

“© 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

A Fuzzy Tree Similarity Based Recommendation Approach for Telecom Products

Dianshuang Wu, Guangquan Zhang, Jie Lu

*Decision Systems & e-Service Intelligence Lab, Centre for Quantum Computation & Intelligent Systems
School of Software, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia*

Dianshuang.Wu@student.uts.edu.au, Guangquan.Zhang@uts.edu.au, Jie.Lu@uts.edu.au

Abstract - Due to the huge product assortments and complex descriptions of telecom products, it is a great challenge for customers to select appropriate products. A fuzzy tree similarity based hybrid recommendation approach is proposed to solve this issue. In this study, fuzzy techniques are used to deal with the various uncertainties existing within the product and customer data. A fuzzy tree similarity measure is developed to evaluate the semantic similarity between tree structured products or user profiles. The similarity measures for items and users both integrate the collaborative filtering (CF) and semantic similarities. The final recommendation hybridizes item-based and user-based CF recommendation techniques. A telecom product recommendation case study is given to show the effectiveness of the proposed approach.

I. INTRODUCTION

Telecom businesses today offer hundreds of different products and services to customers and are constantly exploring new service models. These products have very complex structures and features. There are various service types and service items. With such a vast number of products and so complex description, it is becoming increasingly difficult for customers to find their favorite products quickly and accurately. Recommender systems are designed to resolve this problem by automatically making personalized recommendations to customers. This study aims to develop a recommendation approach to support customers in the selection of the most appropriate telecom products.

There are four main difficulties in telecom product recommendation compared to other industries. First, telecom products have complex descriptions and features and present tree structures [1]. A complete product usually consists of several services and each service consists of several features, which constructs a tree structure. Second, telecom products are updated frequently, but a customer has one product at a time. This results in a lack of rating information on products from customers. Third, the conceptual similarity between the features of different products are usually fuzzy and described by domain experts with linguistic terms, such as ‘very similar’, ‘absolutely different’. Fourth, customers often express their preferences and interests to products using linguistic terms, such as ‘good’, ‘very good’, and ‘interested’.

A fuzzy tree similarity based hybrid recommendation approach is proposed to deal with the above difficulties. A fuzzy tree similarity measure is proposed to evaluate the semantic similarity between the tree structured products or user profiles. To handle the lack of ratings, the semantic and

CF similarities are integrated, and the item-based CF and user-based CF techniques are combined to solve the rating sparsity problem. Fuzzy numbers are used in the approach to deal with fuzzy problems.

The paper makes contributions to both theoretical and practical aspects. At the theoretical level, a fuzzy tree similarity measure is developed. At the practical level, a hybrid recommendation approach is proposed.

The remainder of this paper is organized as follows. In Section II, the related works are expatiated. Section III describes the preliminaries. The fuzzy tree similarity measure is described in Section IV, and the fuzzy tree similarity based hybrid recommendation approach is shown in Section V. In Section VI, a telecom product recommendation case study is presented. Finally, conclusions and future study are given in Section VII.

II. RELATED WORKS

In this section, the related works on recommender systems, tree similarity measure and fuzzy techniques used in recommender systems are reviewed.

A. Recommender Systems

Recommender systems, as an important e-service intelligence method [2], can be defined as programs which attempt to recommend items to users by predicting a user’s interest in a given item based on various types of information, including particulars about items, users and the interactions between users and items [3]. The three main recommendation techniques are CF, content-based (CB) and knowledge-based (KB) techniques [4]. CF technique helps people make choices based on the opinions of other people who share similar interests [5]. It can be further divided into user-based and item-based CF approaches. CB techniques recommend items that are similar to those previously preferred by a specific user [6]. KB techniques offer items to users based on knowledge about the users and items. Each technique has its limitations, such as the item content dependency problem, overspecialization problem for CB [3, 6], the cold start problem and the sparsity problem for CF [3]. The hybrid recommendation approach is a combination of two or more of the aforementioned approaches to emphasize the strengths of these approaches and to achieve the peak performance of a recommender system [3, 4]. It has been proven that the CF recommendation approach, or its combination with another technique, is the most successful and widely used approach for recommender systems [5]. The literature particularly shows

that the combination of user-based CF and item-based CF may achieve good performance in a big-user-set and big-item-set environment [7].

