# Contrastive Max-Sum Opinion Summarization

Makbule Gülçin Özsoy and Ruket Çakıcı

Department of Computer Engineering
Middle East Technical University
Ankara, Turkey
{makbule.ozsoy,ruken}@ceng.metu.edu.tr

**Abstract.** People can reach all kinds of information online including reviews and comments on products, movies, holiday destinations and so on. However, one usually need to go through the reviews to have an objective opinion the positive and the negative aspects of the item reviewed. We aim to provide a method that will extract positive and negative opinions on a specific aspect and compare them in an attempt to ease on the information overflow. Contrastive opinion summarization (COS) aims to solve this issue. COS methods extract representative and comparative sentences in terms of specific aspects of a product. In this paper, we propose a new COS method, namely Contrastive Max-Sum Opinion Summarization (CMSOS). This method considers representativeness and contrastiveness at the same time. For the evaluation, we use an English dataset which was specifically created for COS studies. In addition, we created a new dataset in Turkish and shared it publicly. We provide the results on both datasets with our method.

**Keywords:** Contrastive opinion summarization, Representativeness, Contrastiveness

## 1 Introduction

As the world wide web is becoming a more reliable and more easily available source of information people comment on products the had purchased, movies and locations they had seen, or they read the reviews such as these before they buy something. Today, a very big amount of opinionated text is available for the users. However, finding out the pros and cons of a product itself or a specific aspect of it is not trivial without having to read all the reviews to have an idea. A system which can extract positive and negative opinions on different aspects of a product and compare them is necessary.

Extraction of comparative sentences with contrasting opinions is a recently proposed problem in the literature. Kim and Zhai proposed this problem and named it as *contrastive opinion summarization (COS)* in [8] . In this study, the opinion labels of the sentences are given beforehand, and the system tries to select the most representative and comparable sentences accordingly.

In this paper, a new COS algorithm is proposed and evaluated. A new dataset in Turkish is created and the evaluation results for this new dataset are presented here. Our contributions are as follows: (1) A new algorithm, namely Contrastive Max-Sum Opinion Summarization (CMSOS), which extracts the most representative and comparable opinionated text in parallel. This algorithm gives better results on the dataset of [8] than the ones they report. (2) To our knowledge, there is no publicly available dataset in Turkish that gives aspects of items and opinion of users on these items. We created a new dataset that contains reviews and ratings of the users on movies and aspects that exists in the reviews[1]. (3) We also report the results of applying CMSOS to the newly created Turkish dataset in this paper. We conjecture that these results will aid in providing preliminary results for further studies.

In Section 2, information on related work is given. In Section 3, the proposed algorithm, namely Contrastive Max-Sum Opinion Summarization (CMSOS), is explained. In Sections 4 and 5 information on the datasets and evaluation results are given. Finally, Section 6 gives the conclusion.

## 2 Related Work

Opinion summarization techniques aim to find topics and classify the opinion attached to the topics. They sometimes do this taking into account different aspects of the topic in question. [4], [10] and [12] are the representative works on opinion summarization, that apply data mining and heuristic methods. Heuristics used by these methods vary from word frequencies and noun phrase statistics to information from WordNet. However, these methods do not extract comparative sentences and/or documents.

There are studies that aim to collect comparative sentences, such as [5], [11] and [15]. They detect contradiction in text and/or detect comparative sentences in the documents. They take several different approaches, such as searching words in the documents such as "than" [5], binary classification methods such as SVM [15] and graph representation based methods [11]. Although these methods aim to extract the comparative sentences, they do not aim to find the most representative chunks of text or comparative summaries.

Contrastive opinion summarization (COS) methods aim to collect not only the most representative sentences but also the contrastive ones from the input. This problem is introduced in the literature by [8]. They define COS as follows: "Given two sets of positively and negatively opinionated sentences which are often the output of an existing opinion summarizer, COS aims to extract comparable sentences from each set of opinions and generate a comparative summary containing a set of contrastive sentence pairs." They formulate the problem as an optimization problem and propose two methods each of which focuses on these two issues –representativeness and contrastiveness– in different orders, namely Representativeness First (RF) or Contrastiveness First (CF).

