International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Review on Multi Document Summarization Using Ontology

Rajshree S Hingane[1], Devendra P Gadekar[2]

[1]Pune University, Maharshtra, India

[2]Assistant Professor, Pune University, Maharshtra, India

Abstract: ***In Domain Ontology, as a theoretical model, gives a significant system for semantic representation of textual data. In this paper, we investigate the possibility of utilizing the ontology in illuminating multi-document summarization issues in the domain of disaster management. We give an experimental study of distinctive methodologies in which the ontology has been utilized for summarization tasks. Far reaching analyses on an accumulation of releases relevant applicable to Hurricane Wilma in 2005 exhibit those ontology-based multi-document summarization techniques beat different baselines regarding the summary quality.***

Keywords: Disaster management, multi-document summarization, ontology, query expansion, Generic**.**

1. Introduction

It is well known that hurricane, earthquakes, and other natural calamities causes' physical decimation, loss of life and property around the world. With a specific end goal to productively dissect the pattern of the disasters and minimize the resulting misfortune for future circumstance, successful data gathering methods are essential. In particular, a myriad of news and reports that are identified with the disaster may be recorded as text documents. The area specialists hope to condensed information about the detailed disaster event description, e.g., the developmental inclination of the disaster, the operational status of general society administrations, and the remaking procedure of the estate. In the accompanying, an agent situation is given, in which the data often times researched by a disaster examiner is described. In the area of disaster management, over a great many reports are regularly discharged by the local government then again local emergency offices amid the disaster, which cover most occasions significant to the disaster and the time compass will be days to months, contingent upon how serious the disaster is. The information will be displayed in a configuration of newswire, containing a great deal of routine investigating different parts of the disaster. In such a case, it is to a great extremely difficult for area specialists to rapidly find either the most imperative data in general (generic summarization) or the most applicable data to a predefined query (query/topic-focused summarization). Along these lines, summarization techniques can be utilized to concentrate important data from various reports.

A domain ontology identified with disaster management, depicting the concepts and the relating relations of these concepts, is regularly given by domain experts [1]. Such a ontology contains sample calculated data related to the document set, which may be helpful for users to summarize the reports. A characteristic inquiry is means by which we can use the ontology to acquire high-quality summaries, i.e. speaking to themes with non redundant sentences.

In this paper, we investigate the possibility of utilizing the ontology into multi-document summarization issues in calamity administration domain. We first talk about how to speak to a sentence as a vector utilizing the area ontology. We then derive into the issues from two directions: generic and query-focused summarization. In generic summarization, we give thorough investigations of the centroid-based sentence selection methodologies by utilizing different vector space models; also, investigate the likelihood of using the ontology to attain to the objective of decreasing information redundancy. In query-focused summarization, we advance the last synopsis comes about by utilizing ontology-based query expansion methods into the summarization. We direct analyses on a collection of press release identified with Hurricane Wilma, and the outcomes demonstrate that metaphysics based techniques can give promising execution for summarization.

The rest of the paper is organized as follows:
In Section II, we review the related work. In the wake of presenting the domain ontology connected in our work in Section III, we give a near study on a several ontology-based representations in Section IV. Section VI finishes up the paper with Conclusion and future work.

2. Related Work

A. Generic Summarization

For generic summarization, a saliency score is normally allotted to every sentence, the sentences are ranked agreeing to the saliency score, and afterward the top ranked sentences are chosen as the outline in light of the ranking result.

As of late, both supervised and unsupervised methods have been proposed to break down the data contained in a document set, and concentrate profoundly remarkable sentences into the outline taking into syntactic or statistical features [3]–[7]. For case, MEAD [8] is a usage of the centroid based strategy in which the sentence scores are figured based on sentence-level and inner sentence features.

However, most existing routines disregard the conceptual data in the sentence level. Much of the time, the conceptual data can give clients more meaningful results to summaries.

Licensed Under Creative Commons Attribution CC BY

A few scientists use the express ideas inside sentences to address multi-document summarization [9], [10], e.g., utilizing Wikipedia. Nonetheless, such procedures can't be straight forwardly connected to domain-specific document summarization, since Wikipedia contains an excess of ideas not pertinent to a particular area.

