ARTICLE IN PRESS

Information Sciences xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Automatically generating the weather news summary based on fuzzy reasoning and ontology techniques

Shyi-Ming Chen *, Ming-Hung Huang

Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 7 August 2013
Received in revised form 4 March 2014
Accepted 11 April 2014
Available online xxxx

Keywords:
Weather news summary
Domain ontology
Fuzzy reasoning
Fuzzy rule

ABSTRACT

In this paper, we present a new method for automatically generating the weather news summary based on fuzzy reasoning and ontology techniques, where the weather ontology, the time ontology and the geography ontology are predefined by domain experts. We slice the original weather news articles into a set of terms. Then, we use two ontological features (i.e., the degree of depth of the ontology and the degree of width of the ontology) and one statistical feature (i.e., frequency) as inputs to the system. The values of those features are represented by fuzzy sets. Then, the fuzzy reasoning algorithm infers the score of each sentence. The summary is composed of candidate sentences which have higher scores, where the experimental data are adopted from the weather news website of Taiwan. The experimental results show that the proposed method outperforms the methods presented in [14,15] for automatically generating the weather news summary.

1. Introduction

In recent years, some researchers focused on the research topic of text summarization [1,3,6,8–12,14,15,18–21,23–25,27]. In [1], Abuobieda et al. presented a text summarization features selection method using the pseudo genetic-based model. In [3], Chandra, Gupta and Paul presented a statistical approach for automatic text summarization by extraction. In [8], Chen and Verma presented a user-query-based text summarization system. In [9], Devasena and Hemalatha presented a method for automatic text categorization and summarization using rule reduction. In [10], Jiang presented a method for automatic text summarization based on keyword extraction. In [11], Jiang et al. presented a method for automatic summarization of text based on combined words recognition and paragraph clustering. In [12], Kaikhah presented a machine learning approach which uses artificial neural networks to produce the summaries of arbitrary lengths of news articles. In [18], Manne et al. presented a novel automatic text summarization system with feature terms identification. In [20], Mohamed presented a method for generating summaries by identifying important concepts/entities and relations in the text. In [21], Nguyen et al. studied the impact of user-cognitive styles on the assessment of text summarization. In [23], Thakkar et al. presented graph-based algorithms for text summarization. In [24], Yuan and Sun presented a novel method, named Structured Cosine Similarity, to deal with speech document clustering. In [25], Yang et al. presented a novel text summarization approach using sentence extraction based on kernel words recognition. In [27], Zhang and Li presented a method for automatic text summarization based on sentences clustering and extraction.

* Corresponding author. Tel.: +886 2 27376417; fax: +886 2 27301081.
E-mail address: smchen@mail.ntust.edu.tw (S.-M. Chen).

http://dx.doi.org/10.1016/j.ins.2014.04.027
0020-0255/

In [14], Lee, Chen and Jian presented an ontology-based fuzzy event extraction (OFEE) agent for e-news summarization, where the OFEE agent contains a retrieval agent, a document processing agent and a fuzzy inference agent to perform the event extraction for e-news summarization. The retrieval agent automatically retrieves Internet e-news periodically, stores them into the e-news repository, and sends them to the document processing agent for document processing. Then, the document processing agent utilizes the Part-of-Speech (POS) tagger provided by the knowledge information processing group for processing the retrieved e-news and filters the term set by a term filter. Then, the fuzzy inference agent and the event ontology filter extract the e-news event ontology based on the term set and the domain ontology. Finally, the summarization agent summarizes the e-news by the extracted-event ontology. In [15], Lee, Jian and Huang presented a seven-layer fuzzy reasoning mechanism to generate the weather news summary, where the domain ontology with various events of news is predefined by domain experts. The document preprocessing mechanism generates meaningful terms based on the news corpus and the news dictionary defined by the domain expert. Then, these meaningful terms are classified based on the events of the news by the term classifier. The fuzzy inference mechanism generates the membership degrees for each fuzzy concept of the fuzzy ontology. A news agent based on the fuzzy ontology is developed for news summarization, where the news agent contains a retrieval agent, a document preprocessing mechanism, a sentence path extractor, a sentence generator and a sentence filter for performing news summarization. However, because the performance of the methods presented in [14,15] is not good enough, we must develop a better method to overcome the drawbacks of the methods presented in [14,15].

