Collaborative Recommendation based on Implication Field

Hoang Tan Nguyen[1]
Department of Information
and Communications of
Dong Thap province
Dong Thap, Vietnam

Lan Phuong Phan[2]
College of Information &
Communications Technology
Can Tho University
Can Tho, Vietnam

Hung Huu Huynh[3]
University of
Science and Technology
University of Da Nang,
Da Nang, Vietnam

Hiep Xuan Huynh[4*]
College of Information &
Communications Technology
Can Tho University
Can Tho, Vietnam

Abstract—**Recently, recommender systems has grown rapidly in both quantity and quality and has attracted many studies aimed at improving their quality. Especially, collaborative filtering techniques based on rule mining model combined with statistical implication analysis (SIA) technique also achieved some interesting results. This has shown the potential of SIA to improve the performance of recommender systems. However, it is still not rich and there are several problems to be solved for better results such as the problem of non-binary data processing, dealing with bottleneck case of data partitioning method according to the number of transactions on the very sparse transaction sets during training and testing the model, and not paying attention to exploiting the trend of variation of statistical implication. In order to contribute to solving these problems, the paper focuses on proposing a new data partitioning method, and developing the recommendation model based on equipotential planes mining generated by variation of implication intensity or implication index in the implication field on both binary and non-binary data to improve the recommendations further. Experimental results have shown the success of this new approach through its quality comparison with collaborative filtering recommendation models as well as existing SIA-based ones.**

Keywords—Implication intensity; implication rules; implication field; equipotential surface

I. Introduction

As a result of the available online information and the rapid increase of e-business and e-commercial services, it is difficult for users to make a proper decision without supporting of recommendation engines. And therefore, recommender systems [1], [2], [3], especially those based on collaborative filtering, are more and more popular and become an indispensable part of e-commercial services and others and one of which [2], [5], [15], [16], [17] is recommender system based on association rules mining (ARM).

Although ARM is considered as a popular and effective tool in “market-basket analysis” tasks and developing e-commerce, in recommender systems, the contribution of this technique is limited due to many reasons. Therefore, there have been many studies to improve this technique for recommender systems such as using fuzzy logic [16], [17], binarization real data [7], [8], and some others to refine and improve the rules evaluation measures, etc. While these also obtained certain results [5], [15], [16], [17], it’s so hard to keep up with collaborative filtering others. In recent years, several recommendation models using statistical implicative analysis (SIA) [12], [13], [14] approach to improve the quality and effectiveness of recommender systems [7], [9], [10], [18], [19], [20], [21] by discover the interesting rules using measures like implication intensity, entropy implication intensity, Cohesion, and so on as similarity measures. Almost of which has not paid attention to mining the trend of variation of statistical implication, except for the works [18], [19], [20], [21] that have been published recently.

This paper focuses on three proposals (1) building a recommendation model based on implication field that can deal with both binary and non-binary dataset, (2) proposal data partition method based on rated items on each transaction instead of number of transactions on dataset, and (3) for a more comprehensive assessment of the quality of the proposed models, Item rating-based accuracy metrics are also used in addition to classification-based and prediction-based accuracy metrics to assess the quality of the recommendation listing. Experiment’s results shown that proposed model has performed better than both collaboration filtering ones and existing SIA-based ones, not only on binary data but also on non-binary one.

The paper is organized in five parts. The first one introduces the context and issues to be solved by the present systems as well as proposing our approach. The second part presents a summary of SIA theory and relevant contents about the recommender system. The next one presents proposed solution and its model to improve further the efficiency of recommender systems based on SIA. The fourth part is the experiment and evaluation of the proposed model, which focuses on comparing its performance with previous SIA models and traditional collaborative filtering-based models. Finally, the paper is finished by the conclusion.

II. Literature Review

A. Statistical Implicative Analysis

Statistical implicative analysis (SIA) theory [13], [14], proposed by Régis Gras, studies the implication relationship of data variables. It can be presented as follows.

Let $E = \{e_1, e_2, ..., e_n\}$ be a population of n transactions described by a finite set $I = \{i_1, i_2, ..., i_m\}$ of m variables (attributes, criteria, etc.). Let $e_k \in E, i_v \in I$ where $1 \leq k \leq n$ and $1 \leq v \leq m$. Denote by $\Omega(e_k)$ the set of items taken

[*] Corresponding authors.

from a transaction e_k, and $\Omega(e_k) \subseteq I$. Let a and b be subsets of I. Denote $A = \{e_k \in E | \forall j \in a, j \in \Omega(e_k)\}$ and $B = \{e_k \in E | \forall l \in b, l \in \Omega(e_k)\}$, and $\bar{A}, \bar{B}$ are respectively complementary set A, B in E.

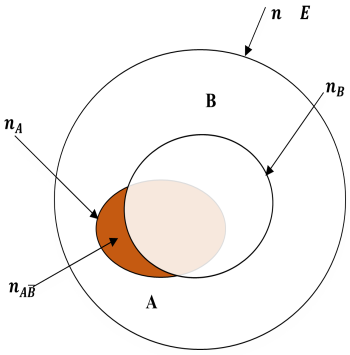


Fig. 1. Illustration of an implication rule $a \rightarrow b$ in SIA by Venn diagram.

An implication relationship (association rule/implication rule) is a pattern of the form $a \rightarrow b$, where a and b are disjoined itemsets ($a \subset I, b \subset I$, and $a \cap b = \emptyset$). In fact, it is relatively common to observe a couple of transactions which contain a and not b instead of having the general trend to have b when a is present. Therefore, in addition to $n = card(E)$ of E, it is necessary to taken into account the number $n_A = card(A)$ of A, $n_B = card(B)$ of B, and $n_{A\bar{B}} = card(A \cap \bar{B})$ of counter-examples $A \cap \bar{B}$ to statistically accept to retain or not the rule $a \rightarrow b$.

The implication relationship between a and b is presented in an implication rule $a \rightarrow b$ could be modeled in the SIA as Fig. 1.

To further illustrate about an implication rule, let's see an example movie transaction data as presented in Table I(a). We can consider it as set $E = \{e_k, | k = 1..9\}$, and let $I = \{Movie_1, Movie_2, Movie_3\}$ an itemset. The set of items $\Omega(e_1) = \{Movie_1\}$, $\Omega(e_2) = \{Movie_1, Movie_2\}$, etc. The movies data in Table I can be represented in a binary format as shown in Table II, where each row corresponds to a transaction and each column corresponds to an movie. A movie can be treated as a binary variable whose value is 1 if the movie is present in a transaction and 0 otherwise. Now, let's consider an implication rule $a \rightarrow b$, where $a = \{Movie_1, Movie_2\}, b = \{Movie_3\}$ then set $A = \{e_2, e_4, e_5, e_6, e_8, e_9\}$, and set $B = \{e_4, e_5, e_6, e_9\}$. Thus, $n = 9, n_A = 6, n_B = 4$, and $n_{A\bar{B}} = 2$, so that rule $a \rightarrow b$ can be represented by $(n, n_A, n_B, n_{A\bar{B}})$ is $(9, 6, 4, 2)$.

More detail, we compare the observed number of counter-examples to a probabilistic model. Let us assume that we randomly draw two subsets X and Y in E which respectively contain n_A and n_B transactions. The complementary sets $\bar{Y}$ of Y and $\bar{B}$ of B in E have the same cardinality $n_{\bar{B}}$. In this case, $N_{X\bar{Y}} = card(X \cap \bar{Y})$ is a random variable and $n_{A \cap B}$ an observed value. The implication rule $a \rightarrow b$ is admissible for a given threshold $1-\sigma$ if σ is greater than the probability that the number of counter-examples in the observations is greater than the number of expected counterexamples in a random drawing [14], i.e. if $Pr(N_{X\bar{Y}}) \leq n_{A\bar{B}}) \leq \sigma$.

