Collaborative Filtering Recommendation in the Implication Field

Hoang Nguyen-Tan, Hung Huynh-Huu, and Hiep Huynh-Xuan

***Abstract*—In the age of information explosion today, the Recommender systems have become increasingly important and popular in supporting human decision-making problems. In the Recommender Systems, Collaborative filtering is one of the most popular and effective techniques available today in the recommender system. However, most of them use symmetric similarity measures. Therefore, the default effect and the role of the pair of users are the same, but in practice this may not be true. In this paper, we propose a method new approach in building the collaborative filtering recommender system in the implication field, uses the asymmetry measures to rank and filter the information to improve accurate precision of the traditional recommender systems.**

***Index Terms*—Implication index, implication intensity, implication field, collaborative filtering, implication rule.**

I. Introduction

The recommender system (RS) has become the background application of many e-commerce applications and information service providers. They help solve information overload and provide appropriate information in the era of information explosion today: Youtube automatically transfers clips related to the clip you are watching or recommends clips that you may like. Amazon will automatically recommends products that can often be bought together, or recommends items that you may like based on your purchase history when you shop on it. Facebook recommends a friend or show to advertise products related to the keyword you just searched. Netflix automatically recommends movies to the user, and many other examples that the internet has the ability to automatically recommend to users the products they may like. In actual practice, more than 65% of movies has been watched by Netflix customers are recommended movies, 35% of sales at Amazon arise from recommended items, 28% of people would like to buy more music on ChoiceStream if they find what they like, and so on. By right-leaning advertising like this, the effectiveness of marketing will also increase. The algorithms behind these applications are Machine Learning algorithms, commonly referred to as recommender systems [1], [2] or Recommendation Systems. The algorithms for the recommender system have attracted the attention of the researchers for practical application. Among them, the collaborative filter algorithms [3] are the most widely used. Most of these algorithms are based on the measure of symmetry for filtering information and recommendations for users. Recently, several solutions have been proposed that use asymmetric similarity to the recommendation system, such as asymmetric similarity for collaborative filtering via matrix factorization [4]-[6]. Recommendation with asymmetric user influence and global importance value [7] to address the asymmetric effects of users in the recommendation systems. Another new trend is the use of statistical implication analysis in the recommendation system, which addresses the problem of asymmetric user influence and solves the problem of assessing the occurrence or functional relationship. Interaction between users and data items in practice, such as the recommender system model based on approach to association rules combined implicative measure [8], to overcome the disadvantage of traditional recommender systems (They only focus on the logic that demonstrates the existence or absence of a priority relationship between the user and the item), in this approach, the authors are particularly interested in the ratio or implicative relationship between the user and the data item in a particular context in order to make recommendations to the user more effective. Another study in the application of statistical implication analysis to the recommender system was the user-based collaborative filtering recommender system using association rules combined implication cohesion measure [9] to calculate the similarity for each pair of users in collaborative filtering. Recently, in [10], we proposed an asymmetric measure for recommender system based on statistical implicative analysis user preferences over time, and in [11] we also proposed a recommendation based on the variance of implication index in implication field to user for solving these issues.

In this paper we also use statistical implication analysis to propose a new approach to collaborative filtering based on threshold value of the equivalence plane in the implication field [12] to continue to solve the problems of asymmetric user influence and the implication relationship between the users in the recommender systems.

The paper is organized in five parts, the first one introduces the context and issues to be solved by the present system as well as proposing our proposed approached, and the second part presents the related contents to the statistical implication analysis and the extended studies in the implication field, the third part presents the model of the recommender system based on the variance of the implied index in the implication field, the next part is the experimental section model with scenarios and finally conclusions.

Manuscript received February 17, 2018; revised May 6, 2018.

Hoang Nguyen-Tan is with the Department of Information and Communications of Dong Thap Province, Viet Nam (e-mail: hoangntdt@gmail.com).

Hung Huynh-Huu is with University of Science and Technology, Da Nang University, Viet Nam (e-mail: hhhung@dut.udn.vn).

Hiep Huynh-Xuan is with Can Tho University, Viet Nam (e-mail: hxhiep@ctu.edu.vn).

doi: 10.18178/ijmlc.2018.8.3.690

II. Implication Statistical Field

A. Implication Statistical Analysis

Statistical implication analysis theory [13]-[15], was proposed by Regis Gras, studies the implication relationship of data variables. Measures in the analysis implicative statistical us implication index (aka Gras implication index) and implication intensity, are used to detect the rule or R-rule (rule of the rule) [16], [17] strong implicative relationship between the two sides of the rule, or to measure the correlation between two variables (individual, attribute ...), these measures are asymmetric. In addition, statistical implication analysis focuses on counter example factor analysis. It can be presented as follows:

Let E be a finite set of binary variables, A and B are two subsets of E, respectively, which contain the elements $a \in A$ such that $A(a) = true$ and $b \in B$, such that $B(b) = true$, sets $\bar{A}$, $\bar{B}$ is the complement of sets A and B respectively, let $n_a = card(A), n_b = card(B)$, $n_{\bar{a}} = n - n_a$, $n_{\bar{b}} = n - n_b$ is the cardinality of A, B, $\bar{A}$, B and $\bar{B}$ respectively, and $n_{a\bar{b}} = card(A \cap \bar{B})$ is the cardinality of the set $A \cap \bar{B}$, that is a set containing the elements that satisfy the properties $a = true$ and $b = false$, $n_{a\bar{b}}$ also called counter-example[13].

The implication relationship between A and B is modeled in the statistical implication analysis as follows (see Fig. 1).

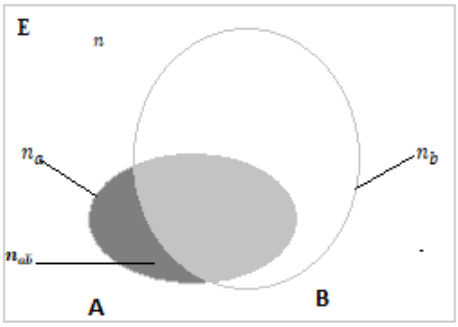

Fig. 1. The illustration of the components of statistical analysis implicated by Venn diagram.

