiment term and is not in association with a subjective sentence by an antonym, then it is marked as objective.

The objective sentences are omitted by the system. Nothing is done about sentences in which user does not express any opinion. Only subjective sentences are passed to the Sentence Segmentation module.

3.3. Sentence segmentation

Most of the aspect extraction methods in literature try to detect the aspect terms in the sentences. This is not the way followed in this study. Instead, the sentences are split into segments such that each segment covers a single aspect. For example, consider the sample sentence below:

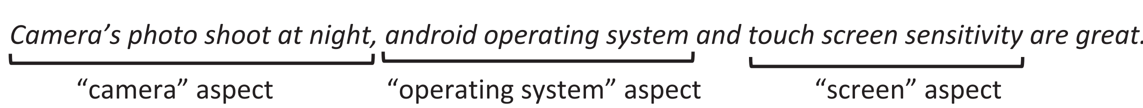
"Camera's photo shoot is great".

Here the user talks about the "camera" aspect of the smartphone. The "camera" is the aspect term and the words "photo", "shoot" are aspect-related words. When an aspect extraction method discovers the term "camera", it is OK for the aspect extraction task. But for SS-LDA, besides the aspect term "camera", the words "photo" and "shoot" which are related with the "camera" aspect are also very important. Because these aspect-related words have great effect on the success of the Grouping Module which contains our LDA adaptation for short texts.

For the segmentation task, we make use of the "frequent nouns" method proposed by Hu and Liu (2004) and the association rules proposed by Agrawal et al. (1993). In their work, Hu et al. stated that in a collection of user reviews, noun type words having high frequency are mostly the words about product aspects and proposed an aspect extraction method based on this proposal. This method has been named as "frequent nouns" method in literature and has been used in many sentiment analysis studies. Although the "frequent nouns" method has been proposed for aspect extraction, in this study, we use it for the sentence segmentation task. Association rules were offered by Agrawal et al. for measuring the association between entities. In association rules high confidence value between X and Y means strong association between them and low confidence value between them means weak association.

The sentence segmentation method will be explained on a sample sentence. For the sake of simplicity, the sample sentence here is quite straightforward and complies with grammar rules. In real world user reviews, the sentences are often very dirty; there is no in compliance with grammar and punctuation rules but the principles of the method does not change.

Sample sentence:



For the "camera" aspect in the sample sentence, "camera" is the aspect term and the words "photo" and "shoot" are the aspect-related words. For the "screen" aspect, "touch" and "sensitivity" words are the aspect-related words.

The words about "camera", "operating system" and "screen" aspects are ordered in the sentence. How will we decide at which points the sentence should be segmented in order to obtain three sentence segments correctly? How will we know that we should cut the first segment just before the word "android" and cut the second segment just before the word "touch"?

First thing to do is to discover which words are aspect and aspect-

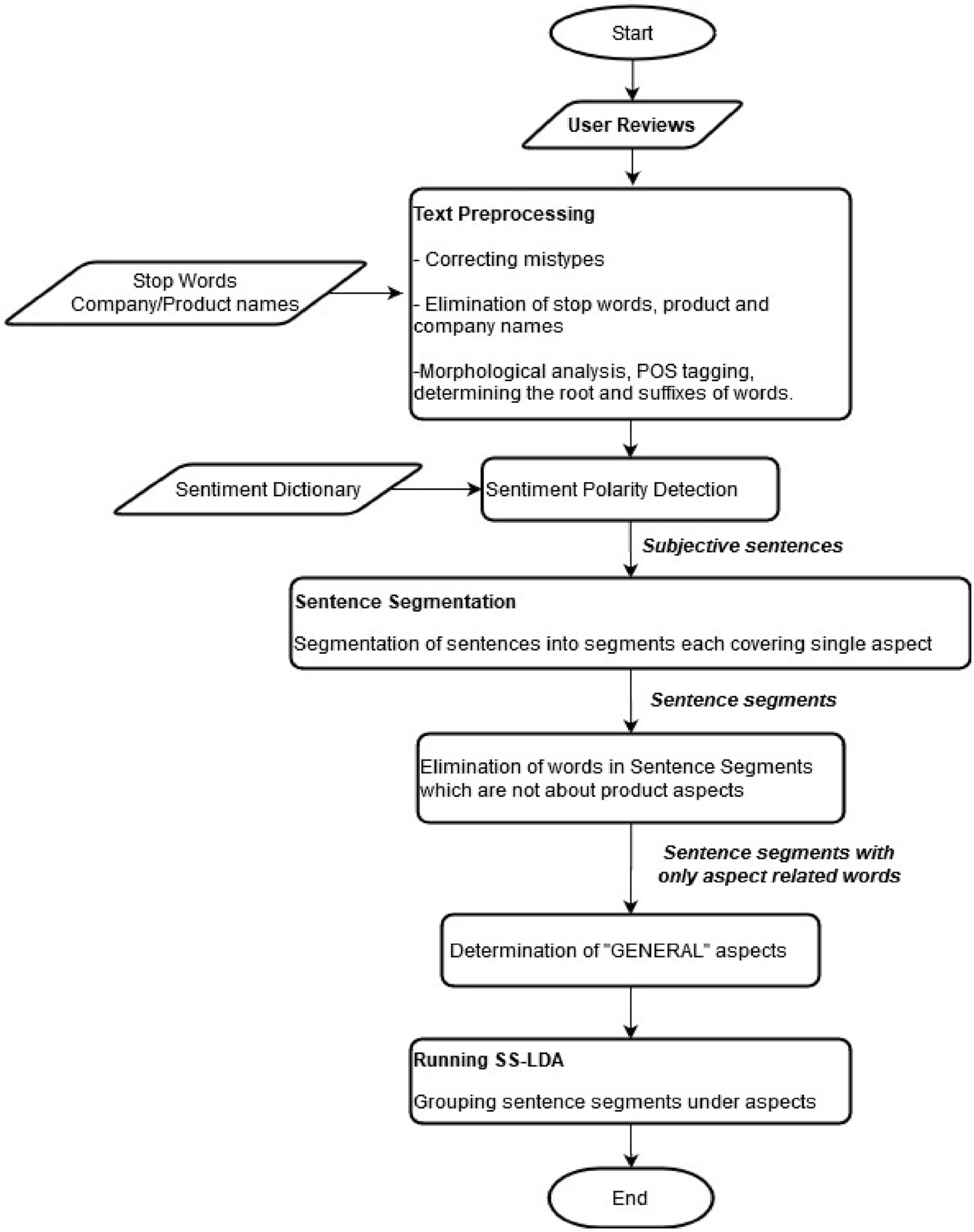


Fig. 1. Outline of proposed method.

related words. Hu et al.'s proposal is true also in our case. In the collection of user reviews, aspect and aspect-related words have high frequency. Hu et al. included only noun type words in their method. However, our inspections on the dataset and on Turkish user reviews revealed that aspect and aspect-related words are mostly nouns (camera, battery), sometimes verbs with some specific conjugations (telefon *ısınıyor* (phone is *heating* up)) and some sentiment terms (elegant, stylish). Therefore, in this study besides noun type words, verbs having specific type of conjugations are also included. These conjugations are *third-person singular present tense*, *third-person singular past tense* and *modal verb* conjugations. In Turkish, thanks to the agglutinative structure of the language, these conjugations can be deducted from the suffixes at the end of the verbs. Hu et al. use the name "frequent nouns" because they include only noun type words. Because we include verbs having some conjugations besides nouns, we use the name as "frequent words".

