CRF-based Arabic Opinion Summarization System

Imen touati, Marwa Graja, Mariem Ellouze, and Lamia Hadrich Belguith

Miracl Laboratory, Arabic Natural Language Processing research Group (ANLP-RG),
University of Sfax, Tunisia
{imen_touati@yahoo.fr,marwa.graja@fsegs.rnu.tn,
mariem.ellouze@planet.tn,l.belguith@fsegs.rnu.tn}

Abstract. This paper presents the study that we have carried out to investigate supervised opinion summarization in Modern Standard Arabic. We use a corpus of news articles. We use conditional random fields (CRF) as machine learning technique. We investigate some features to identify those that allow achieving the best results. Our contribution is to use opinion specific features to summarize Arabic news articles using CRF models. Experimental results show that our proposed approach is very effective for assigning features to sentences.

Keywords: Arabic news article, opinion analysis, arabic opinion summarization

1 Introduction

Opinion summarization task is the overlap of two important fields : opinion mining and text summarization. Opinion mining or interchangeably called sentiment analysis aims to the analysis of people's opinions, judgments, evaluations about a specific entities, individuals, events or topics. Text summarization is an old field since fifties [17] that aims at the extraction of key sentences from a document. Recently, huge corpus have appeared with the growth of Internet. To obtain a comprehensive understanding of detailed opinions in the massive number of product reviews, blogs, news articles, etc., many studies on document summarization of evaluative text, on review mining and summarization have risen ([5],[6],[10]). With opinion summarization, the goal is no longer to produce only a summary of informations in the text, but it's necessary to determine expressed opinions with their semantic orientation (positive, negative) and more narrowly with their semantic categorization [22].

Opinion summarization is challenging and very useful task. Text Analysis Conference (TAC 2008) has dedicated a pilot task to opinion summarization where it's asked to write summaries of opinions from blogs.

Traditional summarization techniques focus on identifying a documents main topics, removing redundancies, and ordering the extracted sentences [11].

In our current work, we address the problem of opinion summarization by considering the creation of simple opinion summaries. Our contribution consists of seeking a more precise definition of the features that can be used effectively in the automatic extraction of opinion summaries.

The remainder of this paper is organized as follows. In section 2, we focus on the most closely related studies on opinion summarization. In section 3, we present experiments performed for detecting opinionated sentences. Finally, conclusion is presented in the last section.

2   Opinion Summarization

In literature, there are different studies with different definitions of what an opinion summary should be. In general, opinion summary may have different forms : a single paragraph, a structured sentence, attribute-value pairs or just consist at an overall score of sentiment conveyed in a document.

To resolve this problem, researchers have studied approaches for automatically summarizing or analyzing opinions expressed in the review data ([15],[20]). [12] classifies existing approaches under two main categories: aspect oriented summarization and non-aspect oriented summarization. Generally, reviews have been the focus of the majority of researches in the sentiment summarization field.

Most existing work in the field of opinion or sentiment summarization is under feature based or aspect based summarization technique umbrella. The key idea of this technique is to identify the features of a product and opinion sentences towards each feature. [16] defines a set of user questions to summarize an English review in order to help customers who want to quickly capture the main idea of a lengthy product review before they read the details. They treat the problem of aligning questions to a review as text summarization problem, with the goal of finding relevant and non-redundant questions for a review.

Other works create a textual sentiment summary based on extraction of relevant sentences. The work of [2] consists of selection of a single passage that reflect the opinion of the document's author. While [19] proposes tracking the sentiment flow within the document to create sentiment summary. So they suggest to choose the sentences at local extrema of the flow (plus the first and last sentence). Other studies are influenced by information extraction methods. They propose to view summary representation as template ([4],[8]). [23] proposes to select a set of most representative review sentences for the nominal features of each product.

[21] are interested by summarizing multiple contrastive viewpoints in opinionated text. The work of [7] defines a novel task of generating entity comparisons from textual corpora in which each document describes one entity at a time. [13] summarizes reviews by choosing complementary reviews and ranking them according to different strategies.

