



Evolutionary Algorithm for Extractive Text Summarization

PDF (Size: 312KB) PP. 128-138 DOI : 10.4236/iim.2009.12019

Author(s)

Rasim ALGULIEV, Ramiz ALIGULIYEV

ABSTRACT

Text summarization is the process of automatically creating a compressed version of a given document preserving its information content. There are two types of summarization: extractive and abstractive. Extractive summarization methods simplify the problem of summarization into the problem of selecting a representative subset of the sentences in the original documents. Abstractive summarization may compose novel sentences, unseen in the original sources. In our study we focus on sentence based extractive document summarization. The extractive summarization systems are typically based on techniques for sentence extraction and aim to cover the set of sentences that are most important for the overall understanding of a given document. In this paper, we propose unsupervised document summarization method that creates the summary by clustering and extracting sentences from the original document. For this purpose new criterion functions for sentence clustering have been proposed. Similarity measures play an increasingly important role in document clustering. Here we've also developed a discrete differential evolution algorithm to optimize the criterion functions. The experimental results show that our suggested approach can improve the performance compared to state-of-the-art summarization approaches.

KEYWORDS

sentence clustering, document summarization, discrete differential evolution algorithm

Cite this paper

R. ALGULIEV and R. ALIGULIYEV, "Evolutionary Algorithm for Extractive Text Summarization," *Intelligent Information Management*, Vol. 1 No. 2, 2009, pp. 128-138. doi: 10.4236/iim.2009.12019.

References

- [1] M. A. Fattah and F. Ren, " GA, MR, FFNN, PNN and GMM based models for automatic text summarization," *Computer Speech and Language*, Vol. 23, No. 1, pp. 126– 144, 2009.
- [2] U. Hahn and I. Mani, " The challenges of automatic summarization," *IEEE Computer*, Vol. 33, No. 11, pp. 29– 36, 2000.
- [3] I. Mani and M. T. Maybury, " Advances in automated text summarization," MIT Press, Cambridge, 442p, 1999.
- [4] J-Y. Yeh, H-R. Ke, W-P. Yang, I-H. Meng, " Text summarization using a trainable summarizer and latent semantic analysis," *Information Processing and Management*, Vol. 41, No. 1, pp. 75– 95, 2005.
- [5] S. Ye, T.-S. Chua, M.-Y. Kan, and L. Qiu, " Document concept lattice for text understanding and summarization, *Information Processing and Management*, 2007, Vol. 43, No. 6, pp. 1643– 1662.
- [6] R. M. Alguliev and R. M. Aliguliyev, " Effective summarization method of text documents," *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI' 05)*, France, pp. 264– 271, 19– 22 September 2005.
- [7] R. M. Alguliev and R. M. Alyguliev, " Automatic text documents summarization through sentences clustering," *Journal of Automation and Information Sciences*, Vol. 40, No. 9, pp. 53– 63, 2008.
- [8] R. M. Alguliev, R. M. Aliguliyev, and A. M. Bagirov, " Global optimization in the summarization of text

- [Open Special Issues](#)
- [Published Special Issues](#)
- [Special Issues Guideline](#)

[IIM Subscription](#)[Most popular papers in IIM](#)[About IIM News](#)[Frequently Asked Questions](#)[Recommend to Peers](#)[Recommend to Library](#)[Contact Us](#)

Downloads: 154,235

Visits: 384,199

[Sponsors, Associates, and Links >>](#)