In RF approach, firstly, the input positive and negative sentence sets ($X$ and $Y$ respectively) are divided into sub-clusters using the representativeness measure. Then, the sentences from each cluster in $X$ and $Y$ are aligned using the representativeness and contrastiveness measures together. Even if multiple sentences in the same cluster have high scores, this approach collects a single sentence from each cluster. In CF approach,the contrastiveness of the sentences in $X$ and $Y$ are calculated first. After that, the pair with the highest contrastiveness is chosen as the first pair of the comparative summary. The other pairs are selected depending on the previously chosen pairs and

[1] The dataset will be made available to download at the following link. http://www.ceng.metu.edu.tr/~e1395383/papers/airs2014/

representativeness and contrastiveness metrics. In this approach, if the first pair is not the optimal choice, the quality of the output comparative summary will be lower.

Paul et. al (2010) also aim to create contrastive summaries from the input set in [13]. They perform a two-stage approach. In the first step they extract multiple viewpoints from the text, and in the second stage they collect both representative and contrastive sentences. For the second stage, they propose Comparative LexRank method, which is based on random walk.

Lerman and McDonald (2009) also name their problem as contrastive summarization in [9]. However, their solution is different than the previous approaches. While previous studies aim to extract different viewpoints of a single product/aspect, this work aims to generate two different summaries on a pair of products. With this approach, the summaries highlight the differences of the products.

[14] focuses on constructing short and comparative summaries of product reviews. Unlike COS methods, they aim to compare two selected products. They produce aligned pairs of sentences related to products on different aspects. The aspects are also selected automatically by their method.

The literature related to opinion summarization in Turkish is limited as a result of the lack of datasets publicly available. Sentiment and/or emotion analysis can be considered as related to opinion summarization. The few works focused on sentiment and emotion analysis in Turkish belong to [1], [2] and [7]. None of these works focus on different aspects of the items and compare the sentiment results on aspects. A very recent work conducted by [6] provides results for feature based summarization of product reviews in Turkish. This study focuses on producing personalized review summaries on multiple products. Unlike our work, they do not return a list of comparative sentences on different aspects of a single product. They conduct their evaluation on a Turkish dataset that they created. Unfortunately, the dataset is not available publicly, to our knowledge.

## 3 Contrastive Max-Sum Opinion Summarization

Traditional opinion summarization techniques aim to select a set of the most representative opinionated text from an input set. Unlike them, contrastive opinion summarization (COS) methods aim to collect not only the most representative sentences but also the ones that have contrastive meaning. This helps users to become aware of the different opinions on different aspects of a chosen item/topic in order to have a better idea by making use of the created comparative summaries.

In this paper, we propose a COS method, namely Contrastive Max-Sum Opinion Summarization (CMSOS). The method creates a list of pairs of the most representative sentences related to a given aspect. Each pair contains a positive sentence and a negative sentence, that have contrastive meaning. For instance, assume that a user wants to find out what other users think about the design of a certain product. The system returns the pair of sentences *"it did an awesome job with the design."* as a positive sentence and *"but my biggest gripe is still the extremely ugly design."* as a negative one[2].

Contrastive Max-Sum Opinion Summarization (CMSOS) is adopted from Max-Sum Diversification method proposed in [3]. This algorithm is proposed as a solution to

[2] These example sentences are from the English dataset.

web-search and aims to get the most relevant and novel document from the input set. In CMSOS, our aim is to get most representative and contrastive sentences. We used document similarity between sentences with the same labels (content similarity) to find out the representative sentences and document similarity between sentences with different labels (contrastive similarity) to find out the contrastive sentences.