The aspect and aspect-related words are sparse in objective sentences, they are often found in subjective sentences. Objective sentences often contain other words than aspect and aspect-related words. In the tests we made, we observed that calculating the word frequencies only in subjective sentences and omitting the objective sentences yields better results for discovering aspect and aspect-related words. So, frequencies of words are calculated in only subjective sentences. In the

Table 2
Sample confidence values calculated from the dataset.

Tuples with high confidence value	Tuples with low confidence value
Confidence (system|operating): 0.82609	*Confidence (android|shoot): 0.00000*
Confidence (operating|system): 0.95000	*Confidence (shoot|android): 0.00000*
Confidence (photo|shoot): 0.34127	*Confidence (android|touch): 0.01794*
Confidence (shoot|photo): 0.36364	*Confidence (touch|android): 0.01653*
Confidence (photo|camera): 0.36208	*Confidence (screen|system): 0.00889*
Confidence (camera|photo): 0.29715	*Confidence (system|screen): 0.02174*
Confidence (touch|screen): 0.26860	*Confidence (cpu|camera): 0.00889*
Confidence (screen|touch): 0.28889	*Confidence (camera|cpu): 0.02174*
Confidence (sd|card): 1.00000	*Confidence (Internet|screen): 0.04592*
Confidence (card|sd): 0.23810	*Confidence (screen|Internet): 0.04000*
Confidence (touch|sensitivity): 0.05785	
Confidence (sensitivity|touch): 0.70000	

sample sentence above, the words “camera”, “photo”, “shoot”, “android”, “operating”, “system”, “touch”, “screen”, “sensitivity” are frequent words according to our dataset and they are discovered as aspect and aspect-related words.

After discovering the aspect and aspect-related words, the next step is to segment the sentences using the association rules. The confidence values between the word tuples of the aspect and aspect-related words of the same aspect are high, because semantic associations between these words are strong and they often coexist together. On the other hand, the confidence values between the aspect and the aspect-related words of different aspects are low, because semantic associations between these words are weak and they do not often coexist together. For example, semantic association between word tuples “camera-photo”, “camera-shoot”, “photo-shoot” are strong, so the confidence values of these tuples are high. On the other hand, the confidence values between word tuples “camera-android”, “camera-operating”, “system-touch”, “system-screen” are low because they seldom coexist together.

In calculating the confidence values, the rule for two words coexist together is that if two words are in the same sentence within a pre-defined distance, then it is assumed that these words coexist together. This distance is defined as 5 in our tests. Therefore, if two words are in the same sentence and the distance between these two words is 5 or less, then we assume that these words coexist together and calculate the confidence values accordingly.

To clarify, sample confidence values calculated from our whole dataset are listed in Table 2. As seen in the table, confidence values between the words of the same aspect are high. On the other hand, confidence values between the words of different aspects are low and some of them are zero.

After discovering the aspect and aspect-related words and calculating the confidence values between these words, the next task is the segmentation of sentences. In the segmentation task, nothing is done about objective sentences, they are directly passed over. If the sentence is subjective, then the elements of the sentence are handled one by one consecutively. Each element is checked whether it is a “frequent word” or not. If it’s a “frequent word”, then:

- If it is the first “frequent word” in the sentence, nothing is done and passed to the next sentence element.
- If it is not the first “frequent word”, then the confidence value between this word and the previous “frequent word” is checked. In this check, both Confidence (previous word | current word) and (current word | previous word) are checked and the greater one is taken.
  - If the confidence is greater than the threshold value then these words are strongly associated with each other. This means that this word and the previous “frequent word” are related words of the same product aspect, so no segmentation is done at this point.
  - If the confidence is below the threshold value, then there is weak association between these words. This means that this “frequent word” and the previous “frequent word” are about different product aspects, so segmentation is done just before this “frequent word”. But when making the segmentation, word relations like possessive construction, adjective-noun, adjective-verb, adverb-verb associations are considered and the segmentation point is shifted one or two words back if necessary.

The formal representation of this algorithm as pseudocode is shown in Algorithm 1 and its flowchart representation is illustrated in Fig. 2.

Algorithm 1. Sentence segmentation algorithm.

```
Input: 1) Sentence 2) Frequent Words List 3) Association Rules 4) Confidence Threshold
Output: Sentence segments
Process:
  firstFrequentWordPassed := false
  previousFrequentWord := null
  for each element e in Sentence
      if e is in Frequent Words List then
         if firstFrequentWordPassed then
            if  confidence(e, previousFrequentWord) >  Confidence Threshold or
               confidence(previousFrequentWord, e) > Confidence Threshold then
                 splitSentenceBeforeElement(e);
            endif
         else
            firstFrequentWordPassed := true
         endif
         previousFrequentWord := e
      endif
  endfor
```

The application of this algorithm is demonstrated on two sample sentences. One is a short sentence in which user talks about a single product aspect and the second one is a relatively long sentence in which user talks about three product aspects. The words like “at”, “and”, “are”, “is” and “its” are stop words and they are eliminated at the text pre-processing phase.

Sample sentence 1:

Its CPU speed is awesome.

The elements of the sentence are passed one by one consecutively and the operations done on each element are as follows (“*Its*” and “*is*” are stop words):

CPU: *Frequent word but it is the first frequent word in the sentence. It is passed over.*

speed: *Frequent word. Confidence value with the previous frequent word (CPU) is high. They are strongly associated and belong to the same product aspect. No segmentation.*

awesome: *It is sentiment term.*

As a result, no segmentation is done on this sentence. The sentence segmentation algorithm returns the whole sentence as a single sentence segment as follows (stop words eliminated):

- *CPU Speed awesome*

Sample sentence 2:

Camera’s photo shoot at night, android operating system and touch screen sensitivity are great.

The elements of the sentences are passed one by one consecutively and the operations done on each element are as follows:

Camera: *Frequent word but it is the first frequent word in the sentence. It is passed over.*

photo: *Frequent word. Confidence value with the previous frequent word (camera) is high. They are strongly associated and belong to the same aspect. No segmentation.*

shoot: *Frequent word. Confidence value with the previous frequent word (photo) is high. They are strongly associated and belong to the same aspect. No segmentation.*

night: *Not a frequent word, it is passed over.*

android: *Frequent word. Confidence value with the previous frequent word (shoot) is low. The association between them is weak. These words are about different product aspects. Segmentation is done just before this word.*

operating: *Frequent word. Confidence value with the previous frequent word (android) is high. They are strongly associated and belong to the same aspect. No segmentation.*

system: *Frequent word. Confidence value with the previous frequent word (operating) is high. They are strongly associated and belong to the same aspect. No segmentation.*

touch: *Frequent word. Confidence value with the previous frequent word (system) is low. The association between them is weak. These words are about different product aspects. Segmentation is done just before this word.*

screen: *Frequent word. Confidence value with the previous frequent word (touch) is high. They are strongly associated and belong to the same aspect. No segmentation.*

sensitivity: *Frequent word. Confidence value with the previous frequent word (screen) is high. They are strongly associated and belong to the same aspect. No segmentation.*

great: *It is sentiment term.*

As a result, the following sentence segments are obtained from the sample sentence (stop words eliminated):

- *Camera’s photo shoot night.*
- *android operating system.*
- *touch screen sensitivity great.*

If the sentence is subjective and contains more than one sentiment term, then the sentence is segmented just after the sentiment terms. By this way, sub-sentences having one sentiment term are obtained. These sub-sentences having one sentiment term are segmented according to the above algorithm. If the sentence does not contain any “frequent word”, then according to the above algorithm no segmentation is done, i.e. one sentence segment is obtained from the sentence.