B. Query-Focused Summarization:

In query-focused summarization, the information identified with a given subject or query should be incepted into summaries, also, the sentences suiting the users announced data need should be separated. Generic summarization for nonexclusive outline can be reached out to consolidate the query information. Saggion et al. [11] displayed robust summarization system grown inside the GATE structural planning that makes utilization of vigorous segments for semantic labeling and co-reference determination gave by GATE. Wei et al. [12] joined the query influence into the shared support affix to adapt with the requirement for query-oriented multi-document summarization. Wan et al. [13] utilized both connections among sentences furthermore, relationships among sentences and relationships by complex positioning. Likelihood models have likewise been proposed with diverse suppositions on the era methodology of the reports and the queries [14], [15].

C. Query Expansion

Query expansion is the process of enlarging the user's query with additional terms in order to improve search results. For example, when we are prepared to pursuit "panther" by some web crawler, we can extend such question by including equivalent words of "panther" to the query, or instance, Daume and Marcu [16] propose a defended queries development strategy in the dialect displaying for IR framework. Nonetheless, it neglects to consider the semantic relatedness between the sentences and the query string.

In paper [17], Crisis Management and Disaster Recovery have increased massive significance in the wake of late man and nature dispensed calamities. A basic issue in an crisis situation is the way to efficiently find, gather, compose, search and scatter real time disaster information. In this paper, we address a few key issues which hinder better data sharing and collaboration effort between both private and open segment members for disaster management and recovery. We design and implement an online model of a Business Continuity Information Network (BCIN) framework using the most recent advances in information mining technologies to make easy to use, Internet-based, data rich services and going about as an essential piece of an organization's business coherence process. In particular, information extraction is used to incorporate the input data from different sources; the substance suggestion motor and the report synopsis module give clients customized and brief perspectives of the disaster information, the group era module creates spatial clustering techniques to help clients assemble dynamic community in disasters.

In paper [18], for semantic representation of text based data, domain ontology will give exceptionally valuable structure. If there should be an occurrence of multi-document summarization issues in the area of disaster management, the practicality of utilizing the ontology is attempted to explore. An empirical study of distinctive methodologies is given where to summarization task, ontology is utilized.

In paper [19], the vast majorities of the multi-document summarization methods decay the document into sentences and work straightforwardly in the sentence space utilizing a term-sentence matrix. Sometimes, the knowledge on the report side, i.e. the topics involved in the records, can help the context understanding and aide the sentence choice in the summarization technique. In this paper, we propose a new Bayesian sentence-based topic model for summarization by making utilization of both the term-record and term-sentence associations. An effective variational Bayesian algorithm is inferred for model parameter estimation. Exploratory results on benchmark information sets demonstrate the adequacy of the proposed model for the multi-record summarization task.

In paper [20], this paper exhibits another measure of semantic comparability in an is-a taxonomy in view of the thought of data substance. Trial assessment proposes that the measure performs enthusiastically well (a correlation of r = 0.79 with a benchmark set of human likeness judgments, with an upper bound of r = 0.90 for human subjects performing the same task), and essentially better than the traditional edge counting methodology (r = 0.66).

3. Disaster Management Domain

A. Domain Description

It is well known that hurricanes, earthquakes, and other natural calamities cause immense physical destruction, loss of life and property as far and wide as possible. The motivation behind the calamity administration project is to upgrade productive coordination and cooperation among open security associations by empowering the interoperable imparting of emergency alerts, incident related information between disparate systems. One of the disaster management systems means to investigate the news and reports identified with the disaster to concise and recapitulative information for area specialists.

B. Domain-Specific Ontology

As a rule, ontology is regularly given by space specialists in disaster management domain [1]. Such ontology gives answers to the inquiries concerning what elements exist in disaster management, and how such elements can be connected inside an hierarchy and subdivided as per similarities and differences among them. The ontology described in this paper is identified with the area of hurricane management, including 109 concept and 326 concept relations. This ontology is gotten from the disaster management venture at Florida International University[21](http://www.bizrecovery.org). The ontology is made for the motivation behind exploration included in this task, and is given by the space specialists from the State Emergency Operations Focus (EOC) 1 of Florida.

