CROSS-PLATFORM REPUTATION GENERATION SYSTEM BASED ON ASPECT-BASED SENTIMENT ANALYSIS

[1] **Dr. K. SAILAJA**

[1] Professor & HOD

[1] Department of Computer Applications

[1] Chadalawada Ramanamma Engineering College College

(Autonomous), Tirupati

[2] **SYED AFREEN**

[2] Student

[2] Department of Computer ApplicationsCollege

[2] Chadalawada Ramanamma Engineering

(Autonomous), Tirupati

ABSTRACT

The active growth of Internet-based applications such as social networks and e-commerce websites leads people to generate a tremendous number of opinions and reviews about products and services. Thus, it becomes very crucial to automatically process them. Over the last ten years, many systems have been proposed to generate and visualize reputation by mining textual and numerical reviews. However, they have neglected the fact that online reviews could be posted by malicious users that intend to affect the reputation of the target product. Besides, these systems provide an overall reputation value toward the entity and disregard generating reputation scores toward each aspect of the product. Therefore, we developed a system that incorporates spam filtering, review popularity, review posting time, and aspect-based sentiment analysis to generate accurate and reliable reputation values. The proposed model computes numerical reputation values for an entity and its aspects based on opinions collected from various platforms. Our proposed system also offers an advanced visualization tool that displays detailed information about its output. Experiment results conducted on multiple datasets collected from various platforms (Twitter, Facebook, Amazon ...) show the efficacy of the proposed system compared with state-of-the-art reputation generation systems. Keywords: Aspect-based sentiment analysis, decision-making, reputation generation, e-commerce, opinion mining.

Keywords: Aspect-based sentiment analysis, decision-making, reputation generation, e-commerce, opinion mining.

1. INTRODUCTION

Having easy access to the web has radically changed the way people interact with brands and products. From physical products to online services, people tend to instantly share their opinions and reviews on various platforms on the Internet. A recent research experiment1 shows that consumers are more willing to share a review when the experience they have had evokes emotions, whether positive or negative. This large volume of consumers' reviews holds insightful information about the quality of the product/service, therefore analyzing them will help consumers make a better judgment toward the targeted item. In the past few years, a new sub- field of natural language processing (NLP) called reputation generation has been well-established as an area of interest. The main focus of reputation generation systems is to produce a numerical value in which an entity is held based

on mining customer reviews and their numerical ratings.

Over the last decade, many reputation generation systems have been proposed to generate and visualize reputation of online products and services based on fusing and mining textual and numerical reviews. However, these systems have not taken into consideration extracting and processing reviews from various platforms, filtering reviews written by potential spammers, generating a numerical reputation value toward each aspect of the target product, and, providing an advanced reputation visualization tool for a better decision-making process. Thereby, we designed and built an upgraded reputation generation model that overcomes the shortcomings of the previous systems in order to compute and visualize the reputation of an entity (product, movie, hotel, restaurant, service) with consistent reliability. The proposed system collects and processes data from both e-commerce and social media platforms. Then, a spam filtering system is applied to eliminate spam reviews and prepare the cleaned output for aspect-based sentiment analysis (ABSA), where aspects of the target entity are extracted from the reviews with their sentiment polarities. Later, the time and popularity features of the reviews are exploited along with the ASBA results to finally generate a reputation value of each aspect of the target entity as well as the overall reputation value using mathematical formulas. The system also proposes an analytical dashboard that displays in-depth information about the reputation of the target entity.

In this manner, this study addresses the following research question: with the consideration of review popularity, review time, spam filtering, and ABSA, can the proposed reputation model offer better results in terms of generating and visualizing reputation than state-of-the-art (SOTA) systems?

This paper is organized as follows. Section 2 presents the related work concerning the previous reputation generation systems as well as the ABSA models. Section 3 presents the preliminaries. Section 4 describes our proposal. Section 5 details the experiments. Section 6 presents the discussion. And finally, Section 7 concludes this paper.