TABLE I. An Example of Movie Transaction Data

E	Items/$\Omega(e_k)$
e_1	$Movie_2$
e_2	$Movie_1, Movie_2$
e_3	$Movie_1$
e_4	$Movie_1, Movie_2, Movie_3$
e_5	$Movie_1, Movie_2, Movie_3$
e_6	$Movie_1, Movie_2, Movie_3$
e_7	$Movie_2$
e_8	$Movie_1, Movie_2$
e_9	$Movie_1, Movie_2, Movie_3$

TABLE II. A Binary Representation of Data in Table I

E	$Movie_1$	$Movie_2$	$Movie_3$
e_1	0	1	0
e_2	1	1	0
e_3	1	0	0
e_4	1	1	1
e_5	1	1	1
e_6	1	1	1
e_7	0	1	0
e_8	1	1	0
e_9	1	1	1

The distribution of random variable $N_{X\bar{Y}}$ depends on the drawing pattern of X and Y. For a certain process of drawing, the random variable $N_{X\bar{Y}}$ follows a Poissonian distribution [14] $P(\lambda)$ with $\lambda = \frac{n_A n_{\bar{B}}}{n}$. For cases where the approximation is justified (e.g. $\lambda \geq 4$), the standardized random variable $\tilde{N}_{X\bar{Y}} = \frac{card(X \cap \bar{Y}) - \lambda}{\sqrt{\lambda}}$ is approximatively $N(0,1)$-distributed. The observed value of $\tilde{N}_{X\bar{Y}}$ is $\tilde{n}_{A\bar{B}} = (\frac{n_{A\bar{B}} - \lambda}{\sqrt{\lambda}})$.

The implication intensity expresses the unlikelihood of counter-examples $n_{A\bar{B}}$ in E. The rule is admitted for a given threshold $1 - \sigma$ if $\varphi(a, b) \geq 1 - \sigma$.

The implication intensity measure $\varphi(a, b)$ [13], [14] of rule $a \rightarrow b$ is defined by equation (1)

$$\begin{aligned} \varphi(a,b) &= \begin{cases} 1 - Pr(\tilde{N}_{X \cap \bar{Y}} \leq \tilde{n}_{A \cap \bar{B}}), & \text{if} \quad n_B < n \\ 0, & \text{otherwise,} \end{cases} \\ &= \begin{cases} 1 - \displaystyle\sum_{s=0}^{n_{A\bar{B}}} \frac{\lambda^s}{s} e^{-\lambda}, & \text{if} \quad n_B < n \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \tag{1}$$

For cases where the approximation is justified, the standardized random variable $\tilde{N}_{X\bar{Y}} = \frac{card(X \cap \bar{Y}) - \lambda}{\sqrt{\lambda}}$ is approximatively N(0,1)-distributed, and $\varphi(a, b)$ is determined as equation (2)

$$\varphi(a,b) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{q(a,\bar{b})}^{\infty} e^{\frac{-t^2}{2}} dt, & \text{if} \quad n_B < n \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

where $q(a, \bar{b})$ is the implication index [14], also known as the Gras implication index, and is determined as follows.

For binary variables [13], [14], the implication index $q(a, \bar{b})$ is defined by equation (3)

$$q(a,\bar{b}) = \frac{n_{A\bar{B}} - \frac{n_A n_{\bar{B}}}{n}}{\sqrt{\frac{n_A n_{\bar{B}}}{n}}} \tag{3}$$

For modal variables [13] $a, b \in [0, 1]$, the implication index $q_p(a, \bar{b})$ is defined by equation (4)

$$q_p(a, \bar{b}) = \frac{\sum_{t \in E} a(t)\bar{b}(t) - \frac{n_A n_{\bar{B}}}{n}}{\sqrt{\frac{(n^2 s_A^2 + n_A^2)(n^2 s_{\bar{B}}^2 + n_{\bar{B}}^2)}{n^3}}} \tag{4}$$

where $a(t)$ is the values of element t^{th} of the a and $\bar{b}(t) = 1 - b(t)$ is complement of element $b(t)$ of b respectively; and s_A, s_B is their standard deviations.

For the frequency variables and the non-negative number variables, in order to use equation (4) they must be normalized [14] in advance by equation (5)

$$\tilde{a}(w) = a(w) / \max_{w \in E} a(w). \tag{5}$$

When $a(t)$ and $\bar{b}(t)$ are binary variables then $q_p(a, \bar{b}) = q(a, \bar{b})$.

The implication rule $a \rightarrow b$ is admissible at the level α if and only if $\varphi(a, b) \geq 1 - \alpha$ [14].

Formula (2) definition of the implication intensity reminds its users, that it is of implication intensity interest only on condition that it is greater than 0.50, that means its $q(a, \bar{b})$ should be negative. It is, therefore, more significant for an implication index that is strongly negative for patterns $a \rightarrow b$.

B. Implication Field

Let's consider the implication index $q(a, \bar{b})$ in the four-dimensional space, in which a point M whose coordinates are the parameters associated with $(n, n_A, n_B, n_{A\bar{B}})$. Then $q(a, \bar{b})$ is a scalar field by applying the mapping from space R^4 to space R. The vector $grad\ q(a, \bar{b})$ containing the partial derivatives of $q(a, \bar{b})$ by the variables $(n, n_A, n_B, n_{A\bar{B}})$ is a special gradient field also known as implication field because it meets the Schwartz criterion for the mixed partial derivatives [12] of $q(a, \bar{b})$ for all pairs of variables $(n, n_A, n_B, n_{A\bar{B}})$. That means for any pair $(n_B, n_{A\bar{B}})$ then the partial derivatives by n_B of partial derivatives of $q(a, \bar{b})$ by $n_{A\bar{B}}$ equal to the partial derivatives by $n_{A\bar{B}}$ of partial derivatives of $q(a, \bar{b})$ by n_B as equation (6)

$$\frac{\partial}{\partial n_B}\left(\frac{\partial q(a, \bar{b})}{\partial n_{A\bar{B}}}\right) = \frac{\partial}{\partial n_{A\bar{B}}}\left(\frac{\partial q(a, \bar{b})}{\partial n_B}\right) = \frac{1}{2}\left(\frac{n_A}{n}\right)^{-\frac{1}{2}}\left(\frac{n_{\bar{B}}}{n}\right)^{-\frac{3}{2}} \tag{6}$$

and similarly for remaining pairs.

In terms of structure, the implication field is the four-dimensional space, consisting of ordered ordinate surfaces corresponding to the successive and ordered values of $q(a, \bar{b})$ with respect to the variation of the cardinalities $(n, n_A, n_B, n_{A\bar{B}})$ [12], [14]. Now, the implication index is considered as a function of four parameters $(n, n_A, n_B, n_{A\bar{B}})$, a line or surface of equipotential in implication field is curve in E. A space along which or in which, point a variable M maintains the same value of potential of $q(a, \bar{b})$. The surface of equipotential is orderly. The curve equation of this surface [12], [14], [18], [19], [20], [21] is shown in equation (7)

$$q(a, \bar{b}) - \frac{n_{A\bar{B}} - \frac{n_A n_{\bar{B}}}{n}}{\sqrt{\frac{n_A n_B}{n}}} = 0. \tag{7}$$

Consequently, on such a curve, the scalar product between gradient of $q(a, \bar{b})$ and partial derivatives of M, $grad\ q(a, \bar{b}).dM$, is zero [12]. This is interpreted as indicating the orthogonality of the gradient with the tangent or the hyperplane tangent to the curve, that is to say the line or the equipotential surface.

By illustrating, the potential S depends only on two variables , for example n_A, n_B, Fig. 2 below shows the orthogonal direction of the gradient for different equipotential surfaces where the potential S does not change on each surface, but it changes from the surface $S = 7$ to $S = 10$. Thus, implication field can be considered a space in which a set of equipotential surfaces corresponding to successive values of $q(a, \bar{b})$ relative to the cardinals $(n, n_A, n_B, n_{A\bar{B}})$ which one would vary. The various gradient fields, true "lines of force", which are associated with them are orthogonal to the surfaces defined by the corresponding values of $q(a, \bar{b})$ [12]. Behind this notion we can imagine a transport of information of variable intensity in a causal universe.