Several interesting and advanced works were performed on English. In contrast and to the best of our knowledge; no work was done about opinion sum-

marization for Arabic language, although that Arabic document summarization is quite a hot topic in the Arabic research community due to its utility for many tasks for NLP ([1],[3]).
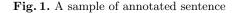
3 CRF-based Arabic Opinion Summarization

3.1 Corpus

The corpus used to perform experiments is a set of news articles from Arabic TreeBank (ATB part3 v3.2) [18] and from some sites of news channels like "Al-jazeera.Net", "BBC Arabic" and "France 24 Arabic". In each article, relevant sentences was annotated manually. Each article presents a set of opinions expressed by different holders about a topic from a political domain. A holder may be a person, an organization, a country, political party, etc. The topic may be a political event, a political person, etc.

Differently from [24], we annotate targets with specification of opinion expressions that are linked to them and their type. Each labeled target has a type (main-topic, part-of-topic and Other-topic). Our corpus has gone through a semantic opinion expression annotation [22].

```
<seg texte="روسيا تشكك بفرص وقف إطلاق النار بسوريا" id="1">
    <token id="1" texte="روسيا" role="Holder" />
    <token category="judgement" type="contest" valence="-1" holderid="1" topicid="1"
    lemma="شك" texte="تشكك" role="ExpOp" />
    <token type="main_topic" id="1" texte="بفرص وقف إطلاق النار بسوريا" role="Topic" />
</seg>
```

Fig. 1. A sample of annotated sentence

3.2 Problem definition

Standard formulation of the problem of opinion summarization assumes a document D composed of a set of sentences $D = \{x_1, ..., x_N\}$ which contains opinions about a specific topic T. The objective is to generate a summary S of the opinions expressed in the document D about T. Each opinion is indicated by an opinion expression, by a holder (the source of the opinion) about a topic T. In this work, we look at an extractive summarization settings where S is built by extracting the most important opinion sentences about the main topic from the document D. We assume that D is a set of candidate sentences of our summary.

3.3 CRF definition

In order to investigate the opinion summarization task in Arabic news articles, we applied a machine learning process, based on the Conditional Random Fields

(CRF) models [14]. CRF, as sequential discriminative probabilistic model, has proved its efficiency in various Natural Language Processing applications, such as named entity identification and morphological tagging. It is also used for many tasks of opinion mining in English, Chinese texts. For Arabic language, CRF has been adopted by [9] for opinion holder extraction.

We can address the problem of opinion summarization as a sequential classification problem where we estimate the conditional probability of a sequence of output values (the class of each lexical unit) $S = y_1...y_N$ given an input sequence (observations) $D = x_1...x_N$. Then,the conditional probability $p(y|x)$ for linear chain CRFs is given as [14]:

$$p(S|D) = \frac{1}{Z(D)} exp(\sum_j \lambda_j t_j(y_{(i-1)}, y_i, D, i) + \sum_k \mu_k s_k(y_i, D, i)) \qquad (1)$$

where Z_D is the normalization factor.

Our implementation of Conditional Random Fields is based on the CRF++ tool[1]. It's used by [14] for sequence labeling classification.

3.4   Features

Detecting sentences which contain an expression of an opinion by a holder about a main-topic of the news articles, is a challenge in Arabic opinion mining. This affects the selection of training features for the considered task. Therefore, we propose, for training, to use a set of opinion specific features as follows:

- **Token**: this feature represents the string of the current token as a feature. This feature introduces lexical information about the domain. We will refer to this feature as **Tok** in the templates table.
- **Opinion Expression**: this feature indicates the existence of an opinion expression in the considered sentence or not. We will refer to this feature as **OpExp** in the templates table.
- **Holder**: this feature is used when there is a holder who expresses an opinion in the considered sentence or not. We will refer to this feature as **Hold** in the templates table.
- **Target**: this feature indicates the existence of span of text representing the target of the opinion conveyed in the considered sentence or not. We will refer to this feature as **Targ** in the templates table.
- **main-topic**: this feature indicates if the target about which is expressed the opinion in the sentence is the main-topic of the news article or not. We will refer to this feature as **maintop** in the templates table.
- **N-gram**: this feature represents bi- and tri-gram expression. We will refer to this feature as **bi-** or **tri-** in the templates table.
- **Tokens in context**: this feature consists in the words preceding and following the considered one which forming a window with variable size (1 and