The equation of CMSOS is written as in Eq.1. In the equation, *S* is the sentence set, *u* and *v* are the sentences in this set, *w(u)* is the representativeness of the sentence *u*, *d(u,v)* represents the contrast between sentences *u* and *v*, and $\lambda$ is the parameter used for setting the trade-off between relevance and similarity. The aim is to maximize *f(S)* in Eq.1.

$$f(S) = \sum_{u,v \in S} (w(u) + w(v) + 2\lambda d(u,v)) \tag{1}$$

We calculate *w(u)* using cosine similarity, Eq.2, which sums the similarity of the sentences to the other sentences and normalizes the result. We apply two different approaches to decide the cosine similarity. The first one is based on word frequencies. In Eq.3, *W(u,v)* contains the common words in the sentences *u* and *v*. The *tf(u,i)* represents the frequency of the word *i* in the sentence *u*. The second measurement is based on tf-idf values of the words in the sentence as shown in Eq.4. The only difference from the first version is that here we use the tf-idf values. Tf-idf is used commonly in information retrieval and text mining to reflect the importance of the word. The tf-idf value increases as the frequency of the word in the input sentence is high while the frequency of the word in the input dataset is low. We calculated the tf and the idf values using the input data. We observe that the input datasets contain repetitive sentences, so we applied the tf-idf version to observe the effect of these sentences on the performance.

$$w(u) = \frac{\sum_{v \in S \wedge u \neq v} cosSim(u,v)}{S-1} \tag{2}$$

$$cosSim_1(u,v) = \frac{\sum\limits_{i \in W(u,v)} tf(u,i)tf(v,i)}{\sqrt{\sum\limits_{i \in W(u,v)} tf(u,i)^2}\sqrt{\sum\limits_{i \in W(u,v)} tf(v,i)^2}} \tag{3}$$

$$cosSim_2(u,v) = \frac{\sum\limits_{i \in W(u,v)} tfIdf(u,i)tfIdf(v,i)}{\sqrt{\sum\limits_{i \in W(u,v)} tfIdf(u,i)^2}\sqrt{\sum\limits_{i \in W(u,v)} tfIdf(v,i)^2}} \tag{4}$$

In order to measure contrastive similarity we first remove the adjectives, that create the contrast. After the removal of adjectives, the cosine similarity between the sentences with different labels are calculated. The idea of the removal of sentiment related words is proposed in [8] in which sentiment related words are defined as negation words and adjectives. In Turkish, the negation is usually constructed by attaching the negation suffix to a verb or using the negation particle. In order to find out if a verb is in the negative form, one needs to apply morphological analysis. In this study, we did not apply any morphological analysis and considered it as a future work.

## 4 Evaluation Settings

### 4.1 Metrics

We use precision and aspect coverage as evaluation metrics, as suggested by [8]. Kim and Zhai in [8] explain that precision represents the contrastiveness of the sentence pairs and the aspect coverage indicates the representativeness of the summary. In all the calculations the responses of each human labeler are analysed separately before reporting the average value for the overall result. For some of the products, the human labelers labeled the sentences with different aspects (some aspects are not common). The uncommon aspects are combined under a new aspect with the name "other".

The precision is calculated by using Eq.5. In this equation, *#agreedPairs* is the number of times the found pairs and the human labels match and $k$ is the total number of selected pairs. The $k$ value is calculated by Eq.6, where $|X|$ and $|Y|$ are the number of positive or negative labeled sentences.

$$Precision = \frac{\text{\#agreedPairs}}{k} \tag{5}$$

$$k = 1 + \log_2(|X| + |Y|) \tag{6}$$

The aspect coverage is calculated by Eq.7. In this equation, the number of unique aspects collected in the summary is divided by the number of unique aspects labeled by human labelers.

$$AspectCov = \frac{\text{\#uniqueAspects(Summary)}}{\text{\#uniqueAspects(LabelledManually)}} \tag{7}$$

### 4.2 Datasets

For evaluation, two datasets will be used: The dataset provided in [8][3] for English, and the newly created Turkish dataset[4] . In the following subsections, information on these datasets are given.

**English Dataset** The English dataset[8] contains reviews on 13 different <product,aspect> sets, 12 of which are collected from Amazon website and 1 of which is a non-product-review text and is about aspartame. In the dataset, the sentences and their polarities are manually labeled by two human labelers. Additionally, a non-product-review is included to show the generality of their method. For this purpose they collected 50 positive and 50 negative sentences about Aspartame using Yahoo! search engine. The number of positive and negative sentences for each <product, aspect> tuple is given in Table 1.