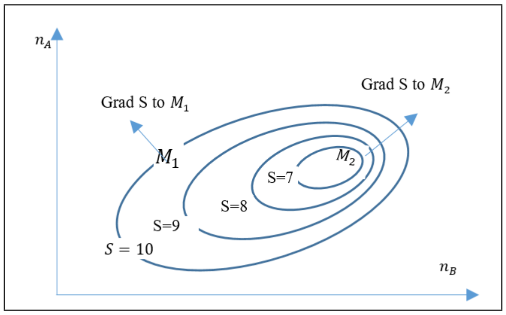


Fig. 2. Illustration on Cartesian Coordinates for an Implication Field.

C. Implicative Recommendation

SIA's original purpose is to analyze data for educational, psychological, and ontological applications [13], [14], etc. In recent years, however, attributable to recognize potential ability of SIA in recommender system techniques, there were several studies in recommender systems to improve their efficiency and have obtained remarkable results.

A typical proposal [7], [8], which has shown the potential of SIA to improve the performance of rule-based collaborative filtering recommender systems, however it also has a several disadvantages need to be addressed as (1) only processing on binary data, which leads to a problem to solve is the combinatorial explosion due to the binarization of non-binary data, (2) for models based on rules mining of these works, SIA is proposed in the post-processing stage of rules mining task, so they have not contributions considerably to limit the outcome rules' combinatorial explosion in large datasets. Another study using SIA to recommender systems [9], [10], which paid the attention to solve the problem on non-binary data and making some new contributions based on the recommendations model with additional SIA metrics such as entropic version of implication intensity, cohesion, contribution, etc. This work is another proof show potential ability of SIA applying on recommender systems.

Most recently, SIA has also been proposed to recommender system models in the works [18], [19], [20], [21]. Accordingly, Nguyen *et al.* has contributed to overcome the shortcomings of previous studies following a new approach on SIA like reducing model's performance time, increasing predictions' accuracy, controlling generated rules set, compared to both traditional collaborative filtering model and former SIA model, by developing a recommender system based on implication rules mining (IRM) using implication variation measures.

In practice, research applying SIA to the development of recommender system models recently have made a positive contribution in this area, it can be seen as a potential research trend. However, in order to further improve the effectiveness of the recommendation models, there are still a few issues that need to be addressed as follows: (1) it is necessary to further mine the unique and outstanding features of the implication field such as equipotential planes, (2) the data sets for the recommendation models are mostly sparse, therefore, using cross k-folds evaluation for recommendation models by partitioning the data set by transactions is not optimal yet, because this will lead to limit number of known items (given items) in test sets, this can significantly affect to model's training quality, (3) using measures of accuracy of the predictions and classifications to evaluate the recommendation models is not enough yet, because in recommender systems, the position of items in the recommendation list is also important, therefore, it is necessary to use additional measures of items' position ranking in recommendation list.

III. Recommendation Model

A. Model

The implication field and its particularly features, as shown in Section II-B, have opened a great potential for the implementation of recommendation models. In this paper, the implication field-based recommender system has been proposed to include the following components as shown in Fig. 3. This model has been experimentally proven on both binary and non-binary datasets to be more efficient than the traditional collaborative filtering models exploiting the association rules on both binary and quantitative data. The main components of the model include the following.

The implication field algorithms include two newly proposed algorithms. The first is responsible for generating the implication field consisting of a set of equipotential surfaces as discussed in Section II-B, This algorithm uses one of the implication variation measures. These measure are presented in Table III. They include four for implication index variation (first four rows) and four for implication intensity variation (four rows later) by n, n_A, n_B, and $n_{A\bar{B}}$. These SIA measures are determined as sum of the implication index (or implication intensity) and the partial derivative of the implication index (or implication intensity) by the variables n, n_A, n_B, and $n_{A\bar{B}}$, and SIA knowledge to generate the implication field, that is composed of a set of equipotential surfaces, from dataset. The second mines frequent implication patterns on equipotential surfaces to provide recommendations to users. It makes personal recommendation by frequent implication pattern mining on equipotential surfaces in given threshold implication index (or implication intensity) for predicting and a recommendation the items or the top k items list to users. These algorithms will be shown in Section III-B.

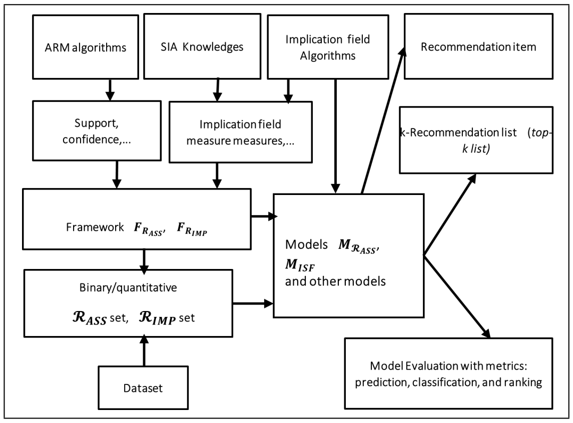

ARM algorithms
SIA Knowledges
Implication field Algorithms
Recommendation item
Support, confidence,...
Implication field measure measures,...
k-Recommendation list (top-k list)
Framework $F_{R_{ASS}}$, $F_{R_{IMP}}$
Models $M_{\mathcal{R}_{ASS}}$, M_{ISF} and other models
Binary/quantitative
$\mathcal{R}_{ASS}$ set, $\mathcal{R}_{IMP}$ set
Dataset
Model Evaluation with metrics: prediction, classification, and ranking

Fig. 3. The Overall Structure of Recommendation Model based on Implication Field.

In Fig. 3, The models $M_{R_{ASS}}$, (Model for Association), and M_{ISF}, (Model for Implication Statistical Field), are recommendation model based on ARM, and IRM respectively in implication field. These models include a set of association rules R_{ASS} or implication rules R_{IMP} and framework for association rules mining $F_{R_{ASS}}$, or implication rules $F_{R_{IMP}}$ correspondingly [20], [21]. They are shown in equations (8) and (9)

$$M_{R_{ASS}} = \{X \mid R_{ASS}, F_{R_{ASS}}\}, \tag{8}$$

$$M_{ISF} = \{X \mid R_{IMP}, F_{R_{IMP}}\}, \tag{9}$$

where R_{ASS} and R_{IMP} are respective association rules set and implication rules set. Each of which is expressed by 4-tuples $(n, n_A, n_B, n_{A\bar{B}})$ meeting the constraints as are defined as equation (10) and (11), respectively, where s (support), s_{min} (minimum support threshold), c (confidence), c_{min} (minimum confidence threshold), imp (a given SIA measures), and imp_{min} (minimum threshold of a given SIA measure) are support, confidence, and one of SIA measures (see Table III), and their respective minimum thresholds.

$$R_{ASS} = \left\{ (n, n_A, n_B, n_{AB}) \left| \begin{array}{r} n_A \leq n,\ n_B \leq n, \\ \min(0, n_A + n_B - n) \\ \leq (n_A - n_{A\bar{B}}) \leq \\ \max(n_A, n_B), \\ s_{min} \leq s, c_{min} \leq c \end{array} \right. \right\}, \tag{10}$$

$$R_{IMP} = \left\{ (n, n_A, n_B, n_{A\overline{B}}) \left| \begin{array}{r} 0 \leq n_A \leq n_B \leq n, \\ 0 \leq n_{A\overline{B}} \leq n_A, \\ s_{min} \leq s, c_{min} \leq c, \\ imp_{min} \leq imp \end{array} \right. \right\}. \tag{11}$$