Implication intensity measure $\varphi(a,b)$ of rule $A \to B$ is defined by [13], [14]:

$$\varphi(a,b) = 1 - P\left(card\left(A \cap \bar{B}\right) > n_{a\bar{b}}\right) = \begin{cases} 1 - \sum_{s=0}^{n_{a\bar{b}}} \frac{\lambda^s}{s!} e^{-\lambda} = \frac{1}{\sqrt{2\pi}} \int_{q(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt, \text{if } n_y < n \\ 0, \quad \text{other wise} \end{cases} \quad (1)$$

where $\boldsymbol{\lambda = \frac{n_a n_{\bar{b}}}{n}}$ and $q(A,\bar{B})$ is implication index and is definded by:

$$q\left(a,\bar{b}\right) = \frac{n_{a\bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}} \quad (2)$$

In terms of approximation (e.g. $\lambda \geq 4$), $q\left(a,\bar{b}\right)$ is the approximation of the normal distribution $N(0,1)$.

The implication rule that A→B is admissible at the confidence level α if and only if $\varphi(A,B) \geq 1 - \alpha$ [13], [14].

B. Implication Index Variation

Let consider small variations in the neighborhood of all four observed values of variables $n, n_a, n_b, n_{a\bar{b}}$. These variables must be considered as real numbers and q as a continuously differentiable function with respect to these variables constrained to respect inequalities: $0 \leq n_a \leq n_b$; $n_{a\wedge\bar{b}} \leq \inf\{n_a, n_b\}$ and $\sup\{n_a, n_b\} \leq n$. The differential of q in Frechet's geometry is expressed in the following way [12]:

$$dq = \frac{\partial q}{\partial n} dn + \frac{\partial q}{\partial n_a} dn_a + \frac{\partial q}{\partial n_b} dn_b + \frac{\partial q}{\partial n_{a\bar{b}}} dn_{a\bar{b}} = gradq.dM \quad (3)$$

where M the point with the coordinates $(n, n_a, n_b, n_{a\bar{b}})$ belong to the scalar vector field C, dM is the differential component vector of the instance variables and grad q is the partial differential vector of the variables.

From (3), the differential of the function q appears as a scalar product between gradient q and the increase of q on the surface representing the variables of the function $q(n, n_a, n_b, n_{a\bar{b}})$. $grad\ q$ denotes the variability of the function of four variables, which is the cardinalities of the sets E, A, B, and $A \cap \bar{B}$, which points to the direction of the function q in four dimensions space. In fact, the value of this differential lies in the estimation of the increase (positive or negative) of q that we note Δq relative to the respective variations $\Delta n, \Delta n_a, \Delta n_b,$ and $\Delta n_{a\bar{b}}$. So we have [12], [13]:

$$\Delta q = \frac{\partial q}{\partial n} \Delta n + \frac{\partial q}{\partial n_a} \Delta n_a + \frac{\partial q}{\partial n_b} \Delta n_b + \frac{\partial q}{\partial n_{a\bar{b}}} \Delta n_{a\bar{b}} + o(\Delta q) \quad (4)$$

where $o(\Delta q)$ is an infinitely small.

To more specific, the partial derivative according to $n, n_a, n_b, n_{a\bar{b}}$ [12], [13] :

$$\frac{\partial q}{\partial n} = \frac{1}{2\sqrt{n}} \left(n_{a\bar{b}} + \frac{n_a n_{\bar{b}}}{n} \right) \quad (5a)$$

$$\frac{\partial q}{\partial n_a} = -\frac{1}{2} \frac{n_{a\bar{b}}}{\sqrt{\frac{n_{\bar{b}}}{n}}} \left(\frac{n}{n_a} \right)^{\frac{3}{2}} - \frac{1}{2} \sqrt{\frac{n_{\bar{b}}}{n_a}} \quad (5b)$$

$$\frac{\partial q}{\partial n_b} = \frac{1}{2} n_{a\bar{b}} \left(\frac{n_a}{n} \right)^{\frac{1}{2}} (n - n_b)^{\frac{3}{2}} + \frac{1}{2} \left(\frac{n_a}{n} \right)^{\frac{1}{2}} (n - n_b)^{\frac{1}{2}} \quad (5c)$$

$$\frac{\partial q}{\partial n_{a\wedge\bar{b}}} = \frac{1}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}} = \frac{1}{\sqrt{\frac{n_{a(n-n_b)}}{n}}} \quad (5d)$$

From (5d), if $n_{a\bar{b}}$ increases, the implication index increased, and thus the intensity implication decreased.

Now, to further examine the relationship between the implication index and implication intensity. Take the primitive of the (1), we have:

$$\frac{d\varphi}{dq} = \frac{1}{\sqrt{2\pi}} e^{\frac{-q^2}{2}} < 0 \quad (6)$$

This confirms that the implication intensity increases as q decreases, but the rate of increase is determined by (6), which allows for a more rigorous study of the variability of φ.

C. Implication Field

Consider the implication index $q(a,\bar{b})$ in the four-dimensional space E, with the point M whose coordinates are the parameters associated with the binary variables a and b are $(n, n_a, n_b, n_{a\bar{b}})$, then $q(a,\bar{b})$ is a scalar field by applying the mapping from space R^4 to space R. For the vector $grad.q$ contains the partial derivatives of q for the variables $n, n_a, n_b, n_{a\bar{b}}$ is a special gradient field is called implication field, because it satisfies the Schwartz criteria in the (7) for the mixed differential, that is, The mixed derivative event of each pair of variables [12], is:

$$\frac{\partial}{\partial n_{a\bar{b}}}\left(\frac{\partial q}{\partial n_b}\right) = \frac{\partial}{\partial n_b}\left(\frac{\partial q}{\partial n_{a\bar{b}}}\right) \quad (7)$$

Similar to each other pairs in the variables $(n, n_a, n_b, n_{a\bar{b}})$, $grad\ q$ is considered to be the potential of q. Vector grad q is performed to change the space of the confidentiality of the case, it's sort of the low value to a higher value. At each point of the gradient, we observe an increase in the implied density of space and to what extent the rate at which it changes under the influence of one or more parameters.