After segmenting the sentences, the next step is to eliminate the words, which are not related with product aspects. Our inspections on our data set, on other Turkish user reviews have revealed that the following types of words are related with product aspects and the other words are not:

- Noun type words.
- Verbs having the following conjugations: *third-person singular present tense*, *third-person singular past tense* and *modal verb* conjugations. Modal verbs are distinct words in English, but in Turkish they are added as a conjugation to the verb.
- Some sentiment terms (elegant, stylish). Not all sentiment terms are related with product aspects. Sentiment terms like “great”, “good”, “bad”, etc. are not related with product aspects, they are general terms. But we cannot know which sentiment terms are related with aspects and which are not. Therefore, we include all of them.

These type of words are left in the sentence segments and the other words are eliminated. The intention here is to pass only the aspect and aspect-related words to the Grouping module. Eliminating the words that are not related with product aspects enhances the success of the Grouping module.

The next task is to determine the “General” aspects. In some sentence segments, the user expresses a positive or a negative opinion about the product without stating any aspect. For example “begendim (*I like it*)”, “Herkese bu telefonu tavsiye ederim (*I recommend this phone to everyone*)” and so on. In the proposed method, these statements are categorized under a special aspect named “General”. Grouping such statements under pre-defined special “General” aspect is first made by Liu (2012). In order to decide whether a sentence segment is about “General” aspect or not, we look for the presence of any “frequent word”. If there is no “frequent word” in the sentence segment, then it is decided that this sentence segment is about “General” aspect. The sentence segments which are determined as “General” are not passed to the Grouping Module.

3.4. Grouping - sentence segment LDA (SS-LDA)

After segmenting the sentences, eliminating the words that are not related with aspects and determining the “General” aspects, we have a collection of sentence segments each covering a single product aspect and each consisting of one or more words. The next step is to group these sentence segments in the way that sentence segments about the same product aspect are grouped under the same category.

For the grouping task, LDA was decided as the most suitable and effective method. However, LDA is not effective on short texts because of the lack of co-occurrence patterns of words in short texts and the data sparsity problem as stated above. So in this study, Sentence Segment LDA (SS-LDA) which is our adaptation of LDA for short text is proposed.

Traditional LDA algorithm assumes the document as Bag of Words (BoW) model. With the BoW model assumption, the sequence of sentence elements and relations between them are not considered. Topic assignment of each word in the document is independent of previous and subsequent words. SS-LDA does not assume the document as a BoW model. Topic assignment of a word in a sentence is not independent of previous and subsequent words.

In fact, in SS-LDA, topic assignment is not made to the words; it is made to the sentence segments. When a sentence segment is assigned to a topic, all words in the sentence segment become assigned to that topic.

The process of document (review) generation in SS-LDA is as follows:

1. *For each topic (aspect)* z
   1.1. *Draw a word distribution*: φ_z *Dirichlet*(β)
2. *For each document (review)* d:

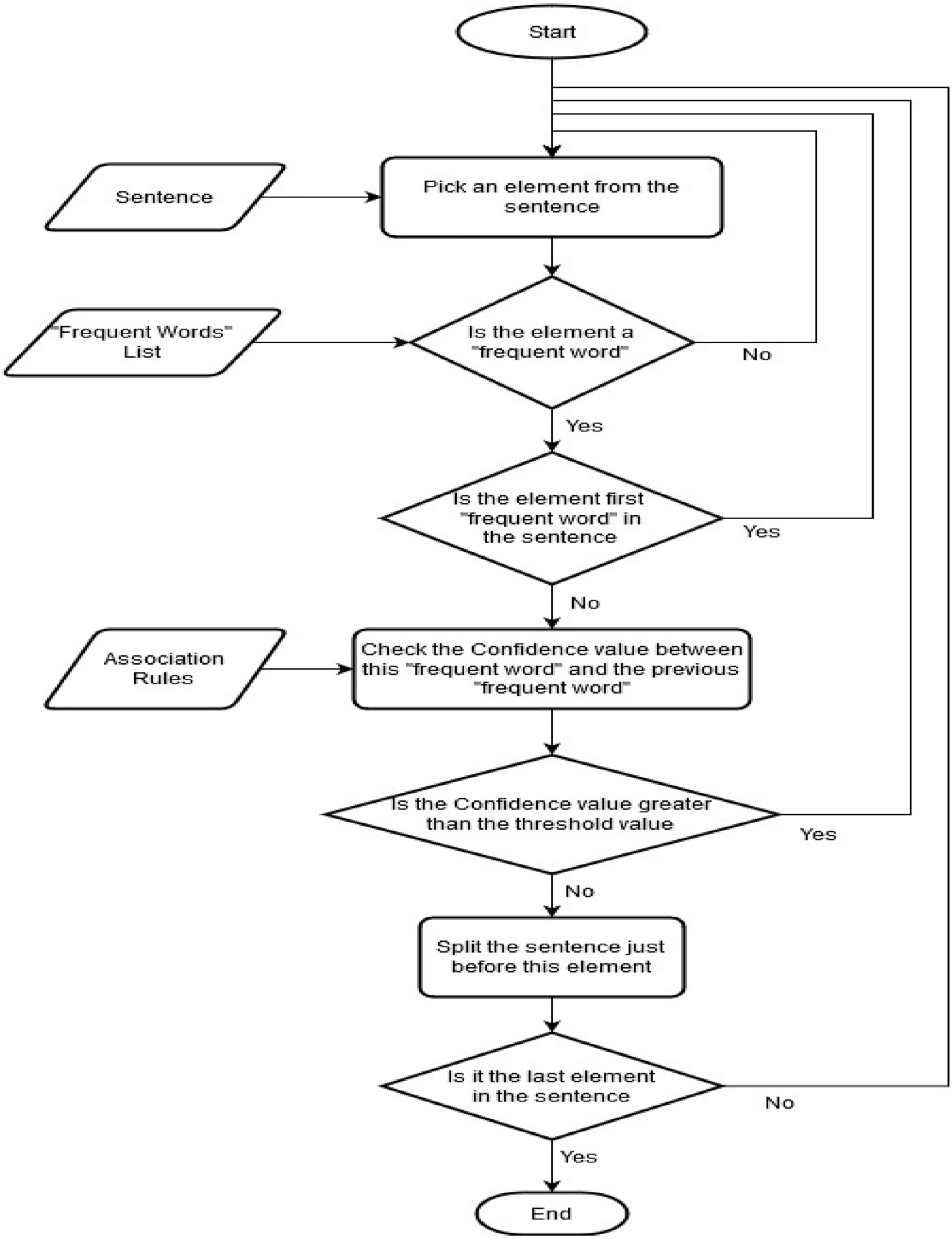


Fig. 2. Flowchart of sentence segmentation algorithm.