- [9] R. M. Aliguliyev, " A novel partitioning-based clustering method and generic document summarization," *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT' 06 Workshops) (WI-IATW' 06)*, Hong Kong, China, pp. 626– 629, 18– 22 December 2006.
- [10] R. M. Aliguliyev, " A new sentence similarity measure and sentence based extractive technique for automatic text summarization," *Expert Systems with Applications*, Vol. 36, No. 4, pp. 7764– 7772, 2009.
- [11] R. M. Aliguliyev, " Clustering techniques and discrete particle swarm optimization algorithm for multi-document summarization, *Computational Intelligence* (accepted).
- [12] Y. Gong and X. Liu, " Generic text summarization using relevance measure and latent semantic analysis," in: *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New Orleans, USA, pp. 19– 25, 9– 12 September, 2001.
- [13] D. R. Radev, H. Jing, M. Stys, and D. Tam, " Centroid- based summarization of multiple documents," *Information Processing and Management*, Vol. 40, No. 6, pp. 919– 938, 2004.
- [14] G. Salton, A. Singhal, M. Mitra, and C. Buckley, " Automatic text structuring and summarization," *Information Processing and Management*, Vol. 33, No. 2, pp. 193– 207, 1997.
- [15] K. M. Svore, L. Vanderwende, and C. J. C. Burges, " Enhancing single-document summarization by combining RankNet and third-party sources," in: *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL' 07)*, Prague, Czech Republic, pp. 448– 457, 28– 30 June 2007.
- [16] D. M. Dunlavy, D. P. O' Leary, J. M. Conroy, and J. D. Schlesinger, " QCS: A system for querying, clustering and summarizing documents," *Information Processing and Management*, Vol. 43, No. 6, pp. 1588– 1605, 2007.
- [17] K. S. Jones, " Automatic summarizing: the state of the art," *Information Processing and Management*, Vol. 43, No. 6, pp. 1449– 1481, 2007.
- [18] D. Zajic, B. J. Dorr, J. Lin, and R. Schwartz, " Multi- candidate reduction: sentence compression as a tool for document summarization tasks," *Information Processing and Management*, Vol. 43, No. 6, pp. 1549– 1570, 2007.
- [19] G. Erkan and D. R. Radev, " Lexrank: Graph-based lexical centrality as salience in text summarization," *Journal of Artificial Intelligence Research*, Vol. 22, pp. 457– 479, 2004.
- [20] D. Radev, E. Hovy, and K. McKeown, " Introduction to the special issue on summarization," *Computational Linguistics*, Vol. 28, No. 4, pp. 399– 408, 2002.
- [21] D. Shen, J.-T. Sun, H. Li, Q. Yang, and Z. Chen, " Document summarization using conditional random fields," *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI' 07)*, Hyderabad, India, pp. 2862– 2867, January 6– 12, 2007.
- [22] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, " Conditional random fields: probabilistic models for segmenting and labeling sequence data," *Proceedings of the 18th International Conference on Machine Learning*, pp. 282– 289, 28 June– 01 July 2001.
- [23] X. Wan, " A novel document similarity measure based on earth mover' s distance," *Information Sciences*, Vol. 177, No. 18, pp. 3718– 3730, 2007.
- [24] S. Fisher and B. Roark, " Query-focused summarization by supervised sentence ranking and skewed word distributions," *Proceedings of the Document Understanding Workshop (DUC' 06)*, New York, USA, 8p, 8– 9 June 2006.
- [25] P. Fung and G. Ngai, " One story, one flow: Hidden Markov story models for multilingual multidocument summarization," *ACM Transaction on Speech and Language Processing*, Vol. 3, No. 2, pp. 1– 16, 2006.
- [26] D. M. McDonald and H. Chen, " Summary in context: Searching versus browsing," *ACM Transactions on Information Systems*, Vol. 24, No. 1, pp. 111– 141, 2006.
- [27] R. Mihalcea and P. Tarau, " A language independent algorithm for single and multiple document summarizations," *Proceedings of the Second International Joint Conference Natural Language*

- [28] J. Li, L. Sun, C. Kit, and J. Webster, " A query-focused multi-document summarizer based on lexical chains," Proceedings of the Document Understanding Conference (DUC' 07), New York, USA, 4p, 26– 27 April 2007.
- [29] X. Wan, " Using only cross-document relationships for both generic and topic-focused multi-document summarizations, Information Retrieval, Vol. 11, No. 1, pp. 25– 49, 2008.
- [30] R. Mihalcea and H. Ceylan, " Explorations in automatic book summarization, Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP- CoNLL' 07), Prague, Czech Republic, pp. 380– 389, 28– 30 June 2007.
- [31] Y. Guo and G. Stylios, " An intelligent summarization system based on cognitive psychology," Information Sciences, Vol. 174, No.1– 2, pp. 1– 36, 2005.
- [32] J. Grabmeier and A. Rudolph, " Techniques of cluster algorithms in data mining," Data Mining and Knowledge Discovery, Vol. 6, No. 4, pp. 303– 360, 2002.
- [33] J. Han and M. Kamber, " Data mining: concepts and technique (2nd Edition), Morgan Kaufman, San Francisco, 800p, 2006.
- [34] A. K. Jain, M. N. Murty, and P. J. Flynn, " Data clustering: A review," ACM Computing Surveys, Vol. 31, No. 3, pp. 264– 323, 1999.
- [35] M. G. H. Omran, A. P. Engelbrecht, and A. Salman, " An overview of clustering methods," Intelligent Data Analysis, Vol. 11, No. 6, pp. 583– 605, 2007.
- [36] K. M. Hammouda and M. S. Kamel, " Efficient phrase- based document indexing for web document clustering," IEEE Transactions on Knowledge and Data Engineering, Vol. 16, No. 10, pp. 1279– 1296, 2004.
- [37] T. Li, " A unified view on clustering binary data," Machine Learning, Vol. 62, No. 3, pp. 199– 215, 2006.
- [38] Y. Li, C. Luo and S. M. Chung, " Text clustering with feature selection by using statistical data," IEEE Transactions on Knowledge and Data Engineering, Vol. 20, No. 20, pp. 641– 652, 2008.
- [39] Y. Li, D. McLean, Z. A. Bandar, J. D. O' Shea, K. Crockett, " Sentence similarity based on semantic nets and corpus statistics," IEEE Transactions on Knowledge and Data Engineering, Vol. 18, No. 8, pp. 1138– 1150, 2006.
- [40] X. Liu, Y. Zhou, and R. Zheng, " Sentence similarity based on dynamic time warping," Proceedings of the First International Conference on Semantic Computing (ICSC' 07), Irvine, USA, pp. 250– 256, 17– 19 September 2007.
- [41] X. Wan, J. Yang, and J. Xiao, " Manifold-ranking based topic-focused multi-document summarization, Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI' 07), Hyderabad, India, pp. 2903– 2908, January 6– 12, 2007.
- [42] D. Bollegala, Y. Matsuo, and M. Ishizuka, " Measuring semantic similarity between words using web search engines," Proceedings of 16th World Wide Web Conference (WWW16), Alberta, Canada, pp. 757– 766, May 8– 12, 2007.
- [43] R. L. Cilibrasi and P. M. B. Vitanyi, " The google similarity distance," IEEE Transaction on Knowledge