UNIVERSITI TEKNOLOGI MARA

A HYBRID OF FUZZY C-MEANS CLUSTERING AND LATENT DIRICHLET ALLOCATION FOR ANALYSING PHILANTHROPIC CORPORATE SOCIAL RESPONSIBILITY ACTIVITIES

NIK SITI MADIHAH BINTI NIK MANGSOR

Thesis submitted in fulfillment
of the requirements for the degree of
Master of Science
(Computer Science)

College of Computing, Informatics, and Media

February 2023

ABSTRACT

To date, philanthropic corporate social responsibility (PCSR) activities are ad-hoc in nature, where assistance is provided more to basic needs with very little attention to activities that can contribute to eradicating poverty. Based on previous related literatures, it is found that there is no proper categorization and documentation of CSR-related activities. Conventional clustering algorithms are able to extract only a single set of flat topics in which local and global topics are mingled and cannot be separated. Therefore, this research aims to develop hybrid fuzzy document clustering techniques to improve attribute cluster membership values of PCSR activities. This study has extended document clustering technique by integrating the traditional document clustering application with topic modeling approach. This integrated approach is able to produce precise results where it can help infer more coherent themes of PCSR activities. The analysis involved five-year data from the annual reports of 19 CSR-award winning companies in Malaysia where they were converted into a structured format, collated and summarized. Then, text pre-processing for data cleaning was performed followed by identification of the best Latent Dirichlet Allocation (LDA) topic modelling technique that was used to integrate document clustering. Next, documents were then clustered using integration technique of K-Means clustering and LDA as the benchmark method before the proposed integration of Fuzzy C-Means (FCM) clustering and LDA was carried out. Findings from the proposed method in this study revealed thirteen clusters that represent thirteen types of PCSR activities performed by the CSR-award winning companies in Malaysia. This finding was then verified and agreed by experts in CSR on the validity of the obtained PCSR cluster which is able to describe the categories of PCSR activities accordingly. Lastly, the performance of the benchmark method and proposed method was evaluated using model evaluation technique. Based on the evaluation, it was found that K-Mean's performance is better than FCM's performance in terms of computational time. However, FCM is better than K-Means in terms of quality, since FCM produces more quality clusters than K-Means which could produce much better in clustering PCSR activities. As a result, the study shows that the proposed integration method produces much better categories of PCSR activities because FCM clustering is applicable to such incomplete external knowledge and it offers the opportunity to deal with data that belong to more than one cluster at the same time. The findings offer an insight to be considered by companies in strategizing the CSR activities, particularly philanthropic responsibility in ensuring optimum impact to innovatively support the society and contribute towards poverty mitigation.

ACKNOWLEDGEMENT

In the name of Allah, Most Gracious, Most Merciful. All praises be to Allah, the Cherisher and Sustainer of the world, for giving me the strength and opportunity to embark on my MSc and for completing this long and challenging journey successfully.

Firstly, my gratitude and thanks go to my supervisor Dr Syerina Azlin Md Nasir and also my co-supervisor PM Dr Shuzlina Abdul Rahman and Dr Zurina Ismail, for providing me the opportunity to do the study, and giving me all support and guidance, which was essential in completing my MSc. I am extremely thankful to them for providing a pleasant and nice support and guidance.

Finally, this thesis is dedicated to my beloved family especially my parents also my friends for the vision and determination to educate me throughout this journey. This piece of victory is dedicated to all of them. Alhamdulilah.

Thank you.

TABLE OF CONTENTS