In ISA, the four-dimensional space forms an implication field, consisting of ordered ordinate planes corresponding to the sequential successive values of q relative to the cardinalities $(n, n_a, n_b, n_{a\bar{b}})$ that would be varied [12]. Consider the implication index as a function of four variables $q(n, n_a, n_b, n_{a\bar{b}})$. A line or plane of equipotential in implication field is curved in E, an 4-dimensionals space along which or at which point a variable M maintains the same value of potential of q. The plane of equipotential is orderly. The equation of this curve is shown in [12]:

$$q(a,\bar{b}) - \frac{n_{a\bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}} = 0 \quad (8)$$

III. Recommendation Based on Equipotential Plane in Implication Field

A. Implication Statistical Rules

Let D is a dataset that consists of T transactions, each transaction Ti consists of objects or items that are objects that appear in transactions such as (products, services, ...). The itemset I is a set consisting of m items.The implication rule is an extension of association rules. It is a rule of the form $X \rightarrow Y$, where $X \cap Y = \emptyset$; $X, Y \subset I$ are itemsets or other rule (In this article we only limit X,Y are itemsets.), X is called antecedent, Y is the consequent. To measure the quality of the rules in addition to the use of quality measures such as association rules such as support, confidence, lift, the implication rule also uses the implication index $q(a,\bar{b})$ and the implication intensity $\varphi(a,b)$ measure as outlined in Section II.A. These measures expressed the degree to implication which association rules are not.

To model the recommender system, the implication rule is expressed as follows: assume I is the set of m items, $A \subset I$ that is the set of items rated by the user u_a; $\bar{A}$ is the complement of A. The set $B \subset I$ is the set of items rated by the user u_b; $\bar{B}$ is the complement of B; $n_a = card(A)$ is the number of data items rated by the user u_a, which is the number of elements of the set A); $n_b = card(B)$ is the number of data items rated by user u_b (number of elements of set B); $n_{a\bar{b}} = card(A \cap \bar{B})$ is the number of items rated by the user u_a but not rated by the user u_b., An implication rule expressed by a set of four variables $(n, n_a, n_b, n_{a\bar{b}})$. They are called the cardinalities of the implication rules. In other words, the relationship between the user u_a and u_b is the relationship between the item set A is liked by the user u_a and the item set B is liked by the user u_b. The implication rule is represented by the set of four elements $(n, n_a, n_b, n_{a\bar{b}})$.

B. Threshold Implication Variability between Equipotential Planes

As discussed in the previous section, statistical implication analysis focuses on counter example factor to infer and analysis the implication rule that represent data relationships. Thus, the variation of counter-examples that affect how for implication rule.

It is difficult to replace an original rule by another rule when a few counter examples (unlikelihood) appear, only when the counter example higher the confidence of the rule decreases and the rule can be denied. However, when the number of example (likelihood) is numerous and the number of counter examples is rarer, the rule becomes stronger and is recognized. For example, let's look at the rules that are acceptable. “Ferrari cars are red." Even if one or two of the counter examples appear (Ferrari cars are not red), this rule is maintained, and it will be even confirmed once again by the release of new examples. Thus, contrary to mathematics, where rules are not allowed to have any exceptions, the rules here considered are still acceptable when the number of counter-examples remains in the "acceptable" threshold, because in these situations rules are still active and effective. In data analysis, the problem is to define a consensus standard, thereby quantifying the confidence threshold of the rule according to user requirements.

In this section, we propose a recommendation based on the variation of the implication index depending on the variation of the counter example in the implication field for determining the equipotential plane of implication index and implication intensity, from there, the item (or k-top items list) consultant is suitable for the user with a definite implication threshold.

1) Determines the variation threshold of the implication index

The threshold θ is the tolerance value of q in the same equipotential plane, θ is defined by byFactor. To determine θ, it is necessary to consider how the dependent variable $q(a,b)$

varies when an element x is added to (or removed from) the sample data, with four occurrences.

Let $(\lambda_1, \lambda_2), (q_1, q_2)$ and (φ_1, φ_2), corresponding to λ, implication index q and implication intensity φ (of rule $A \to B$) are related to the original and the extended data sample as Table I (Value ± 1 corresponds to 1 in the additional case and -1 when removing the x in the dataset).

TABLE I: THE VARIATION OF $n_a, n_b, n_{a\bar{b}}$ AND q WHEN ADDING OR DELETING AN ITEM FROM THE DATASET

x	a	b	Δn_a	$\Delta n_{\bar{b}}$	$\Delta n_{a\bar{b}}$	$\Delta q = q_2 - q_1$
(i)	0	0	0	±1	0	$\frac{n_{a\bar{b}} - \frac{n_a(n_{\bar{b}}+1)}{n+1}}{\sqrt{\frac{n_a(n_{\bar{b}}+1)}{n+1}}} - \frac{n_{a\bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}}$
(ii)	0	±1	0	0	0	$\frac{n_{a\bar{b}} - \frac{n_a n_{\bar{b}}}{n+1}}{\sqrt{\frac{n_a n_{\bar{b}}}{n+1}}} - \frac{n_{a\bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}}$
(iii)	±1	0	±1	±1	±1	$\frac{(n_{a\bar{b}}+1) - \frac{(n_a+1)(n_{\bar{b}}+1)}{n+1}}{\sqrt{\frac{(n_a+1)(n_{\bar{b}}+1)}{n+1}}} - \frac{n_{a\bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}}$
(iv)	±1	±1	±1	0	0	$\frac{n_{a\bar{b}} - \frac{(n_a+1)n_{\bar{b}}}{n+1}}{\sqrt{\frac{(n_a+1)n_{\bar{b}}}{n+1}}} - \frac{n_{a\bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}}$

For determine the variation threshold θ of implication index on the equipotential planes, let $\frac{\partial q}{\partial \xi}$ and $\frac{\Delta q}{\Delta \xi}$ respectively partial derivatives and increment of q according to ξ, where $\xi \in \{n, a, b, a\bar{b}\}$. A variation of q from the addition (or eliminate) of an individual on the dataset can change the number of k implication rules based on the dataset, this leads to an increase in threshold $\theta = k\frac{\Delta q}{\Delta \xi}$, it mean:

$$\frac{\partial q}{\partial \xi} = k\frac{\Delta q}{\Delta \xi} + o(q) \tag{9}$$

where $o(q)$ is an infinitely small. $\frac{\partial q}{\partial \xi}$, $\frac{\Delta q}{\Delta \xi}$ are definded with formulas from (5a) to (5d) and from case (i) to (iv) in Table I. Threshold θ is defined as $k\frac{\Delta q}{\Delta \xi}$ from (9).