---

[1] https://taku910.github.io/crfpp/

2). To determine the best window size, we performed a set of experiments with different window size on our data. We will refer to this feature as **+1** or **+2** to designate tokens preceding or following the current token in the templates table.

Table 1. Different Features combinations

Template0	Tok
Template1	Tok+(+1)+(-1)
Template2	Tok+(+1)+(-1)+(bi-)
Template3	Tok+(+1)+(-1)+(bi-)
Template4	Tok+OpExp
Template5	Tok+OpExp+Hold
Template6	Tok+OpExp+Hold+Targ
Template7	Tok+OpExp+Hold+Targ
Template8	Tok+OpExp+Hold+Targ+(+1)+(-1)
Template9	Tok+OpExp+Hold+Targ+maintop
Template10	Tok+OpExp+Hold+Targ+maintop+(+1)+(-1)

3.5   Experiments

Usually, the evaluation process consists in comparing the result file of the test step with a carefully annotated file. We carried out an evaluation of our proposed system in terms of three evaluation metrics: precision (P) , recall (R) and F-measure metrics. The precision value evaluates the noise of a system while recall value evaluates its coverage. These metrics are often combined using the well-known weighted harmonic F-measure. As evaluation of our system to run the task, we have to verify if the system comes to guess all the sequence of words that are composing the opinionated sentences or not. All reported experiments in this work are performing using simple validation (Table 2). For all templates, the best performance is in bold.

3.6   Discussions

We have carried out standard training and evaluation. We find that main-topic feature have a key role in selecting opinionated sentences for the summary, since we obtained the best result (F-measure 90.77 %) with Template 9. Experiments show that using bigram feature made considerable increase in comparison with Template0. But forming bigram by the words following or preceding the considered one as done respectively in Template2 and Template3. However, the inclusion of opinion expression feature in Template4 has low F-measure compared to Template0.

The evaluation shows that such features encourage the inclusion of sentences in the summary that preserves the overall opinion distribution expressed across

Table 2. Simple Validation

Template	Precision(%)	Recall(%)	F-measure(%)
0	60.12	32.62	42.29
1	95.48	44.95	61.13
2	95.55	45.74	61.86
3	95.33	45.74	61.82
4	98.77	27.13	42.57
5	97.84	76.23	85.70
6	97.85	76.57	**85.91**
7	97.95	74.89	84.88
8	100	69.84	82.24
9	96.82	85.43	**90.77**
10	100	80.16	**88.98**

the original document. We conclude that the proposed new features offers improvements over traditional summarization features of opinionated text.

4   Conclusion

We have studied summarization in the field of sentiment analysis with the objective of producing opinion summaries in standard Arabic. Experiments show that our study focuses on the problem of automatically extracting opinionated sentences from Arabic news article in order to form a summary of evoked opinions. After determining opinion words, their holders, their targets, the main-topic, our summarization system, based on CRF models, generates an easily readable summary for the considered news article.

