A Customer Complaint Analysis Tool for Mobile Network Operators

Feyzullah Kalyoncu, Engin Zeydan, Ibrahim Onuralp Yigit, and Ahmet Yildirim

Türk Telekom Labs, Istanbul, Turkey.

Email: {feyzullah.kalyoncu, engin.zeydan, ibrahimonuralp.yigit}@turktelekom.com.tr, ahmet.yildirim3@boun.edu.tr

***Abstract*—Mobile Network Operators (MNOs) are eager to learn more about complaint behaviour of their subscribers. In this demo, we study topic modeling approach for extracting relevant problems experienced by subscribers of MNOs in Turkey and visualize the topic distributions using LDAvis data analytics tool. For building topic models using Latent Dirchlet Allocation (LDA), we have built customer complaint text dataset of subscriber complaints for each MNOs from Turkey's largest customer complaint website. The proposed analysis tool can be used as customer complaint analysis service by MNOs in Turkey to gain more insight. We have also validated our generated topic model using another dataset obtained from Turkey's largest online community website. Our results indicate similar and dissimilar topics of complaints as well as some of the distinctive problems of MNOs in Turkey based on their subscriber's experiences and feedback.**

***Index Terms*—MNOs, text analytics, topic modeling, complaints, subscribers.**

I. Introduction

There has been a global rise in amount of text based customer complaints in the last few years. Social media and online forums are one of the great sources of customer complaint receiving platforms. Using online media, subscribers of Mobile Network Operators (MNOs) can easily convey their customer complaints and get answers easily. However, most of the textual data of complaints are collected as unstructured, huge documents. These documents need to be summarized for humans, hence analyzed using appropriate platforms that are designed for extracting the summary of the text, i.e. the relevant topics of the documents.

Many models exist in the literature for generating topic models such as Latent Dirichlet Allocation (LDA), Hierarchical Dirichlet Process (HDP), Spherical HDP, Latent Semantic Analysis, Latent Semantic Indexing, pLSA, etc. Moreover, lately there have been different efforts on visualizing the output of topics models that are generated with LDA [1]. Eye balling models such as [2], [3], [4] can be used for visualizing topic model and top topic terms for easier analysis. LDAExplore which is a tool to visualize a document corpus is given in [5]. Different techniques such as principles coordinates analysis, metric multi-dimensional scaling can be used for visualizing the similarity of topics in two-dimensions. In our recent work of [6], we have analyzed the customer call center data records using LDA method and have shown top topic distributions of the calls using the real records of a customer call center company. Besides textual analysis, there also exist previous works that compare end-to-end network performance of MNOs [7], [8].

Different than above works, in this demo paper, our main focus is on analysis of customer complaints received from online data sources regarding the performance of three major MNOs in Turkey. The main contributions of the demo paper can be summarized as follows: (i) We have crawled all MNO specific complaints from a well-known customer complaint website as well as the largest online forum in Turkey. (ii) We have built a LDA based topic model using the previously collected customer complaints and used the generated LDA model on online forum data. (iii) Finally, we have visualized the obtained topic models with LDAvis and present topic distributions of complaints of MNOs' subscribers.

II. A Tool for Customer Complaint Analysis

We perform visual analysis of topics of customer analysis complaints data which are inserted into our proposed platform. Gensim library [9], which is built on top of open-source Python stack: NumPy, SciPy and Pyro [10], is used for automatic extraction of semantic topics from documents. In our analysis, we use LDAvis [4] which is an interactive visualization that is used to extract useful information from a topic model.

Dataset: Our first dataset of customer complaint, that is used for topic model generation using LDA, is crawled for dates from 27 February 2017 to 27 February 2018 spanning one year from Turkey's largest customer complaint website [11]. The website is organized according to their topic and sectors. Total number of unique count is 43,734 with total number of words $5,244,140$. The total number of cases for MNO-1, MNO-2 and MNO-3 are 33034, 53889 and 26574 respectively. We have tested our previously generated topic model with the second dataset collected from one of the largest online community website in Turkey [12]. Online community website is mainly used for information exchange on various topics ranging from science to ordinary life issues. This second dataset is collected from January 2016 to May 2018 spanning

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain
978-1-5386-6051-5/18/$31.00 © 2018 IEEE