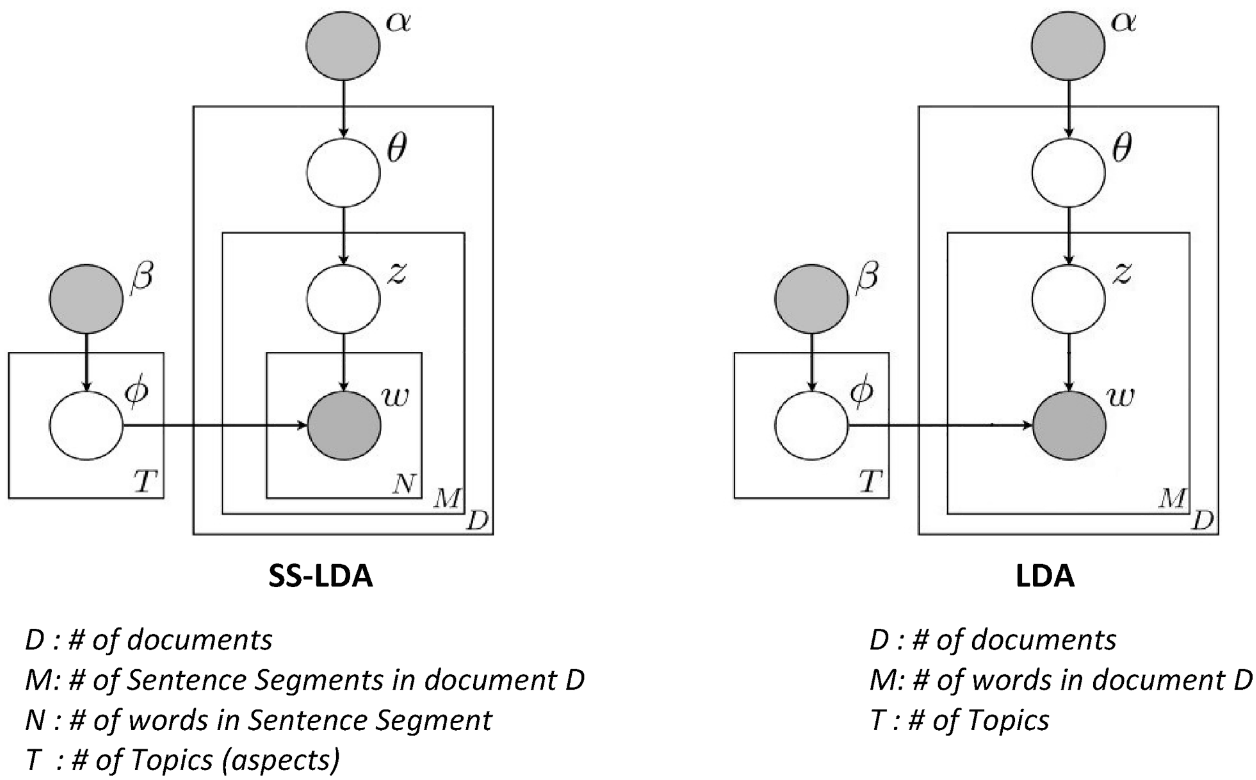


Fig. 3. Plate notations of SS-LDA and LDA.

2.1. *Draw the topic (aspect) distribution of the review:* $\theta_{\tilde{d}}$ *Dirichlet*(α).
2.2. *For each Sentence Segment SS,*
  2.2.1. *A topic (aspect) z is chosen for SS*
  2.2.2. *For each word in SS*
    2.2.2.1. *A word is chosen from z's word distribution.*

The Plate notation of SS-LDA is shown in Fig. 3. To make a comparison and to see the difference, plate notation of traditional LDA is also illustrated in Fig. 2. As seen, there is Sentence Segment generation process in SS-LDA, which is not present in LDA.

The topic assignment formula in Gibbs Sampling iterations of SS-LDA should also be modified accordingly. But before, we will first look at the word topic assignment formula of traditional LDA in each step of the iteration and then we will look how it is modified in SS-LDA. In each Gibbs Sampling iteration of traditional LDA, while choosing a new topic for word W in document D, the probability of choosing topic Z for word W is calculated with Eq. (2).

$$\mathrm{P}(\mathrm{Z}|W,D)=\left(\begin{matrix} number\,of\,words\,in\,D \\ that\,are\,assigned\,to\,Z \end{matrix}+\alpha\right)*\frac{number\,of\,word\,W\,in\,topic\,Z+\beta}{total\,number\,of\,words\,in\,Z+\beta sum} \tag{2}$$

Table 3
Notations.

Notation	Description
N_z^{SS}	Total number of sentence segments that are assigned to topic (aspect) z in all documents
N_{zw}^{W}	Number of word w assigned to topic (aspect) z
N_z^{W}	Total number of words assigned to topic (aspect) z
D	Number of documents (reviews)
N_{SSW}	Number of words in sentence segment (SS)
α	Dirichlet Parameter for distribution of review on aspects
β	Dirichlet Parameter for distribution of topic on words
β_{sum}	β * Vocabulary size

The logic behind this formula is that when choosing a new topic Z for a word in a document, two issues are considered:

- How common is topic Z in this document (left side of multiplication). If topic Z is common in this document, then the probability that this word will be assigned to topic Z will increase.
- How often does word W appear in topic Z elsewhere? (right side of multiplication). If word W is often assigned to topic Z in all documents, than the probability that this instance of W will be assigned to Z will increase.

In the Gibbs Sampling iterations of SS-LDA, we pass through sentence segments (not words as in LDA) and assign a new topic for each sentence segment. The above formula and the logic behind are modified according to document generation process of SS-LDA and according to the consideration that user reviews are very short texts. These modifications are as follows:

1. Modifying "How common is topic (aspect) Z in this document (review)": This criteria is meaningless in a user review because a product aspect often appears only once in a user review that consists of a few sentences. Instead, we consider "How common is aspect Z in all reviews". If all other users are often talking about aspect Z, then the probability that this sentence segment will be about aspect Z will increase.

2. Modifying "How often does word W appear in topic Z elsewhere?": Here, we don't have a single word, we have a sentence segment consisting of some number of words. The criteria "How often does word W appear in topic Z elsewhere?" is considered for each word in the sentence segment. Therefore, we calculate probability for each word in the sentence segment using the right side of the multiplication in Eq. (2) and then we multiply these probabilities. So each word in the sentence segment in considered in calculating the right side of the Eq. (2).

Consequently, we get the formula in Eq. (3) for the Gibbs Sampling

Table 4
Inferred aspects.