2) Determines the variation threshold of the implication intensity

For determine the variation threshold θ of the implication intensity, we consider the variability of λ, Table II shows that $\Delta\lambda < 0$ where b is changed and a is not, other cases $\Delta\lambda > 0$.

TABLE II: THE VARIATION OF $n_a, n_b, n_{a\bar{b}}$ AND λ WHEN ADDING OR DELETING AN ITEM FROM THE DATASET

x	a	b	Δn_a	$\Delta n_{\bar{b}}$	$\Delta n_{a\bar{b}}$	$\Delta\lambda = (\lambda_2 - \lambda_1)$
(i)	0	0	0	±1	0	$\frac{n_a(n_{\bar{b}}+1)}{n+1} - \frac{n_a n_{\bar{b}}}{n}$
(ii)	0	±1	0	0	0	$\frac{n_a n_{\bar{b}}}{n+1} - \frac{n_a n_{\bar{b}}}{n}$
(iii)	±1	0	±1	±1	±1	$\frac{(n_a+1)(n_{\bar{b}}+1)}{n+1} - \frac{n_a n_{\bar{b}}}{n}$
(iv)	±1	±1	±1	0	0	$\frac{(n_a+1)n_{\bar{b}}}{n+1} - \frac{n_a n_{\bar{b}}}{n}$

According to the ISA theory [13], [18], we have

$$\Delta\varphi(a,b) = P\left(card\left(A_2 \cap \overline{B_2}\right) > n_{a\bar{b}}\right) - P\left(card\left(A_1 \cap \overline{B_1}\right) > n_{a\bar{b}}\right) \tag{10}$$

For cases $\Delta\lambda > 0$ include the cases $(i), (iii)$ and (iv), according to Pilar Orus et al [18], we have:

$$\Delta\varphi(a,b) \in \left[e^{-\lambda_1}(\lambda_2 - \lambda_1)(1 - F_1(n_{a\bar{b}} - 1)), e^{-\lambda_1}(\lambda_2 - \lambda_1)(1 - F_2(n_{a\bar{b}} -))\right] \tag{11}$$

where F_i is the Poisson cumulative distribution function of parameter λ_i (for i=1, 2):

$$F_i = 1 - \sum_{i=n_{a\bar{b}}+1}^{\infty} \frac{i\lambda_1^{i-1}}{i!} \tag{12}$$

For case (ii): $\Delta\lambda < 0$, according to Pilar Orus *et al.* [18]

$$\Delta\varphi(a,b) \in \left[e^{-\lambda_2}(\lambda_2 - \lambda_1)(1 - F_1(n_{a\bar{b}} - 1)) - e^{-\lambda_2}\frac{\lambda_2^{n_{a\bar{b}}+1}}{(n_{a\bar{b}}+1)!}, ?\right.$$
$$\left. e^{-\lambda_1}(\lambda_2 - \lambda_1)(1 - F_2(n_{a\bar{b}} -)) - e^{-\lambda_2}\frac{\lambda_2^{n_{a\bar{b}}+1}}{(n_{a\bar{b}}+1)!}\right] \tag{13}$$

where F_i is the Poisson cumulative distribution function of parameter λ_i (for i=1, 2) and is defined by (12):

To determine the variation threshold θ of implication intensity on the equipotential planes, let $\frac{\partial \varphi}{\partial \xi}$ and $\frac{\Delta \varphi}{\Delta \xi}$ respectively partial derivatives and increment of φ according to ξ, where $\xi \in \{n, a, b, a\bar{b}\}$. A variation of φ from the addition (or eliminate) of an individual on the dataset can change the number of k implication rules based on the dataset, this leads to an increase in threshold $\theta = \max_{\Delta\xi}(\frac{\Delta\varphi}{\Delta\xi})$, it mean:

$$\frac{\partial \varphi}{\partial \xi} = \max_{\Delta\xi}(\frac{\Delta\varphi}{\Delta\xi}) + o(q) \tag{14}$$

where $o(q)$ is an infinitely small. $\frac{\partial \varphi}{\partial \xi}$, $\frac{\Delta \varphi}{\Delta \xi}$ are definded with formulas from (5a) to (5d) and from case (i) to case (iv) in table I. Threshold θ is defined as $\max_{\Delta\xi}(\frac{\Delta\varphi}{\Delta\xi})$ from (14).

C. Recommendation Based on Variation Implication Index with Threshold Value of Equipotential Plane

Continuing to develop the model of the recommender system by applying the statistical implication theory we have implemented in some previous work [8], [10], [11], in this paper, we study the variation of the implied index and the implication in Section II as the basis for proposing a recommender system model: Collaborative filtering recommendation with threshold value of the equipotential plane in implication field, This model consists of two main algorithms for generating implication rules (IRG) and defining set of equipotential planes as the basis for predicting and recommending to users the appropriate items (RBEP) as the following :

Algorithm 1. **IRG** (Implication Rules Generator)
Input: set of transactions
Output: implicative rule set and their cardinality $(n, n_a, n_b, n_{a\bar{b}})$.

Step 1: generate rules set from set of transactions by using data mining algorithms (such as apriori, eclat, etc).