[3] http://sifaka.cs.uiuc.edu/ir/data/cos/

[4] http://www.ceng.metu.edu.tr/~e1395383/papers/airs2014/

| Id | Product:Aspect Name | #Pos | #Neg |
|---|---|---|---|
| 1 | Apex AD2600 Progressive-scan DVD player:player | 44 | 56 |
| 2 | MicroMP3:batterylife | 9 | 7 |
| 3 | MicroMP3:design | 8 | 6 |
| 4 | MicroMP3:headphones | 7 | 6 |
| 5 | MicroMP3:software | 7 | 9 |
| 6 | Nokia 6600:battery-life | 7 | 8 |
| 7 | Creative Labs Nomad Jukebox Zen Xtra 40GB:navigation | 9 | 8 |
| 8 | Creative Labs Nomad Jukebox Zen Xtra 40GB:software | 37 | 41 |
| 9 | Creative Labs Nomad Jukebox Zen Xtra 40GB:size | 15 | 11 |
| 10 | Creative Labs Nomad Jukebox Zen Xtra 40GB:weight | 7 | 7 |
| 11 | Creative Labs Nomad Jukebox Zen Xtra 40GB:transfer | 9 | 7 |
| 12 | Hitachi router:adjustment | 7 | 6 |
| 13 | aspartame:safety | 50 | 50 |

**Table 1.** Number of positive and negative sentences for the English dataset

**Turkish Dataset** To our knowledge, there is no publicly available dataset in Turkish with explicit information on the aspects of items, and the opinions and ratings of users on these items. We created a new dataset that contains reviews and ratings of the users on movies and on specific aspects of each movie that exist in the reviews.

Firstly, we collected the reviews and ratings of the users for the movies from a website, namely beyazperde[5]. On this website, users are asked to write reviews for movies and rate them in the range [1,5] with increments of 0.5. We considered the reviews with 3.5 or more points as positive and the rest as negative. We queried 1000 pages and collected only 107 of them since only these contain any review information.

Secondly, we asked two human labelers to label the reviews in terms of aspects. We created a form for the labelers with six different aspects; namely scenario, acting, music, visuals, director and general. The labelers are also allowed to write down their own aspects if they find any significant ones. Each labeler is assigned to different movies, so that the reviews of each movie are labeled by a single labeler. We could not use multiple labelers for each movie, since the time and the number of the labelers were limited.

Lastly, we removed movies whose positive or negative sentence counts were less than $k$, given in Eq.6. At the end, we obtained 31 movies with aspects, reviews and ratings in Turkish. The numbers of positive and negative sentences for each movie in this dataset are given in Table 2.

We share this dataset on our web-site and we conjecture that this new dataset will be useful to many researchers who needs data that contains aspect, rating and/or review information in Turkish.

## 5 Evaluation Results

We report results for cosine similarity and tf-idf usage on precision and aspect coverage metrics for both datasets. In Eq. 1, $\lambda$ is used for setting the trade-off between relevance

[5] www.beyazperde.com

| Id | Movie Name | #Pos | #Neg | Id | Movie Name | #Pos | #Neg |
|---|---|---|---|---|---|---|---|
| 548 | Yedinci Mühür | 27 | 5 | 263 | Paris, Texas | 15 | 4 |
| 260 | Otomatik Portakal | 143 | 32 | 29 | Serseri Aşıklar | 16 | 8 |
| 305 | Stranger Than Paradise | 4 | 3 | 309 | Yokedici | 61 | 7 |
| 308 | Taksi Şoförü | 91 | 33 | 290 | Siyam Balığı | 26 | 5 |
| 297 | Yedi Samuray | 39 | 6 | 62 | Yaratık | 47 | 5 |
| 339 | Gremlinler | 22 | 5 | 337 | Sekiz Buçuk | 15 | 6 |
| 88 | Annie Hall | 25 | 8 | 363 | Can Dostum | 105 | 8 |
| 140 | Birdy | 20 | 5 | 142 | Brazil | 16 | 8 |
| 151 | Köpekler | 10 | 5 | 437 | Hayalet Avcıları | 19 | 6 |
| 188 | Şeytanın Ölüsü | 49 | 22 | 176 | Dune | 7 | 6 |
| 183 | Eraserhead | 21 | 12 | 180 | Fil Adam | 47 | 7 |
| 467 | Enter The Dragon | 15 | 4 | 462 | Korku Burnu | 41 | 11 |
| 448 | Geleceğe Dönüş | 162 | 9 | 505 | 13 | 20 | 10 |
| 226 | Lolita | 14 | 6 | 255 | Nostalji | 5 | 4 |
| 253 | New York'tan Kaçış | 10 | 5 | 248 | Monty Python and the Holy Grail | 9 | 7 |
| 243 | Geceyarısı Ekspresi | 21 | 39 | | | | |