B. Tree Similarity Measure

The research on tree similarity measure has attracted great attention due to the ubiquitousness of tree-structured data in many application fields [8-10]. In previous research, trees have been compared from both structural and semantic aspects. The tree edit distance model [11] is the most widely used method for comparing the structures of ordered or unordered labeled trees. It measures the degree of similarity between two trees by the minimum cost of the edit operation sequences that convert one tree into another. The edit operations give rise to an edit distance mapping which is a graphical specification of which edit operations apply to each node in the two labeled trees [11]. Considering structural constraints, constrained edit distance [12] requires that disjoint sub-trees be mapped to disjoint sub-trees. Because tree structures reflect the semantic meanings of the objects, these structural constraints are necessary in many applications. The semantic or conceptual similarity between attributes is also taken into account when comparing two trees [9]. Only conceptually similar attributes can be mapped or transformed. In a business environment, the data are more complex. Besides tree structures and attribute concepts, node values and weights are also considered [1]. However, the previous tree similarity measure models cannot deal with the uncertain weights or attribute similarity described with linguistic terms. In this study, a comprehensive fuzzy tree similarity measure will be developed.

C. Fuzzy Techniques in Recommender Systems

In practical situations, customers like to express their preferences for items in linguistic terms, such as ‘very interested’, or ‘not interested’ for the features of a product. Therefore, recommendations are often generated on the basis of uncertain or vague information [13]. The similarities between items or between users are naturally fuzzy, which attracts many researchers to apply fuzzy set theory, fuzzy logic and fuzzy relations to recommender systems in an attempt to achieve more accurate and effective recommendations. For example, a fuzzy relational approach was applied to event recommendation [14] and trade exhibition recommendation [15]. Porcel et al. [16] developed a fuzzy linguistic-based recommender system based on both CB filtering and fuzzy linguistic modelling techniques.

III. PRELIMINARIES

A. Fuzzy Number

In this section, we present some basic definitions related to the fuzzy number [17] and linguistic variable theory [18, 19].

Definition 1. A fuzzy number \tilde{a} is a fuzzy subset on the space of real number R that is both convex and normal.

Definition 2. If \tilde{a} is a fuzzy number and $0 < a_\lambda^L \leq a_\lambda^R \leq 1$, for any $\lambda \in (0,1]$, then \tilde{a} is called a normalized positive fuzzy number, where a_λ^L and a_λ^R are the lower and upper bounds of the λ -cut set of \tilde{a} .

Definition 3. A triangular fuzzy number \tilde{a} can be defined by a triplet (a_0^L, a, a_0^R) and the membership function $\mu_{\tilde{a}}(x)$ is defined as:

$$\mu_{\tilde{a}}(x) = \begin{cases} (x - a_0^L)/(a - a_0^L), & a_0^L \leq x \leq a, \\ (a_0^R - x)/(a_0^R - a), & a \leq x \leq a_0^R, \\ 0, & \text{others.} \end{cases} \quad (1)$$

Let $F^*(R)$ be the set of all finite fuzzy numbers on R .

Definition 4. Let $\tilde{a}, \tilde{b} \in F^*(R)$, then the quasi-distance function of \tilde{a} and \tilde{b} is defined as:

$$d(\tilde{a}, \tilde{b}) = \left(\int_0^1 \frac{1}{2} [(a_\lambda^L - b_\lambda^L)^2 + (a_\lambda^R - b_\lambda^R)^2] d\lambda \right)^{\frac{1}{2}}. \quad (2)$$

Definition 5. A linguistic variable is a variable whose values are linguistic terms. [19]

In this study, a set of five linguistic terms $\{\textit{Strongly Interested (SI)}, \textit{More Interested (MI)}, \textit{Interested (I)}, \textit{Less Interested (LI)}, \textit{Not Interested (NI)}\}$ are used to describe the user ratings. Fuzzy numbers are applied to deal with these linguistic terms. The related fuzzy numbers to these linguistic terms are shown in Table I.

TABLE I
LINGUISTIC TERMS AND RELATED FUZZY NUMBERS FOR RATINGS

Linguistic terms	Triangular Fuzzy Numbers
<i>Strongly Interested (SI)</i>	(4,5,5)
<i>More Interested (MI)</i>	(3,4,5)
<i>Interested (IN)</i>	(2,3,4)
<i>Less Interested (LI)</i>	(1,2,3)
<i>Not Interested (NI)</i>	(1,1,2)

B. Features of Tree-Structured Data

Definition 6. [20] A tree is defined as a directed graph $T=(V, E)$ where the underlying undirected graph has no cycles and there is a distinguished root node in V , denoted by $root(T)$, so that for any node $v \in V$, there is a path in T from $root(T)$ to node v .

In real applications, the definition is usually extended to represent practical objects. In this research, a tree-structured data model for business data is proposed by adding the following features to the definition.

1) A domain attribute term set A , which is a set of symbols to specify semantic meanings to nodes, is introduced. An attribute assignment function $a: V \rightarrow A$ is defined so that each node in the tree is assigned an attribute.