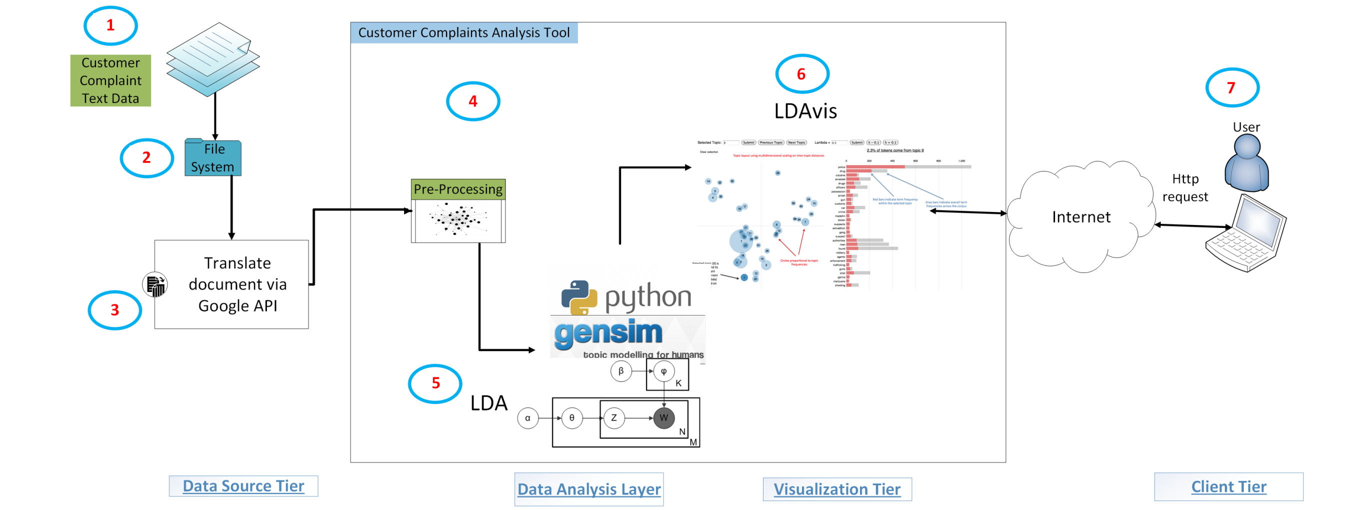

Fig. 1: Demonstration Setup

more than two years. The total number of cases for MNO-1, MNO-2 and MNO-3 are 1679, 1377 and 1600 respectively.

The proposed tool processes customer complaints of MNOs data that are crawled a priori from the public websites for each MNOs. The tool is composed of four main modules: **(a) Data Source Tier**, **(b) Data Analysis Tier**, **(c) Visualization Tier**, **(d) Client Tier**. The solution integrates the received customer complaints of MNOs' subscribers with the text analytics software. Fig. 1 demonstrates the general architecture of our demonstration solution.

A. Demonstration WorkFlow

During the experiment, we demonstrate how topics of multiple MNOs can be visualized via an interactive web-based visualization from a previously fitted LDA topic model in steps (1)–(7) given in Fig. 1. In **Data Source Tier** marked as steps (1), (2) and (3) in Fig. 1, after customer complaint website data is crawled as in step (1), the data is stored in a *File System* as CSV in step-(2). Later, a translation stage from Turkish to English using Google Translator APIs is followed due to smaller corpus size of English language compared to Turkish as shown in step-(3). For example, a corpus of 10 million word token contains four times as many word types as a similarly sized English corpus [13]. Fig. 2(a) and Fig. 2(b) show details for the execution stages for building topic models using LDA with customer complaint data as well as the topic building and visualization steps using the existing LDA model with both customer complaint and online forum data. In **Data Analysis Tier**, after language translation stage is performed, pre-processing stage is followed marked as step-(4) in Fig. 1. In *Pre-Processing* stage, processes such as removing newline characters, tokenizing words and cleaning up words, removing stop-words, lemmatization and creating a dictionary and corpus needed for the topic modeling are performed. After pre-processing step, topic model generation using LDA is performed in step-(5). For this, a document-term matrix is constructed using Gensim library [9]. In **Visualization Tier**, marked as step-(6) in Fig. 1 , we utilize LDAvis which is a combination of R and D3 packages for interactive topic model visualization. In **Client Tier**, marked as step-(7) in Fig. 1, a user is interacting with the customer complaint topic modeling tool via the user interface and interprets the topics based on different MNOs.