TABLE III. FAMILY OF VARIATION IMPLICATION MEASURES OF IMPLICATION RULES MINING FRAMEWORK

No	SIA Measures	Formulas
1	$q_n(a,\bar{b})$	$q(a,\bar{b})+\frac{1}{2\sqrt{n}}(n_{A\bar{B}}+\frac{n_A n_{\bar{B}}}{n})$
2	$q_{n_A}(a,\bar{b})$	$q(a,\bar{b})-\frac{1}{2}\frac{n_{A\bar{B}}}{\sqrt{\frac{n_{\bar{B}}}{n}}}(\frac{n}{n_A})^{\frac{3}{2}}-\frac{1}{2}\sqrt{\frac{n_{\bar{B}}}{n_A}}$
3	$q_{n_B}(a,\bar{b})$	$q(a,\bar{b})+\frac{1}{2}n_{A\bar{B}}(\frac{n_A}{n})^{-\frac{1}{2}}(n-n_B)^{-\frac{3}{2}}+$ $\frac{1}{2}(\frac{n_A}{n})^{-\frac{1}{2}}(n-n_B)^{-\frac{1}{2}}$
4	$q_{n_{A\bar{B}}}(a,\bar{b})$	$q(a,\bar{b})+\frac{1}{\sqrt{\frac{n_A(n-n_B)}{n}}}$
5	$\varphi_n(a,b)$	$\varphi(a,b)+\frac{1}{\sqrt{2\pi}}\int_{q(a,\bar{b})}^{q_n(a,\bar{b})} e^{\frac{-t^2}{2}}dt$
6	$\varphi_{n_A}(a,b)$	$\varphi(a,b)+\frac{1}{\sqrt{2\pi}}\int_{q(a,\bar{b})}^{q_{n\,A}(a,\bar{b})} e^{\frac{-t^2}{2}}dt$
7	$\varphi_{n_B}(a,b)$	$\varphi(a,b)+\frac{1}{\sqrt{2\pi}}\int_{q(a,\bar{b})}^{q_{n_B}(a,\bar{b})} e^{\frac{-t^2}{2}}dt$
8	$\varphi_{n_{A\bar{B}}}(a,b)$	$\varphi(a,b)+\frac{1}{\sqrt{2\pi}}\int_{q(a,\bar{b})}^{q_{n\,A\bar{B}}(a,\bar{b})} e^{\frac{-t^2}{2}}dt$

$F_{R_{ASS}}$ is framework of ARM [20], [21], including famous ARM algorithm, *apriori*, was proposed by [11], and measures support (s), confidence (c) meeting the constraints $s \geq s_{min}$, and $c \geq c_{min}$ as are defined as equation (12) and $F_{R_{IMP}}$ is frameworks of IRM [20], [21], including IRM algorithms (see details in Section III-B following), and measures support, confidence, and ASI measures as defined in Table III meeting the constraints as are defined as equation (13)

$$F_{RASS}=\left\{\begin{pmatrix} ARM\,algs, \\ supp\,s, conf\,c\end{pmatrix}\middle| \begin{array}{r} n_A \leq n, n_B \leq n, \\ \min(0, n_A+n_B-n) \\ \leq (n_A - n_{A\bar{B}}) \\ \leq \max(n_A, n_B), \\ s_{min} \leq s, c_{min} \leq c \end{array}\right\} \tag{12}$$

$$F_{RIMP}=\left\{\begin{pmatrix} IRM\,algs, \\ supp\,s, conf\,c, \\ SIA\,measure\,imp\end{pmatrix}\middle| \begin{array}{r} 0 \leq n_A \leq n_B \leq n, \\ 0 \leq n_{A\overline{B}} \leq n_A, \\ min \leq s, c_{min} \leq c, \\ imp_{min} \leq imp \end{array}\right\}. \tag{13}$$

The evaluation models are designed for testing and evaluating recommendation models and they will be presented details in section III-C following, and then they are used in experiments of sections IV-D and IV-E.

B. Algorithms

To generate the list of recommendations, the implication field-based recommendation system model focuses on the IRM algorithm including two phases. The first is to generate the implication field from the dataset, in this phase the IFGEN algorithm will be used to generate produce a set of equipotential planes for potential implication values, based on the variation of one of the four variables $(n, n_A, n_B, n_{A\bar{B}})$. The second uses MAKEIFREC algorithm to mine implication patterns of equipotential planes to generate a list of recommended items for the user.

Algorithm IFGEN (Implication fields Generator)

Input: a dataset; the thresholds of confidence, support and an implication field measure; type of data (binary/quantitative).

Output: Implication rule set.

Step 1: Constructing implication field measure, defined as in the Table III.

Step 2: Generating the implication rules set from the dataset using a data mining algorithm (such as Apriori, Eclat, etc.) and the thresholds of support, confidence and implication field measure that is defined in step 1. Note that: if data is in binary form, $q(a,\bar{b})$ is computed by equation (2); if the data is in quantitative form, $q(a,\bar{b})$ is computed by equation (4) and (3).

Step 3: Presenting each implication rules by four values n, n_A, n_B, and $n_{A\bar{B}}$ as well as its values according to the measures such as support, confidence, implication index, implication intensity, and implication field measures as shown in Table III.

With the algorithm IFGEN, the generated implication rules will be more accurate because of the high examples (from support /confidence measures) and low counterexamples (from the statistical implication measure). This will be confirmed in the Section IV.

Algorithm MAKEIFREC (making implication field recommendation).

Input: a dataset; the thresholds of confidence, support, and an implication field measure; type of data (binary/quantitative).

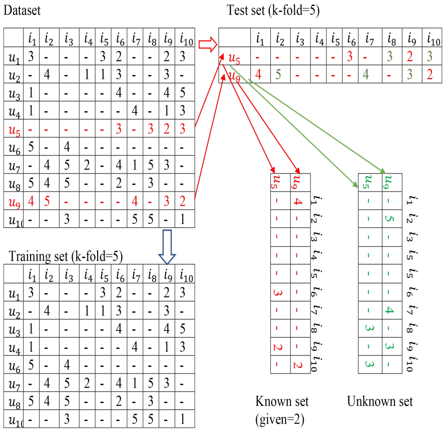
Output: predicting item or the list of top k items to be recommended to users.

Step 1: calling the **IFGEN** algorithm for generating the set of equipotential surfaces in implication rules.

Step 2: mining frequent implication patterns in equipotential surfaces for predicting and returning the recommendation result (1 item or k items) to users.

C. Evaluation

Normally, to evaluate a machine learning model, evaluating procedures divide the dataset into a training set and test set based on transactions. In recommender systems, however, that can get a weaknesses for sparse datasets which have some transactions with very few users' rating, this leads to the maximal number of items to keep (*given*) on known set is limited considerable because it cannot be greater than the number of rated items on any transaction in dataset, see example in Fig. 4, because there are only two rating in transaction u_6, we cannot set $given \geq 2$. This could lead to limit learning ability considerably and therefore quality of recommendation model will be not good. Moreover, some of the proposed models only focus on processing the binary data and try to binarize the non-binary datasets [7], [8], [9]. This could be the main cause to the accuracy of recommendation models to be affected.