References

1. Al-Saleh, A.B., Menai, M.E.B.: Automatic Arabic text summarization: a survey. Artif Intell Rev. vol. 45, 203–234 (2016)
2. Beineke, P., Hastie, T., Manning, C., Vaithyanathan, S.: An exploration of sentiment summarization. In: Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications, Stanford, US, (2004)
3. Belguith, L., Ellouze, M., Maaloul, M., Jaoua, M., Jaoua, F., Blache, P. : Automatic summarization. In: Zitouni I (ed) Natural language processing of semitic languages, theory and applications of natural language processing. pp 371408. Springer, Berlin (2014)
4. Cardie, C., Wiebe, J., Wilson, T., Litman, D.: Combining Low-Level and Summary Representations of Opinions for Multi-Perspective Question Answering. In: Proceedings of the AAAI Spring Symposium on New Directions in Question Answering, pp. 20–27. (2003)
5. Carenini, G., Cheung, J.C.K.: Extractive vs. NLG-based abstractive summarization of evaluative text: The effect of corpus controveriality. In: Proceedings of the 5th International Natural Generation Conference. (2008)

6. Carenini, G., Cheung, J.C.K., Pauls, A.: Multi-Document Summarization of Evaluative Text. Computational intelligence. vol. 29, 545–576 (2012)
7. Contractor, D., Singla, P., Mausam,.: Entity-balanced Gaussian pLSA for Automated Comparison. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 69–79. (2016)
8. Dini, L., Mazzini, G.: Opinion classification through Information Extraction. In: Proceedings of the Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields (Data Mining), pp. 299–310. (2002)
9. Elarnaoty, M., AbdelRahman, S., : A machine learning approach for opinion holder extraction in arabic language. International Journal of Artificial Intelligence and Applications. vol. 3(2) (2012)
10. Di Fabbrizio, G., Aker, A., Gaizauskas, R.: Summarizing on-line product and service reviews using aspect rating distributions and language modeling. IEEE Intelligent Systems. vol. 28, 28-37 (2013)
11. Goldstein, J., Kantrowitz, M., Mittal, V., Carbonell, J.: Summarizing text documents: sentence selection and evaluation metrics. In SIGIR, pp. 121–128. (1999)
12. Kim, H.D., Ganesan, K., Sondhi,P., Zhai,C.: Comprehensive Review Of Opinion Summarization. Computer Science research and tech Reports. (2011)
13. Krestel, R., Dokoohaki, N.: Diversifying customer review rankings. Neural Networks. vol. 66, 36 - 45 (2015)
14. Lafferty, J.D., McCallum, A., Pereira, F. C. N.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proceedings of the Eighteenth International Conference on Machine Learning (ICML '01), (2001)
15. Liu, B.: Sentiment Analysis and Opinion Mining. CA: Morgan & Clay-pool, San Rafael (2012)
16. Liu, M., Fang, Y., Park, D. H., Hu, X., Yu, Z.: Retrieving Non-Redundant Questions to Summarize a Product Review. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 385–394. New York (2016)
17. Luhn, H.P.: The Automatic Creation of Literature Abstracts. IBM Journal of Research development. vol. 2, 159–165 (1958)
18. Maamouri, M., Bies, A., Kulick, S., Krouna, S., Gaddeche, F., Zaghouani, W.: Arabic TreeBank (ATB): Part 3 Version 3.2. Linguistic Data Consortium. Catalog No: LDC2010T08 (2010)
19. Mao, Y., Lebanon, G.: Sequential Models for Sentiment Prediction. In: Proceedings of the ICML Workshop: Learning in Structured Output Spaces Open Problems in Statistical Relational Learning Statistical Network Analysis: Models, Issues and New Directions, (2006)
20. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retr. vol. 2, pp. 1-135., (2008)
21. 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. (2010)
22. Touati, I., Graja, M., Ellouze, M., Hadrich Belguith, L.: Arabic Fine-Grained Opinion Categorization Using Discriminative Machine Learning Technique. In: Proceedings of the International Conference on Advanced Intelligent Systems and Informatics, pp. 104–113. Cairo (2016)
23. Wang, D., Zhu, S., Li, T.: SumView: A Web-based engine for summarizing product reviews and customer opinions. Expert Syst. Appl. vol. 40. pp. 27-33 (2013)

24. Farra, N., Mckeown, K., Habash, N.: Annotating Targets of Opinions in Arabic using Crowdsourcing. In: Proceedings of the Second Workshop on Arabic Natural Language Processing, pp. 89–98 (2015)