Abstract

The need for summarizing texts evolves from the large amount of data present in electronic channels which leads to distraction of users and wastage of their time. There are generally two major techniques for text summarization: extractive method and abstractive method. The extractive method has proven to be quite reliable and involves extracting the key sentences from the document to form a summary. In this paper, an unsupervised text mining model is developed for clustering and summarizing texts. The model is deployed into a web-based system for summarizing large documents. Using the informational criteria of redundancy, coherence, speed and information coverage, our approach chooses ‘not likely’, ‘high’, ‘fast’, and ‘medium’ as semantic dimensions values for the criteria respectively.

References

1. Tohalino, J. V., & Amancio, D. R. (2017). Extractive multi-document summarization using dynamical measurements of complex networks. Paper presented at the 2017 Brazilian
Conference on Intelligent Systems (BRACIS).
2. Rananavare, L. B., & Reddy, P. V. S. (2017). An Overview of Text Summarization. Int. J. Comput. Appl, 171(10), 1-17.
3. Rautray, R., & Balabantaray, R. C. (2017). Bio-inspired approaches for extractive document summarization: A comparative study. Karbala International Journal of Modern Science, 3(3), 119-130.
4. Nikolov, N. I., Pfeiffer, M., & Hahnloser, R. H. (2018). Data-driven summarization of scientific articles. arXiv preprint arXiv:1804.08875.
5. Prateek Joshi. (2018). An Introduction to Text Summarization using the TextRank Algorithm. from https://www.analyticsvidhya.com/blog/2018/11/introduction-text-summarization-textrank-python/
6. Nallapati, R., Zhou, B., & Ma, M. (2016). Classify or select: Neural architectures for extractive document summarization. arXiv preprint arXiv:1611.04244.
7. Vahed, Q., Radev, D., Mohammad, S.M., Dorr, B., Zajic, D., Whidby, M., & Moon, T. (2013). Generative Extractive summaries of scientific paradigms. Journal of artificial Intelligent Research, 46, 165-201.
8. Verberne, S., Kraher, E., Hendrickx, I., Wubben, S., & van Den Bosch, A. (2018). Creating a reference data set for the summarization of discussion forum threads. Language Resources and Evaluation, 52(2), 461-483.
9. Sanchan, N., Aker, A., & Bontcheva, K. (2017). Gold standard online debates summaries and first experiments towards automatic summarization of online debate data. Paper presented at the International Conference on Computational Linguistics and Intelligent Text Processing.
10. Tohalino, V. J., Diego, R. & Amancio. (2017). Extractive multi document summarization using dynamical measurements of complex networks. arXiv:1708.01769
11. Verberne, S., Krahmer, E., Hendrickx, I., Wubben, S., & van Den Bosch, A. (2018). Creating a reference data set for the summarization of discussion forum threads. Language Resources and Evaluation, 52(2), 461-483.
12. Pontes, E. L., Huet, S., & Torres-Moreno, J.-M. (2018). A multilingual study of compressive cross-language text summarization. Paper presented at the Mexican International Conference on Artificial Intelligence.
13. Yadav, C. S., & Sharan, A. (2015). Hybrid approach for single text document summarization using statistical and sentiment features. International Journal of Information Retrieval Research (IJIRR), 5(4), 46-70.
14. El-Refaiy, A., Abas, A.R. & Elhenawy, I. (2018). Review of recent techniques for extractive text summarization. Journal of Theoretical and Applied Information Technology, 96, 7739-7759.
15. Badry, R. M., Eldin, A. S., & Elzanfally, D. S. (2013). Text summarization within the latent semantic analysis framework: comparative study. International Journal of Computer Applications, 81(11), 40-45.

Index Terms

Computer Science

Information Sciences
Keywords

Extractive summaries, text clustering, web application, sentence clustering.