2. LITERATURE SURVEY

2.1 DIFFERENT AUTHORS DISCUSSION:

Over the last decade, many reputation generation systems have been proposed to generate and visualize reputation of online products and services based on fusing and mining textual and numerical reviews.

2.2 DOMAIN DESCRIPTION:

Therefore, we developed a system that incorporates spam filtering, review popularity, review posting time, and aspect-based sentiment analysis to generate accurate and reliable reputation values. The proposed model computes numerical reputation values for an entity and its aspects based on opinions collected from various platforms. Our proposed system also offers an advanced visualization tool that displays detailed information about its output. Experiment results conducted on multiple datasets collected from various platforms (Twitter, Facebook, Amazon . . .) show the efficacy of the proposed system compared with state-of-the-art reputation generation systems

3. PROBLEM STATEMENT

3.1 EXISTING SYSTEM

Poria presented the first deep learning approach for the AE task in opinion mining. The authors employed a 7-layer deep convolutional neural network to tag each word in the textual opinions as either aspect or non-aspect word. The authors also proposed a set of heuristic linguistic patterns and integrated them with the deep learning classifier which significantly improves the accuracy compared with the previous SOTA methods. In the authors proposed an attention-based long short-term memory (LSTM) for aspect-level sentiment classification. The idea is to learn aspect embeddings and let aspects participate in computing attention weights.

The proposed model can focus on different parts of a sentence when different aspects are given so that they are more competitive for aspect-level classification. The proposed model achieved better results compared with the standard LSTM on the SemEval 2014 Task 4 dataset In Wei and Toi improved the deficiencies of the previous LSTM approaches by proposing convolutional neural networks and gating mechanisms (GCAE) based model, which has been proved to be more accurate and efficient. The novel Gated Tanh-ReLU Units can selectively output the sentiment features according to the provided aspect or entity. The architecture of the proposed model is much simpler than the attention layer used in the previously existing models.

The experiments on SemEval datasets show a performance improvement compared with the LSTM based models. The authors inproposed an interactive multi-task learning network (IMN) capable of jointly learning multiple related tasks simultaneously at both the token-level as well as the document-level. The IMN introduces a message passing mechanism that allows informative interactions between tasks, enabling the correlation to be better exploited. Experiments on three benchmark datasets, taken from SemEval2014 and

SemEval 2015 show that IMN outperforms other baselines by large margins. Since most existing methods ignore the position information of the aspect when encoding the sentence, authors in proposed a hierarchical attention-based position-aware network (HAPN), which includes position embeddings to learn the position-aware representations of sentences to generate the target-specific representations of contextual words. HAPN achieved the SOTA performance on SemEval 2014 dataset compared with the previous methods.

Xu presented a review reading comprehension (RRC) task where they adopted BERT as a base model, and proposed a joint post-training and fine-tuning approach for ATE, APC. Experimental results show that the proposed post-training approach is very effective. Later in the authors proposed a novel architecture called BERT Adversarial Training (BAT) to employ adversarial training for AE and APC by generating artificial data which is carried out in the embedding space. The proposed model outperforms the standard BERT as well as the in-domain post-trained BERT in both AE and APC tasks. In, the authors exploit domain-specific BERT language model fine tuning in addition to supervised task-specific fine tuning to produce a new SOTA performance on the SemEval 2014 Task 4 restaurants dataset. The authors also showed that cross-domain adapted BERT model performs better than strong baseline models such as XLNet-base and vanilla BERT-base. In the authors compared the induced trees from pre-trained models and the dependency parsing trees on various popular models for the ABSA task.