In this paper, we present a new method for automatically generating the weather news summary based on fuzzy reasoning and ontology techniques, where the weather ontology, the time ontology and the geography ontology are predefined by domain experts. We use traditional ontology techniques [2,5,7,19,22,33,28] for knowledge representation, where the values of the attributes of concepts are crisp values. We slice the original weather news articles into a set of terms. We use two ontological features (i.e. the degree of depth of the ontology and the degree of width of the ontology) and one statistical feature (i.e. frequency) as inputs to the system. Then, the fuzzy reasoning algorithm infers the score of each sentence in each news item, where Mamdani's Max–Min operations for fuzzy reasoning [13] are used based on 27 fuzzy rules. Based on the weather ontology, the time ontology and the geography ontology, the depth, the width and the frequency of each sentence can be obtained. Then, the depth, the width and the frequency of the sentences are used as the input values of the fuzzy rules to perform fuzzy reasoning to get the score of each sentence. If the score of a sentence in a news item is larger than or equal to a predefined threshold value determined by the domain expert, then it is a candidate sentence. After performing the fuzzy reasoning process, we can get a set of candidate sentences which have higher scores as a news summary. The experimental data are adopted from the weather news website of Taiwan (http://tol.chinatimes.com/CT_NS/CTSearch.aspx). The experimental results show that the proposed method outperforms the methods presented in [14,15] for automatically generating the weather news summary. The novelty of the proposed method outperforming the methods presented in [14,15] is that it is based on fuzzy reasoning and ontology techniques, where a fuzzy reasoning algorithm is proposed and 27 fuzzy rules are used for fuzzy reasoning to infer the score of each sentence in a news item. The advantage of the proposed method is that it gets a better performance for weather news summary than the methods presented in [14,15].

The rest of this paper is organized as follows. In Section 2, we briefly review the concepts of ontology and the fuzzy set theory [26]. In Section 3, we present a new method for weather news summarization based on fuzzy reasoning and the ontology techniques. In Section 4, we make a comparison of the experimental results of the proposed method with the ones of the methods presented in [14,15]. The conclusions are discussed in Section 5.

2. Preliminaries

In the following, we describe the domain ontologies used in this paper. Fig. 1 shows the Weather news ontology. It is composed of three domain ontologies, i.e., the Weather ontology, the Geography ontology and the Time ontology. Fig. 2, Fig. 3 and Fig. 4 show the Weather ontology, the Geography ontology and the Time ontology, respectively, which are predefined by the domain expert. There are two categories of nodes in a domain ontology. The node on the top of the ontology is the domain name. The other nodes are the concepts in each domain. Each concept contains attributes which are utilized to extend the terms. Sometimes, we utilize different words to describe the same concept for keyword expansion. For example, the concept "Taipei" contains three attributes "Taipei City", "City of Taipei" and "Taipei" in the Geography ontology, which describe the same place in Taiwan. The other type of keyword expansion is that they have different meaning but the same degree. For instance, the concept "Year" has three attributes, "Last Year", "This Year" and "Next Year", respectively. Three kinds of relationships are utilized to express relationships between nodes. The generalization denotes the "is-kind-of" relationship; the aggregation denotes the "is-part-of" relationship; the arrow associated with the term "has" denotes the "possess" relationship.

In 1965, Zadeh proposed the theory of fuzzy sets [26]. Fig. 5 shows a triangular fuzzy set A in the universe of discourse U with the membership function μ_A parameterized by a triplet (a, b, c), where $A = (a, b, c)$.

Let A and B be two fuzzy sets in the universe of the discourse U and let the membership functions of the fuzzy sets A and B be μ_A and μ_B, respectively. The intersection between the fuzzy sets A and B is defined as follows:

$$\mu_{A\cap B}(x) = \min\{\mu_A(x), \mu_B(x)\}, \forall x \in U. \quad (1)$$

Please cite this article in press as: S.-M. Chen, M.-H. Huang, Automatically generating the weather news summary based on fuzzy reasoning and ontology techniques, Inform. Sci. (2014), http://dx.doi.org/10.1016/j.ins.2014.04.027