Forming Communities in Web-based Educational Systems through Users´ Preferences and Interest Measuring

Reginaldo A. Gotardo, Cesar A. C. Teixeira, Sergio D. Zorzo
reggotardo@gmail.com, cesar@dc.ufscar.br, zorzo@dc.ufscar.br

Abstract - Service customizing in Web-based Educational Systems aims at directing content and teaching strategies towards students´ individual and group needs. This paper presents a virtual learning community forming approach, which helps knowledge exchange among its members. The approach here presented uses implicit and explicit information collection of users´ interests and preferences and relates the values in a correlation among the users. The correlation measuring ends up in groups with distinct characteristics. The proposal presented is validated by case study and it shows that from correlation values among the users in the interest and preference items, a group algorithm results in the formation of the intended groups. The results among interest and preference correlation were compared and assessed.

Index Terms – Personalization, Recommendation Systems, Web-based Educational Systems.

1 - INTRODUCTION

As the internet becomes more widespread, Web based Educational Systems (WbE-S) demand more and more attention. Another very important area on computing deals with Web systems personalization. The personalization applied to WbE-S is the focus of several works in research fields like Data Mining, Web Mining, User Modeling, Adaptive Hypermedia, Intelligent Tutoring Systems and Recommendation Systems [1].

The virtual communities or groups for learning used as a tool to content dissemination and to improve learning process is an important theme to distance education and content personalization [2].

The user modeling, or student modeling, has been one of the major areas of studies and challenges for Web based Adaptive Educational Systems. To model and adapt the user profile, information about the user’s behavior is needed. This was implicitly observed or explicitly asked to the user.

In this work, we present the use of Collaborative Filtering techniques, a subset of Recommendation Systems, to forming student communities to the WbE-S. The selection of information for the Collaborative Filtering was performed in a traditional way: through the explicit classification of preferences in the content by the user in a WbE-S. However, this classification isn’t sufficient to well-formed interest groups and we used implicit information about user’s navigation, also called usage mining. Commonly, Collaborative Filtering uses explicit evaluation of the users about the system’s resources, and it doesn’t considerer the implicit relationship about them.

The implicit information collected measure the interaction of the user with the system resources. The information bellows are divided in three types: Access Total Time, Most Recently Used and Most Frequently Used.

However, these information are weighted by a tutor or specialist that knows the domain, in order to measure the user choice by his real behavior in the system.

Through of measures of three user interaction values with the system, it is presented a different approach to obtain interests groups in a WbE System.

This work is organized as it follows: section 2 presents the traditional Recommender Systems Approach, Web-based Education uses and Forming Communities information; section 3 presents how transform web usage in interest measures; the section 4 presents how preferences was measured; the section 5 presents our Approach to Forming Communities using User’s Interest and Preferences; section 6 discusses the case study results and section 5 presents the conclusions and future works.

2 - PERSONALIZATION, RECOMMENDATION SYSTEMS, WEB-BASED EDUCATION AND FORMING COMMUNITIES

Personalization is an important issue in e-learning as it might help to improve both student performance and use experience. The common e-learning systems usually don’t allow the courses personalization to the student profile, but only propose standard courses. They are limited to direct appropriate learning to the student and don’t consider a student learning model. [3]

E-learning systems, which uses user modeling and personalization, can allow the user receive more interest information and more adequate content to learn.

There are several methods for e-learning personalization as Intelligent Tutoring System, Adaptive Hypermedia and the Recommendation System - method traditionally used in e-commerce applications. A Recommender System is a new approach to help users find relevant information. Basically, all the Recommender Systems do the same: they try to identify the most important items to the users and then recommend these items [4].

October 22 – 25, 2008, Saratoga Springs, NY
38th ASEE/IEEE Frontiers in Education Conference
F4E-15
Traditional Recommender Systems use three techniques to make suggestions to the users: Collaborative Filtering, Content-Based Filtering and both. In the Collaborative Filtering the recommendation is based on user's preferences whose profiles are similar to target-users. The Content-Based Filtering is based on the relation among the contents. Besides that, these two techniques can be combined and it's called Hybrid Method.

These techniques of recommendation are used in several areas. Frequently, they are used in e-commerce Web sites (for example, Amazon.com, eBay, MovieFinder.com). They aim at increasing their sales and improving the client's treatment on the internet.

