International Journal of Engineering In Advanced Research Science and Technology
ISSN: 2278-2566
VOLUME-2
ISSUE-5
NOV-2016

AN EFFICIENT RECOMMENDER SYSTEM BASED ON ONLINE RATINGS USING REGRESSION

Varada Naresh[1], V.Sudhakar[2]

[1]M.Tech Student, Dept.Of CSE, Sarojini Institute of Technology, Telaprolu(V), Unguturu (M), Gannavaram, krishna (D). A.P

[2]Assistant Professor, Dept.Of CSE, Sarojini Institute of Technology, Telaprolu(V), Unguturu (M), Gannavaram, krishna (D). A.P

ABSTRACT: Recommender systems have been shown to help users find items of interest from among a large pool of potentially interesting items. Influence is a measure of the effect of a user on the recommendations from a recommender system. Influence is a powerful tool for understanding the workings of a recommender system. Experiments show that users have widely varying degrees of influence in ratings-based recommender systems. Proposed influence measures have been algorithm-specific, which limits their generality and comparability. We propose an algorithm-independent definition of influence that can be applied to any ratings-based recommender system. We show experimentally that influence may be effectively estimated using simple, inexpensive metrics.

1 INTRODUCTION:

Sociologists have long tried to characterize the influence of a person in a social network of many people [1]. Identifying the influential people can bring twin advantages to those who study group dynamics: (1) The influential people can be directly studied, yielding insight since their choices may be predictive of group choices; or (2) The influential people may be influenced to change the behavior of the group. Many social networks are formed and maintained through informal, qualitative, and unobserved interactions. Capturing data about these interactions is difficult, and the act of capturing those data may change the social interactions themselves. Collaborative Filtering (CF) recommender systems [2, 3, 4] base their decisions on the opinions of users. In contrast to other social networks, recommender systems capture interactions that are formal, quantitative, and observed. The social network can be analyzed directly through data already captured in the computer system. Past research has demonstrated that analyzing the social network can provide leverage in influencing the group [5]. The analysis performed in these studies is based on a deep investigation of the characteristics of one particular recommender algorithm, the wellknown user-user nearest neighbor algorithm [2]. Careful analysis of this type has many advantages, but one key disadvantage: it is tied closely to the details of the algorithm. In principle, similar techniques could be applied to other algorithms, but doing so would be laborious, and the resulting influence measure only applies to algorithms that work precisely according to the details of the analysis. Since many commercial operators tweak the operation of the recommender in many ways

to fit the needs of their business, this analysis may not apply in practice. Further, the resulting measures of influence would be unlikely to be comparable between different algorithms, since they have been produced through very different techniques. A key goal of the present research is to identify a measure of influence for recommender systems that is applicable to any ratings-based recommender system, independent of the particulars of the algorithm. Such a measure would allow for consistent, black-box analysis of influence.

2 RELATED WORK

2.1 Recommender Systems. Resnick, et al. [2] introduced an automatic collaborative filtering algorithm based on a k-nearest neighbors (kNN) algorithm among users; this algorithm is now called user-user CF. The user-user algorithm we use in this paper is a version of the original kNN algorithm, tuned to achieve best known performance. Sarwar et al. [4] proposed an alternative kNN CF algorithm based on similarity among items. This variant is often called item-item CF. Breese et al. [3] have divided a number of CF algorithms into two classes: memory-based algorithms and model-based algorithms. Over the years many other algorithms were proposed including ones based on SVD, clustering, Bayesian Networks [3]. We focus on the user-user and item-item algorithms in this paper because they are the most common in existing systems.

2.2 Social Networks And Influence.

A Social network is a form of graph delineating relationships and interactions among individuals. Finding the important nodes in such graphs has been an object of interest to sociologists for a long time. One proposed measure for importance is centrality [1]. Two examples of "centrality" measures are "degree centrality", which treats high degree nodes as important, and "distance central ity",

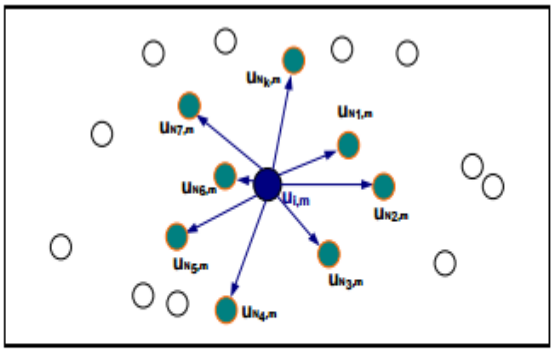

Figure 1: Showing the notion of *in-links* for the k closest neighbors of u_i. Here, prediction is being computed for the (user, item) pair, (u_i, m).

which treats nodes with short paths to many other nodes as important [1] . Kleinberg's HITS [6], and Brin and Page's PageRank [7] algorithms for ordering nodes in a graph of web are based on social network principles. Domingos et al. [5] have studied the problem of choosing influential users for marketers who wish to attract attention to their products. They show that selecting the right set of users for a marketing campaign can make a big difference. Kempe et al. [8] focus on a collection of models widely studied in social networks, as well as the models in [5], under the categories: Linear

Threshold Models, and Independent Cascade Models.