Step 2: Calculating cardinalities of implication rules, Details are as follows: Count number of transactions n. Generating two binary (True/False) matrixes $lhsRules$, $rhsRules$, with true value if item j belong to left hand side for $lhsRules$ (respectively, right hand side for $rhsRules$). Then, for each rule$[i]$:

$$lhsProduct = lhsRules \times (data)^{T\,[1]}$$
$$na[i] = rowSum(lhsProduct[i])$$
$$rhsProduct = rhsRules \times (data)^{T}$$
$$nb[i] = rowSum(rhsProduct[i])$$

The calculation $n_{ab}[i]$ is the same as $na[i]$, $nb[i]$ but on both the left and right sides.

$$n_{a\bar{b}}[i] = n_a[i] - n_{ab}[i]$$

Step 3: Return $(n[i], n_a[i], n_b[i], n_{a\bar{b}}[i])$

Algorithm 2. RBEP (Recommendation by Equipotential Plane)

Input: $\boldsymbol{dataset}, threshold\ \theta, ind, byFactor$

Output: recommendation: item/ top k item list

Step 1: call IRG (dataset) for generating rules set and calculating $n, n_a, n_b, n_{a\bar{b}}$.

Step 2: With each $rule(i)$, calculate implication index $q(i)$ according to (2). After that, calculate partial derivatives of $q(i)$ follow byFactor according to (5).

Step 3: Determine the set of *recSet* containing q on the same equipotential plane follows$byFactor$: if $ind = true$, then threshold θ is defined by (9) such that: $(|\Delta q(a, \bar{b})| \leq \theta)$, else threshold θ is defined by (14) such that: $(|\Delta\varphi(a, b)| \leq \theta)$.

Step 4: return recommended: item or *k* items from *recSet* set.

The algorithms described above serves as the basis for a recommendation model for a statistical-based recommendation system ISF (Implication statistical Field) as Fig. 2.

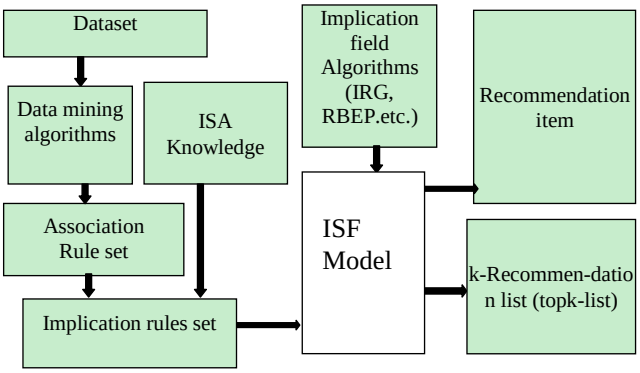

Fig. 2. ISF Recommender system model.

IV. Experiment

A. Dataset

With the collaborative filtering based on implicationfield recommendation model has been built above, we conducted experiments proposed model on both user-based and item-based on the dataset MovieLens was collected by learning GroupLens from page MovieLens site, of which approximately 100,000 ratings are from around 1682 films. It was made by 943 users the ratings range from 1-5 corresponding to the films rated from the lowest to the highest. The data set is preprocessed to serve the experiment to be more accurate, by:

- Standardization of data: Users who rank high (or low) for all their films 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 also to speed up computation, by not interested in the film has had only a few times, because the ratings of these films may be subject to bias due to lack of data, and users rated only a few films because their ratings may be biased.

The dataset has been preprocessed to avoid overfitting problems, as well as to get better accuracy. We conducted experiments in k-fold cross validation mode.

B. Experimental Tools

Experiments were conducted based on the $implicativefield$ toolkit developed by our group using the R language, it includes statistical analysis tools for our proposed algorithms.

C. Scenario 1: The Recommendation Is Based on the Implication Field Model

1) The experimental results of the user-based Implication field recommendation model

The experimental results of our proposed model in the dataset described in the preceding paragraph, the rules set was generated (with conditional support = 0.4 and confident = 0.4. A total of 119 rules, after eliminating meaningless rules (the left side of the rule by nil), and satisfying the implied magnitude greater than 0.5, the remaining 84 rules, With the threshold θ = 0.337565 we obtain the updated values of the implication index q in terms of the variation of any element of $(n, n_a, n_b, n_{a\bar{b}})$, in this scenario byFactor= $n_{a\wedge\bar{b}}$ and collected 27 set of equipotential planes 3 dimensions (n, n_a, n_b), These hyperplanes have the potentials of the implication index of unevenness, listed in Table III. The equipotential plane of the equation is composed of the rules set number 1 {138, 90, 226, 86}, the rules set number 2 {38, 29, 15}, etc., and the rules set number 27 {150, 128, 125, 135, 211}. The sets on each hyperplane have implication index values that are the same with an approximation of θ as Table IV.

TABLE III: The Intensity of Implication Field on Equipotential Planes and Their Implications by *byFactor* =$n_{a\bar{b}}$ (Based-on User)

Eq. plane	quantity of rule	*q*	Eq. plane	quantity of rule	*q*	Eq. plane	quantit of rule	*q*
1	4	-9.0434	10	1	-6.73528	19	1	-3.83082
2	3	-8.80657	11	5	-5.97182	20	1	-3.5129
3	1	-8.69112	12	5	-5.70381	21	2	-3.13998
4	2	-8.1998	13	4	-5.43779	22	5	-2.80447
5	5	-7.75697	14	4	-5.1421	23	5	-2.58721
6	4	-7.45843	15	4	-4.8574	24	2	-2.36298
7	4	-7.26654	16	2	-4.59922	25	2	-2.16484
8	3	-6.97816	17	4	-4.27338	26	3	-1.78044
9	1	-6.75977	18	2	-3.9652	27	5	-1.3093

Trends variability implication factorial *byFactor*, here, elements $n_{a\bar{b}}$ have a major role in strengthening or reject a rule (in theory implicative statistics mentioned in the previous paragraphs), this factor increases have increased the

[1] The transition matrix of data.

implication index value, is synonymous with strength reduction implication intensity, however not significant reduction, so set the rule on equality of treatment implication remains previous level.