To recommend items to each user, first the system collects several users' information. It is represented in a user model to try to find relevant information about user's preferences and the system contents. These information can be collected using explicit (users provide them deliberately) or implicit methods (the user's behavior in the system is collected) [4].

In Educational Systems, these methods are used in the identification of the sequence navigation to each student in the system, as the case-based reasoning methods used in learning interactive environments [5].

Groups or communities of individuals, organized informally or semi-formally, are instrument of knowledge management and sharing [2].

The term “community” is a socio-scientific collective term for specific types of social groups [2]. Although there is no generally accepted definition, we use the term in this paper to describe a group of users that have interest and preferences in common.

We have used the personalization techniques, more specifically system recommendation techniques, to forming communities using user's navigation data and user's classification about visited resources.

In the next section we present how we have transformed user's data in interest measured.

3 - TRANSFORMING WEB USAGE IN USER’S INTERESTS MEASURES

The Collaborative Filtering uses students’ contributions to system classification. The system input is a matrix like which is shown in table 1 and which lines are users and columns are items. The matrix data are users’ ratings about items. Besides that, this technique is called “k-nearest-neighbor” and the similarity measuring can use coefficients as Pearson’s one [4; 5; 6]. The interests’ measurement is performed considering the explicit classification the users supply to the visited content.

The explicit user classification about the resources has a problem: the user can use the resource and he/she could not classify it. However, the system wouldn’t have information about the user and resource relationship.

To solve this problem, we propose a different approach to measure the user interaction with the system resources.

The interaction between the user and the resources is measured using the characteristics of known algorithms used in Operating Systems for replacement of memory pages. The measures are three: Most Frequently Used, Most Recently Used and Access Total Time.

The MFU (Most Frequently Used) Algorithm, inversion of the LFU (Least Frequently Used) algorithm, is used to verify items whose user access is more frequent, measuring it through the number of visits to the resource.

\[
MFU(u, Ri_u) = \frac{\text{frequency}(u, Ri_u)}{\max \text{Frequency}(Ri)}
\]  

Thus, the MFU measure of a user “u” in relating to a resource “Ri” is computed by the value of the visitation frequency of u to the resource Ri. Dividing this value by the frequency of maximum visitation to the same resource (this value is calculated considering all the users and respective frequencies of visitation to the same resource) we have a normalized MFU relating to all the users. This normalization will be important for the posterior use of MFU.

The MRU (Most Recently Used) Algorithm, inversion of the LRU (Least Recently Used) algorithm, is used to verify items whose user access is more recent, using the last access values for this.

The MRU measure is obtained by the value of the last access to the resource Ri, for the user u. This measure is normalized, dividing it by the last access maximum value, that is, the most recent access to the resource Ri.

\[
MRU(u, Ri_u) = \frac{\text{lastAccess}(u, Ri_u)}{\max \text{lastAccess}(Ri)}
\]  

Moreover, the access time is used as reference of the user interest, therefore, many times; the frequency in short times necessarily does not represent the user interest.

\[
ATT(u, Ri_u) = \frac{\text{ATT}(u, Ri_u)}{\max \text{ATT}(Ri)}
\]  

The values maxFrequency, maxlastAccess and maxATT are used to normalize the measurement of these variables between 0 and 1.

After the measurement of these three values, according to user’s navigation, the user’s interest can be calculated weighting each value by a knowledge domain specialist or a tutor, as is shown in figure 1. For instance, in some course topics can be more interesting the value of recent access (MRU) than the access total time in a resource (ATT).
The Collaborative Filtering uses students’ contributions to system classification. The system input is a matrix, which is depicted in Table 1. The lines are users and the columns are items. The matrix data are users’ ratings about items. Besides that, this technique is called “k-nearest-neighbor” and the similarity measure can use coefficients as Pearson’s one [4; 5; 6].

The preferences’ measurement is performed considering the explicit classification the users supply to the visited content.

4 - MEASURING PREFERENCES

The Collaborative Filtering uses students’ contributions to system classification. The system input is a matrix, which is depicted in Table 1. The lines are users and the columns are items. The matrix data are users’ ratings about items. Besides that, this technique is called “k-nearest-neighbor” and the similarity measure can use coefficients as Pearson’s one [4; 5; 6].