**Table 2.** Number of positive and negative sentences for the Turkish dataset

and contrast, such that larger $\lambda$ gives more importance to contrastiveness. In the experiments $\lambda$ is set between 0 and 1 with 0.1 increments and the results for the datasets are presented in the following section.

### 5.1 Results for English Dataset

The precision and the aspect coverage results of CMSOS for the English dataset are given in Table 3 and Table 4. These values are obtained by taking the average of the results obtained for all <product,aspect> tuples.

| lambda | Cosine similarity tf | Cosine similarity tf-idf |
|---|---|---|
| 0.00 | 0.576 | 0.544 |
| 0.10 | 0.572 | 0.558 |
| 0.20 | 0.588 | 0.517 |
| 0.30 | 0.582 | 0.562 |
| 0.40 | 0.624 | 0.562 |
| 0.50 | 0.630 | 0.562 |
| 0.60 | 0.630 | 0.574 |
| 0.70 | **0.649** | 0.574 |
| 0.80 | **0.649** | **0.594** |
| 0.90 | 0.630 | **0.594** |
| 1.00 | 0.630 | 0.584 |

**Table 3.** Results on precision of CMSOS for English dataset

| lambda | Cosine similarity tf | Cosine similarity tf-idf |
|---|---|---|
| 0.00 | 0.910 | **0.960** |
| 0.10 | 0.901 | 0.877 |
| 0.20 | 0.918 | 0.935 |
| 0.30 | 0.910 | 0.935 |
| 0.40 | **0.916** | 0.935 |
| 0.50 | 0.897 | 0.928 |
| 0.60 | 0.897 | 0.928 |
| 0.70 | 0.897 | 0.941 |
| 0.80 | 0.887 | 0.941 |
| 0.90 | 0.895 | 0.941 |
| 1.00 | 0.895 | 0.941 |

**Table 4.** Results on aspect coverage of CMSOS for English dataset

The results in Table 3 show that in terms of precision, cosine similarity with tf performs better than cosine similarity with tf-idf. Actually cosine similarity with tf-idf is better for inputs where there are many repeated sentences. For example, we notice that the subject <aspartame, safety> includes sentences "asparteme is safe" and "asparteme is dangerous" several times. While using cosine similarity with tf, we observe that the resulting summary contains these sentences frequently. But when we use cosine similarity with tf-idf, we obtain better results which are not repetitive. The results obtained by our method using cosine similarity with tf or cosine similarity with tf-idf are given in Table 5 and Table 6 [6]. As expected, tf-idf usage inhibits the effects of repetitive terms, and can produce more informative summaries.

Table 4 shows that cosine similarity with tf and cosine similarity with tf-idf performs similarly as lambda value gets larger for aspect coverage. Aspect coverage results stay nearly balanced around 0.90-0.95 for different $\lambda$ values.

| **Id(+)** | **Sent.(+)** | **Id(−)** | **Sent.(−)** |
|---|---|---|---|
| 4 | aspartame is safe | 4 | aspartame is dangerous |
| 21 | that aspartame is safe | 20 | that aspartame is dangerous |
| 20 | aspartame is safe | 2 | aspartame is dangerous |
| 25 | aspartame is safe | 44 | - aspartame is dangerous |

**Table 5.** Outputs for <aspartame, safety> with cosine similarity

We compare the results of our method to the results reported in [8], in Table 7. Two different methods proposed in [8], namely Representativeness First (RF) and Contrastiveness First (CF), are given with labels “Kim et al-RF” and “Kim et al-CF” respectively in the table. We also show the results of cosine similarity with tf-idf and cosine