TABLE IV: IMPLICATION RULES AND IMPLICATION INDEX EQUIPOTENTIAL PLANE NO.1

ID of rule	Description of the rule	Implication index
138	{Star Wars (1977),Empire Strikes Back, The (1980)} => {Raiders of the Lost Ark (1981)}	-9.149508
90	{Star Wars (1977),Raiders of the Lost Ark (1981),Return of the Jedi (1983)} => {Empire Strikes Back, The (1980)}	-9.053185
226	{Star Wars (1977),Empire Strikes Back, The (1980),Return of the Jedi (1983)} => {Raiders of the Lost Ark (1981)}	-9.000696
86	{Empire Strikes Back, The (1980),Return of the Jedi (1983)} => {Raiders of the Lost Ark (1981)}	-8.970471

TABLE V: ERROR INDEXES OF ISF (USER-BASED), IBCF AND UBCF MODEL

Model	RMSE	MSE	MAE
ISF	0.9434059	0.8900147	0.7419290
IBCFcosine	1.2372211	1.5307160	0.9264473
UBCFcosine	0.9857491	0.9717012	0.7785217
IBCFPearson	1.2204847	1.4895830	0.9094559
UBCFPearson	0.9987563	0.9975141	0.7919161

The density of the implication field an unequal distribution, the high implicative density in the equipotential plane has a slightly more variable indicator value and is more concentrated than the 5, 11, 12, 22, 23 and 27. The density of the implication field the least and the minimum of such aspects as the equipotential plane 3, 9, 10, 19 and 20, as shown in Table III. This shows the suitability of the rule. With the variation of the implication index, where the implication index a certain amount of variability, where the rule is not accepted at a specified threshold, it will move to another equipotential plane whose implication threshold more appropriate. And so, it will help to recommendation users of the item with the most appropriate level of implication.

A target user will be recommended a movie or a list of films that he or she would like to follow corresponding content based on previous movies they viewed, as shown in Table IV. It is possible to recommendation movie "Raiders of the Lost Ark (1981)" for users who have seen movies "Empire Strikes Back, The (1980),Return of the Jedi (1983)", (rule No.86).

2) The experimental results of the item-based implication field model

The experimental results of the item-based Implication field recommendation model in the data set described in the preceding paragraph, the collective rules set was generated (with conditional support = 0.4 and confident = 0.4), a total of 309 rules, after eliminating meaningless rules (the left side of the law by nil), and satisfying the implied magnitude of greater than 0.5, the remaining 230 sets. With the threshold θ = 0.5 we obtain the updated values of the implication index q in terms of the variation of any element of $(n, n_a, n_b, n_{a\bar{b}})$, in this scenario byFactor= $n_{a\wedge\bar{b}}$ and collected 11 set of equipotential planes 3 dimensions (n, n_a, n_b), these hyperplanes have the potentials of the implication index of unevenness, listed in Table V. The equipotential plane of the equation is composed of the rules set number 1 $\{156, 114, 78, 112\}$, the rules set number 2 $\{94, 113, 157, 115\}$, the rules set number 3 $\{138, 90, 226, 86, 182, 144, 88, 305\}$, etc. The sets on each hyperplane have implication index values that are the same with an approximation of θ as Table VI:

TABLE VI: THE INTENSITY OF IMPLICATION FIELD ON EQUIPOTENTIAL PLANES AND THEIR IMPLICATIONS BY *BYFACTOR* =$n_{a\bar{b}}$ (BASED-ON ITEM)

Eq. plane	Nom of rule	Implication index q	Eq. plane	Nom of rule	Implication index q
1	4	-4.57047781	7	46	-1.881271404
2	4	-4.23720113	8	38	-1.445796054
3	8	-3.78085887	9	20	-0.971536794
4	25	-3.35525911	10	16	-0.474975571
5	21	-2.87787274	11	5	-0.09723695
6	43	-2.36529753			

TABLE VII: IMPLICATION RULES AND EQUIPOTENTIAL PLANE NO.3

ID rule	Description of the rule	Implication index
138	{222} => {276}	-4.091659843
90	{130} => {276}	-3.929653633
226	{7} => {450}	-3.885708699
86	{151} => {450}	-3.74297594
182	{682} => {276}	-3.673668845
144	{178} => {416}	-3.662406743
88	{271} => {450}	-3.646012841
305	{308,450} => {429}	-3.614784456

TABLE VIII: ERROR INDEXES OF ISF (ITEM-BASED), IBCF AND UBCF MODEL

Model	RMSE	MSE	MAE
ISF	0.9316129	0.8679026	0.7342853
IBCFcosine	1.349152	1.820211	1.025243
UBCFcosine	0.9523739	0.9070161	0.7540841
IBCFPearson	1.3085552	1.7123168	0.9726921
UBCFPearson	0.9678600	0.9367529	0.7674181

Trends variability implication factorial *byFactor*, here, elements $n_{a\bar{b}}$ have a major role in strengthening or reject a rule (in theory implicative statistics mentioned in the previous paragraphs), this factor increases have increased the implication index value, is synonymous with strength reduction implication intensity, however not significant reduction, so set the rule on equality of treatment implication remains previous level.

The density of the implication field an unequal distribution, the high implicative density in the equipotential plane has a slightly more variable indicator value and is more concentrated than the 4, 5, 6, 7, 8, 9 and 10. The density of the implication field the least and the minimum of such aspects as the equipotential plane 1, 2 and 11, as shown in Table VI. This shows the suitability of the rule. With the variation of the implication index, where the implication index a certain amount of variability, where the rule is not accepted at a specified threshold, it will move to another equipotential plane whose implication threshold more appropriate. And so, it will help to recommendation users of the item with the most appropriate level of implication.

A target user will be recommended a movie or a list of films that he or she would like to follow corresponding content based on previous movies they viewed, as shown in table VII. It is possible to recommendation movie has ID="450" for users who have seen movies has ID= "151" or ID= "7".

D. Scenario 2: Comparison with Other User-Based Collaborative Filtering Model

1) The experimental results of the user-based Implication field recommendation model

To compare the accuracy of the proposed model with user-based collaborative filter models (UBCF) using the Cosine and Pearson measures, the experiment in this scenario is also carried out with the results recorded in the Fig. 3 and Fig. 4, the user-based ISF model has better results for model UBCF use Pearson measure but it is slightly less than model UBCF using Cosine measure. Table V shows that the error indexes of ISF model are low more then other UBCF models.

