Computers & Industrial Engineering 91 (2016) 30–41

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Feature extraction using rough set theory in service sector application from incremental perspective

Chun-Che Huang [a], Tzu-Liang (Bill) Tseng [b,*], Chia-Ying Tang [a]

[a] Department of Information Management, National Chi Nan University, Taiwan, No. 1, University Road, Puli, Nantou 545, Taiwan, ROC
[b] Department of Industrial Manufacturing and Systems Engineering, The University of Texas at El Paso, 500 West University Avenue, El Paso, TX 79968, United States

ARTICLE INFO

Article history:
Received 15 April 2014
Received in revised form 16 August 2015
Accepted 21 September 2015
Available online 28 September 2015

Keywords:
Rough set theory
Incremental algorithm
Incremental object
Rule induction
Feature selection
Service science

ABSTRACT

In service industry application, there is vague and qualitative information required to be processed properly, for example, to identify customer preferences in order to provide adequate services. From literature, Rough Set Theory (RST) has been indicated to be one of promising approaches to cope with vagueness in a large scale database. Basically, the rough set approach integrates learning-from-example techniques, extracts rules from a data set of interest, and discovers data regularities. Most of the existing RS based approaches are able to implement rule induction but it is very time consuming from computation perspective particularly from a large database. To date, there is a demand to generate and analyze business decision rules based on dynamical data sets and conclude such rules on the daily basis in the service industry. Therefore, in this study, an Incremental Weight Incorporated Rule Identification (IWIRI) algorithm is proposed to fulfill such demand. The proposed approach is proficient to efficiently process in-coming data (objetcs) and generate updated decision rules without re-computation efforts in the database. Identification of features based on the customer's preference and implementation of the proposed algorithm are summarized in the case study. This paper forms the basis for solving many other similar problems that occur in service industries.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Over the past decades, services have become the largest part of most industrialized nations' economies (Spohrer, Maglio, Bailey, & Gruhl, 2007). The explosive growth of services in economies around the world has vast implications for business practice, academic knowledge creation and education. Service can be defined as the application of competences for the benefit (Lusch & Vargo, 2006), and broadly considered as an application of specialized knowledge, skill, and experience, performed for co-creation of respective values of both consumer and provider (Lusch & Vargo, 2006; Qiu, Fang, Shen, & Yu, 2007; Spohrer et al., 2007).

The real value of a delivered service lies in its ability to satisfy customer's needs which are not simply and strictly shown from technical characteristics of service industry. It is essential to meet customer's needs from competitive business perspective. Particularly, a service significantly affects customer's satisfaction. The understanding is also a necessity for a systemic decision-making on how the service system should be transformed to improve customer satisfaction or for a competitive advantage (Qiu, 2009).

For the enterprise to provide their services effectively, they must first understand the customer's thinking and preferences. However, with the great number of customers, how do we understand and identify their interests? The answer to this question is to build personalized service to explore possible solutions (Yu, 1999) and it can be practiced through understanding of customer preference. For example, in aviation service industry, through customer's satisfaction investigation including preference identification relevant to cabin facility, ticket price, safety record, connection service, food and beverage etc., it can be conducted to elicit particular features when they plan to book the flight for travel.

Obviously, feature selection is a core and effective approach for exploring the critical features from customers. And the information relevant to the customer-preferred features is collected through research, interviews, meetings, questionnaires, sampling, and other techniques. These types of data are often discretized and are frequently in "qualitative" format (e.g. salary level, preference level, etc.). Numerous approaches have been applied to feature selection, e.g. generic algorithms (da Silva, Traina, Ribeiro, Batista, & Traina, 2009; Lee, Shin, & Chung, 2006; Van Coillie,

* Corresponding author.
E-mail addresses: cchuang@ncnu.edu.tw (C.-C. Huang), btseng@utep.edu (T.-L. (Bill) Tseng), s94213012@ncnu.edu.tw (C.-Y. Tang).

http://dx.doi.org/10.1016/j.cie.2015.09.011
0360-8352/© 2015 Elsevier Ltd. All rights reserved.

Frieke, Verbeke, & Wulf, 2007), artificial neural network (ANN) (Chakrabarti, 2008; Glezakos et al., 2009; Saravanan & Ramachandran, 2010), Tabu Theory (Wang, Guo, & Wang, 2010, 2009), branch-and-bound algorithm (Chen, 2003; Nakariyakul & Casasent, 2007), Fuzzy C-means Algorithm (Francesco, 2003) and SOM (Laaksonen, Koskela, & Oja, 2004). However, these approaches are not used for processing qualitative information. They are not suitable for feature selection based on customer's preference when enterprise wish to service them because the aforementioned methodologies are population-based approaches which may require several statistical assumptions and they have limitations in handling qualitative data (Kusiak, 2001).