2) An attribute conceptual similarity measure within the domain attribute term set A is defined by domain experts with linguistic terms. It is represented as a set of mappings $sc: A \times A \rightarrow S$, in which $S = \{\textit{Absolutely different (AD)}, \textit{Very different (VD)}, \textit{Different (D)}, \textit{Medium (M)}, \textit{Similar (S)}, \textit{Very similar (VS)}, \textit{Absolutely similar (AS)}\}$. Each mapping denotes the conceptual similarity between the two attributes. The related fuzzy numbers to the linguistic terms in S are shown in Table II.

TABLE II
LINGUISTIC TERMS AND RELATED FUZZY NUMBERS FOR ATTRIBUTE SIMILARITY AND WEIGHTS

Linguistic terms	Fuzzy numbers	Linguistic terms	Fuzzy numbers
------------------	---------------	------------------	---------------

<i>AD</i>	(0,0,0.1)	<i>VL</i>	(0,0,0.1)
<i>VD</i>	(0,0.1,0.3)	<i>L</i>	(0,0.1,0.3)
<i>D</i>	(0.1,0.3,0.5)	<i>ML</i>	(0.1,0.3,0.5)
<i>M</i>	(0.3,0.5,0.7)	<i>M</i>	(0.3,0.5,0.7)
<i>S</i>	(0.5,0.7,0.9)	<i>MH</i>	(0.5,0.7,0.9)
<i>VS</i>	(0.7,0.9,1.0)	<i>H</i>	(0.7,0.9,1.0)
<i>AS</i>	(0.9,1.0,1.0)	<i>VH</i>	(0.9,1.0,1.0)

3) Each attribute $b \in A$ is associated with a value domain D_b and a value similarity measure $s_b : D_b \times D_b \rightarrow [0,1]$.

4) A weight function $w:V \rightarrow W$ assigns a weight to each node to represent its importance degree to its siblings, in which $W = \{ \text{Very low (VL), Low (L), Medium low (ML), Medium (M), Medium high (MH), High (H), Very high (VH)} \}$. The related fuzzy numbers to the linguistic terms in W are shown in Table II.

IV. FUZZY TREE SIMILARITY MEASURE

A fuzzy tree similarity measure is presented in this section. In the measure, the most conceptual corresponding node pairs among two trees are identified. Then, the conceptual similarity and the value similarity between two trees are evaluated, and the final similarity measure is assessed as a weighted sum of their conceptual and value similarities. The following symbols are used to represent trees and nodes. Suppose that we have a numbering for each tree. Let $t[i]$ be the i th node of the tree T in the given numbering, $T[i]$ be the sub-tree rooted at $t[i]$, $F[i]$ be the unordered forest obtained by deleting $t[i]$ from $T[i]$, and $t[i_1], t[i_2], \dots, t[i_{n_i}]$ be the children of $t[i]$.

A. Weights Normalization

As mentioned in Section III, nodes are assigned with linguistic weights which are expressed by fuzzy numbers. These weights should be normalized first by:

$$w_{t[i_k]}^* = w_{t[i_k]} / \sum_{j=1}^{n_i} w_{t[i_j]}^R, \quad (3)$$

where $w_{t[i_k]}^*$ is the normalized weight of node $t[i_k]$.

B. Conceptual Similarity

Because the attribute conceptual similarity and the node weights are described by linguistic terms, and represented as fuzzy numbers, the conceptual similarity between two trees is fuzzy. Given two trees $T_1[i]$ and $T_2[j]$ to be compared, their conceptual similarity is calculated as:

$$sc_T(T_1[i], T_2[j]) = \begin{cases} sc(a(t_1[i]), a(t_2[j])), & F_1[i] = \phi, F_2[j] = \phi \\ \alpha \cdot sc(a(t_1[i]), a(t_2[j])) \\ + (1 - \alpha) \cdot \sum_{r=1}^{n_j} w_{j_r} \cdot sc_T(T_1[i], T_2[j_r]), & F_1[i] = \phi, F_2[j] \neq \phi \\ \alpha \cdot sc(a(t_1[i]), a(t_2[j])) \\ + (1 - \alpha) \cdot \sum_{r=1}^{n_i} w_{i_r} \cdot sc_T(T_1[i_r], T_2[j]), & F_1[i] \neq \phi, F_2[j] = \phi \\ \alpha \cdot sc(a(t_1[i]), a(t_2[j])) \\ + (1 - \alpha) \cdot sc_F(F_1[i], F_2[j]), & F_1[i] \neq \phi, F_2[j] \neq \phi \end{cases} \quad (4)$$

where $a(t_1[i])$ and $a(t_2[j])$ represent the attributes of $t_1[i]$ and $t_2[j]$ respectively, w_{j_r} and w_{i_r} are the normalized weights of $t_2[j_r]$ and $t_1[i_r]$ respectively, and α is the influence factor of