Aspect	Inferred Words
Ekran (*Screen*)	Ekran, dokunmatik, hassas, cozunurluk, boyut, hassasiyet, parlaklik, algilama buyuk, hizli. (*Screen*, *touch*, *sensitive*, *resolution*, *size*, *sensitivity*, *brightness*, *sense*, *large*, *fast*)
Kamera-Coklu Ortam (*Camera-Multimedia*)	Kamera, kalite, goruntu, fotograf, ses, cekim, guzel, video, resim, harika, cek, mp, flas, makine, net, yeterli, isik, zoom, hd. (*Camera*, *quality*, *good*, *image*, *photo*, *sound*, *shoot*, *nice*, *video*, *photo*, *wonderful*, *shoot*, *mp*, *flash*, *machine*, *enough*, *light*, *zoom*, *hd*)
Hiz-Internet Hizi (*Speed-Internet Speed*)	Hizli, hiz, Internet, islemci, uygulama, wifi, baglanti, flash, calisiyor, program, menu, video. (*Fast*, *speed*, *Internet*, *CPU*, *application*, *wifi*, *connection*, *flash*, *works*, *software*, *menu*, *video*)
Isletim Sistemi (*Operating System*)	Android, sistem, isletim, uygulama, market, kullanisli, guncelleme, ad, ucret, yuk, program. (*Android*, *system*, *operating*, *application*, *market*, *useful*, *update*, *name*, *price*, *load*, *software*)
Hafiza-Bellek (*Memory-Storage*)	Hafiza, ram, oyun, dahili, kart, dusuk, mb, gb, yetersiz, uygulama, program, sd. (*Memory*, *ram*, *game*, *internal*, *card*, *insufficient*, *mb*, *gb*, *application*, *software*, *sd*)
Tasarim-Kasa (*Design-Body*)	Sik, kullanim, tasarim, hafif, zarif, renk, menu, gorunum, kibar, gorunus, rahat, dur, kullanma, gorsel, harika, estetik, kullanis, hos, yorum, beyaz, kilif, durus, tur, kasa, pratik, dizayn, boyut. (*stylish*, *usage*, *design*, *light*, *elegant*, *color*, *menu*, *appearance*, *gracious*, *look*, *ease*, *look*, *usage*, *visual*, *wonderful*, *stylish*, *usage*, *nice*, *comment*, *white*, *cover*, *look*, *type*, *body*, *practical*, *design*, *size*)
Batarya (*Battery*)	Sarj, gun, git, sure, batarya, pil, gitme, kullanim, bit, omur, isinma, durum, kullan, dayan. (*charge*, *day*, *stand*, *battery*, *battery*, *go*, *usage*, *finish*, *life*, *state*, *use*, *stand*)
Fiyat (*Price*)	Fiyat, uygun, performans, para, piyasa, pahali, ucuz, android, makul, cazip, kullanisli, marka, garanti, ideal, butce, hesapli. (*Price*, *proper*, *performance*, *money*, *market*, *expensive*, *cheap*, *android*, *reasonable*, *inviting*, *useful*, *brand*, *guarantee*, *ideal*, *budget*, *economic*)
Kargo-Siparis (Shipping-Ordering)	hizli, gun, kargo, site, siparis, hediye, ulas, gel, gonderi, gec, teslimat, sure, cik, yil, teslim, guvenilir, kullanisli, alis, hizmet (*fast*, *shipping*, *site*, *order*, *gift*, *arrive*, *come*, *post*, *late*, *delivery*, *time*, *leave*, *year*, *deliver*, *reliable*, *useful*, *buying*, *service*)

iteration steps of SS-LDA.v

$$P(Z|SS, D) = \left(\begin{matrix} \textit{number of SS in all documets} \\ \textit{that are assigned to } \mathbf{Z} \end{matrix} + \alpha \right) * \prod_{w=1}^{N_{SSW}} \frac{\textit{number of word } w \textit{ in topic } Z + \beta}{\textit{total number of words in } Z + \beta_{sum}} \quad (3)$$

The formal representation of this formula is given in Eq. (4) and the notations are described in Table 3.

$$P(Z|SS, D) = \left(N_z^{SS} + \alpha\right) * \prod_{w=1}^{SS_{nw}} \left(\frac{N_{zw}^{W} + \beta}{N_z^{W} + \beta_{sum}} \right) \quad (4)$$

Table 5
Success level of the proposed method in extracting product aspects.

Metric	Value
Number of reviews	1292
Number of aspects annotated	4014
Number of aspects extracted by SS-LDA	4116
Number of aspects truly extracted	3349
Precision	%81.36
Recall	%83.43
F-Score	%82.39

Table 6
Success level of the proposed method for each product aspect

Product aspect	Precision	Recall	F-Score
"General"	%91.49	%84.15	%87.67
Ekran (*Screen*)	%80.72	%88.16	%84.28
Kamera-Çoklu Ortam (*Camera-Multimedia*)	%72.32	%85.76	%78.47
Hiz-Internet Hizi (*Speed-Internet Speed*)	%79.68	%77.18	%78.41
Isletim Sistemi (*Operating System*)	%73.91	%54.48	%62.73
Hafiza-Bellek (*Memory-Storage*)	%63.35	%93.25	%75.45
Tasarim-Kasa (*Design-Body*)	%78.19	%74.53	%76.31
Batarya (*Battery*)	%66.23	%84.07	%74.09
Fiyat (*Price*)	%81.37	%93.91	%87.19
Kargo-Siparis (*Shipping-Ordering*)	%62.35	%82.79	%71.13

In SS-LDA, making topic assignments to sentence segments rather than single words significantly improves the grouping success. In traditional LDA, the association relations between words and the coexistence patterns of words are not considered. However, in SS-LDA the coexistence patterns of words and association relations between them are considered. For example, consider the words "camera" and "shoot". The "shoot" often exists with "camera", so it is often in the same sentence segment with "camera". When they are in the same sentence segment, they are assigned to the same topic. And then when the "shoot" word exists somewhere without "camera", the probability that SS-LDA will assign it to the same topic with "camera" will be quite high because the "shoot" words that are in the same sentence segment with "camera" are all assigned to that topic. This is how the coexistence patterns of words significantly enhances the grouping success in SS-LDA.

An important parameter in SS-LDA is the number of topics (aspects). We specified the following 9 product aspects for smartphones:

1. Ekran (*Screen*)
2. Kamera-Coklu Ortam (*Camera-Multimedia*)
3. Hiz-Internet Hizi (*Speed-Internet Speed*)
4. Isletim Sistemi-Yazilim (*Operating System-Software*)
5. Hafiza-Bellek (*Memory-Storage*)
6. Tasarim-Kasa (*Design-Body*)
7. Batarya (*Battery*)
8. Fiyat (*Price*)
9. Kargo-Siparis (*Shipping-Ordering*)

The reviews in dataset have been manually annotated by us according to above 9 aspects and the "General" aspect. As seen in Fig. 1, the determination of "General" aspects is done before the Grouping Module, which contains the SS-LDA implementation. This means that the sentence segments which are decided as about "General" aspect are not passed to the Grouping Module.

4. Experimental results

The dataset in this study is a collection of 1292 user reviews in Turkish language about smartphones. The user reviews in the dataset are taken from www.hepsiburada.com, which is a popular e-commerce web site in Turkey. This dataset has been manually annotated by us. Besides our dataset, the proposed method has also been tested on SemEval-2016 Task-5 Turkish restaurant reviews to make a literature comparison.

Both the dataset in this study and SemEval-2016 Task-5 Turkish restaurant reviews dataset are real world data. SS-LDA has been tested and validated on real word user reviews.