Dataset

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9	i_{10}
u_1	3	-	-	-	3	2	-	-	2	3
u_2	-	4	-	1	1	3	-	-	3	-
u_3	1	-	-	-	-	4	-	-	4	5
u_4	1	-	-	-	-	-	4	-	1	3
u_5	-	-	-	-	-	3	-	3	2	3
u_6	5	-	4	-	-	-	-	-	-	-
u_7	-	4	5	2	-	4	1	5	3	-
u_8	5	4	5	-	-	2	-	3	-	-
u_9	4	5	-	-	-	-	4	-	3	2
u_{10}	-	-	3	-	-	-	5	5	-	1

Test set (k-fold=5)

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9	i_{10}
u_5	-	-	-	-	-	3	-	3	2	3
u_9	4	5	-	-	-	-	4	-	3	2

Training set (k-fold=5)

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9	i_{10}
u_1	3	-	-	-	3	2	-	-	2	3
u_2	-	4	-	1	1	3	-	-	3	-
u_3	1	-	-	-	-	4	-	-	4	5
u_4	1	-	-	-	-	-	4	-	1	3
u_6	5	-	4	-	-	-	-	-	-	-
u_7	-	4	5	2	-	4	1	5	3	-
u_8	5	4	5	-	-	2	-	3	-	-
u_{10}	-	-	3	-	-	-	5	5	-	1

u_5	u_9	
-	4	i_1
-	-	i_2
-	-	i_3
-	-	i_4
-	-	i_5
3	-	i_6
-	-	i_7
-	-	i_8
2	-	i_9
-	2	i_{10}

Known set
(given=2)

u_5	u_9	
-	-	i_1
-	5	i_2
-	-	i_3
-	-	i_5
-	-	i_5
-	-	i_6
-	4	i_7
3	-	i_8
-	3	i_9
3	-	i_{10}

Unknown set

Fig. 4. An Example about Typical Partition Data Method.

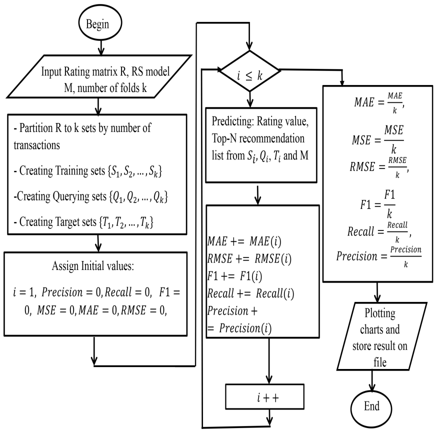

Begin
Input Rating matrix R, RS model M, number of folds k
- Partition R to k sets by number of transactions
- Creating Training sets $\{S_1, S_2, \dots, S_k\}$
-Creating Querying sets $\{Q_1, Q_2, \dots, Q_k\}$
- Creating Target sets $\{T_1, T_2, \dots, T_k\}$
Assign Initial values:
$i = 1,\ Precision = 0, Recall = 0,\ F1 = 0,\ MSE = 0, MAE = 0, RMSE = 0,$
$i \leq k$
Predicting: Rating value, Top-N recommendation list from S_i, Q_i, T_i and M
$MAE \mathrel{+}= MAE(i)$
$RMSE \mathrel{+}= RMSE(i)$
$F1 \mathrel{+}= F1(i)$
$Recall \mathrel{+}= Recall(i)$
$Precision + = Precision(i)$
$i++$
$MAE = \frac{MAE}{k},$
$MSE = \frac{MSE}{k}$
$RMSE = \frac{RMSE}{k},$
$F1 = \frac{F1}{k}$
$Recall = \frac{Recall}{k},$
$Precision = \frac{Precision}{k}$
Plotting charts and store result on file
End

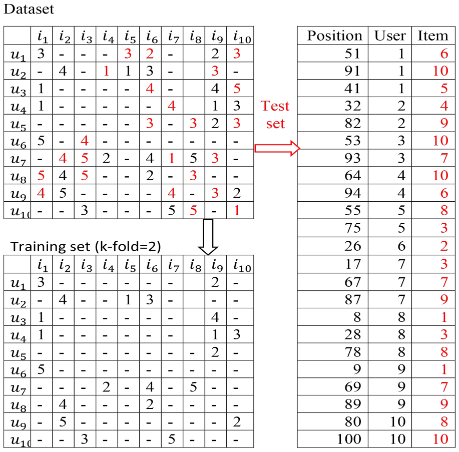
Fig. 5. Flowchart of Algorithm for Evaluation Recommendation Model.

In addition, see flowchart of evaluation model algorithm in Fig. 5, the recommended evaluation measures used in [9], [10] focus on only two main groups, the first is the accuracy of the ratings such as MAE, MSE and $RMSE$ [4], [6] to determine the accuracy of the prediction ratings are missing and the second group is predictive predictions such as $precision, recall$, and $F1 - score$ [4], [6], which focus on introducing items that are useful to the user and helping them make the right decision. These metrics, however, have a major downside: they are concerned with the entire dataset rather than top-N recommendation lists. Therefore, it is not easy to assess accurately the recommender systems when comparing the list of items recommended to the list of relevant items, because the metrics do not focus on the identification of rank and position of an item in the list.

In practice, good recommender systems are not only interested on how many relevant results they give, they also want to give users with a good order. They need to be able to put relevant items remarkably high up the list of recommendations. Most probably, the users will not scroll through hundreds of items to find favorite item they like. Take now famous searching engines like Google [*], Bings [†] as an example, it is evident that they prioritize relevant query results in some sort of descending order. Therefore, it is necessary to need metrics of ranking awareness properly to select recommender systems that can solve major aims: (1) Where position of item that recommender system suggests is in list of recommendation result, (2) how good recommender system could solve in modeling users' relative preference.

To overcome this shortcoming of the recommender system model, two suggestions are given for improving the evaluation quality of the recommender system.

The first, in terms of dataset partitioning method for evaluating, the dataset that has n transactions and m items can be partitioned into two sets of training and testing based on number of items ranked per transaction instead based on numbers of transaction on dataset.



Dataset

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9	i_{10}
u_1	3	-	-	-	3	2	-		2	3
u_2	-	4	-	1	1	3	-		3	-
u_3	1	-	-	-	-	4	-		4	5
u_4	1	-	-	-	-	-	4		1	3
u_5	-	-	-	-	-	3	-	3	2	3
u_6	5	-	4	-	-	-	-	-	-	-
u_7	-	4	5	2	-	4	1	5	3	-
u_8	5	4	5	-	-	2	-	3	-	-
u_9	4	5	-	-	-	-	4	-	3	2
u_{10}	-	-	3	-	-	-	5	5	-	1

Test set

Position	User	Item
51	1	6
91	1	10
41	1	5
32	2	4
82	2	9
53	3	10
93	3	7
64	4	10
94	4	6
55	5	8
75	5	3
26	6	2
17	7	3
67	7	7
87	7	9
8	8	1
28	8	3
78	8	8
9	9	1
69	9	7
89	9	9
80	10	8
100	10	10

Training set (k-fold=2)

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9	i_{10}
u_1	3	-	-	-	-	-	-		2	-
u_2	-	4	-	-	1	3	-		-	-
u_3	1	-	-	-	-	-	-		4	-
u_4	1	-	-	-	-	-	-		1	3
u_5	-	-	-	-	-	-	-	-	2	-
u_6	5	-	-	-	-	-	-	-	-	-
u_7	-	-	-	2	-	4	-	5	-	-
u_8	-	4	-	-	-	2	-	-	-	-
u_9	-	5	-	-	-	-	-	-	-	2
u_{10}	-	-	3	-	-	-	5	-	-	-

Fig. 6. An Example of Proposed Data Partition Method.

Accordingly, for each transaction, items will be randomly partitioned into k folds: $n \times (k-1)/k$ items for training set and n/k items for testing set. In Fig. 6, dataset's items are partitioned in 2 folds randomly (for simplicity in illustration), in which items for test set are presented in red, (remained items in black for training set). In this way the training and testing sets are formed from all transactions, which means that all transactions are involved in both the training and testing set, and the number of items per sets are $(k-1) \times n/k$ and n/k percent of items' dataset correspondingly. A problem in test set's presentation is that it will be very sparse compared to the dataset. In order to save memory and time for manipulation, a good suggestion that using one-dimension array namely positions for storing position of all items in test set where $0 < position[k] < n\times$ m. Accordingly, the row i and column j of an item are defined respectively as follows the quotient +1 and the remainder in $\frac{position}{n}$, if $quotient = m$ then $(i, j) = (n, m)$. For instance, in Fig. 6 position[1]= 51 means that transaction of user u_1 rated item i_6 (since the remainder of 51/10 is 1, and the quotient of 51/10 plus 1 is 6). It is apparent that items in the training and testing set are

[*] https://www.google.com/
[†] https://www.bing.com/

extracted from all transactions of the dataset, which makes better model testing and training. Moreover, this approach also fixes the shortcoming of k-fold partitioning based on transaction because the number of items retained to build a training model that is no longer limited to the minimum number of items per transaction in dataset.