The preferences’ measurement is performed considering the explicit classification the users supply to the visited content.

4 - MEASURING PREFERENCES

Five levels of preferences had been stipulated, related to the user’s approval to the visited resource as it’s showed in figure 2. For example, a user who visits a page with information that he believes to be interesting, can classify it with the maximum value (five). If this same user visits a forum of the same subject and does not find it interesting, he could classify it with the minimum value (one).

To classify the relationship among the user and objects of the system, all the related objects are considered Resources. In this way, in a conventional WbE-S, the resources would be grouped in: learning objects, tools, users and groups (or communities). Thus, a resource R can be a User (U), a Community (C), a System Tool (T), or a Learning Object (O).

To classify the relationship among the user and objects of the system, all the related objects are considered Resources. In this way, in a conventional WbE-S, the resources would be grouped in: learning objects, tools, users and groups (or communities). Thus, a resource R can be a User (U), a Community (C), a System Tool (T), or a Learning Object (O).

Where

\[ \alpha \] MFU weight
\[ \beta \] MRU weight
\[ \gamma \] ATT weight

And

\[ \alpha + \beta + \gamma = 10 \]

To classify the relationship among the user and objects of the system, all the related objects are considered Resources. In this way, in a conventional WbE-S, the resources would be grouped in: learning objects, tools, users and groups (or communities). Thus, a resource R can be a User (U), a Community (C), a System Tool (T), or a Learning Object (O).

The Collaborative Filtering uses students’ contributions to system classification. The system input is a matrix, which is depicted in Table 1. The lines are users and the columns are items. The matrix data are users’ ratings about items. Besides that, this technique is called “k-nearest-neighbor” and the similarity measure can use coefficients as Pearson’s one [4; 5; 6].

The preferences’ measurement is performed considering the explicit classification the users supply to the visited content.

4 - MEASURING PREFERENCES

The preferences’ measurement is performed considering the explicit classification the users supply to the visited content.

To classify the relationship among the user and objects of the system, all the related objects are considered Resources. In this way, in a conventional WbE-S, the resources would be grouped in: learning objects, tools, users and groups (or communities). Thus, a resource R can be a User (U), a Community (C), a System Tool (T), or a Learning Object (O).

Where

\[ \alpha \] MFU weight
\[ \beta \] MRU weight
\[ \gamma \] ATT weight

And

\[ \alpha + \beta + \gamma = 10 \]

To classify the relationship among the user and objects of the system, all the related objects are considered Resources. In this way, in a conventional WbE-S, the resources would be grouped in: learning objects, tools, users and groups (or communities). Thus, a resource R can be a User (U), a Community (C), a System Tool (T), or a Learning Object (O).

The Collaborative Filtering uses students’ contributions to system classification. The system input is a matrix, which is depicted in Table 1. The lines are users and the columns are items. The matrix data are users’ ratings about items. Besides that, this technique is called “k-nearest-neighbor” and the similarity measure can use coefficients as Pearson’s one [4; 5; 6].

The preferences’ measurement is performed considering the explicit classification the users supply to the visited content.

5 - NEIGHBOR CALCULATION

The Collaborative Filtering uses students’ contributions to system classification. The system input is a matrix, which is depicted in Table 1. The lines are users and the columns are items. The matrix data are users’ ratings about items. Besides that, this technique is called “k-nearest-neighbor” and the similarity measure can use coefficients as Pearson’s one [4; 5; 6].

The preferences’ measurement is performed considering the explicit classification the users supply to the visited content.

The Collaborative Filtering uses students’ contributions to system classification. The system input is a matrix, which is depicted in Table 1. The lines are users and the columns are items. The matrix data are users’ ratings about items. Besides that, this technique is called “k-nearest-neighbor” and the similarity measure can use coefficients as Pearson’s one [4; 5; 6].

The preferences’ measurement is performed considering the explicit classification the users supply to the visited content.

To classify the relationship among the user and objects of the system, all the related objects are considered Resources. In this way, in a conventional WbE-S, the resources would be grouped in: learning objects, tools, users and groups (or communities). Thus, a resource R can be a User (U), a Community (C), a System Tool (T), or a Learning Object (O).