The inferred aspects by SS-LDA under the above 9 aspect categories are given in Table 4. The results in Table 4 reveal that SS-LDA is quite successful in grouping the words under the aspects. It is observed that

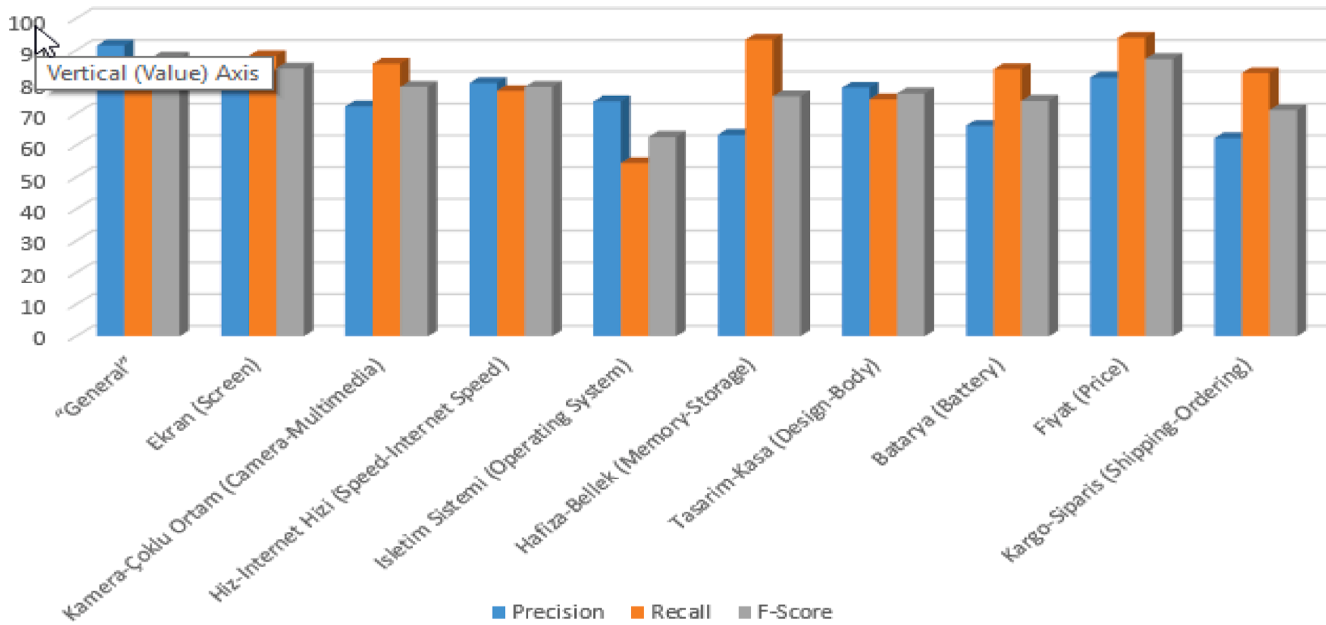


Fig. 4. Success level of the proposed method for each product aspect.

Table 7
F-Score of SS-LDA and submissions in SemEval-2016.

Method/Submission	F-Score
UFAL/U	%61.029
basel./C	%58.896
IIT-T./U	%56.627
IIT-T./C	%55.728
INSIG./C	%49.123
SS-LDA	%62.250

different words pointing to same product aspect are truly grouped together. For example, SS-LDA groups the words "screen", "touch", "sensitive", "resolution" under the "Screen" category, which is a quite satisfactory result.

To measure the success of the proposed method in extracting the product aspects, three measuring metrics have been used: Precision, Recall and F-Score. We manually annotated 4014 product aspects in these 1292 user reviews. Table 5 shows the evaluation results of the proposed method in extracting the product aspects from reviews.

The evaluation results for each product aspect were also measured. The aspect extraction results of the proposed method for each aspect including the "General" aspect are given in Table 6 and Fig. 4. The success level for each product aspect are not the same. This is related with the complexity of statements and characteristics of coexistence patterns in statements about the aspects. For example, the precision and recall of "Fiyat (*Price*)" aspect is high because the users rarely use implicit expressions when talking about price aspect. They often directly use the word "Fiyat (*Price*)". Fiyat (*Price*) is the most frequent word in the reviews by the way.

On the other hand, the precision in "Hafiza-Bellek (*Memory-Storage*)" aspect is relatively low. There are two reasons for this: First, the users use often use varying words and statements (gb, ram, sd card, game, etc.) when talking about this aspect. Second reason is that the abundance of this aspect in reviews is quite low. Both sentence segmentation task and SS-LDA depend on statistical distributions (coexistence and co-occurrence patterns) of words. When the abundance of an aspect is

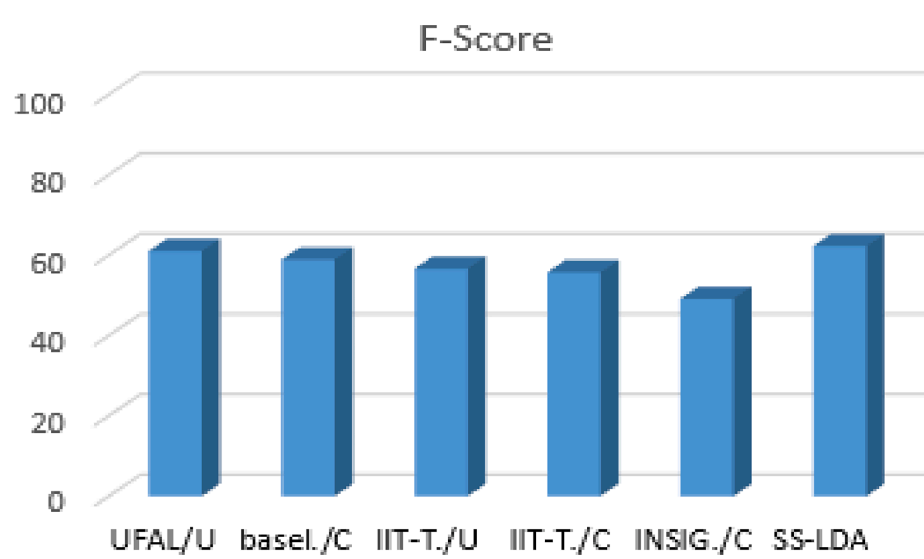


Fig. 5. F-Score of SS-LDA and submissions in SemEval-2016.

Table 8
Comparison of SS-LDA, Sentence-LDA, Biterm Topic Model and LDA, in extracting product aspects

Metric	SS-LDA	Sentence-LDA	Biterm Topic Model	LDA
Number of Aspects extracted	4116	2353	4266	2603
Number of Aspects truly extracted	3349	1846	2740	1279
Precision	%81.36	%78.48	%64.23	%49,16
Recall	%83.43	%58.62	%68.26	%31.88
F-Score	%82.39	%67.11	%66.18	%39.16

relatively low, the statistical distributions about that aspect becomes relatively inadequate.

Measuring the success level of polarity detection in ABSA is a little bit tricky. Each review is annotated with aspect#polarity tuples. For a polarity to be discovered by the system, first the aspect with which it is associated must be extracted by the system. So if a product aspect cannot be truly extracted by the system, also its polarity cannot be detected. Therefore, it is not very clear how the precision, recall and F-score values will be calculated for sentiment polarity detection in ABSA. The correctness of detected polarities on the aspects extracted by the system have been measured and it has been observed that %96.79 of them are correct. However, the polarities in unextracted aspect#polaritiy tuples are not included in this percent.