Licensed Under Creative Commons Attribution CC BY

4. Summarization Approaches

To address the summarization issues in the domain of hurricane management, we first guide most sentences in the record set onto the domain ontology, and after that take advantage of the intrinsic properties of the ontology to speak to each sentence. In this segment, we investigate the impact of the ontology in multi-document summarization assignments from two headings: generic summarization and query-focused summarization**.**

A. Sentence Mapping

Ontology in disaster management domain gives us rich theoretical and semantic information, which may encourage the system of multi-document summarization. To use the ontology for better comprehension the documents, we at first deteriorate the gathering of space particular reports into sentences and after ward outline sentence to the ontology order. For every idea of the ontology hierarchy, a gathering of keywords (i.e., noun) are assigned by the specialists for the purpose of sentence mapping. In this methodology, we compute the word set covering between a sentence and the keyword set allocated to every idea as the measure of relatedness, and rank the scores to choose the most related concept. Since distinctive concept in the ontology have diverse unambiguous agent noun sets allocated by area specialists, it is improbable that the same noun will show up in additional than one concept.

B. Sentence Representation

A key question in multi-document summarization utilizing the ontology is the way to speak to the sentences we have mapped onto the ontology. We analyze a few approaches to model a sentence into a vector, including term frequency(TF) model [22], term frequency opposite sentence recurrence (TFISF) model [22], term frequency-inverse sentence frequency (TFICF) model, concept hierarchy (CH) model [23], [24], what's more, the linear combinations of these models. The vector space models specified above give diverse experiences to archive summarization. In the accompanying, we subtle elements of these models with the end goal of comparison**.**

C. Generic Summarization

We apply the centroid-based methods to choose vital sentences as the summery. To do so, we run the standard k-Means on the sentence set, where the cluster number k is indicated as the number of concept in the main level of the ontology. The instinct is that the differing qualities of themes in the entire record accumulation ought to be confined in the idea set of the first level of the ontology. By and large talking, in the space of disaster management, the investigators frequently perform post occasion examination on diverse parts of a solitary disaster that they may be occupied with. By utilizing the ontology, we are capable to independent the sentences into diverse gatherings. Furthermore, the data not mapped straightforwardly to the ontology is separated out by the ontology of sentence mapping, which is not vital as far as post occasion investigation. To register the likeness of sentence pairs, we utilize the cosine similarity [25] taking into account the vector space models we talked about in Section IV-B.

5. Conclusion and Future Work

In this paper, we gave an exact study on a several approaches that use the ontology to illuminate diverse multi document summarization problems. For generic summarization, we utilized distinctive vector space models to speak to sentences in the document gathering, and investigated the achievability of diverse combination of the VSMs. At that point the centroid-based methods were used to group the sentence pairs and the imperative sentences close to the centroids of the sentence groups are removed. The last summary was produced by decreasing information redundancy and ranking sentences. For query summarization, we delved into the impact of query extension in summarization undertakings. The ontology is rich in applied data identified with the particular domain. We will continue chipping away at the issue of ontology-based multi document summarization, particularly document summarization tasks, i.e., redesign summarization and relative summarization. Another intriguing course is to investigate profoundly how to use the various leveled relationships in the cosmology to further enhance the nature of the outline and to perform various hierarchical text categorization [26], [27]. Furthermore, we will attempt to utilize data extraction methods to further enhance summarization results. We are additionally interested in extending our proposed strategy to the summarization utilizing public ontologies, for example, WordNet and Wikipedia.

References

[1] Lei Li and Tao Li "An Empirical Study of Ontology-Based Multi-Document Summarization in Disaster Management", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, VOL. 44, and NO. 2, FEBRUARY 2014
[2] E. Klien, M. Lutz, and W. Kuhn, "Ontology-based discovery of geographic information services—an application in disaster management, Comput" *Environ. Urban Syst.*, vol. 30, no. 1, pp. 102–123, 2006.
[3] H. Hsu, C. Tsai, M. Chiang, and C. Yang, "Topic generation for web document summarization," in *Proc. IEEE SMC*, 2008, pp. 3702–3707.
[4] X. Yong-dong, W. Xiao-long, L. Tao, and X. Zhi-ming, "Multi-document summarization based on rhetorical structure: Sentence extraction and evaluation," in *Proc. IEEE SMC*, 2008, pp. 3034–3039.
[5] D. Wang, T. Li, S. Zhu, and C. Ding, "Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization, "in Proc. *SIGIR*, 2008, pp. 307–314.
[6] G. Erkan and D. Radev, "Lexpagerank: Prestige in multi-document text summarization," in *Proc. EMNLP*, vol. 4, 2004, pp. 365–371.
[7] X. Wan and J. Yang, "Multi-document summarization using cluster based link analysis," in *Proc. SIGIR*, 2008, pp. 299–306.