Where

\[ \alpha \] MFU weight
\[ \beta \] MRU weight
\[ \gamma \] ATT weight

And

\[ \alpha + \beta + \gamma = 10 \]

To classify the relationship among the user and objects of the system, all the related objects are considered Resources. In this way, in a conventional WbE-S, the resources would be grouped in: learning objects, tools, users and groups (or communities). Thus, a resource R can be a User (U), a Community (C), a System Tool (T), or a Learning Object (O).

The Collaborative Filtering uses students’ contributions to system classification. The system input is a matrix, which is depicted in Table 1. The lines are users and the columns are items. The matrix data are users’ ratings about items. Besides that, this technique is called “k-nearest-neighbor” and the similarity measure can use coefficients as Pearson’s one [4; 5; 6].

The preferences’ measurement is performed considering the explicit classification the users supply to the visited content.

The Collaborative Filtering uses students’ contributions to system classification. The system input is a matrix, which is depicted in Table 1. The lines are users and the columns are items. The matrix data are users’ ratings about items. Besides that, this technique is called “k-nearest-neighbor” and the similarity measure can use coefficients as Pearson’s one [4; 5; 6].

The preferences’ measurement is performed considering the explicit classification the users supply to the visited content.

The Collaborative Filtering uses students’ contributions to system classification. The system input is a matrix, which is depicted in Table 1. The lines are users and the columns are items. The matrix data are users’ ratings about items. Besides that, this technique is called “k-nearest-neighbor” and the similarity measure can use coefficients as Pearson’s one [4; 5; 6].

The preferences’ measurement is performed considering the explicit classification the users supply to the visited content.

The Collaborative Filtering uses students’ contributions to system classification. The system input is a matrix, which is depicted in Table 1. The lines are users and the columns are items. The matrix data are users’ ratings about items. Besides that, this technique is called “k-nearest-neighbor” and the similarity measure can use coefficients as Pearson’s one [4; 5; 6].

The preferences’ measurement is performed considering the explicit classification the users supply to the visited content.

The Collaborative Filtering uses students’ contributions to system classification. The system input is a matrix, which is depicted in Table 1. The lines are users and the columns are items. The matrix data are users’ ratings about items. Besides that, this technique is called “k-nearest-neighbor” and the similarity measure can use coefficients as Pearson’s one [4; 5; 6].

The preferences’ measurement is performed considering the explicit classification the users supply to the visited content.

The Collaborative Filtering uses students’ contributions to system classification. The system input is a matrix, which is depicted in Table 1. The lines are users and the columns are items. The matrix data are users’ ratings about items. Besides that, this technique is called “k-nearest-neighbor” and the similarity measure can use coefficients as Pearson’s one [4; 5; 6].

The preferences’ measurement is performed considering the explicit classification the users supply to the visited content.
dissimilarity. In equation 1, it is described the Pearson Correlation calculation according to user’s interests.

\[
W_{a,u} = \frac{\sum\sum (i_{a,i} - \overline{i}_a)(r_{u,i} - \overline{r}_u)}{\sqrt{\sum(i_{a,i} - \overline{i}_a)^2} \times \sqrt{\sum(i_{u,i} - \overline{i}_u)^2}}
\]  

\(a\) – target user
\(u\) – neighbor user
\(W_{a,u}\) – correlation between user “a” and user “u”
\(i_{a,i}\) – interest of user “a” to resource “r”
\(i_{u,i}\) – interest of user “u” to resource “r”
\(\overline{i}_a\) - average of every common interest between user “a” and user “u”
\(\overline{i}_u\) - average of every common interest between user “u” and user “a”

After this calculation we have a correlation matrix (CI) about users’ interests, like it shows in table 2.

<table>
<thead>
<tr>
<th>Users</th>
<th>Users</th>
<th>Users</th>
<th>Users</th>
<th>Users</th>
<th>Users</th>
<th>Users</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>i = 1</td>
<td>i = 2</td>
<td>…</td>
<td>i = N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>j = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>j = 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>j = N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For preferences calculation we have another similar matrix (CP), but the calculation is done based in classification that the users insert about resources.

Both matrix (CI and CP) will be used to forming communities.

6 - FORMING COMMUNITIES

This paper presents a virtual learning community forming approach, which helps knowledge exchange among its members. The approach uses implicit and explicit information collection of users’ interests and preferences and relates the values in a correlation among the users. The correlation measuring ends up in groups with distinct characteristics.