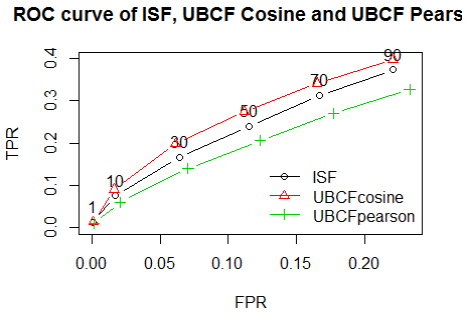

Fig. 3. The ROC curve compares the ISF and other UBCF modes.

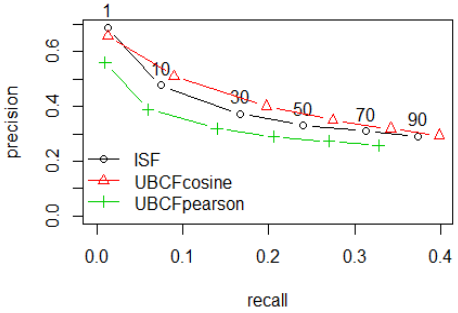

Fig. 4. Precision and recall comparison between ISF and other UBCF modes.

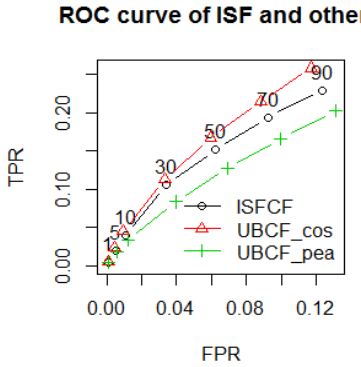

Fig. 5. The ROC curve compares the item-based ISF and other UBCF modes.

2) The experimental results of the item-based Implication field recommendation model

In order to compare the accuracy of the proposed model with User-Based Collaboration (UBCF) models using Cosine and Pearson measures, the experiment in this scenario is also carried out with the results recorded in the Fig. 5 and Fig. 6, the ISF model has better results than the IBCF model using Pearson's but less well than a few the UBCF model using the Cosine measurement. Table V shows that the error indexes of ISF model are low more then other UBCF models.

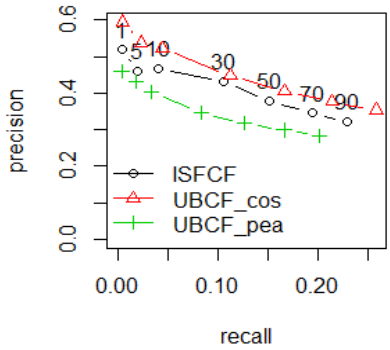

Fig. 6. Precision and Recall comparison between item-based ISF and other UBCF modes.

E. Scenario 3: Comparison with Other Item-Based Collaborative Filtering Model

1) The experimental results of the user-based implication field recommendation model

In this scenario, the user-based ISF is compared to the IBCF using the Pearson and Cosine index over the ROC and the Precision-Recall curves in Fig. 7 and Fig. 8 show that the ISF model is better the Cosine and Pearson IBCF models. Table VIII also shows that the error indexes of ISF model are low more then other IBCF models.

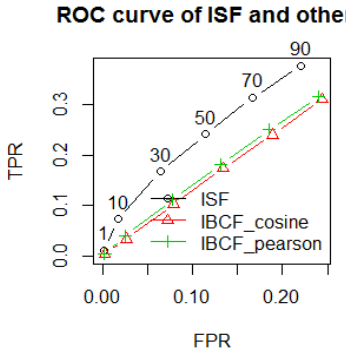

Fig. 7. The ROC curve compares the ISF and other UBCF modes.

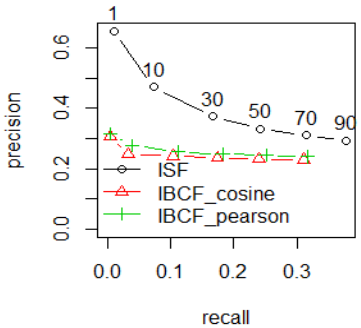

Fig. 8. Precision and Recall comparison between ISF and other UBCF modes.

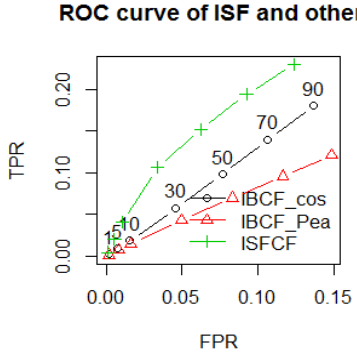

Fig. 9. The ROC curve compares the ISF and other IBCF modes.

2) The experimental results of the item-based Implication field recommendation model

In this scenario, the item-based ISF is compared to the IBCF using the Pearson and Cosine measures over the ROC curve and the Precision-Recall in Fig. 9 and Fig. 10: the item-based ISF model is better IBCF model using Cosine and Pearson. Table VIII also shows that the error indexes of ISF model are low more then other IBCF models.

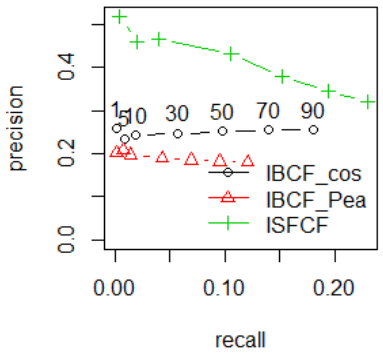

Fig. 10. Precision and Recall comparison between ISF and other IBCF modes.

V. Conclusions

Approach to variation in the implication index and implication intensity was applied to modeling ISF recommender system, we have experimented on both user-based ISF model and item-based ISF Model (in this paper only presents the experimental results of the model based on the variance of the implication index in the implication field, for the model's experiments based on the variation of the implication intensity in the implication field will also be quite similar results). The proposed model was also tested on the Movilens 100K and the implicativefield toolkit for user recommendation and evaluation with traditional models using symmetry similarity measures, the results were mostly good more than those using common symmetry. These contributions are intended to increase the effectiveness of recommendations (to improve the accuracy of ranking predictions, and to show trends in rules).