the parent node. According to the condition of whether $t_1[i]$ and $t_2[j]$ are leaves, four situations are listed in Formula (4). In the first situation, $t_1[i]$ and $t_2[j]$ are both leaves, and their conceptual similarity is equivalent to the conceptual similarity of their attributes. In the second and third situations, one node is a leaf and the other is an inner node. As the concept of a tree is dependent not only on its root's attribute, but also on its children's attributes, the children of the inner node are also considered in the formula. In the last situation, both $t_1[i]$ and $t_2[j]$ have children. Their children construct two forests $F_1[i]$ and $F_2[j]$, which are compared with the forest similarity measure $sc_F(F_1[i], F_2[j])$.

To calculate $sc_F(F_1[i], F_2[j])$, the conceptual corresponding sub-trees are first identified based on both their concepts and structures, and are then compared separately. Finally, these local similarities are weight aggregated. To identify the conceptual corresponding node pairs, a bipartite graph $G_{ij} = (V_{1,i} \cup V_{2,j}, E)$ is constructed, in which $V_{1,i} = \{t_1[i_1], t_1[i_2], \dots, t_1[i_{n_i}]\}$, $V_{2,j} = \{t_2[j_1], t_2[j_2], \dots, t_2[j_{n_j}]\}$. For any $t_1[i_p] \in V_{1,i}$ and $t_2[j_q] \in V_{2,j}$, a weight is assigned to edge $(t_1[i_p], t_2[j_q])$ based on $sc_T(T_1[i_p], T_2[j_q])$. Let a fuzzy positive-ideal value $r^+ = 1$, and a fuzzy negative-ideal value $r^- = 0$. For edge $(t_1[i_p], t_2[j_q])$, its weight is defined as $w_{p,q} = \frac{1}{2}(d^- + (1 - d^*))$, where $d^+ = d(sc_T(T_1[i_p], T_2[j_q]), r^+)$, $d^- = d(sc_T(T_1[i_p], T_2[j_q]), r^-)$, and $d(\cdot, \cdot)$ is the distance between two fuzzy numbers. To find most corresponding node pairs between $V_{1,i}$ and $V_{2,j}$, a maximum weighted bipartite matching (MWBM) problem [21] of G_{ij} is resolved. A MWBM of $V_{1,i}$ and $V_{2,j}$, M_{ij} is constructed. The conceptual similarity between $F_1[i]$ and $F_2[j]$ is calculated as:

$$sc_F(F_1[i], F_2[j]) = \sum_{(t_1[i_p], t_2[j_q]) \in M_{ij}} w \cdot sc_T(T_1[i_p], T_2[j_q]), \quad (5)$$

where $w = \frac{1}{2}(w_{t_1[i_p]}^* + w_{t_2[j_q]}^*)$.

During the computation process of the conceptual similarity between two trees, the maximum weighted bipartite matching results are recorded. Based on the records, the most corresponding nodes among two trees can be identified. The roots of two trees are corresponding node pairs. Then the corresponding nodes in the children of two roots are identified based on two roots' children's maximum weighted bipartite matching. Other corresponding nodes are identified in the same way.

C. Value Similarity

The values of tree nodes are considered in this sub-section. Given two trees $T_1[i]$ and $T_2[j]$ to be compared, a maximum conceptual similarity tree mapping M_s has been constructed. According to whether $t_1[i]$ and $t_2[j]$ in different situations are assigned values or not, the value similarity between $T_1[i]$ and $T_2[j]$, $sv_T(T_1[i], T_2[j])$, is calculated in the following three cases.

$$sv_T(T_1[i], T_2[j]) = \begin{cases} s_b(v(t_1[i]), v(t_2[j])), & v(t_1[i]) \neq null, v(t_2[j]) \neq null \\ s_b(v(t_1[i]), v(T_2[j])), & v(t_1[i]) \neq null, v(t_2[j]) = null \\ \sum_{(t_{1p}, t_{2j_q}) \in M_s} w \cdot sv_T(T_1[i_p], T_2[j_q]), & v(t_1[i]) = null, v(t_2[j]) = null \end{cases} \quad (6)$$

where $w = \frac{1}{2}(w_{t_{1p}}^* + w_{t_{2j_q}}^*)$. In the first case, both roots are assigned values; their values are compared directly. In the second case, one root is not assigned value. The values of the sub tree are aggregated. In the third case, neither root is assigned a value. The values of matching sub trees are compared separately, and then the similarities are aggregated.