There are not many ABSA researches in Turkish language. In their research, Dehkharghani et al. (Dehkharghani et al., 2016a) stated that they could not find any ABSA work in Turkish to make a direct comparison.

We made literature comparison of SS-LDA in three ways. First, we made a comparison with the works in SemEval-2016 workshop (Pontiki et al., 2016). In SemEval-2016, there are different tasks about sentiment analysis and Task 5 is about aspect extraction. There are datasets in several different languages and the submissions on same languages are compared among themselves. There are four submissions about Task 5 on Turkish restaurant reviews dataset. Additionally, there is a baseline implementation and a baseline score for each task in SemEval-2016. The submissions are compared according to F-Score values.

SS-LDA has been run on SemEval-2016 Turkish restaurant reviews dataset. The results are %65.84 for precision, %59.04 for recall and %62.25 for F-Score. The F-Score values of baseline, four submissions and SS-LDA method are shown in Table 7 and Fig. 5. The %62.25 F-Score value is slightly better than the best score which is %61.029.

In comparing the results, there is an important point to mention: In SemEval-2016, with the training data supplied, all submissions implement supervised machine learning algorithms. However, SS-LDA is unsupervised. Supervised machine learning based methods make use of training data and this provides a great advantage for them. In their works, Kennedy and Inkpen (2006) and Hailong et al. (2014) also reveal that supervised machine learning methods are more successful than unsupervised methods. Nevertheless, supervised machine learning methods have a great disadvantage that they require annotated training data, which is not always possible. With this fact in mind, when we compare our method's F-Score value with others, we think that unsupervised SS-LDA method is quite successful in SemEval-2016 Turkish dataset.

Second, SS-LDA has been compared with traditional LDA and two other LDA adaptations for short texts in the literature. They are Sentence-LDA proposed by Jo and Oh (2011), and Biterm Topic Model proposed by Yan et al. (2013). Biterm Topic Model's source code is shared on github repository by the author. After getting the source code from github, the input has been prepared from our dataset according to format it requires and Biterm Topic Model has been successfully run on our dataset. Sentence-LDA is similar to SS-LDA since it also handles the reviews at sentence level, but its LDA adaptation for short text is different from SS-LDA. Therefore, some code changes have been made on SS-LDA and Sentence-LDA has been implemented. Consequently, SS-LDA, Sentence-LDA, Biterm Topic Model and LDA have been run on our

Table 9
Comparison of SS-LDA with method by Bajpai et al. in extracting product aspects.

Metric	SS-LDA	Method by Bajpai et al.
Number of reviews	339	39
Number of aspects extracted	1453	130
Number of aspects truly extracted	953	71
Precision	%65.588	%54.615
Recall	%59.229	%49.305
F-Score	%62.250	%51.825

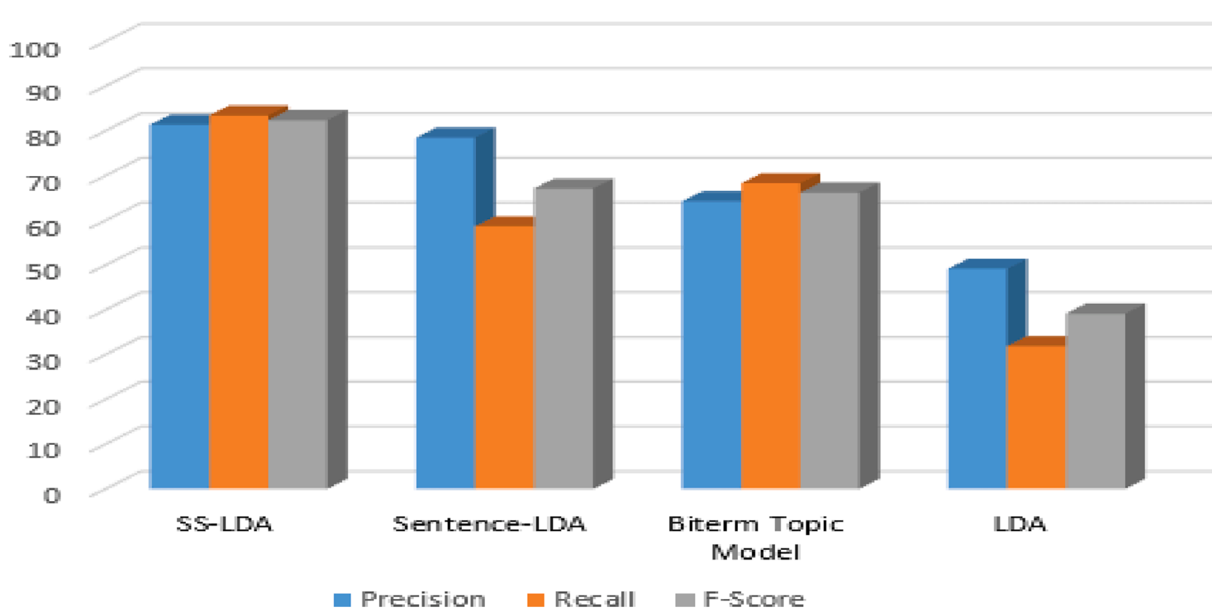


Fig. 6. Comparison of SS-LDA, Sentence-LDA, Biterm Topic Model and LDA, in extracting product aspects.

smartphone reviews dataset and their capability in aspect extraction have been compared. The results are illustrated on Table 8 and Fig. 6. SS-LDA has performed the best scores. Sentence-LDA also performed a good precision score but its recall value is relatively low. LDA's performance in aspect extraction is not satisfactory. Before including LDA in this comparison, we were predicting that its performance in aspect extraction would be low. At the end, the results verified our prediction.

In recent years, deep learning and word embeddings based approaches have been quite popular in sentiment analysis studies. Therefore, as the third literature comparison, SS-LDA has been compared with the method proposed by Bajpai et al. (Bajpai et al., 2019). In the proposed method, they use deep convolutional neural network and word embeddings approaches for aspect extraction. The source code of the aspect extraction method is available on github repository. To run the method, a word embeddings file in word2vec format is required. In Turkish language, the most comprehensive and the largest word embeddings file is the one provided by Akoksal (Akoksal) which has been constructed with Wikipedia dumps. By using this word embeddings file, we have run the method on SemEval-2016 Turkish restaurant reviews dataset. Since SS-LDA has also been tested on SemEval-2016 Turkish restaurant reviews dataset before, the comparison has been carried out on this dataset.

The method proposed by Bajpai et al. is a supervised method, which requires training data. So it has been trained with SemaEval-2016 training dataset (300 reviews) and has been tested with SemEval-2016 test dataset (39 reviews). On the other hand, SS-LDA is unsupervised, it requires no training data. Therefore, SS-LDA has been tested on SemEval-2016 all datasets (training + test). The test results are illustrated in Table 9. The method proposed by Bajpai et al. has performed % 51.825 as the F-Score value. The F-Score value of SS-LDA is %62.250. As a result, SS-LDA has revealed better performance in aspect extraction from SemEval-2016 Turkish restaurant reviews dataset.