One of individual object model-based approaches to deal with qualitative information is the rough set approach (Kusiak, 2001). The rough set theory (RST) is fundamental importance in artificial intelligence (AI) and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, expert systems, decision support systems, inductive reasoning, and pattern recognition (Pawlak, 1991). The rough set approach is suitable for processing qualitative information that is difficult to analyze by standard statistical techniques (Heckerman, Mannila, Pregibon, & Uthurusamy, 1997). It integrates learning-from-example techniques, extracts rules from a data set of interest, and finds data regularities (Komorowski & Zytkow, 1997). To select desired features, the rough set approach attempts to eliminate as many features as possible in the problem domain, and still obtain useful and meaningful outcomes with acceptable accuracy. Having a minimal number of features often leads to establish simple models that can be more easily interpreted. Tseng and Huang (2007) proposed the integer-programming methodology and WIRI algorithm to support rule induction more effectively. Moreover, it is also able to handle several conditions, like weights assigned to different features with different given objects. Through the weight analysis, the features are highly correlated to customers' characteristics and the preferences can be are identified.

To date, the current rough set based approaches, such as the WIRI algorithm and other non-classic RS-based algorithm, for example Dominance-based Rough Set Approach (Greco, Matarazzo, & Slowinski, 2001, 2002) and its extension Believable Rough Set Approach (Chai & Liu, 2014) are capable to generate a set of decision rules efficiently, but they are not able to generate rules incrementally when new data (i.e., objects) are provided. In practice, the volume of data, big data (Lynch, 2008; Manyika et al., 2011) is frequently increased dynamically (Guo, Wang, Wu, & Yan, 2005). To obtain updated decision rules from new data sets (i.e., the original and additional data), we have to elaborate computation efforts to re-calculate reducts based on the new data sets. It should consume significant computation time and memory space, and therefore efficiency of these algorithms is very low (Liu, Xu, & Pan, 2004). Since this is a practical issue in the real world,
an efficient incremental technique is desired to solve this problem without re-implementing the algorithm from a dynamic database.

From literatures, it is concluded that (1) the traditional RS approaches required to re-implement the algorithm through analyzing the whole database including the incremental and original data is frequently observed and (2) inefficiency that the traditional RS method often inducts features with too many rules without focus, and cannot produce rules containing preference order. Therefore, this paper proposes models of the integer-programming and an incremental weight incorporated rule identification (IWIRI) method. The methodology is able to achieve the following objectives:

- Search the minimal number of feature rules for decision marking when the new data is add-in without re-compute the entire data set.
- Aggregate the weight of the feature and the frequency of the object to search for the optimal rules under the situation when new objects are given.
- Identify outcomes and significant features simultaneously.
- Validate the superiority of the proposed approach by a case study related to the aviation domain. In this case study, the IWIRI algorithm is used and allows managers handle and classify the new customers' data more efficient.

The remainder of this paper is organized as follows: Section 2 presents the integer-programming models to the incremental rough set-based rule induction problem due to provision of new objects. Particularly, confliction and converge associated with the original rules are incorporated to various solution cases. In Section 3, the WIRI algorithm and the IWIRI algorithm are developed to resolve limitation of the IP models. A case study in Section 4 is illustrated application of market analysis of the airline customer through the use of the proposed methodology. Section 5 concludes the paper.

2. The incremental rough set-based rule induction problem

Given by a set of data and their associated reducts, this section models the incremental problem when a new set of data are added in. Through implementing the exhaustive search procedure, five types of results are classified:

Case I The object's increment does not cause any conflict with original rules, and the original rules could cover the new object.

Case II The new object's increment does not cause any conflict with original rules, but the original rules could not cover this new object.

Case III Increment of the new object causes conflict with the original rules, but it could be covered by one of the original reducts/rules.

Case IV Increment of the new object causes conflict with original rules, and the new object cannot be covered by one of the original reducts.

Case V Increment of the new object causes conflict with the original rule, and new object's features are identical with one of the original objects, but the two objects have different output.

Note that "conflict" refers to that the new object has same feature as the original rules, but the new object has different output with the original rules. "The original rule covers the new object" refers to that the new object fits to the original rules in the condition (i.e., premise) part.

Table 1
The incremental object in Case V.

Object no.	F_1	F_2	F_3	F_4	Output
1	0	2	1	2	2
2	1	2	1	0	2
3	1	0	2	0	1
4	0	1	2	1	0
5	0	1	2	1	1

Download English Version:

https://daneshyari.com/en/article/1133492

Download Persian Version:

https://daneshyari.com/article/1133492

Daneshyari.com