They found that the induced tree from fine tuned RoBERTa (FT-RoBERTa) outperforms the parser-provided tree. The experiments show that the RoBERTa-based model can outperform or approximate the previous SOTA performances on six datasets across four languages including SemEval 2014 task 4. Recently, authors in proposed a multi-task learning model named LCF-ATEPC for ABSA based on the multi-head self-attention and the local context focus (LCF) mechanisms. The proposed model is multilingual and applicable to the classic English review SA task, such as the SemEval-2014 task4. The proposed model can automatically extract aspects and determine their sentiment polarities. Since the LCF-ATEPC model currently achieves SOTA performance on AE and APC tasks,2 it was selected to be employed in this paper.

3.2 DISADVANTAGE OF EXISTINTG SYSTEM:

An existing system not implemented Aspect term extractor which performs the basic token-level classification for each token, which means that each token will be given a label, and a classification is performed to predict the aspects in the sentence. An existing system is not implemented Local Context Focus in which Local context is a new technique that can be adapted to most ne-grained NLP tasks.

4.PROPOSED SYSTEM

4.1 PROPOSED SYSTEM

This system aims at generating a reputation value toward online entities (movies, hotels, restaurants, services, etc.) and computing a satisfaction score toward each aspect of the target entity by processing textual and numerical data collected from multiple platforms. Proposed system presents its architecture. First, we start by collecting users' reviews from different platforms such as Twitter, Amazon, YouTube, etc. Next, an automatic spammers filtering system is employed to detect and eliminate unwanted spam reviews. Then, we apply a SOTA ABSA model to users' textual reviews in order to compute a score based on the sentiment orientation of the extracted aspects from those reviews. Further, we calculate a popularity score and a time score based on statistical features extracted with the textual reviews. Finally, we compute a reputation value based on the previously calculated scores, and we propose a new user-friendly visualization interface that displays in-depth details about the reputation of the target entity.

One of the important features of the proposed system is the ability to collect and process data from various platforms. Previous reputation generation systems gather necessary data from either e-commerce websites such as Amazon, TripAdvisor, or social media platforms such as Twitter and Facebook. In this work, we decided to normalize the features of all platforms in order to create a single merged dataset by classifying the platforms on the Internet into two types: the first type provides the accessibility of extracting the textual review with the number of likes received for that review such as Amazon, YouTube, etc. The second type provides the accessibility of extracting the textual review with the number of likes received for that review along with the number of times the review was shared among the network such as Twitter, Facebook, etc.

4.2 ADVANTAGE OF PROPOSED SYSTEM:

Multi-Head Self-Attention (MHSA): The multi-head attention mechanism helps the model to learn the words' relevant information in different presentation subspaces. MHSA is based on multiple scale-dot attention that can be used to extract deep semantic features in the context. Aspect Polarity Classifier: To perform the sentiment polarity classification, the LCF-ATEPC model combines the local context features and the global context features. Then, the aspect polarity classifier performs a head-pooling on the learned concatenated context features from the feature interactive learning layer.

5.IMPLEMENTATION

5.1 Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Attack Status, View Attack Status Ratio, Download Trained Data Sets, View Attack Status Ratio Results, View All Remote Users.

5.2 View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

5.3 Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT ATTACK STATUS TYPE, VIEW YOUR PROFILE.

