A personalized trustworthy seller recommendation in an open market

Seungsup Lee [a], Keunho Choi [b,*], Yongmoo Suh [b]

[a] LS Global Inc., Sanbon 1-Dong, Gunpo-Si, Gyeonggi-Do, Republic of Korea
[b] Business School, Korea University, Anam-Ro 145, Seongbuk-Gu, Seoul 136-701, Republic of Korea

ARTICLE INFO

ABSTRACT

Although more and more customers are buying products on online stores, they have a difficulty in selecting a both *trustworthy* and *suitable* seller who sells a product they want to buy since there is a plenty number of sellers who sell the same product with different options. Therefore, the objective of this research is to propose a personalized trustworthy seller recommendation system for the customers of an open market in Korea. To that end, we first developed a module which classifies sellers into trustworthy one or not using a classification technique such as decision tree, and then developed another module which makes use of the content-based filtering method to find best-matching top *k* sellers among the selected trustworthy sellers. Experimental results show that our approach is worthwhile to take. This study makes a contribution at least in that to our knowledge it is the first attempt to recommend sellers, not products as done in most other studies, to customers.

Crown Copyright © 2012 Published by Elsevier Ltd. All rights reserved.

1. Introduction

The wide spread of Internet infrastructure facilitated e-transactions all around the world since early 1990s. Especially in Korea, this type of phenomenon has been observed since then. According to OECD reports, the percentage of the households which have access to the internet is 80%, the deployment ratio of DSL (Digital Subscriber Line) is about 99.5%, and that of broadband cable is approximately 57%, in Korea. Consequently, such high coverage of infrastructure stimulated e-commerce transactions more than in other countries.

Behind the ever-expanded e-commerce industry, there have been new types of problem. That is, relatively low cost of starting a new business in e-commerce would cause more fierce competition among sellers, while providing more choices to the customers (Choi, Stahl, & Whinston, 1997). Therefore, the participants of online transactions can commonly face with significant problems such as "trust" (Chen & Barnes, 2007; Hoffman, Novak, & Peralta, 1999; McKnight & Chervany, 2001) and "information overload" (Mooney & Roy, 2000; Yu, Schwaighofer, Tresp, Xu, & Kriegel, 2004).

In order to mitigate the trust problem in online shopping, several approaches have been proposed. For example, the online shopping service providers such as eBay and Amazon suggested reputation systems to promote more honest transactions between sellers and customers (Dellarocas, 2003). The reputation systems allow customers to share their opinion and experiences on products, services, and the online shopping mall through a large-scale of word-of-mouth (Dellarocas, 2006). However, the reputation systems are so simple that most of them fail to consider the collusive attempts by some fraudulent sellers who increase their own ratings (Wang & Chiu, 2008).

In order to deal with information overload problem, on the other hand, fundamental techniques have been studied to develop recommendation systems, including content-based filtering (CBF) (Lang, 1995; Mooney & Roy, 2000; Pazzani & Billsus, 1997) and collaborative filtering (CF) (Yu et al., 2004). Those systems were developed mainly to recommend products. However, recommending sellers who are *trustworthy* and *suitable* to each customers has been addressed by few researches, although it is usual that customers have difficulty in finding trustworthy sellers suitable to them in e-malls. Therefore, it would be very helpful to customers if a few trustworthy sellers suitable to each of them are recommended when they want to buy a product, especially in an open market.

The objective of this research is two-fold. One is to evaluate the sellers registered on an open market by classifying them as either trustworthy or not using a classification technique such as decision tree, and to select only trustworthy sellers. The other is to propose a recommendation system which recommends best-matching top *k* sellers to a target customer among the selected trustworthy sellers based on the CBF method.

The rest of this paper is organized as follows. Section 2 presents literature review regarding classification model, CBF, and seller recommendation. Section 3 describes the seller recommendation system proposed in this paper. Section 4 shows our experimental results from the recommendation system. The last section concludes this paper with a summary, implications and limitations of this paper.

* Corresponding author. Tel.: +82 2 3290 1945; fax: +82 2 922 7220.
E-mail address: keunho@korea.ac.kr (K. Choi).

2. Previous works

2.1. Classification model

A number of classification techniques have been developed and used for various purposes such as classification, estimation or prediction, including decision tree, artificial neural network, Bayesian network, support vector machine, etc. They have been applied to various domains including finance, marketing, health, and so on.