D. Final Similarity

Based on the conceptual similarity and value similarity of two trees, the final fuzzy similarity measure between T_1 and T_2 is defined as follows:

$$s(T_1, T_2) = \alpha_1 \cdot sc_T(T_1, T_2) + \alpha_2 \cdot sv_T(T_1, T_2), \quad (7)$$

where $\alpha_1 + \alpha_2 = 1$. The similarity measure is a normalized fuzzy number. In applications it needs to be defuzzified. Let a fuzzy positive-ideal similarity value $s^+ = 1$ and a fuzzy negative-ideal similarity value $s^- = 0$. The defuzzified similarity measure is defined as:

$$s_T(T_1, T_2) = \frac{1}{2}(d^- + (1 - d^*)), \quad (8)$$

where $d^* = d(s(T_1, T_2), s^+)$ and $d^- = d(s(T_1, T_2), s^-)$.

V. A FUZZY TREE SIMILARITY BASED HYBRID RECOMMENDATION APPROACH

The proposed approach takes the user-item linguistic rating matrix, the product trees, the user usage profile trees and requirement trees as input. The approach combines item-based CF and user-based CF predictions. It integrates CF and semantic similarities when calculating both the item and user similarities. The recommendation process is described in eleven steps as follows.

Step 1: Calculate the fuzzy CF item similarity

The Pearson correlation is selected for measuring the similarities between the two items x and y . Since the ratings are represented as fuzzy numbers, the following fuzzy similarity measure is given:

$$s_{CFi}(x, y) = \frac{\sum_{s \in S_{x,y}} \int_{\frac{1}{2}}^1 (r_{x,s,\lambda}^L - \bar{r}_{x,\lambda}^L + r_{x,s,\lambda}^R - \bar{r}_{x,\lambda}^R) \times \frac{1}{2} (r_{y,s,\lambda}^L - \bar{r}_{y,\lambda}^L + r_{y,s,\lambda}^R - \bar{r}_{y,\lambda}^R) d\lambda}{\sqrt{\sum_{s \in S_{x,y}} \int_{\frac{1}{2}}^1 [(r_{x,s,\lambda}^L - \bar{r}_{x,\lambda}^L + r_{x,s,\lambda}^R - \bar{r}_{x,\lambda}^R)]^2 d\lambda} \times \sqrt{\sum_{s \in S_{x,y}} \int_{\frac{1}{2}}^1 [(r_{y,s,\lambda}^L - \bar{r}_{y,\lambda}^L + r_{y,s,\lambda}^R - \bar{r}_{y,\lambda}^R)]^2 d\lambda}} \quad (9)$$

where $S_{x,y}$ represents the set of users that both rated items x and y , \bar{r}_x and \bar{r}_y are the average ratings of users in $S_{x,y}$ on x and y respectively.

Step 2: Calculate the item semantic similarity

Let items x and y be represented by two trees T_x and T_y respectively. The semantic similarity between x and y is calculated by

$$s_{semi}(x, y) = s_T(T_x, T_y). \quad (10)$$

Step 3: Integrating semantic similarity with CF item similarity
The total similarity between items x and y is computed by integrating the two similarity measures computed in the last two steps.

$$s_I(x, y) = \beta \times s_{semi}(x, y) + (1 - \beta) \times s_{CFi}(x, y), \quad (11)$$

where $\beta \in [0, 1]$ is a semantic combination parameter specifying the weight of similarity in the integrated measure, $s_{CFi}(x, y) = (1 + s_{CFi}(x, y)) / 2$.

Step 4: Item neighbors selection

The top-N most similar items are selected as neighbors to predict ratings.

Step 5: Calculate the fuzzy item-based CF prediction

In this step, all the unrated ratings are calculated using the item-based CF method and all the empty cells in the user-item rating table will be filled. The prediction is calculated as:

$$Fp_{x,s} = \frac{\sum_{y=1}^c \tilde{r}_{y,s} * s_I(x, y)}{\sum_{y=1}^c s_I(x, y)} \quad (12)$$

$$= \bigcup_{\lambda \in [0, 1]} \lambda \left[\frac{\sum_{y=1}^c r_{y,s,\lambda}^L * s_I(x, y)}{\sum_{y=1}^c s_I(x, y)}, \frac{\sum_{y=1}^c r_{y,s,\lambda}^R * s_I(x, y)}{\sum_{y=1}^c s_I(x, y)} \right]$$

where $Fp_{x,s}$ refers to the predicted rating of user s on item x , c is the number of selected neighbours, $\tilde{r}_{y,s}$ is the rating of user s on item y , and $s_I(x, y)$ is the similarity between item x and item y .

