Information Sciences 180 (2010) 2095–2123

Contents lists available at ScienceDirect

Information Sciences

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

A probabilistic reputation model based on transaction ratings ☆

François Fouss [a,*], Youssef Achbany [b], Marco Saerens [b]

[a] Management Department – LSM, Facultés Universitaires Catholiques de Mons (FUCaM), Belgium
[b] Information Systems Research Unit (ISYS) – LSM, Université Catholique de Louvain (UCL), Belgium

ARTICLE INFO

Article history:
Received 4 August 2008
Received in revised form 22 September 2009
Accepted 14 January 2010

Keywords:
Reputation
Trust
Rating prediction
Ratings aggregation

ABSTRACT

This work introduces a probabilistic model allowing to compute reputation scores as close as possible to their intrinsic value, according to the model. It is based on the following, natural, consumer–provider interaction model. *Consumers* are assumed to order items from *providers*, who each has some intrinsic, latent, "*quality of service*" score. In the basic model, the providers supply the items with a quality following a normal law, centered on their intrinsic "quality of service". The consumers, after the reception and the inspection of the item, rate it according to a linear function of its quality – a standard regression model. This regression model accounts for the *bias* of the consumer in providing ratings as well as his *reactivity* towards changes in item quality. Moreover, the *constancy* of the provider in supplying an equal quality level when delivering the items is estimated by the standard deviation of his normal law of item quality generation. Symmetrically, the *consistency* of the consumer in providing similar ratings for a given quality is quantified by the standard deviation of his normal law of ratings generation. Two extensions of this basic model are considered as well: a model accounting for truncation of the ratings and a Bayesian model assuming a prior distribution on the parameters. Expectation-maximization algorithms, allowing to estimate the parameters based on the ratings, are developed for all the models. The experiments suggest that these models are able to extract useful information from the ratings, are robust towards adverse behaviors such as cheating, and are competitive in comparison with standard methods. Even if the suggested models do not show considerable improvements over other competing models (such as Brockhoff and Skovgaard's model [12]), they, however, also permit to estimate interesting features over the raters – such as their reactivity, bias, consistency, reliability, or expectation.

1. Introduction

1.1. General introduction

The Internet has created a lot of new opportunities to interact with strangers. These interactions involve various kinds of applications, such as on-line markets. After just three years of business, eBay had already conducted over one billion dollars in transactions in 1998, and by 2008 [24] this figure had climbed to $59.6 billion. Moreover, a total of 405.3 million users either bid or listed an item on eBay during 2008. With the growth of on-line markets comes an increasing need for bidders and sellers to engage in transactions with counterparts with whom they have had little or no previous interaction.

☆ This manuscript is the authors' original work and has not been published nor has it been submitted simultaneously elsewhere. All authors have checked the manuscript and have agreed to the submission.

* Corresponding author. Tel.: +32 65 32 32 16.
E-mail addresses: francois.fouss@fucam.ac.be (F. Fouss), youssef.achbany@ucLouvain.be (Y. Achbany), marco.saerens@ucLouvain.be (M. Saerens).

0020-0255/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2010.01.020

This new type of market has introduced a new risk dimension to traders: the winner of the auction might not deliver payment, the seller might not deliver the item, or the delivered item might not be as the seller described [36,63]. These risks are an important restraint to the growth of on-line markets. One of the principal means by which on-line auction sites try to mitigate these risks associated with exchange among strangers is to use electronic reputation or feedback mechanisms (see, e.g., [64] for a detailed description of the eBay reputation system; see [84,85] for a discussion of desiderata that are specific for on-line reputation systems). In other words, reputation or feedback mechanisms aim at providing the type of information available in more traditional close-knit groups, where members are frequently involved in one another's dealings.