6. SYSTEM ARCHITECTURE
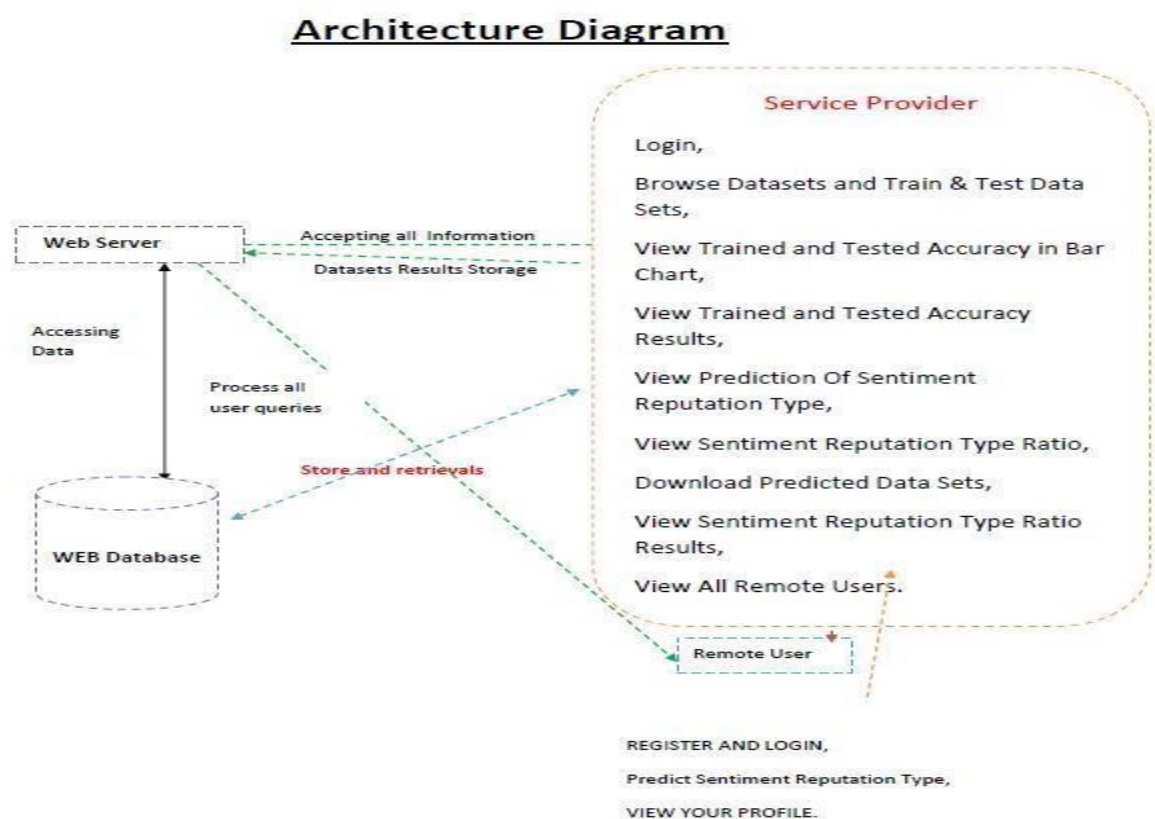
8.CONCLUSION

In this paper, we proposed a reputation system capable of generating numerical reputation values for a specific item (product, movie, service, hotel, etc.) and its aspects based on opinions and reviews expressed online. The contribution of this work revolves around four components that were not exploited in previous systems. The first one is cross-platform compatibility, where the proposed system can collect and process opinions from different platforms (Face book, Amazon, Twitter, Trip Advisor, etc.) as well as managing and standardizing those platforms' features. The second one is opinion spam filtering, where the spam opinions are detected and eliminated based on spammers' behavior features, keeping only authentic opinions. The third one is employing a SOTA aspect-based sentiment-analysis model named LCF-ATEPC in order to extract and analyze the aspects within the textual opinions. Finally, we incorporated the previous results with a calculated

review time score and review popularity score using mathematical formulas to obtain a reputation value for the targeted entity as well as the reputation values of the entities' aspects. In addition a holistic reputation visualization is provided within the system that displays the detailed output results of the reputation generation process. To assess the effectiveness of our reputation system, we invited 32 participants and 3 experts to choose the best performing system out of four SOTA reputation systems by giving numerical satisfaction scores to each system. Our reputation system achieved the highest average satisfaction scores from both users and experts. In the future, we propose to investigate the effectiveness of our proposed system by attempting to generate more than the numerical reputation values, such as extending the system to automatically generate a textual summary of the benefits and drawbacks of the targeted entity. Also, we intend to extend this system to be used in multilingual content.