Step 6: Calculate the fuzzy CF user similarity

The fuzzy CF user similarity between users s and t is calculated as:

$$s_{CFu}(s, t) = \frac{\sum_{x \in I_{s,t}} \int_{\frac{1}{2}}^1 (r_{x,s,\lambda}^L - \bar{r}_{s,\lambda}^L + r_{x,s,\lambda}^R - \bar{r}_{s,\lambda}^R) \times \frac{1}{2} (r_{x,t,\lambda}^L - \bar{r}_{t,\lambda}^L + r_{x,t,\lambda}^R - \bar{r}_{t,\lambda}^R) d\lambda}{\sqrt{\sum_{x \in I_{s,t}} \int_{\frac{1}{2}}^1 [(r_{x,s,\lambda}^L - \bar{r}_{s,\lambda}^L + r_{x,s,\lambda}^R - \bar{r}_{s,\lambda}^R)]^2 d\lambda} \times \sqrt{\sum_{x \in I_{s,t}} \int_{\frac{1}{2}}^1 [(r_{x,t,\lambda}^L - \bar{r}_{t,\lambda}^L + r_{x,t,\lambda}^R - \bar{r}_{t,\lambda}^R)]^2 d\lambda}} \quad (13)$$

where $I_{s,t}$ represents the set of items rated by both user s and t , \bar{r}_s and \bar{r}_t are the average of all ratings from users s and t respectively.

Step 7: Calculate the user semantic similarity

The user usage profiles or buying requests are described by trees. Let users s and t are represented as T_s and T_t respectively. The semantic similarity between s and t is calculated as:

$$s_{semu}(s, t) = s_T(T_s, T_t). \quad (14)$$

Step 8: Integrating semantic similarity with CF user similarity
The total similarity between users s and t is computed as:

$$s_U(s, t) = \beta \times s_{semu}(s, t) + (1 - \beta) \times s_{CFu}(s, t), \quad (15)$$

where $\beta \in [0, 1]$, $s_{CFu}(s, t) = (1 + s_{CFu}(s, t)) / 2$.

Step 9: User neighbors selection

The top-N most similar users are selected as neighbors to predict ratings.

Step 10: Calculate the fuzzy user-based CF prediction

This step is to predict the ratings of every unrated telecom products for target users using user-based CF. The new predicted ratings will replace the ratings predicted in Step 5.

$$r_{x,s} = \frac{\sum_{t=1}^c \tilde{r}_{x,t} * s_U(s,t)}{\sum_{t=1}^c s_U(s,t)}, \quad (16)$$

$$= \bigcup_{\lambda \in [0,1]} \lambda \left[\frac{\sum_{t=1}^c r_{x,s,\lambda}^L * s_U(s,t)}{\sum_{t=1}^c s_U(s,t)}, \frac{\sum_{t=1}^c r_{x,s,\lambda}^R * s_U(s,t)}{\sum_{t=1}^c s_U(s,t)} \right]$$

where $r_{x,s}$ is the predicted rating of item x from user s , c is the number of neighbors selected in Step 7, $\tilde{r}_{x,t}$ is the rating of item x from user t , and $s_U(s,t)$ is the similarity between user s and user t .

Step 11: Generate recommendations

Let the fuzzy positive-ideal rating and the fuzzy negative-ideal rating be p^* and p^- respectively. The ranking coefficient of product x is calculated as:

$$c_{x,s} = \frac{1}{2}(c^- + (c - c^*)), \quad (17)$$

where $c^* = d(p_{x,s}, p^*)$, $c^- = d(p_{x,s}, p^-)$ and $c = d(p^-, p^*)$.

The unrated products of s are ranked by the coefficient, and the top-K products are recommended.

VI. A TELECOM PRODUCT RECOMMENDATION CASE STUDY

A telecom product recommendation case study is given to illustrate the proposed approach.

A telecom company has five product packages which are described by trees in Fig. 1 and there are five business users and their usage profiles are described in Fig. 2. The rating matrix, by linguistic terms, is depicted in Table III.

The conceptual similarities between the attributes are defined as follows:

$sc(\$59\text{Complete}, \$79\text{Complete})=VS$, $sc(\$59\text{Complete}, \$49\text{Complete})=VS$,
 $sc(\$49\text{Complete}, \$79\text{Complete})=S$, $sc(\$29\text{Smart}, \$39\text{Smart})=VS$,
 $sc(\$49\text{Smart}, \$39\text{Smart})=VS$, $sc(\$29\text{Smart}, \$49\text{Smart})=S$,
 $sc(\$29\text{Smart}, \$79\text{Complete})=VD$, $sc(\$59\text{Complete}, \$29\text{Smart})=VD$,
 $sc(\$49\text{Complete}, \$29\text{Smart})=D$, $sc(\$39\text{Smart}, \$79\text{Complete})=VD$,
 $sc(\$59\text{Complete}, \$39\text{Smart})=VD$, $sc(\$49\text{Complete}, \$39\text{Smart})=D$,
 $sc(\$49\text{Smart}, \$79\text{Complete})=VD$, $sc(\$59\text{Complete}, \$49\text{Smart})=VD$,
 $sc(\$49\text{Complete}, \$49\text{Smart})=M$, $sc(\$45\text{LAD}, \$55\text{LAD})=VS$,
 $sc(\$75\text{LAD}, \$55\text{LAD})=VS$, $sc(\$30\text{LAD}, \$45\text{LAD})=VS$,
 $sc(\$30\text{LAD}, \$55\text{LAD})=S$, $sc(\$30\text{LAD}, \$75\text{LAD})=M$, $sc(\$70\text{LAD}, \$45\text{LAD})=S$.