For example, credit scoring, which has been widely used especially in finance domain, was designed to classify a loan applicant' credit status into either "good" or "bad" using such information as marital status, income, and age (Chen & Huang, 2003). It could be adopted to make a decision on loaning (West, 2000). During the early years in that domain, linear discriminant analysis was the first model used for credit scoring (Hsieh, 2004; Hsieh, 2005). Due to its shortcomings (Chen & Huang, 2003), many studies have been later conducted to develop more sophisticated ways of modeling such as logistic regression, kernel density estimation, k-nearest neighbor, decision tree, and artificial neural network (West, 2000). Among them, data mining techniques such as decision tree and artificial neural network have been widely used for the purpose of credit evaluation because of their accuracy of classifying (Witten & Frank, 2005).

There are a number of researches in which data mining techniques were used for the purpose of credit evaluation in different domains such as loan, bank customer evaluation, and bankruptcy prediction (Hsieh, 2004; 2005; Hu, Patuwo, & Indro, 1999; Ong, Huang, & Tzeng, 2005; Zhang, Hu, Patuwo, & Indro, 1999). Zhang et al. (1999) built artificial neural network model for bankruptcy prediction, and compared its performance with that of logistic regression model. Ong, Huang, and Tzeng (2005) suggested a credit scoring model based on genetic algorithm.

We built a classification model using a data mining technique, for the purpose of classifying sellers as either trustworthy or not, which is similar, in some sense, to the credit scoring method. Since decision tree algorithm has been widely used in many domains due to its simplicity and accuracy, we used it to build a classification model to identify trustworthy sellers. We then recommend some of the sellers to each customer using CBF technique. That is, we recommend them, if they deal with the product the customer wants to buy and if historical records of sellers and those of customers show high similarity in terms of some characteristics of products.

2.2. Content-based filtering

With the wide spread of Internet technology in early 1990's, web-based recommendation systems have been developed mostly taking either collaborative filtering (CF) or content-based filtering (CBF) approach. CF approach recommends items based on the ratings of like-minded people, while CBF recommends items based on the similarity between items to recommend and items already purchased. Thus, CF requires ratings of customers on items and CBF requires item profiles, showing their characteristics. Some recommendation systems take advantages of both approaches. Since our data does not satisfy the requirements for CF, we decided to take CBF approach to recommending sellers to customers.

So far, there have been many recommendation systems developed taking CBF approach. They were used to recommend books (Mooney & Roy, 2000), music (Kodama et al., 2005), news (Lang, 1995), web pages (Pazzani & Billsus, 1997) and supermarket product (Lawrence, Almasi, Kotlyar, Viveros, & Duri, 2001). Most of these systems recommend various kinds of products, but few systems recommend sellers of open markets in spite of the growing importance of seller recommendation systems in the e-commerce environment, where customers usually do not have trust in some sellers.

2.3. Seller recommendation

To our knowledge, there have been only a few studies on recommending sellers. Xu and Zhang (2009) suggested seller credit evaluation method based on analytic hierarchy process (AHP) and set pair analysis (SPA). The advantage of their method is the fact that they suggested a way of using both qualitative and quantitative indicators to define a unified score. This method, however, has a limitation in that it cannot control the malicious evaluations on sellers caused by several customers. And recently, Wang and Chiu (2008) described a way of recommending trustworthy sellers using social network analysis. By the structure of social relationship of sellers which was made up of historical transactional data between traders, they suggested a collaboration-based recommendation system and attempted to lessen the problem of collusive or fraudulent rating for sellers.

As mentioned above, although a few studies have been conducted to recommend trustworthy sellers, few studies have been conducted to develop a system which recommends *trustworthy* sellers *suitable to each customer*. In this paper, we intend to develop such a recommendation system, using CBF technique, in an open market in Korea.

3. The proposed seller recommendation system

This section presents the overview of the recommendation system proposed in this paper, followed by a detailed description of each step of the system.

3.1. System overview

The overall framework of our proposed seller recommendation system consists of two main modules (i.e., classification and recommendation modules), as shown in Fig. 1.

First, in classification module, we build a classification model using the features selected from a table which integrated customer table, transaction table and seller table, and then identify trustworthy sellers using the classification model, with the assumption that the degree of customer's purchase satisfaction determines whether the sellers are trustworthy or not. That is, sellers are believed to be trustworthy if they have provided customers with good purchase satisfaction, since high satisfaction forms a customer's trust in sellers (Kim, Ferrin, & Rao, 2003).

In recommendation module, profiles of both customers and the identified trustworthy sellers are built, and then, the similarity between the profiles of a target customer and each seller who deals with items the target customer want to buy are calculated. Finally, the top k sellers whose profiles are most similar to that of target customer are recommended. Detailed explanation on each module is provided in following sections.

3.2. Classification module

Our proposed system aims to recommend trustworthy sellers who provides product or services that each customer wants. To that end, we first built a classification model which classifies sellers into one of two classes, trustworthy and untrustworthy.

In order to select features relevant to our classification analysis, we reviewed the previous researches which examine the factors affecting the trust of e-sellers. Through the literature review, we