B. Analysis of Demonstration Results

To obtain optimal number of topics in customer complaint dataset and coherent topics, we use coherence value as our metric [14]. Fig. 3 shows the coherence score values with increasing topic number. The number of topics of 18 with the highest coherence score for the existing dataset is selected for LDA model generation with 100 number of passes. Fig. 4 shows the LDAvis dashboard of all topics for all MNOs and Table I gives the total list of topics and top 6 topic terms. We can observe that the topic distribution has captured the internal structure of the customer complaints of MNO subscribers with distinct topics. The importance of topics is represented by the size of the circle where topic # 1, 2 and 3 have bigger circles. In this figure, we have marked topic #5 which is on billing complaints. The top 6 topic terms for this topic are “TL”, “bill”, “month”, “fee”, “pay”, “invoice”. The remaining top 6 topic terms for all the other topics are given in Table I which are understandable.

Fig. 5 shows the customer complaint distribution over the topics using the same model generated by LDA. We can observe that most of the complaints are over topic #3 which is on customer services. This is inline with the intuition that as the customers cannot be able to get satisfactory service after contacting with call center center, the trend is to share their experiences with online communities.

Fig. 6 shows the distribution of topics with respect to complaints and non-complaints for each MNO using the online community data [12]. We can observe that the largest percentage of topics that falls into one of the categories of Table I is for MNO-2 followed by MNO-1 and MNO-

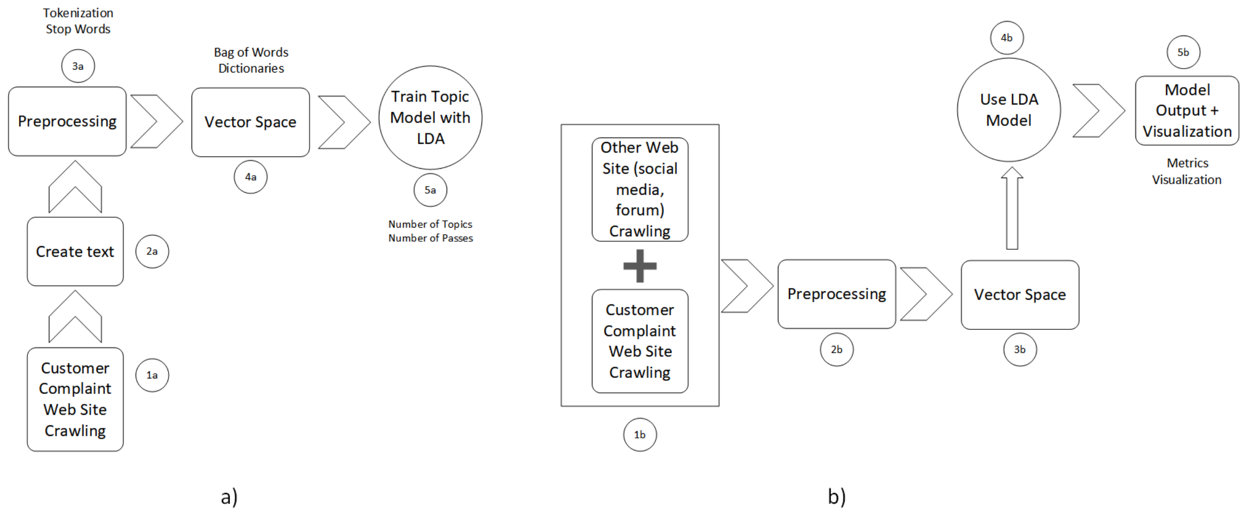

Fig. 2: The execution stages for (a) building topic model with LDA using customer complaint data (b) utilizing existing LDA model for topic building and visualization with both customer compliant and online forum datasets.

TABLE I: Topic Number and Top 6 Topic Terms obtained from Turkey's largest customer complaint website.

Topic #	Topic Terms	Topic #	Topic Terms
1: Appointment	said, day, called, service, customer, told	10: Temporary Problems	2017, 11, 00, 12, 2018, date
2: Contract	tariff, campaign, contract commitment, price, discount	11: Infrastructure	infrastructure, internet, building, port, apartment, fiber
3: Customer Service	customer, service,company call, complaint, reach	12: Service Cancellation	membership, channel, watch, subscription, facebook, geveze
4: Technical Problem	problem, modem, fault, internet, solve, record	13: Mobile Package	package, GB, minute, gift, internet, 1000
5: Billing	TL, bill, fee, invoice, amount, month	14: Mobile Payment	phone, message, received, sent, mobile, sms
6: Repetitive Calls	bad, shame, tired, say, money, every	15: Speed & Quota	speed, mbps, quota, fair, slow, internet
7: Cancellation	cancellation, petition, process request, mail, subscription	16: Fixed Line	network, telephone, line, home, fixed, phone
8: Problematic Line Termination	line, dept, closed, paid, bill, credit, payment	17: Device	device, bought, product, dealer, prime, smart
9: Mobile Connection	time, year, months, village, 3G, 4.5G	18: Maintanence	soon, possible, ping, resolved, station, victimization