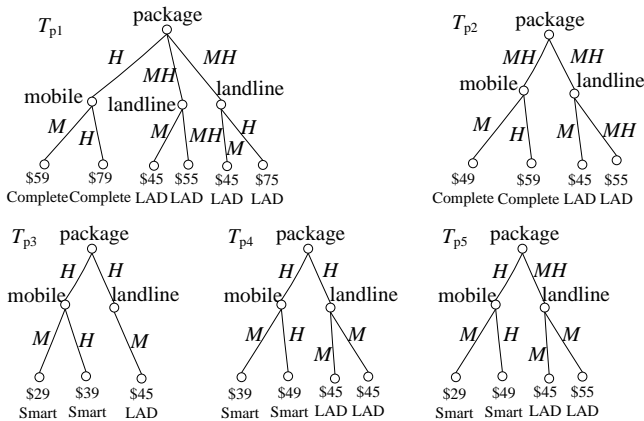


Fig. 1 Telecom product packages

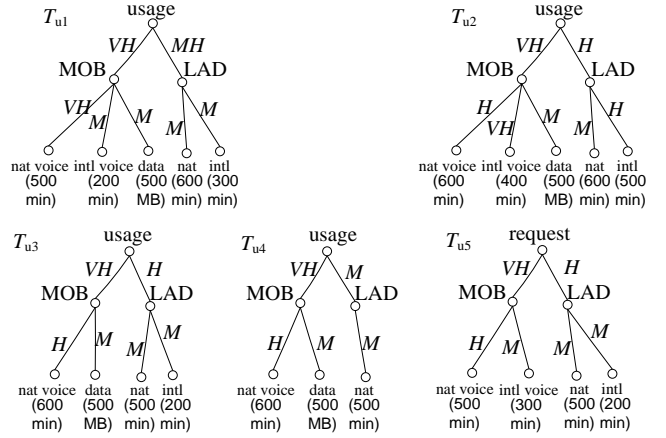


Fig. 2 User usage profiles or requests

TABLE III
USER PRODUCT RATING MATRIX

	P1	P2	P3	P4	P5
U1	MI	SI		MI	
U2	SI	MI		IN	
U3		LI	IN	IN	SI
U4	NI		SI		MI
U5		IN		SI	MI

Using the proposed recommendation approach, the CF and semantic similarities between items are calculated, shown in Table IV. Because of the sparsity of rating matrix, most of the CF similarities cannot be calculated. The unrated ratings can be calculated by the item-based CF prediction. The user CF and semantic similarities are then calculated, shown in Table V. The final predictions of the unrated ratings are calculated by the user-based CF prediction, and the ranking coefficients are computed, which are shown in Table VI. The products can be ranked and recommended accordingly.

TABLE IV
ITEM CF AND SEMANTIC SIMILARITY

CF	Sem	P1	P2	P3	P4	P5
P1			0.937	0.435	0.435	0.760
P2				0.440	0.458	0.775
P3					0.901	0.790
P4						0.936
P5						

TABLE V
USER CF AND SEMANTIC SIMILARITY

CF	Sem	U1	U2	U3	U4	U5
U1			0.951	0.799	0.719	0.971
U2		0.085		0.978	0.974	0.526
U3		-0.444	-0.004		0.975	0.898
U4		0.074	-0.758	0.227		0.525
U5		-0.797	-0.660	0.404	0.430	

TABLE VI
PREDICTED RANKING COEFFICIENT

	P1	P2	P3	P4	P5
U1			3.9		4.2
U2			3.7		4.3
U3	3.2				

U4		3.3		3.6	
U5	3.0		3.7		

From the case study, it can be seen that the fuzzy tree similarity measure can effectively evaluate the semantic similarity between the telecom products or between the user profiles. This makes the similarity between items or users more meaningful and also solves the problem of lack of ratings. The combination of item-based CF and user-based CF techniques takes advantage of both the horizontal and vertical information in the rating matrix, which also solves the sparsity problem that is common in telecom product recommendation.