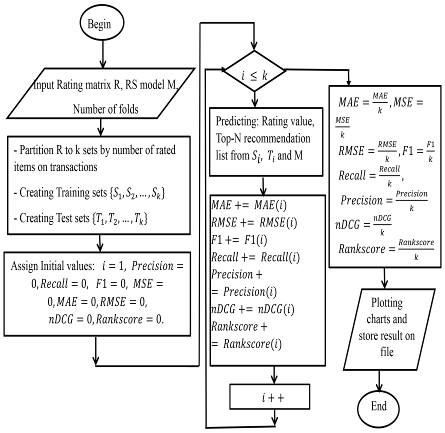

Begin
Input Rating matrix R, RS model M, Number of folds
- Partition R to k sets by number of rated items on transactions
- Creating Training sets $\{S_1, S_2, \ldots, S_k\}$
- Creating Test sets $\{T_1, T_2, \ldots, T_k\}$
Assign Initial values: $i = 1$, $Precision = 0, Recall = 0, F1 = 0, MSE = 0, MAE = 0, RMSE = 0, nDCG = 0, Rankscore = 0.$
$i \leq k$
Predicting: Rating value, Top-N recommendation list from S_i, T_i and M
$MAE += MAE(i)$
$RMSE += RMSE(i)$
$F1 += F1(i)$
$Recall += Recall(i)$
$Precision += Precision(i)$
$nDCG += nDCG(i)$
$Rankscore += Rankscore(i)$
$i++$
$MAE = \frac{MAE}{k}, MSE = \frac{MSE}{k}$
$RMSE = \frac{RMSE}{k}, F1 = \frac{F1}{k}$
$Recall = \frac{Recall}{k}$,
$Precision = \frac{Precision}{k}$
$nDCG = \frac{nDCG}{k}$
$Rankscore = \frac{Rankscore}{k}$
Plotting charts and store result on file
End

Fig. 7. Proposed Algorithm Flowchart to Improve Recommendation Model.

In addition, due to the way of partitioning the dataset by items on each transaction, it is not necessary to create the unknown and know set from the test set as in transaction-based dataset partitioning way.

The second, in relation to using of evaluating measures, besides predicting-based and classifying-based precision evaluating metrics, the ranking-based ones are also suggested to evaluate more comprehensively the recommendation list's quality. Although there are many metrics of this type like $MRR, MAP, nDCG$ and $Rankscore$ [4], [6], only $nDCG$ and $Rankscore$ are suggested because they can be used to deal with both binary and none-binary dataset.

Finally, these proposals are presented in the algorithm as shown in Fig. 7.This is a revision of algorithm in Fig. 5 based on two major changes, that is using a new method for dataset partition based on number of ranked items per transaction and adding measures nDCG and Rankscore measures. These are the major changes for improving the efficiency and quality of the recommended system as discussed in the following experiments.

IV. Experiment

A. Datasets

Using collaborative filtering based-on implication field recommendation model described above, we conduct experiments on both the binary dataset (MSWeb)[‡] and the quantitative dataset (MovieLens)[§]. The MSWeb dataset is created by sampling and processing the www.microsoft.com logs of 38.000 anonymous, randomly selected users in one-week timeframe. For each user, the dataset lists all the areas of the web site (Vroots) that user visited in a one-week timeframe in February 1998. This dataset contains 32710 valid users and 285 Vroots. The MovieLens dataset collected by GroupLens consists of 100.000 ratings made by 943 users for 1.682 films. The ratings range from 1 to 5 corresponding to from the lowest to the highest.

To serve the experiment to be more accurate, the datasets are preprocessed by:

Normalization of data: Users who rank high (or low) for all their films/Vroots depending on the individual can lead to bias. Eliminate this effect by normalizing the data so that the average rating of each user is the same scale.

Selecting relevant data: Ignoring data can lead to bias and to speed up computation, by not interested in the films/Vroots has had only a few times, because the ratings of these films/Vroots may be subject to bias due to lack of data, and users rated only a few films because their ratings may be biased.

Using k-fold cross validation method (with k=5 for this paper): to avoid overfitting problems as well as to get better accuracy as for each model evaluation. The dataset (MovieLens or MSWeb) is split into equal sized k-fold to build training set (using k-1 fold) and test set (using remaind fold) by the number of ratings on transaction instead of by number of transactions on dataset to overcome the limitations as analyzed in Section III-B.

B. Tool

The experiments were performed on implication field RS tools developed in the R language [¶]. This tool is developed for making, performing, and evaluating models of recommender system based on implication field as described in Section III-A. In addition, it can build and run other collaborative filtering-based recommender systems for mutual comparison and evaluation. The SIA measure is used for $F_{R_{IMP}}$ is $\varphi_{n_{A\bar{B}}}(a, b)$.

C. Analyze Equipotential Surfaces in Implication Field

1) Experiment Description: To analyze the implication field as a set of equipotential planes, In this experiment, the Implication field-based recommender system model was performed on the Movielens non-binary dataset that was described in Section IV-A, on the F_{imp} (minsup = 0.1, minconf = 0.3, min$\varphi_{n_{A\bar{B}}}(a, b) = 0.5)$.

2) Results and Discussions: Results are presented in Fig. 8 and 9, they are presented in the form of 3D scatter and 3D graph, representing equipotential surfaces with a warm color (red) that is common in the implied intensity range of 0.8 to 1.0, and the remaining scattered is the equipotential surfaces have the implication intensity decreasing by the gradual cold color (blue).

In Fig. 10, contour form, accordingly, the implication field with equipotential surfaces has a variable value spectrum of implication intensity concentrated in the range 0.8 to 1 represented by the gray spectrum and the remainder is

[‡] https://grouplens.org/datasets/movielens/100k/

[§] https://kdd.ics.uci.edu/databases/msweb/msweb.html.

[¶] https://www.r-project.org/about.html

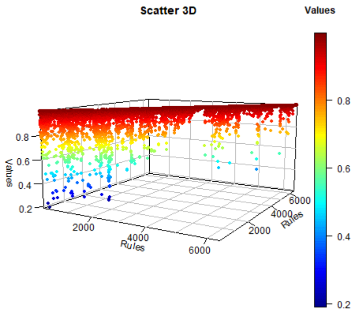


Fig. 8. Implication Field and Its Equipotential in Scatter 3D.

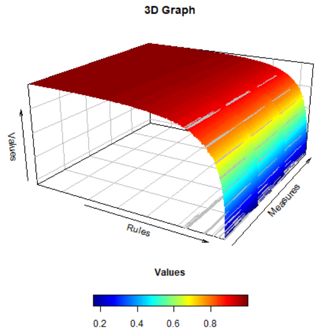


Fig. 9. Implication Field and Its Equipotential in Graph 3D.

represented by the gradual transition color spectrum green. Fig. 11, the implication intensity variation on the equipotential surfaces presented in 3-dimensional form, it is easy to see that the implication samples with high implication intensity are concentrated on warm colored equipotential surfaces and rapidly decrease. in the low intensity region is represented by blue. The common recommendations will be filtered on high intensity equipotential surfaces, whereas those for specific, rare items will be provided in low implication equipotential surfaces.

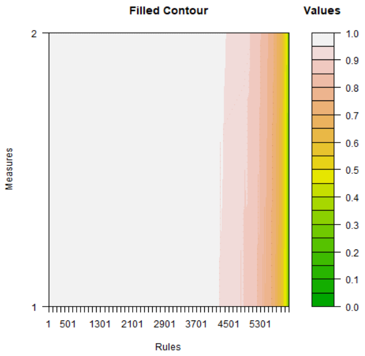


Fig. 10. Implication Field and Its Equipotential in Contour.