3. IMPLEMENTATION:

False-Reputation Module:

In an online rating system, it is almost impossible to obtain the ground-truth data because there is no way of knowing which users have caused a false reputation in a real-life database. We artificially establish various situations in which a false reputation may occur and test the performance of the proposed algorithm in these situations. In order to claim that the generated situations are likely to occur in real-life online rating systems, we list various scenarios involving a false reputation and categorize them according to the types of user and situations.In this section, we define dangerous users who cause a false reputation and dangerous situations leading to a false reputation. Using the definitions of dangerous users and dangerous situations,

Dangerous users. Dangerous situations.

False Reputation Scenarios.

TABLE I

FALSE-REPUTATION SCENARIOS

	Product launch phase	Unpopular products
Planned attacker	Hired planned attackers manipulate the reputation of a product during the product launch phase	Hired planned attackers manipulate the reputation of an unpopular product
Unplanned attacker	Extremists give biased rat-ings or don't-carers give meaningless ratings to a product during product launch phase	The product is unpopular and attracts unplanned at-tackers who give distorted ratings

Robustness:

In order to enhance the robustness of recommendation systems, it is imperative to develop detection methods against shilling attacks. Major research in shilling attack detection falls into three categories:

1) classifying shilling attacks according to different types of attacks.

2) extracting attributes that represent the characteristics of the shilling attacks and quantifying the attributes

3) developing robust classification algorithms based on the quantified attributes used to detect shilling attacks Strategies for improving the robustness of multi agent systems can be classified into two categories. The first group of strategies is based on the principle of majority rule. Considering the collection of majority opinions (more than half the opinions) as fair, this group of strategies excludes the collection of minority opinions, viewed as biased, when calculating the reputation.

Unfair Ratings:

The trustworthiness of a reputation can be achieved when a large number of buyers take part in ratings with honesty .If some users intentionally give unfair ratings to a product, especially when few users have participated, the reputation of the product could easily be manipulated. In this paper, we define false reputation as the problem of a reputation being manipulated by unfair ratings. In the case of a newly-launched product, for example, a company may hire people in the early stages of promotion to provide high ratings for the product. In this case, a false reputation adversely affects the decision making of potential buyers of the product. a reputation based on the confidence scores of all ratings, the proposed algorithm calculates the reputation without the risk of omitting ratings by normal users while reducing the influence of unfair ratings by abusers. We call this algorithm, which solves the false reputation problem by computing the true reputation, TRUE-REPUTATION. Our framework for online rating systems and the existing strategies in multi agent systems serve the same purpose in that they are trying to address unfair ratings by abusers.

Buyer Modules:

Numerous studies have been conducted to improve the trustworthiness of online shopping malls by detecting abusers who have participated in the rating system for the sole purpose of manipulating the information provided to potential buyers (e.g., reputations of sellers and recommended items). Especially in the fields of multiagent and recommendation systems, various strategies have been proposed to handle abusers who attack the vulnerability of the system. In online rating systems, on the other hand, a buyer can give only a single rating per item. Thus, the relationship between buyers and items is significantly different from the relationship between buyers and sellers; as such, the graph structure of an online rating system is very different from that of a multi agent system. This paper uses an approach that considers the relation between buyers and items.

Regression:-

Linear regression is the most basic and commonly used predictive analysis. Regression estimates are used to describe data and to explain the relationship between one dependent variable and one or more independent variables.

At the center of the regression analysis is the task of fitting a single line through a scatter plot. The simplest form with one dependent and one independent variable is defined by the formula y = c + b*x, where y = estimated dependent, c = constant, b = regression coefficients, and x = independent variable.

Linear regression is the most basic and commonly used predictive analysis. Regression estimates are used to describe data and to explain the relationship between one dependent variable and one or more independent variables.

At the center of the regression analysis is the task of fitting a single line through a scatter

plot. The simplest form with one dependent and one independent variable is defined by the formula y = c + b*x, where y = estimated dependent, c = constant, b = regression coefficients, and x = independent variable.

Sometimes the dependent variable is also called a criterion variable, endogenous variable, prognostic variable, or regressand. The independent variables are also called exogenous variables, predictor variables or regressors.

However linear regression analysis consists of more than just fitting a linear line through a cloud of data points. It consists of 3 stages – (1) analyzing the correlation and directionality of the data, (2) estimating the model, i.e., fitting the line, and (3) evaluating the validity and usefulness of the model.