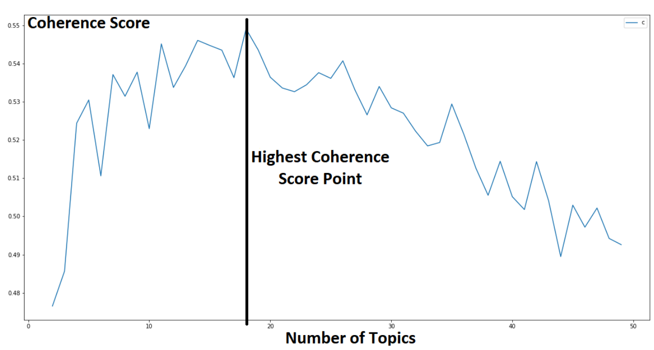

Fig. 3: Selecting the optimal topic model using coherence score.

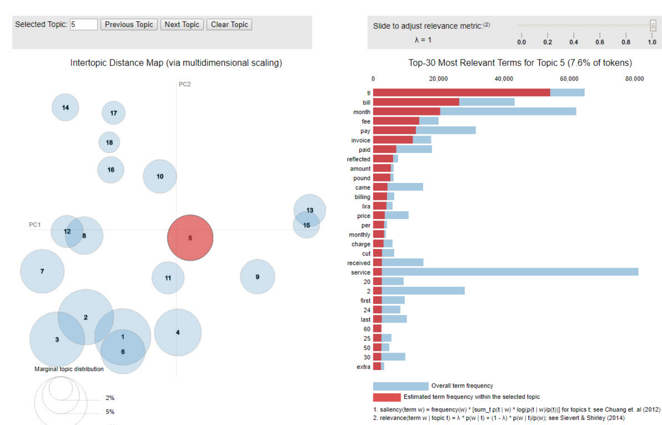

Fig. 4: Dashboard illustrating the topic models using LDAvis for all MNOs using dataset of Turkey's largest customer complaint site.

3 with 21.71%, 21.62% and 17.81% respectively. Using the same dataset, Fig. 7 shows MNO based customer complaint distribution of topics of Table I. We can observe that MNO-3 has received the highest percentage of complaints on topic#3 (customer service), whereas MNO-2 has received the highest

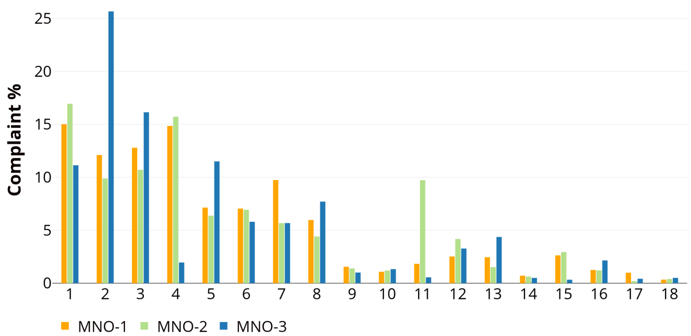

Fig. 5: Topic based percentage distributions of customer complaints for each MNO using dataset of Turkey's largest customer complaint site.

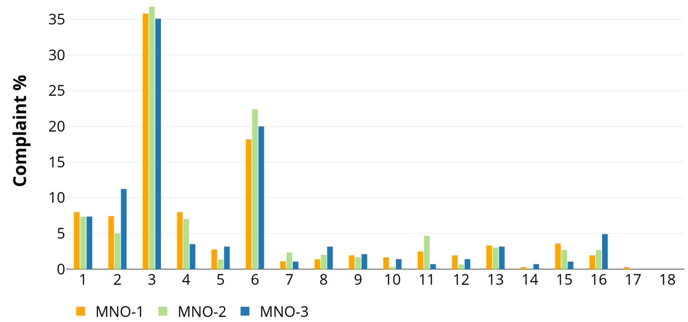

Fig. 7: Topic based percentage distributions of customer complaints for each MNO using dataset of Turkey's largest online community site.

on topic#11 (Infrastructure) with respect to other MNOs.