The users’ interests and preferences values are acquired through measuring the relationship between the users and systems resources – learning objects, tools, users and groups (or communities). To calculate the user interest in a resource, the interaction between them is measured using known algorithms features, which are commonly used in Operational Systems for substituting memory pages. The user’s interest on a given object is acquired through pondered adding of these measures, where pondering may be variable and related to the domain in question. We have, therefore, a I(UxR) matrix which relates the users’ interests values (U) to resources (R). The Pearson’s correlation among users is calculated from this matrix.

The preference values are acquired by explicit classification which users provide for the accessed resources. From this classification we have a P(UxR) matrix which relates users’ preferences values (U) and resources (R). Again, the Pearson correlation among the users from this matrix is calculated.

In the approach, described in figure 1, the accessed and classified resources are registered in a data base of collected information (1). The calculation of the user’s interest is done (2) according to the weight which the tutor insert in the system (3). After this, the tutor inserts again weights to Interests and Classification which will be used to forming user’s group. Besides that, the neighboring inference is calculated (4) using k-means algorithm. To conclude, the Tutor can use the formed groups to future works and homework.

The neighboring inference (4) is done, using correlation matrix M1 and M2 (users’ interests and preferences). The Tutor can insert a weight for each matrix and, based on this weight the calculation is done.

7 - EXPERIMENTAL CASE STUDY

The proposal presented is validated by case study and it shows that from correlation values among the users in the interest and preference items, a group algorithm results in the formation of the intended groups. The results among interest and preference correlation were compared and assessed. To test the model efficiency we evaluated a scene with 10 pupils and 8 resources using values provided by a specialist as a hypothetical scenario. The values for the interaction of the user with the resources had been measured, considering the three previously described variables (MFU, MRU and TTA). For the considered scene, it was considered the respective weights to these variables: 1, 3 and 6. Besides that, we had the user’s classification about resources.
Table 3 demonstrates the measured users’ interest values which were pondered for the weights of variables MFU, MRU and ATT.

<table>
<thead>
<tr>
<th>Resources</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>User 1</td>
<td>5.8</td>
<td>6.3</td>
<td>6.5</td>
<td>7.1</td>
<td>10.0</td>
<td>2.8</td>
<td>3.1</td>
<td>9.4</td>
</tr>
<tr>
<td>User 2</td>
<td>3.3</td>
<td>4.0</td>
<td>1.8</td>
<td>1.6</td>
<td>5.2</td>
<td>6.3</td>
<td>7.3</td>
<td>1.8</td>
</tr>
<tr>
<td>User 3</td>
<td>9.1</td>
<td>9.4</td>
<td>8.5</td>
<td>9.1</td>
<td>8.8</td>
<td>3.2</td>
<td>3.3</td>
<td>10.0</td>
</tr>
<tr>
<td>User 4</td>
<td>3.1</td>
<td>2.0</td>
<td>2.9</td>
<td>3.0</td>
<td>6.6</td>
<td>5.3</td>
<td>3.7</td>
<td>3.6</td>
</tr>
<tr>
<td>User 5</td>
<td>6.0</td>
<td>6.0</td>
<td>7.7</td>
<td>8.3</td>
<td>10.0</td>
<td>2.5</td>
<td>3.7</td>
<td>8.5</td>
</tr>
<tr>
<td>User 6</td>
<td>4.5</td>
<td>2.3</td>
<td>2.6</td>
<td>3.5</td>
<td>7.0</td>
<td>2.7</td>
<td>3.2</td>
<td>3.0</td>
</tr>
<tr>
<td>User 7</td>
<td>3.2</td>
<td>4.3</td>
<td>2.6</td>
<td>3.4</td>
<td>6.4</td>
<td>3.8</td>
<td>3.0</td>
<td>3.7</td>
</tr>
<tr>
<td>User 8</td>
<td>1.1</td>
<td>1.3</td>
<td>1.3</td>
<td>1.5</td>
<td>5.4</td>
<td>8.8</td>
<td>6.6</td>
<td>0.9</td>
</tr>
<tr>
<td>User 9</td>
<td>0.0</td>
<td>2.3</td>
<td>2.9</td>
<td>1.2</td>
<td>5.0</td>
<td>9.3</td>
<td>7.3</td>
<td>1.5</td>
</tr>
<tr>
<td>User 10</td>
<td>1.0</td>
<td>0.0</td>
<td>2.3</td>
<td>0.8</td>
<td>4.6</td>
<td>8.3</td>
<td>5.6</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Remember that in the table 3 the values are already weighted and we have a pondered sum using MRU, MFU and ATT. After that, we calculated the correlation matrix about user’s interest, using table 3 values. Table 4 shows the interest correlation matrix calculated using the values of matrix 3 and the formula 5. The correlation between a user “i” and himself is 1 as definition.