[6] We removed punctuations and showed only the first four results

| **Id(+)** | **Sent.(+)** | **Id(−)** | **Sent.(−)** |
|---|---|---|---|
| 21 | that aspartame is safe | 20 | that aspartame is dangerous |
| 37 | conclusively determined that aspartame is safe | 12 | conclusively that aspartame is dangerous |
| 29 | aspartame is safe - scientists support aspartame safety | 44 | - aspartame is dangerous |
| 46 | can we tell if aspartame is safe | 32 | i can tell you from personal experience that aspartame is dangerous |

**Table 6.** Outputs for <aspartame, safety> with tf-idf

similarity with tf results in the table. CMSOS configurations generally give the best results for precision and aspect coverage on this dataset. In CF if the initially chosen pair of sentences is not optimal this affects the quality of the output summary badly. CMSOS optimizes for representativeness and contrastiveness at the same time. It is able to choose multiple sentences from an aspect and it is not affected by any previously chosen sentence pairs.

| **Method** | **Prec.** | **AspectCov.** |
|---|---|---|
| Kim et al-RF | 0.503 | 0.737 |
| Kim et al-CF | 0.537 | 0.804 |
| CMSOS(tf & $\lambda$=0.80) | **0.649** | 0.887 |
| CMSOS(tf-idf & $\lambda$=0.80) | 0.594 | **0.941** |
| CMSOS(tf & $\lambda$=0.10) | 0.572 | 0.901 |
| CMSOS(tf-idf & $\lambda$=0.10) | 0.558 | 0.887 |

**Table 7.** Comparison of results

Our method gives the same priority to representativeness and contrastiveness, and it performs better than the methods that give precedence to one over the other. Cosine similarity with tf usage generally performs better than cosine similarity with tf-idf. However, for data with repeating sentences tf-idf usage is more helpful.

### 5.2 Results for Turkish Dataset

The average precision and aspect coverage results obtained for the Turkish dataset are given in Table 8 and Table 9, respectively.

In terms of precision, cosine similarity with tf and tf-idf perform similarly. The best precision result is 0.927 and 0.917 for tf and tf-idf metrics respectively. The best aspect coverage values are 0.928 and 0.901 for cosine similarity with tf and tf-idf metrics respectively.

Comparison of English and Turkish results show that the CMSOS algorithm performs equally well for both datasets, in terms of aspect coverage. However, in terms of

| lambda | Cosine similarity tf | Cosine similarity tf-idf |
|---|---|---|
| 0.00 | 0.913 | 0.862 |
| 0.10 | **0.927** | 0.875 |
| 0.20 | 0.890 | 0.894 |
| 0.30 | 0.924 | 0.896 |
| 0.40 | 0.911 | 0.891 |
| 0.50 | 0.911 | 0.908 |
| 0.60 | 0.904 | 0.916 |
| 0.70 | 0.899 | 0.905 |
| 0.80 | 0.893 | 0.905 |
| 0.90 | 0.893 | 0.905 |
| 1.00 | 0.893 | **0.917** |

**Table 8.** Results on precision of CMSOS for Turkish dataset

| lambda | Cosine similarity tf | Cosine similarity tf-idf |
|---|---|---|
| 0.00 | 0.923 | 0.899 |
| 0.10 | **0.928** | 0.896 |
| 0.20 | 0.923 | **0.901** |
| 0.30 | 0.919 | **0.901** |
| 0.40 | 0.927 | **0.901** |
| 0.50 | 0.926 | **0.901** |
| 0.60 | 0.926 | **0.901** |
| 0.70 | 0.926 | **0.901** |
| 0.80 | 0.918 | 0.891 |
| 0.90 | 0.926 | 0.891 |
| 1.00 | 0.926 | 0.891 |

**Table 9.** Results on aspect coverage of CMSOS for Turkish dataset

precision the algorithm performs worse for the English dataset (about 0.650) than the Turkish dataset (about 0.920). This is the result of difference between the number of aspects labeled by users. The average number of aspects for the items in English dataset is about 2.80, whereas, it is about 4.45 for the Turkish dataset. Therefore, it is easier to find an agreed pair for Turkish set than the English set.

We present the first results on the Turkish dataset. We conjecture that these results will be useful preliminary results for further research on this problem.