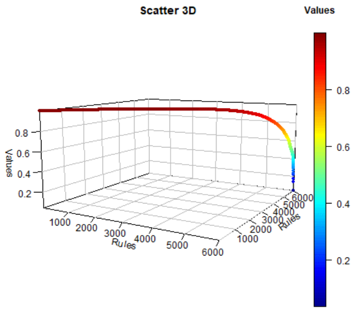


Fig. 11. The Implication Variation in the Implication Field.

D. Scenario 1. Comparing with Traditional Recommendation Models

1) Experiment Description: In this experimental scenario, the recommender system model based on the implication statistical field (ISFRS), is compared with the traditional collaborative filtering recommendation models based on user for both Cosine (UBCF_cRS) and Pearson measures (UBCF_psRS), and collaborative filtering recommendation models based on item for both Cosine (IBCF_cRS) and Adjusted Cosine measures (IBCF_acRS), The data set used in this experiment is the Movielens non-binary data set described in Section IV-A. For the collaborative filtering models to have good results, a problem needs to face is how to choose the number of neighbors best, we try to experiment on many neighbor k parameters for these models including k=2, 5, 10, and 15, and finding that k = 15 is better than other values. Moreover, dataset partitioning for training and testing is conducted based on number of items in transactions instead of numbers of transaction. Recommendation models were experimented on two groups' measure: classification and ranking.

2) Results and Discussions: The first, models were experimented on classification measures, on ROC curve, precision/ recall, F1, The results are shown in Fig. 12 to 14. As a result, the ISFRS model is the best, next is the collaborative filtering model based on user using both Pearson and cosine measures, and finally weakest model is item-based collaborative filtering model (in case of both Pearson and adjusted cosine measures).

The second, models were experimented on ranking measures, on $nDCG$ and $Rankscore$. The results, have presented in Fig. 15 and 16, also show the preeminence of the ISFRS model over the collaborative filtering model, which is the same as the case of the group of classification measures that is discussed above.

These result in this experiment shows the contribution of both the proposed ISF RS model and the proposed data partitioning method to evaluation in improving the model's classification and ranking capability and training quality compared to the recommended models based on traditional collaborative filtering.

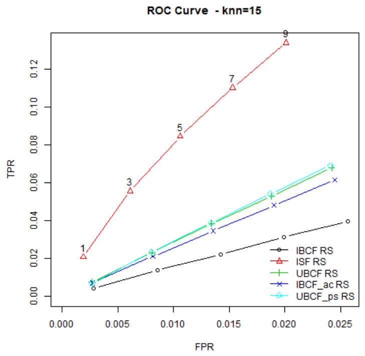


Fig. 12. ROC Curve of ISF Model and CF others, k=15.

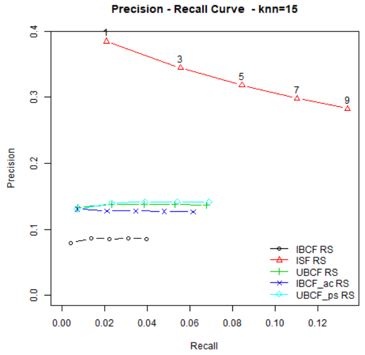


Fig. 13. Precision/ Recall of ISF Model and CF others, k=15.

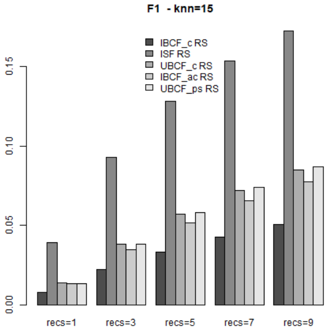


Fig. 14. F1 of ISF Model and CF others, k=15.

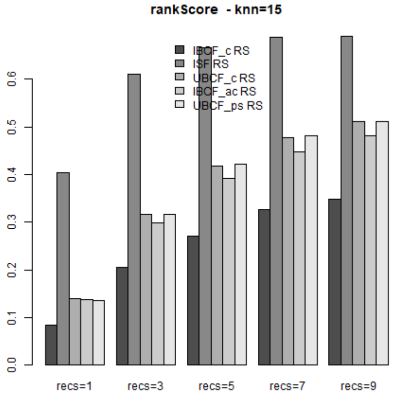


Fig. 15. Rankscore of ISF Model and CF others, k=15.

E. Scenario 2. Comparing with Implicative Recommendation Models

1) Experiment Description: In this experimental scenario, the MSWeb binary data set is used to compare the implication statistical field recommender system (ISFRS) model with two other SIA models that was reviewed in Section II-B including works in [7], [8] (Implication index and intensity - IIIRS) and [9], [10] (Phi-Cohesion- Gamma- PCGRS) on two types of measure. Reason that MSWeb is chosen to use in this experiment instead of Movielens as in previous one is attributable to models in [7], [8] was designed and performed on binary dataset only as mentioned in Section II-C. In addition, to get more precision results, dataset partitioning for training and testing sets is conducted based on number of rated items in transactions instead of numbers of transaction.

2) Results and Discussions: The first is classification measures including precision/ recall, ROC, F1, experimental results show the preeminence of IFS RS recommendation model compared to PCG RSmodel and IIIRS model, in which the weakest is the model IIIRS on all 3 evaluation measures, as shown in Fig. 17 (for Recall/Precision), Fig. 18 (for ROC curve), Fig. 19 (for F1).

The second is ranking measures, the experimental results, were shown in Fig. 20 (for Rankscore) and Fig. 21 (for nDCG), are quite similar to the results on the group of classification measures. This means that the ISFRS model has the best results ranking items according to the nDCG and Rankscore indicators, followed by the PCGRS model and the worst is the IIIRS model.

This indicates that recommender system based on the implication statistical field has ability better on both classification and ranking compared to previous recommendation model based on SIA applying. The experiment proofed that proposed ISFRS resolved three problems of recommender systems based on applying SIA previously as mentioned in Section II-B. Therefore, it is apparent that it is a new and promising trend in applying statistical implication analysis theory to the recommender systems domain.

V. Conclusion

In order to ensure relevance and novelty for recommender systems, its proposal has to be personalized enough to meet the user's personal preferences and deep enough to make a pleasant surprise for the user. In this regard, the paper has proposed a novel recommendation model based on the implication field to significantly improve the quality of the recommender system compared to the traditional collaborative filtering-based recommender systems. The second contribution of the paper is to propose a new data set partitioning method to build training and test sets to build and train the recommendation model,

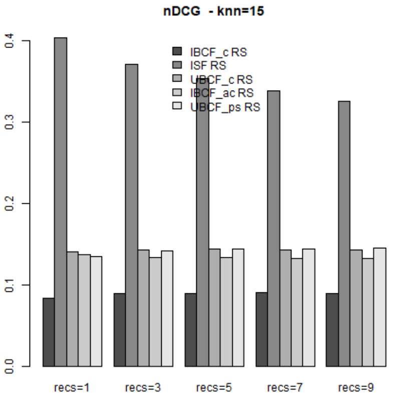


Fig. 16. nDCG of ISF Model and CF others, k=15.

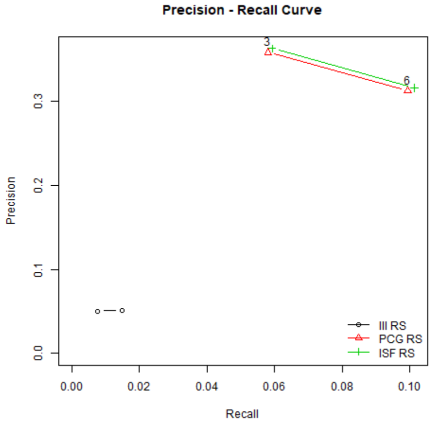


Fig. 17. Precision/Recall of ISF Models and others in SIA.