9. FUTURE ENHANCEMENT

One of the important features of the proposed system is the ability to collect and process data from various platforms. Previous reputation generation systems gather necessary data from either e-commerce websites such as Amazon, TripAdvisor, or social media platforms such as Twitter and Facebook. In this work, we decided to normalize the features of all platforms in order to create a single merged dataset by classifying the platforms on the Internet into two types: the first type provides the accessibility of extracting the textual review with the number of likes received for that review such as Amazon, YouTube, etc. The second type provides the accessibility of extracting the textual review with the number of likes received for that review along with the number of times the review was shared among the network such as Twitter, Facebook, etc.

10. REFERENCES

[1] Abdel-Hafez, Y. Xu, and D. Tjondronegoro, ''Product reputation model: An opinion mining based approach,'' in *Proc. 1st Int. Workshop Sentiment Discovery Affect. Data*, vol. 917, London, U.K., Jun. 2013, pp. 16–27. [Online]. Available: https://eprints.qut.edu.au/58118/

[2] U. Farooq, A. Nongaillard, Y. Ouzrout, and M. A. Qadir, ''A featurebased reputation model for product evaluation,'' *Int. J. Inf. Technol. Decis. Making*, vol. 15, no. 6, pp. 1521–1553, Nov. 2016, doi: 10.1142/S0219622016500358.

[3] Z. Yan, X. Jing, and W. Pedrycz, ''Fusing and mining opinions for reputation generation,'' *Inf. Fusion*, vol. 36, pp. 172–184, Jul. 2017, doi: 10.1016/j.inffus.2016.11.011.

[4] A. Benlahbib and E. H. Nfaoui, ''A hybrid approach for generating reputation based on opinions fusion and sentiment analysis,'' *J. Organizational Comput. Electron. Commerce*, vol. 30, no. 1, pp. 9–27, 2020, doi: 10.1080/10919392.2019.1654350.

[5] E. I. Elmurngi and A. Gherbi, ''Building sentiment analysis model and compute reputation scores in E-commerce environment using machine learning techniques,'' *Int. J. Organizational Collective Intell.*, vol. 10, no. 1, pp. 32–62, Jan. 2020.

[6] A. Benlahbib and E. H. Nfaoui, ''Aggregating customer review attributes for online reputation generation,'' *IEEE Access*, vol. 8, pp. 96550–96564, 2020, doi: 10.1109/ACCESS.2020.2996805.

[7] A. Gupta, S. Priyani, and R. Balakrishnan, ''Customized reputation generation of entities using sentiment analysis,'' *World J. Eng.*, vol. 18, no. 4, pp. 596–605, Jul. 2021, doi: 10.1108/WJE-09-2020-0470.

[8] A. Boumhidi and E. Nfaoui, ''Leveraging lexicon-based and sentiment analysis techniques for online reputation generation,'' *Int. J. Intell. Eng. Syst.*, vol. 14, no. 6, pp. 274–289, 2021, doi: 10.22266/ijies2021.1231.25.

[9] V. M. Pradhan, J. Vala, and P. Balani, ''A survey on sentiment analysis algorithms for opinion mining,'' *Int. J. Comput. Appl.*, vol. 133, no. 9, pp. 7–11, Jan. 2016.

[10] A. Tripathy, A. Anand, and S. K. Rath, ''Document-level sentiment classification using hybrid machine learning approach,'' *Knowl. Inf. Syst.*, vol. 53, no. 3, pp. 805–831, 2017.

[11] B. Liu, ''Sentiment analysis and subjectivity,'' in *Handbook of Natural Language Processing*, 2nd ed. Boca Raton, FL, USA: Taylor & Francis, 2010.

[12] K. Schouten and F. Frasincar, ''Survey on aspect-level sentiment analysis,'' *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 813–830, Mar. 2016.

[13] A. Da'u and N. Salim, ''Aspect extraction on user textual reviews using multi-channel convolutional neural network,'' *PeerJ Comput. Sci.*, vol. 5, p. e191, May 2019.