In the future, we will improve the proposed model to achieve better results than traditional user-based collaborative filter models using cosine and we will also develop a proposed model based on variability in other factors, as well as on the variability of the composite from a variety of factors and the experience will also be conducted on a other data sets such as 1M Movielens, Groceries, MSWeb, and so on. This will evaluate Approach to variation in the implication index was applied to modeling ISF recommender system, we have experimented on both user-based ISF model and item-based ISF Model. The proposed model was also tested on the Movilens 100K and the implicativefield toolkit for user recommendation and evaluation with workflow models using symmetry similarity measures, the results were mostly good more than those using common symmetry. These contributions are intended to increase the effectiveness of recommendations (to improve the accuracy of ranking predictions, and to show trends in rules).

References

[1] A. Gediminas and T. 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] A. Gediminas and T. Alexander, *Context-Aware Recommender Systems*, Springer US, pp. 217-253, 2011.

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

[4] B. Cao, Q. Yang, J.-T. Sun, and Z. Chen, "Learning bidirectional asymmetric similarity for collaborative filtering via matrix factorization," *Data Mining and Knowledge Discovery*, vol. 22, issue 3, pp. 393–418, 2011.

[5] R. Katarya and O. P. Verma, "Effective collaborative movie recommender system using asymmetric user similarity and matrix factorization," in *Proc. The 2016 IEEE International Conference on Computing, Communication and Automation*, 2016, pp. 1-12.

[6] M. Deshpande and G. Karypis, "Item-based top-N recommendation algorithms," *ACM Transaction on Information Systems*, vol. 22, no. 1, pp. 143–177, 2004.

[7] Z.-L. Zhao C.-D. Wang, and J.-H. Lai, "AUI&GIV Recommendation with asymmetric user influence and global importance value," *Public Library of Science ONE*, 2016.

[8] N. Q. Phan, K. M. Nguyen, H. T. Nguyen, and H. X. Huynh, "Recommender system based approach combining associationrule and implicative statistical measure," in *Proc. the VIII National Conference on Fundamental and Applied IT Research*, Ha Noi, 2015.

[9] L. P. Phan, T. U. Tran, H. H. Huynh, and H. X. Huynh, "The user-based collaborative filtering recommeder system using associaion rules combined implication statistical cohension measure," in *Proc. the IX National Conference on Fundamental and Applied IT Research*, Cần Thơ, 2016.

[10] H. T. Nguyen, H. H. Huynh, and H. X. Huynh, "Recommender system based on analysis Implicative statistical user preferences over time," in *Proc. IX International Conference A.S.I. Analyse Statistique Implicative – Statistical Implicative Analysis (ASI9)*, French, pp. 493-507, 2017.

[11] H. T. Nguyen, H. H. Huynh, and H. X. Huynh, "Recommendation based on the variance of implication index in statistical implication field," in *Proc. the X National Conference on Fundamental and Applied IT Research (FAIR'17)*, Da Nang, 2017, pp. 938-950.

[12] R. Gras, P. Kuntz, and N. Greffard, "Notion de champ implicatif en analysis statistique implicative," in *Proc. the 8th International Meeting on Statistical Implicative Analysis*, Tunisia, 2015, pp. 1-21.

[13] R. Gras, E. S. F. Guillet, and F. Spagnolo, *Statistical Implicative Analysis, Theory and Application*, Springer Verlag Berlin Heidelberg, 2008.

[14] R. Gras and R. Couturier, "Spécificités de l'Analyse Statistique Implicative (A.S.I.) par rapport à d'autres mesures de qualité de règles d'association," *Quaderni di Ricerca in Didattica - GRIM*, pp. 19-57, 2010.

[15] L.-R. Dominique, *Didactics of Mathematics and Implicative Statistical Analysis, Statistical Implicative Analysis - Studies in Computational Intelligence*, pp. 277-298, 2008.

[16] R. Gras and L.-R. Dominique, "Duality between variables space and subjects space of the statistic implicative analysis, Dualite entre espace des variables et espace des sujets en analyse statisticque implicative," in *Proc. the VI International Conference on Analyse statistique implicative- Implicative statistical Analysis Caen*, France, 2012, pp. 1-28.

[17] R. Gras and P. Kuntz, "Discovering R-rules with a directed hierarchy," *Journal Soft Computing - A Fusion of Foundations, Methodologies and Applications*, vol. 10, issue 5, Springer-Verlag, pp. 453-460, 2006.

[18] P. Orús and P. Gregori, "Fictitious pupils and implicative analysis: A case study," *Statistical Implicative Analysis: Theory and Applications*, Springer-Verlag Berlin Heidelberg, pp. 325-349, 2008.

Hoang Nguyen-Tan received the computer engineer degree from Ho Chi Minh City University of Technology, in 1996, Vietnam and holds a master's degree in information system from Can Tho University, in 2010, Viet Nam, and continue studing doctoral degree in computer science from Danang University of Science and Technology, Da Nang University, Viet Nam. He is currently working at the Department of Information and Communications of Dong Thap province, Viet Nam. His general research interest is in data

mining, recommender system, statistical implicative analysis.

Hung Huynh-Huu received the B.S. degree in computer science from the IPH, Vietnam, in 1998, the M.Sc.A. degree in computer science in 2003, and the Ph.D. degree in computer science from the Aix-Marseille University in 2010. He now is a lecturer at Danang University of Science and Technology (DUT), Da Nang University, Vietnam. His current research interests include computer vision and and health care systems.

Hiep Huynh-Xuan holds a master's degree in information technology from Institut de la Francophonie pour l'Informatique, Viet Nam, and a doctoral degree in data mining from Nantes University, France. He is currently an associate professor, vice dean of College of information technology and communication at Can Tho University, Viet Nam. His general research interests are in data mining, artificial intelligence, statistical implicative analysis, wireless sensor network.