Licensed Under Creative Commons Attribution CC BY

[8] D. Radev, H. Jing, M. Sty, and D. Tam, "Centroid-based summarization of multiple documents," *Inf. Process. Manage.* vol. 40, no. 6, pp. 919–938, 2004.
[9] V. Nastase, "Topic-driven multi-document summarization with encyclopedic Knowledge and spreading activation," in *Proc. EMNLP*, 2008, pp. 763–772.
[10] C. Lee, Z. Jian, and L. Huang, "A fuzzy ontology and its application to news summarization," *IEEE Trans. Syst., Man, Cybern., B Cybern.*, vol. 35, no. 5, pp. 859–880, Oct. 2005.
[11] H. Saggion, K. Bontcheva, and H. Cunningham, "Robust generic and Query-based summarization," in *Proc. ECAL*, 2003, pp. 235–238.
[12] F. Wei, W. Li, Q. Lu, and Y. He, "Query-sensitive mutual reinforcement chain and its application in query-oriented multi-document summarization," in *Proc. SIGIR*, 2008, pp. 283–290.
[13] X. Wan, J. Yang, and J. Xiao, "Manifold-ranking based topic focused Multi-document summarization," in *Proc. IJCAI*, 2007, pp. 2903–2908.
[14] J. Tang, L. Yao, and D. Chen, "Multi-topic based query-oriented summarization," in *Proc. SDM*, 2009.
[15] A. Highlight and L. Vanderwende, "Exploring content models for multi-document summarization," in *Proc. HLT-NAACL*, 2009, pp. 362–370.
[16] H.Daum´e and D. Marcu, "Bayesian query-focused summarization," in *Proc. ACL*, vol. 44, no. 1. 2006, p. 305.
[17] Li Zheng, Chao Shen, Liang Tang, Tao Li, Steve Luis, Shu-Ching Chen, Vagelis Hristidis," Using Data Mining Techniques to Address Critical Information Exchange Needs in Disaster Affected Public-Private Networks", 11200 S.W. 8th Street, Miami, Florida, 33199, U.S.A
[18] Pranjali Avinash Yadav-Deshmukh, R. Ambekar," Survey on Multi-Document Summarization in Disaster Management based on Ontology", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Impact Factor (2012): 3.358
[19] Dingding Wang, Shenghuo Zhu Tao Li, Yihong Gong," Multi-Document Summarization using Sentence-based Topic Mode" NEC Laboratories America, Cupertino, CA 95014, USA.
[20] Philip Resnik," Using Information Content to Evaluate Semantic Similarity in a Taxonomy", Sun Microsystems Laboratories Two Elizabeth Drive Chelmsford, MA 01824-4195 USA.
[21] L. Zheng, C. Shen, L. Tang, T. Li, S. Luis, S. Chen, and V. Hristidis,"Using data mining techniques to address critical information exchange needs in disaster affected public-private networks," in *Proc. SIGKDD*,2010, pp. 125–134.
[22] D. Jurafsky, J. Martin, A. Kehler, K. Vander Linden, and N. Ward, *Speech and Language Processing*. Englewood Cliffs, NJ, USA: Prentice- Hall, 2000.
[23] S.-T. Yuan and J. Sun, "Ontology-based structured cosine similarity in speech document summarization," in *Proc. WI*, 2004, pp. 508–513.
[24] S. Yuan and J. Sun, "Ontology-based structured cosine similarity in document summarization: With applications to mobile audio-based knowledge management," *IEEE Trans. Syst., Man, Cybern, B Cybern*, vol. 35, no. 5, pp. 1028–1040, Oct. 2005.
[25] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Boston, MA, USA: Pearson Addison Wesley, 2006.
[26] T. Li, S. Zhu, and M. Ogihara. "Text categorization via generalized discriminant analysis," *Inf. Process. Manage.*, vol. 44, no. 5, pp. 1684–1697,2008.

Licensed Under Creative Commons Attribution CC BY