## 6 Conclusion

Traditional opinion summarization techniques do not output contrasting sentences as a feature. In order to deal with this problem a new kind of opinion summarization problem, namely contrastive opinion summarization (COS), is introduced in the literature.

In this paper, we presented a new COS method, namely Contrastive Max-Sum Opinion Summarization(CMSOS), for this purpose. We considered representativeness and contrastiveness at the same time, and applied cosine similarity with tf and cosine similarity with tf-idf measures.

For the evaluation, we used a known English dataset, and we compared our results to the ones of [8], who have the initial results on this dataset. We obtained better results than their results. We observed that using cosine similarity with tf for calculations performed better than tf-idf usage. However, for data with a lot of repeated sentences we conjecture that tf-idf usage is more helpful. In addition, we created a new Turkish dataset for the COS purposes, which can also be used as data for further research that needs rating, review and/or aspect labeled data in Turkish. We evaluated the CMSOS method on the new Turkish dataset and reported the results in this paper.

Future work includes plans for increasing the size of the Turkish dataset, making the evaluation step automatic by automatic extraction of aspect names, and also applying CMSOS to the multiple product review summarization problem.

## References


1. Boynukalın, Z., Karagöz, P.: Emotion analysis on Turkish texts. In: Gelenbe, E., Lent, R. (eds.) Information Sciences and Systems 2013, Lecture Notes in Electrical Engineering, vol. 264, pp. 159–168. Springer International Publishing (2013)
2. Eroğul, U.: Sentiment Analysis in Turkish. Master's thesis, Middle East Technical University, Turkey (2009)
3. Gollapudi, S., Sharma, A.: An axiomatic approach for result diversification. In: Proceedings of the 18th International Conference on World Wide Web. pp. 381–390. WWW '09, ACM, New York, NY, USA (2009)
4. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 168–177. KDD '04, ACM, New York, NY, USA (2004)
5. Jindal, N., Liu, B.: Identifying comparative sentences in text documents. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 244–251. SIGIR '06, ACM, New York, NY, USA (2006)
6. Kavasoğlu, Z., Gündüz Öğüdücü, S.: Personalized summarization of customer reviews based on user's browsing history. International Journal on Computer Science and Information Systems 8, 147–158 (2013)
7. Kaya, M., Fidan, G., Toroslu, H.: Transfer learning using twitter data for improving sentiment classification of Turkish political news. In: Gelenbe, E., Lent, R. (eds.) Information Sciences and Systems 2013, Lecture Notes in Electrical Engineering, vol. 264, pp. 139–148. Springer International Publishing (2013)
8. Kim, H.D., Zhai, C.: Generating comparative summaries of contradictory opinions in text. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management. pp. 385–394. CIKM '09, ACM, New York, NY, USA (2009)
9. Lerman, K., McDonald, R.: Contrastive summarization: An experiment with consumer reviews. In: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers. pp. 113–116. NAACL-Short '09, Association for Computational Linguistics, Stroudsburg, PA, USA (2009)

10. Liu, B., Hu, M., Cheng, J.: Opinion observer: Analyzing and comparing opinions on the web. In: Proceedings of the 14th International Conference on World Wide Web. pp. 342–351. WWW '05, ACM, New York, NY, USA (2005)
11. Malouf, R., Mullen, T.: Taking sides: User classification for informal online political discourse. Internet Research (2008)
12. Mei, Q., Ling, X., Wondra, M., Su, H., Zhai, C.: Topic sentiment mixture: Modeling facets and opinions in weblogs. In: Proceedings of the 16th International Conference on World Wide Web. pp. 171–180. WWW '07, ACM, New York, NY, USA (2007)
13. Paul, M.J., Zhai, C., Girju, R.: Summarizing contrastive viewpoints in opinionated text. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. pp. 66–76. EMNLP '10, Association for Computational Linguistics, Stroudsburg, PA, USA (2010)
14. Sipos, R., Joachims, T.: Generating comparative summaries from reviews. In: CIKM. pp. 1853–1856 (2013)
15. Thomas, M., Pang, B., Lee, L.: Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. pp. 327–335. EMNLP '06, Association for Computational Linguistics, Stroudsburg, PA, USA (2006)