There are 3 major uses for regression analysis – (1) causal analysis, (2) forecasting an effect, (3) trend forecasting. Other than correlation analysis, which focuses on the strength of the relationship between two or more variables, regression analysis assumes a dependence or causal relationship between one or more independent and one dependent variable.

Firstly, it might be used to identify the strength of the effect that the independent variable(s) have on a dependent variable. Typical questions are what is the strength of relationship between dose and effect, sales and marketing spend, age and income.

Secondly, it can be used to forecast effects or impacts of changes. That is regression analysis helps us to understand how much will the dependent variable change, when we change one or more independent variables. Typical questions are how much additional Y do I get for one additional unit X.

Thirdly, regression analysis predicts trends and future values. The regression analysis can be used to get point estimates. Typical questions are what will the price for gold be in 6 month from now? What is the total effort for a task X?

There are several linear regression analyses available to the researcher.

• Simple linear regression
1 dependent variable (interval or ratio), 1 independent variable (interval or ratio or dichotomous)

• Multiple linear regression
1 dependent variable (interval or ratio) , 2+ independent variables (interval or ratio or dichotomous)

• Logistic regression
1 dependent variable (binary), 2+ independent variable(s) (interval or ratio or dichotomous)

• Ordinal regression
1 dependent variable (ordinal), 1+ independent variable(s) (nominal or dichotomous)

• Multinominal regression
1 dependent variable (nominal), 1+

independent variable(s) (interval or ratio or dichotomous)

• Discriminant analysis
1 dependent variable (nominal), 1+ independent variable(s) (interval or ratio)

When selecting the model for the analysis another important consideration is the model fitting. Adding independent variables to a linear regression model will always increase the explained variance of the model (typically expressed as R^2). However adding more and more variables to the model makes it inefficient and over fitting occurs. Occam's razor describes the problem extremely well – a model should be as simple as possible but not simpler. Statistically if the model includes a large number of variables the probability increases that the variables test statistically significant out of random effects.

The second concern of regression analysis is under fitting. This means that the regression analysis' estimates are biased. Under fitting occurs when including an additional independent variable in the model will reduce the effect strength of the independent variable(s). Mostly under fitting happens when linear regression is used to prove a cause-effect relationship that is not there. This might be due to researcher's empirical pragmatism or the lack of a sound theoretical basis for the model.

4. CONCLUSION:

In this paper I have studied the problems in E-Marketplaces and various solutions how to overcome some of the problems. There are more factors (other than those addressed in this paper) known to be elemental in assessing the trust of users in the field of social and behavioral sciences. I plan to study how to incorporate them into our model to compute the reputation of items more accurately. In the e-market place such as Amazon.com and eBay.com, buyers give ratings on items they have purchased. I note, however, that the rating given by a buyer indicates the degree of his satisfaction not only with the item (e.g., the quality) but also with its seller (e.g., the promptness of delivery). In a further study, I plan how to develop an approach to accurately separate an item score and a seller score from a user rating. Separating the true reputation of items and that of sellers would enable customers to judge items and sellers independently.

REFERENCES:

[1] A Trust-Aware System for Personalized User Recommendations in Social Networks, MagdaliniEirinaki, Malamati D. Louta,Member, IEEE,andIraklisVarlamis, Member, IEEE, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, VOL. 44, NO. 4, APRIL 2014

[2] Using Machine Learning to Augment Collaborative Filtering of Community Discussions, Michael Brennan, Stacey Wrazien, Rachel Greenstadt, Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)

[3] Can You Trust Online Ratings? A Mutual Reinforcement Model for

Trustworthy Online Rating Systems, HyunKyo Oh, Sang-Wook Kim, Member, IEEE ,Sunju Park, and Ming Zhou, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS ,YEAR 2015

[4] Detecting Product Review Spammers using Rating Behaviors,Ee-Peng Lim, Viet-An Nguyen, Nitin Jindal, CIKM'10,October 26–30, 2010

[5] Shin: Generalized Trust Propagation with Limited Evidence, Chung-Wei Hang, Zhe Zhang, and Munindar P. Singh, EEE Computer Society 2013.

[6] iCLUB: An Integrated Clustering-Based Approach to Improve the Robustness of Reputation Systems, Siyuan Liu, Jie Zhang, Chunyan Miao, Yin-LengTheng, Alex C. Kot, Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011).

[7] Preventing Shilling Attacks in Online Recommender Systems, Paul-AlexandruChirita, Wolfgang Nejdl, CristianZamfir, WIDM'05,November 5, 2005.

[8] HySAD: A Semi-Supervised Hybrid Shilling Attack Detector for Trustworthy Product Recommendation, Zhiang Wu, JunjieWu,JieCao, Dacheng Tao, KDD'12, August 12–16, 2012