VII. CONCLUSIONS AND FURTHER STUDY

A fuzzy tree similarity based hybrid recommendation approach is presented in this paper and its possible application in telecom products is designed and discussed. A fuzzy tree similarity measure is developed to comprehensively compare both the concepts and values of the tree structured telecom product data and user profile data. In the hybrid recommendation approach, the item and user similarity measures both integrate the semantic and CF similarity. The approach combines the item-based CF and user-based CF prediction techniques. This can deal with the data sparsity and cold start problems. The telecom product recommendation case study shows that the proposed approach can make recommendations in the situation that lacks of rating information and has fuzzy data, and it is well-suited to the telecom product recommendation.

The proposed recommendation approach is being implemented into an online system. It will be tested and compared with existing ones in the further study.

ACKNOWLEDGMENT

The work presented in this paper was supported by the Australian Research Council (ARC) under discovery grant DP110103733.

REFERENCES

[1] D. Wu, J. Lu, and G. Zhang, "A hybrid recommendation approach for hierarchical items," in *Proceedings of the 2010 International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, 2010, pp. 492-497.

[2] J. Lu, D. Ruan, and G. Zhang, *E-service intelligence: methodologies, technologies and applications* vol. 37: Springer Verlag, 2006.

[3] G. Adomavicius and A. Tuzhilin, "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, pp. 734-749, 2005.

[4] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Modeling and User-Adapted Interaction*, vol. 12, pp. 331-370, 2002.

[5] J. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative filtering recommender systems," in *The Adaptive Web*. vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds., ed: Springer-Verlag, Berlin Heidelberg, 2007, pp. 291-324.

[6] M. Pazzani and D. Billsus, "Content-based recommendation systems," in *The Adaptive Web*. vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds., ed: Springer-Verlag, Berlin Heidelberg, 2007, pp. 325-341.

[7] X. Y. Shi, H. W. Ye, and S. J. Gong, "A personalized recommender integrating item-based and user-based collaborative

filtering," in *ISBIM '08. International Seminar on Business and Information Management*, 2008, pp. 264-267.

[8] V. C. Bhavsar, H. Boley, and L. Yang, "A weighted-tree similarity algorithm for multi-agent systems in e-business environments," *Computational Intelligence*, vol. 20, 2004.

[9] Y. Xue, C. Wang, H. Ghenniwa, and W. Shen, "A tree similarity measuring method and its application to ontology comparison," *Journal of Universal Computer Science*, vol. 15, pp. 1766-1781, 2009.

[10] C. Wang, J. Lu, and G. Zhang, "Integration of Ontology Data through Learning Instance Matching," in *IEEE/WIC/ACM International Conference on Web Intelligence, 2006. WI 2006.*, 2006, pp. 536-539.

[11] P. Bille, "A survey on tree edit distance and related problems," *Theoretical Computer Science*, vol. 337, pp. 217-239, 2005.

[12] K. Zhang, "A constrained edit distance between unordered labeled trees," *Algorithmica*, vol. 15, pp. 205-222, 1996.

[13] R. R. Yager, "Fuzzy logic methods in recommender systems," *Fuzzy Sets and Systems*, vol. 136, pp. 133-149, 2003.

[14] C. Cornelis, X. Guo, J. Lu, and G. Zhang, "A fuzzy relational approach to event recommendation," in *the Second Indian International Conference on Artificial Intelligence (IICAI-05)*, Pune (pronounced as PooNay), INDIA 2005, pp. 2231-2242.

[15] C. Cornelis, J. Lu, X. Guo, and G. Zhang, "One-and-only item recommendation with fuzzy logic techniques," *Information Sciences*, vol. 177, pp. 4906-4921, 2007.

[16] C. Porcel, A. G. López-Herrera, and E. Herrera-Viedma, "A recommender system for research resources based on fuzzy linguistic modeling," *Expert Systems with Applications*, vol. 36, pp. 5173-5183, 2009.

[17] G. Zhang and J. Lu, "Using general fuzzy number to handle uncertainty and imprecision in group decision-making," *Intelligent Sensory Evaluation: Methodologies and Applications*, pp. 51-70, 2004.

[18] G. Zhang and J. Lu, "An integrated group decision-making method dealing with fuzzy preferences for alternatives and individual judgments for selection criteria," *Group Decision and Negotiation*, vol. 12, pp. 501-515, 2003.

[19] G. Zhang and J. Lu, "A linguistic intelligent user guide for method selection in multi-objective decision support systems," *Information Sciences*, vol. 179, pp. 2299-2308, 2009.

[20] G. Valiente, *Algorithms on Trees and Graphs* Springer-Verlag, New York, 2002.

[21] D. Jungnickel, *Graphs, Networks, and Algorithms*, 3rd ed.: Springer-Verlag, Berlin Heidelberg, 2008.