5. Conclusions

In this paper, SS-LDA, a topic modeling based method for aspect extraction for aspect based sentiment analysis is proposed. SS-LDA method is unsupervised and does not require any annotated training data. It requires a sentiment dictionary as a resource. The dataset in this study is a collection of smartphone reviews in Turkish language. Turkish is an agglutinative and a very flexible language on which sentiment analysis and NLP works are quite hard to implement. The proposed method works on Turkish language but with some little revisions, it can work on other languages as well.

SS-LDA has been tested on smartphone reviews and the experimental results have shown that it is quite successful in extracting the product aspects. To make a literature comparison, SS-LDA has been tested on SemEval-2016 Turkish restaurant reviews and outperformed all the works on that dataset. Second, SS-LDA has been compared with traditional LDA and with two other LDA adaptations for short text in literature. Among these LDA implementations, SS-LDA performed the best scores in aspect extraction. Deep learning and word embeddings based approaches have been popular in recent years. Therefore, SS-LDA has also been compared with a study in literature, which uses deep learning and word embeddings approaches for aspect extraction.

Some improvements can be made on the proposed method. Especially the first phase of the method where the sentence segmentation task is performed is open for improvement. The segmentation the sentences can be improved or different algorithms can be used there. Correct segmentation of sentences is very important because it is determinant in the overall success of the proposed method.

CRediT authorship contribution statement

Baris Ozyurt: Conceptualization, Methodology, Software, Validation, Investigation, Resources, Writing - original draft. **M. Ali Akcayol:** Conceptualization, Methodology, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Agrawal, R., Imielinski, T., & Swami, A. N. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '93)* (pp. 207–216).

Akbas, E. (2012). *Aspect based opinion mining on Turkish tweets.* Ms. Thesis, Bilkent University, Department of Computer Engineering. Turkey: Ankara.

Akoksal. (2020). Turkish word embeddings file. https://github.com/akoksal/Turkish-Word2Vec. (Accessed 22 October 2020).

Bajpai, R., Hazarika, D., Singh, K., Gorantla, S., Cambria, E., & Zimmerman, R. (2019). *Aspect-Sentiment Embeddings for Company Profiling and Employee Opinion Mining, arXiv preprint, arXiv*, 1902.08342.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research, 3*, 993–1022.

Cetin, M., & Amasyali, M. F. (2013). Supervised and traditional term weighting methods for sentiment analysis. *Signal Processing and Communications Applications Conference (SIU)*, 1–4.

Dehkharghani, R., Saygin, Y., Yanikoglu, B., & Oflazer, K. (2016a). SentiTurkNet: A Turkish polarity lexicon for sentiment analysis. *Language Resources and Evaluation, 50*, 667–685. https://doi.org/10.1007/s10579-015-9307-6

Dehkharghani, R., Saygin, Y., Yanikoglu, B., & Oflazer, K. (2016b). Sentiment analysis in Turkish at different granularity levels. *Natural Language Engineering, 23*, 535–559. https://doi.org/10.1017/S1351324916000309

García-Pablos, A., Cuadros, M., & Rigau, G. (2018). W2VLDA: Almost unsupervised system for aspect based sentiment analysis. *Expert Systems with Applications, 91*, 127–137. https://doi.org/10.1016/j.eswa.2017.08.049

Hailong, Z., Wenyan, G., & Bo, J. (2014). Machine learning and lexicon based methods for sentiment classification: A survey. In *Proceedings of the 11th Web Information System and Application Conference (WISA'14)* (pp. 262–265).

Hong, L., & Davison, B. (2010). Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics (SOMA '10)* (pp. 80–88).

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of international conference on knowledge discovery and data mining (KDD '04)* (pp. 168–177).

Jo, Y., & Oh, A. (2011). Aspect and sentiment unification model for online review analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining (WSDM '11)* (pp. 815–824).

Kaya, M., Fidan, G., & Toroslu, I. H. (2012). Sentiment analysis of Turkish political news. In Proceedings of WI-IAT'12 IEEE/WIC/ACM international joint conferences on web intelligence and intelligent agent. *Technology*, 174–180.

Kennedy, A., & Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence, 22*, 110–125. https://doi.org/10.1111/j.1467-8640.2006.00277.x

Liao, C., Fengn, C., Yang, S., & Huang, H. (2016). Topic-related Chinese message sentiment analysis. *Neurocomputing, 210*, 237–246. https://doi.org/10.1016/j.neucom.2016.01.110

Lin, C., & He, Y. (2009). Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management (CIKM '09)* (pp. 375–384).

Liu, B. (2012). *Sentiment analysis and opinion mining.* Morgan & Claypool Publishers.

Liu, Q., Gao, Z., Liu, B., & Zhang, Y. (2016). Automated rule selection for opinion target extraction. *Knowledge-Based Systems, 104*, 74–88. https://doi.org/10.1016/j.knosys.2016.04.010

Maharani, W., Widyantoro, D. H., & Khodra, M. L. (2015). Aspect extraction in customer reviews using syntactic pattern. *Procedia Computer Science, 59*, 244–253. https://doi.org/10.1016/j.procs.2015.07.545

Meral, M., & Diri, B. (2014). Sentiment analysis on twitter. *IEEE 22nd Signal Processing and Communications Applications Conference (SIU)*, 690–693.

Mowlaei, M. E., Abadeh, M. S., & Keshavarz, H. (2020). Aspect-based sentiment analysis using adaptive aspect-based lexicons. *Expert Systems with Applications, 148*. https://doi.org/10.1016/j.eswa.2020.113234

Nimala, K., Magesh, S., & Arasan, R. T. (2018). Hash tag based topic modelling techniques for twitter by tweet aggregation strategy. *Journal of Advanced Research in Dynamical and Control Systems, 3*, 571–578.

Oflazer, K., & Bozsahin, H. C. (1994). *Turkish natural language processing initiative. In Proceedings of the Third Turkish symposium on artificial intelligence and artificial neural networks.* Turkey: Ankara.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., … Eryigit, G. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (Semeval-2016)* (pp. 19–30).

Rezaeinia, S. M., Rahmani, R., Ghodsi, A., & Veisi, H. (2019). Sentiment analysis based on improved pre-trained word embeddings. *Expert Systems with Applications, 117*, 137–149. https://doi.org/10.1016/j.eswa.2018.08.044

Tang, F., Fu, L., Yao, B., & Xu, W. (2019). Aspect based fine-grained sentiment analysis for online reviews. *Information Sciences, 488*, 190–204. https://doi.org/10.1016/j.ins.2019.02.064

Weng, J., Lim, E., Jiang, J., & He, Q. (2010). Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining (WSDM '10)* (pp. 261–270).

Xianghua, F., Guo, L., Yanyan, G., & Zhiqiang, W. (2013). Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowledge-Based Systems, 37*, 186–195. https://doi.org/10.1016/j.knosys.2012.08.003

Xiong, S., Wang, K., Ji, D., & Wang, B. (2018). A short text sentiment-topic model for product reviews. *Neurocomputing, 297*, 94–102. https://doi.org/10.1016/j.neucom.2018.02.034

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A Biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web (WWW '13)* (pp. 1445–1456).

Zemberek NLP. (2020). http://zembereknlp.blogspot.com/. (Accessed 17 February 2020).