[14] C. Feng, Y. Rao, A. Nazir, L. Wu, and L. He, ''Pre-trained language embedding-based contextual summary and multi-scale transmission network for aspect extraction,'' *Proc. Comput. Sci.*, vol. 174, pp. 40–49, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920315672

[15] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, ''Sentic LSTM: A hybrid network for targeted aspect-based sentiment analysis,'' *Cogn. Comput.*, vol. 10, pp. 639–650, Mar. 2018.

[16] M. Hu and B. Liu, ''Mining and summarizing customer reviews,'' in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2004, pp. 168–177, doi: 10.1145/1014052.1014073.

[17] R. Agrawal and R. Srikant, ''Fast algorithms for mining association rules,'' in *Proc. 20th Int. Conf. Very Large Data Bases*. San Francisco, CA, USA: Morgan Kaufmann, 1994, pp. 487–499. [Online]. Available: http://dl.acm.org/citation.cfm?id=645920.672836

[18] S. Poria, E. Cambria, and A. Gelbukh, ''Aspect extraction for opinion mining with a deep convolutional neural network,'' *Knowl.Based Syst.*, vol. 108, pp. 42–49, Sep. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705116301721

[19] Y. Wang, M. Huang, X. Zhu, and L. Zhao, ''Attention-based LSTM for aspect-level sentiment classification,'' in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Austin, TX, USA, Nov. 2016, pp. 606–615. [Online]. Available: https://aclanthology.org/D16-1058

[20] S. Hochreiter and J. J. Schmidhuber, ''Long short-term memory,'' *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[21] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, ''SemEval-2014 task 4: Aspect based sentiment analysis,'' in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, Dublin, Ireland, 2014, pp. 27–35. [Online]. Available: https://aclanthology.org/S14-2004

[22] W. Xue and T. Li, ''Aspect based sentiment analysis with gated convolutional networks,'' in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, Melbourne, VIC, Australia, Vol. 1, Jul. 2018, pp. 2514–2523. [Online]. Available: https://aclanthology.org/P18-1234

[23] Y. LeCun and Y. Bengio, *Convolutional Networks for Images, Speech, and Time Series*. Cambridge, MA, USA: MIT Press, 1998, pp. 255–258.

[24] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, ''An interactive multitask learning network for end-to-end aspect-based sentiment analysis,'' in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, Florence, Italy, 2019,pp. 504–515.[Online].Available:https://aclanthology.org/P19-1048

[25] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, ''SemEval-2015 task 12: Aspect based sentiment analysis,'' in *Proc. 9th Int. Workshop Semantic Eval. (SemEval)*, Denver, CO, USA, 2015,pp. 486–495.[Online].Available:https://aclanthology.org/S15-2082

[26] L. Li, Y. Liu, and A. Zhou, ''Hierarchical attention based position-aware network for aspect-level sentiment analysis,'' in *Proc. 22nd Conf. Comput. Natural Lang. Learn.*, Brussels, Belgium, 2018, pp. 181–189. [Online]. Available: https://aclanthology.org/K18-1018

[27] H. Xu, B. Liu, L. Shu, and P. Yu, ''BERT post-training for review reading comprehension and aspect-based sentiment analysis,'' in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, Minneapolis, MN, USA, vol. 1, Jun. 2019, pp. 2324–2335. [Online]. Available: https://aclanthology.org/N19-1242

[28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ''BERT: Pre-training of deep bidirectional transformers for language understanding,'' in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, Minneapolis, MN, USA, vol. 1, Jun. 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423

[29] A. Karimi, L. Rossi, and A. Prati, ''Adversarial training for aspect-based sentiment analysis with BERT,'' in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 8797–8803.

[30] A. Rietzler, S. Stabinger, P. Opitz, and S. Engl, ''Adapt or get left behind: Domain adaptation through bert language model finetuning for aspect-target sentiment classification,'' in *Proc. 12th Lang. Resour. Eval. Conf.*, Marseille, France, May 2020, pp. 4933–4941. [Online]. Available: https://aclanthology.org/2020.lrec-1.607