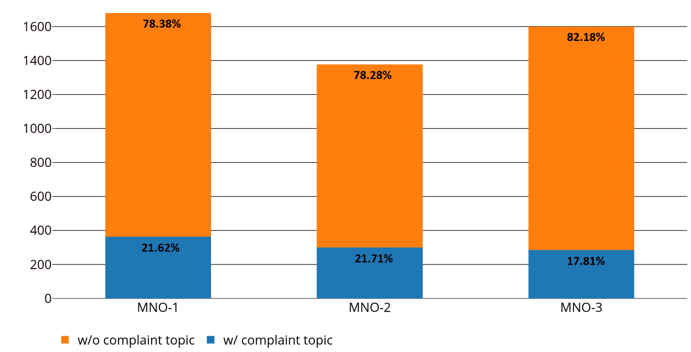

Fig. 6: Complaint and non-complaint percentage distributions of customer complaints for each MNO using dataset of Turkey's largest online community site.

A key observation of the demonstration results shows that the some topics with respect to MNOs regarding their performances are more prominent in customer compliant website data of Fig. 5 compared to online community website data of Fig. 7. This is due to the fact that the LDA model is trained using the customer complaint dataset which results in more decisive performance comparisons on topics.

III. Conclusions

In this demonstration, we analyze customer complaints of major MNOs in Turkey using dataset of Turkey's largest customer complaint and online community websites. The analysis results are visualized using LDAvis visualization toolbox. Our results indicate the existence of different similar and dissimilar topics of MNOs and the availability of some of the distinctive problems of MNOs in Turkey based on their customer's experiences and feedback.

Acknowledgment

This work has been performed in the framework of the ITEA project 16037 PAPUD and TÜBİTAK TEYDEB 1509 project numbered 9170063.

References

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[2] J. Chuang, C. D. Manning, and J. Heer, "Termite: Visualization techniques for assessing textual topic models," in *Proceedings of the international working conference on advanced visual interfaces*, pp. 74–77, ACM, 2012.

[3] J. Chuang, S. Gupta, C. Manning, and J. Heer, "Topic model diagnostics: Assessing domain relevance via topical alignment," in *International Conference on Machine Learning*, pp. 612–620, 2013.

[4] C. Sievert and K. Shirley, "Ldavis: A method for visualizing and interpreting topics," in *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, pp. 63–70, 2014.

[5] A. Ganesan, K. Brantley, S. Pan, and J. Chen, "Ldaexplore: Visualizing topic models generated using latent dirichlet allocation," *arXiv preprint arXiv:1507.06593*, 2015.

[6] I. O. Yigit, E. Zeydan, and A. A. Feyzi, "An application for detecting network related problems from call center text data," in *Black Sea Conference on Communications and Networking (BlackSeaCom), 2017 IEEE International*, pp. 1–4, IEEE, 2017.

[7] I. O. Yigit, G. Ayhan, E. Zeydan, F. Kalyoncu, and C. O. Etemoglu, "A performance comparison platform of mobile network operators," in *Network of the Future (NOF), 2017 8th International Conference on th*, pp. 144–146, IEEE, 2017.

[8] A. Nikravesh, D. R. Choffnes, E. Katz-Bassett, Z. M. Mao, and M. Welsh, "Mobile network performance from user devices: A longitudinal, multidimensional analysis.," in *PAM*, vol. 14, pp. 12–22, Springer, 2014.

[9] "Gensim: topic modeling for humans." https://radimrehurek.com/gensim/index.html, 2017. [Online; accessed 09-May-2018].

[10] R. Rehurek and P. Sojka, "Gensimstatistical semantics in python," 2011.

[11] "Sikayetvar: Turkey's largest customer complaint website." www.sikayetvar.com.tr, 2018. [Online; accessed 24-May-2018].

[12] "Eksi Sozluk: Turkey's largest online community website." www.eksisozluk.com, 2018. [Online; accessed 16-May-2018].

[13] D. Jurafsky and J. H. Martin, *Speech and language processing*, vol. 3. Pearson London:, 2014.

[14] "Gensim Topic Modeling: A guide to building best LDA models." https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/, 2018. [Online; accessed 16-May-2018].