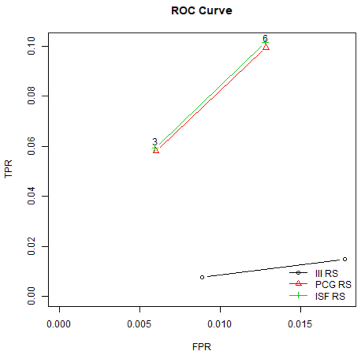


Fig. 18. ROC Curve of ISF Models and others in SIA.

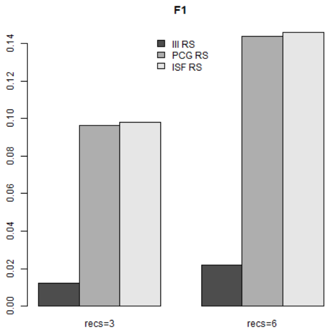


Fig. 19. F1 of ISF Models and others in SIA.

based on the rating ratio per transaction instead of based on the number of transactions, which overcomes the limitation of sparse datasets in recommender systems, making them more likely to recommend accurately. Another contribution to this paper is to propose metrics that provide a more in-depth assessment of the quality of recommendations. In addition to the metrics for precision of classification like precision, recall, and F1, metrics for rank score were also added that evaluate the relevance of recommendations like nDCG and Rankscore and using them to compare among different models. This helps evaluate outputs' quality of recommendation models more comprehensively as shown in the experimental results. Finally, this paper also aggregates and compare the effectiveness of the works existing SIA-based recommendation systems, experiment's results showing that the application of the implication variation tendency in the implication field is the most satisfactory result in all these recommender systems.

From these results, it is clear that the exploitation of the relationships between variables (objects/ individuals/ attributes/ items) in the form implication rules in the implication field has achieved positive results for recommender systems. Therefore, the study to extend further these relationships in the form between rules or/and between rules and variables in the implication field and exploitation them to further improve the effectiveness of the recommendation system is a promising one in the future.

REFERENCES

[1] Adomavicius Gediminas, Tuzhilin Alexander, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, *IEEE Transactions on Knowledge and Data Engineering, Vol.17 No.6*, pp.734–749, 2005.

[2] Ahmed Mohammed K. Alsalama, "A Hybrid Recommendation System Based On Association Rules", *International Science Index, Computer and Information Engineering Vol:9, No:1*, pp.55-62, 2015.

[3] Francesco Ricci, Lior Rokach and Bracha Shapira, *Introduction to Recommender Systems Handbook*, Springer-Verlag and Business Media LLC, pp.1-35, 2011.

[4] Guy Shani, Asela Gunawardana, "Evaluating Recommendation Systems",*Recommender System Handbook*, Springer, pp.257-297, 2010.

[5] Gavin Shaw, Yue Xu and Shlomo Geva, "Using Association Rules to Solve the Cold-Start Problem in Recommender Systems", *Advances in Knowledge Discovery and Data Mining*, pp.340-347, 2010.

[6] Herlocker J.L et al. "Evaluating collaborative filtering recommender systems". *ACM Transactions Information System, vol. 22, no. 1*, pp.5–53, 2004.

[7] Nghia Quoc Phan, Phuong Hoai Dang, Hiep Xuan Huynh, "Collaborative recommendations based on statistical implication rules", *Journal of Computer Science and Cybernetics, Vol. 33, No. 3*, pp.247-262, 2017.

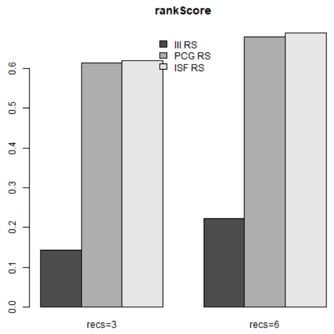


Fig. 20. Rankscore of ISF Model and others in SIA.

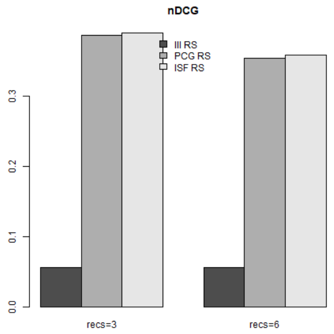


Fig. 21. nDCG of ISF Models and others in SIA.

[8] Nghia Quoc Phan, Ky Minh Nguyen, Hoang Tan Nguyen, Hiep Xuan Huynh, "The recommendation system is based on association rules and statistical implication measures". *Proceedings of The 8th National Conference on Fundamental and Applied IT Research – (FAIR'15)*; Natural Science and Technology Publishing House, pp.297-308, 2015.

[9] Lan Phuong Phan, Hung,Huu Huynh, Hiep , Xuan Huynh, "Recommendation using Rule based Implicative Rating Measure", *International Journal of Advanced Computer Science and Applications (IJACSA)*, pp.176-181, 2018.

[10] Lan Phuong Phan, Hung Huu Huynh, Hiep Xuan Huynh, "Recommender systems based-on implication intensity and contribution measure", *Proceedings of the X National Conference on Fundamental and Applied IT Research (FAIR18)*; Natural Science and Technology Publishing House, pp.256-274, 2017.

[11] Rakesh Agrawal, Ramakrishnan Srikant, "Fast algorithms for mining association rules", *Proceedings of the 20th International Conference on Very Large Data Bases*, p.487-499, 1994.

[12] Régis Gras, Pascale Kuntz and Nicolas Greffard, "Notion of implicative field in implicative statistical analysis", *The 8th International Meeting on Statistical Implicative Analysis*, Tunisia, pp.1-21, 2015. (in French)

[13] Regis Gras, Raphael Couturier, "Specificities of Implicative Statistical Analysis (A.S.I.) compared to other quality measures of association rules", *Quaderni di Ricerca in Didattica - GRIM (ISSN on-line 1592-4424)*, pp.19-57, 2010. (in French)

[14] Régis Gras, Einoshin Suzuki Fabrice Guillet, Filippo Spagnolo (Eds.), *Statistical Implication Analysis - Theory and Application*, Springer Verlag, 2008.

[15] Timur Osadchiy, Ivan Poliakov, Patrick Olivier, Maisie Rowland, Emma Foster,"Recommender system based on pairwise association rules", *Expert Systems with Applications 115*, pp.535–542. 2018.

[16] Tzung-Pei Hong, Chun-Hao Chen, Yeong-Chyi Lee, and Yu-Lung Wu., "Trade-off between computation time and number of rules for fuzzy mining from quantitative data", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, pp.587-604,2001.

[17] Tzung-Pei Hong, Chun-Hao Chen, Yeong-Chyi Lee, and Yu-Lung Wu, "Genetic-Fuzzy Data Mining with Divide-and-Conquer Strategy", *IEEE Transactions on Evolutionary Computation*, pp.252-265, 2008.

[18] Hoang Tan Nguyen, Hung Huu Huynh, and Hiep Xuan Huynh, "Collaborative filtering recommendation with threshold value of the equipotential surface in implication field", *Second ACM International Conference on Machine Learning and Soft Computing*, pp.39-44,2018.

[19] Hoang Tan Nguyen, Hung Huu Huynh, and Hiep Xuan Huynh, "Collaborative Filtering Recommendation in the Implication Field", *International Journal of Machine Learning and Computing, Volume 8 Number 3*, pp.214-222, 2018.

[20] Hoang Tan Nguyen, Lan Phuong Phan, Hung Huu Huynh, Hiep Xuan Huynh, "Recommendation with quantitative implication rules", *EAI Endorsed Transactions on Context-aware Systems and Applications, Volume 6 , Issue 16*, pp.1-8, 2019.

[21] Hoang Tan Nguyen, Lan Phuong Phan, Hung Huu Huynh, Hiep Xuan Huynh , "Improved collaborative filtering recommendations using quantitative implication rules mining in implication field", *ICMLSC 2019: Proceedings of Third ACM International Conference on Machine Learning and Soft Computing*, pp.110–116, 2019.