More generally, as discussed for instance in [8,20,80], on-line markets are often asymmetric information problems: the service provider has private information about product characteristics whereas consumers only have a vague idea. Indeed, consumers cannot directly observe the product quality – they only observe it when buying and receiving it. This asymmetric information may result in undesired behavior, such as a breakdown of the market or the so-called lemons problem [3,36]. Therefore, providers (or firms) have to set up signaling instruments/devices to overcome this issue and to establish a trust relationship with the consumer. Such instruments have been studied in marketing and economics – they try to facilitate the communication of the product quality or increase its reputation. They are of various types [8], like advertising, providing free products that consumers can test, contracting a certification label (ISO), offering warranties, branding, providing a reputation mechanism, etc. Indeed, as already observed in the middle age, the sharing of reputation seriously lowers the ability of the dishonest agent to make profit in the future [32,80]. In other words, reputation mechanisms help the community learn the true quality of the product by acting as signaling devices [20]. This is the main incentive for designing on-line reputation systems allowing to compute, using ratings provided by consumers, providers' reputation scores that are *as close to their intrinsic values as possible* – this is precisely the aim of the present work. Moreover, as discussed later in this general introduction (see *Reputation in on-line markets* paragraph), there is empirical evidence [9] showing that a reputation mechanism induces quite a substantial improvement in transaction efficiency in on-line markets. Notice however that there remain technical issues with reputation mechanisms in digital markets. For instance, there is usually no real incentive for on-line traders to share one's own experience-based information with others and on-line traders can change their identities as often as they wish [80].

Two fundamental aspects of reputation systems are the presence of communication protocols (allowing participants to provide ratings about transaction partners as well as to obtain reputation scores of potential transaction partners), and a reputation computation model to derive aggregated scores for each participant, based on received ratings and possibly also on other information. Our work is devoted to the definition of new reputation computation models. The last part of this section introduces (1) various fields where reputation mechanisms have been studied before being applied in on-line markets, (2) important considerations of on-line market reputation systems, and (3) several works related to the development of reputation models (Section 1.2).

1.1.1. Before on-line markets

Before being studied in such on-line markets, reputation and trust have become important topics of research in many fields, as shown in this section.

As already mentioned in the general introduction, reputation has long been of interest to *economists* [8,20,51]. Kreps et al. use reputation to explain the cooperation observed in experimental studies of prisoners' dilemma game [45]. Economic theory indicates that there is a balance between the cost of establishing a good reputation and the financial benefit of having a good reputation, leading to an equilibrium [41]. Variations in the quality of services or goods can be a result of deliberate management decisions or uncontrolled factors, and whatever the cause, the changes in quality will necessarily lead to variations in reputation. Although a theoretic equilibrium exists, there will always be fluctuations, and it is possible to characterize the conditions under which oscillations can be avoided or converge towards the equilibrium [71].

Scientometrics [57], referring to the study of measuring research outputs such as journal impact factors, also used the notion of reputation. In this context, reputation scores are related to the number of cross citations that a given author or journal has accumulated over a period of time.

In the field of *social science*, reputation as a quantitative concept is often studied as a network parameter associated with a society of agents [15,76]. Reputation or prestige is often measured by various centrality measures. An example of such a measure is provided by Katz [43], taking into account not only the number of direct links between agents but, also, the number of indirect links (going through intermediaries) between agents.

Another consideration, closely related to the work done in social sciences, about reputation systems concerns its *collaborative* aspect. Indeed, reputation systems could also be called collaborative-filtering systems [41] to reflect their collaborative nature (notice that Katz' measure has recently been rediscovered in the context of collaborative recommendation [37] and kernel methods where it is known as the von Neumann kernel [72]). Collaborative-filtering systems (also called recommender systems) try to provide people with recommendations of items they will appreciate (depending on purchase profitability [16], and/or on purchase probability – see, e.g., [2,27,35]), based on their past preferences (evaluations), history of purchase, and demographic information. Jøsang points out [41] similarities between collaborative-filtering and reputation systems (in that both collect ratings from members in a community) as well as fundamental differences: (1) in recommender systems different people have different tastes, and rate things differently according to subjective taste while in reputation systems, all members in a community should judge the performance of a transaction partner or the quality of a product