A Framework for Semantic Clustering of News Articles Based on Fuzzy

Nidhi Dewan[1] and Shagufta Farzana[2]

Research Scholar, Department of Computer Science and Engineering[1]
Assistant Professor, Department of Computer Science and Engineering[2]
Dr. C. V. Raman University, Bilaspur, India

Abstract: *Text mining is a process that uses data mining approaches to extract valuable information held in the hidden form in textual data. In this paper, we are proposing a framework for fuzzy clustering of news articles. These news articles originate on different news portals on the web. The data obtained need to be stored in a central database and then pre-processing reduces the noise. The keyword extraction is used to extract keywords from the text and then word-frequency vector is generated. On these vectors, distance measure or similarity measure function is used to find the similarity between articles. One article may belong to more than one cluster so fuzzy context vector must be generated. Mutual Information can be used to find fuzzy membership values. The threshold values are required for the identification of clusters. The proposed framework shows that fuzzy clustering does not restrict each news article to belong exactly to one cluster. Therefore this framework when applied to informationretrieval systems or other application systems, system gives better performance and relevance to the users.*

Keywords: News article, Clustering, Fuzzy, Text Mining

REFERENCES

[1]. HTTPS://EN.WIKIPEDIA.ORG/WIKI/UNSTRUCTURED_DATA
[2]. Zadeh, L. A. 1965. Fuzzy Set. Information And Control,8(3):338-353.
[3]. Bellman, R., Kalaba, R. and Zadeh L. 1966. Abstraction and Pattern Classification. Journal of Mathematical and Applications,13(1):1-7.
[4]. Smeaton, A. F., Burnett, M., Crimmins, F. and Quinn, G. 1997. An Architecture for Efficient Document Clustering and Retrieval on a Dynamic Collection of Newspaper Text. Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval, 31(SI):74-81.
[5]. Yang, Y., Pierce, T. and Carbonell, J. 1998. A study on retrospective and on-line event detection. Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, 28-36.
[6]. Toda, H. and Kataoka, R. 2005. A Clustering Method for News Articles Retrieval. ACM, 988-989.
[7]. Corso, G.M.D., Gullf, A. and Romani, F. 2005. Ranking a Stream of News. International World Wide Web Conference Committee (IW3C2) ACM, 97-106.
[8]. Aiello, M. and Pegorett, A. 2006. Textual Article Clustering in Newspaper Pages. Applied Artificial Intelligence, 20(9):767- 796.
[9]. Newman, D. Chemudugunta, C. Smyth, P. and Steyvers, M. 2006. Analyzing Entities and Topics in News Articles Using Statistical Topic Models. Intelligence and Security Informatics LNCS. 3975:93-104.
[10]. Sekiya, H., Kondo, T., Hashimoto, M. and Takagi, T. 2007. Context representation using word sequences extracted from a news corpus. International Journal of Approximate Reasoning, 45(3):424-438.
[11]. Pera, M.S. and Ng, Y.K. 2008. Utilizing Phrase-Similarity Measures for Detecting and Clustering. Informative RSS News Clustering. Journal integrated Computer-Aided Engineering, 15(4):331-350.
[12]. Pera, M. S. and Ng, Y. K. 2009. Synthesizing Correlated RSS News Articles Based on a Fuzzy Equivalence Relation. International Journal of Web Information Systems (IJWIS), 5(1);77-109.
[13]. Bouras, C. and Tsogkas, V. 2010. W-kmeans: Clustering News Articles Using WordNet, Knowledge-Based and intelligent Information and Engineering Systems, 4278:379-388

[14]. Vadrevu, S., Teo, C. H., Rajan, S., Punera, K., Dom, B., Smola, A., Chang, Y. and Zheng, Z. 2011. Scalable Clustering of News Search Results. Proceedings of the fourth ACM international conference on Web search and data mining, 675-684.

[15]. Pera, M.S. and Ng, Y.K.D. 2012. Using maximal spanning trees and word similarity to generate hierarchical clusters of non- redundant RSS news article. Journal of Intelligent Information Systems, 39(2):513-534.

[16]. Svadas, T. and Jha, J. 2012. Document Cluster Mining on Text Documents. International Journal of Computer Science and Mobile Computing, 4(6):778-782.

[17]. Krstajic, M., Araghi, M.N., Mansmann, F. and Keim, D.A. 2013. Story Tracker: Incremental visual text analytics of news story development. Information Visualization, 12:308-323.

[18]. Bouras, C. and Tsogkas, V. 2014. Improving news articles recommendations via user clustering. ,8(1):223-237.

[19]. Soloshenko, A.N., Orlova, Y.A., Rosaliey, V.L. and Zotova, A.V.Z. 2014. Thematic Clustering Methods Applied to News Texts Analysis. Knowledge-Based Software Engineering, 466:294-310.

[20]. Dangre, N., Bodke, A., Date, A., Rungta, S. and Pathak, S.S. 2016. System for Marathi News Clustering. Procedia Computer Science, 92:18-22

[21]. Marshall, J.A. and Rafsky, L.C. 2017. Exact clustering in linear time. arXiv:1702.05425[cs.DS].

[22]. Nicholls,T. and Bright, J. 2018. Understanding news story chains using information retrieval and network clustering techniques.Social and Information Network, Information Retrieval. arXiv:1801.07988[cs.SI].

[23]. Ahmed, R. and Ahmad, T. 2018. Fuzzy Concept Map Generation from Academic Data Sources. Internationl Conference on Signals, Machines and Automation NSIT.

[24]. Kon B. .2022. Research on the Fusion of Hybrid Fuzzy Clustering Algorithm and Computer Automatic Test Paper Composition Algorithm, School of Information Engineering, Volume, Article ID-4264144, pp.12.

[25]. Jan M. et al. 2022. Interest-Based Content Clustering for Enhancing Searching and Recommendations on Smart TV , Wireless Communications and Mobile Computing , Volume , Article-ID 3896840, pp. 14.