<table>
<thead>
<tr>
<th>Users</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0</td>
<td>0.7</td>
<td>0.8</td>
<td>0.7</td>
<td>0.9</td>
<td>0.7</td>
<td>0.7</td>
<td>0.4</td>
</tr>
<tr>
<td>2</td>
<td>0.7</td>
<td>1.0</td>
<td>0.5</td>
<td>0.8</td>
<td>0.5</td>
<td>0.8</td>
<td>0.9</td>
<td>0.7</td>
</tr>
<tr>
<td>3</td>
<td>0.8</td>
<td>0.5</td>
<td>1.0</td>
<td>0.3</td>
<td>0.9</td>
<td>0.4</td>
<td>0.3</td>
<td>-0.1</td>
</tr>
<tr>
<td>4</td>
<td>0.7</td>
<td>0.8</td>
<td>0.3</td>
<td>1.0</td>
<td>0.4</td>
<td>1.0</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>5</td>
<td>0.9</td>
<td>0.5</td>
<td>0.9</td>
<td>0.4</td>
<td>1.0</td>
<td>0.4</td>
<td>0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>6</td>
<td>0.7</td>
<td>0.8</td>
<td>0.4</td>
<td>1.0</td>
<td>0.4</td>
<td>1.0</td>
<td>0.9</td>
<td>0.8</td>
</tr>
<tr>
<td>7</td>
<td>0.7</td>
<td>0.9</td>
<td>0.3</td>
<td>0.9</td>
<td>0.3</td>
<td>0.9</td>
<td>1.0</td>
<td>0.9</td>
</tr>
<tr>
<td>8</td>
<td>0.4</td>
<td>0.7</td>
<td>-0.1</td>
<td>0.9</td>
<td>0.0</td>
<td>0.8</td>
<td>0.9</td>
<td>1.0</td>
</tr>
<tr>
<td>9</td>
<td>0.5</td>
<td>0.7</td>
<td>0.0</td>
<td>0.8</td>
<td>0.1</td>
<td>0.6</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>10</td>
<td>0.4</td>
<td>0.6</td>
<td>-0.2</td>
<td>0.9</td>
<td>0.0</td>
<td>0.8</td>
<td>0.8</td>
<td>1.0</td>
</tr>
</tbody>
</table>

To group (or community) forming we have used the Interest Matrix Calculation (table 3) and a K-means algorithm. We have used the Weka Tool as implementation of K-means algorithm. The k-means is an algorithm to cluster n objects based on attributes into k partitions, where k < n. It assumes that the object attributes form a vector space. The objective it tries to achieve is to minimize total intra-cluster variance, or, the squared error.

\[
V = \sum_{i=1}^{k} \sum_{x_j \in S_i} (x_j - \mu_i)^2
\]

In the function 6 there are k clusters \( S_i \), \( i = 1, 2, \ldots, k \) and \( \mu_i \) is the centroid or mean point of all the points \( x_j \in S_i \).

We could use only Interest Matrix Calculation as input to a k-means algorithm and find interest groups as is shown in figure 4.

In this case, we’d have 3 distinct groups based on interest data. Figure 4 (b) shows each group with each participant.

Table 5 which contains the users’ interest data. Figure 4 (b) shows each group with each participant.

<table>
<thead>
<tr>
<th>Resources</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>User 1</td>
<td>5.0</td>
<td>2.0</td>
<td>4.0</td>
<td>1.0</td>
<td>3.0</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>User 2</td>
<td>4.0</td>
<td>3.0</td>
<td>4.0</td>
<td>2.0</td>
<td>4.0</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>User 3</td>
<td>4.0</td>
<td>3.0</td>
<td>5.0</td>
<td>2.0</td>
<td>3.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>User 4</td>
<td>2.0</td>
<td>1.0</td>
<td>2.0</td>
<td>5.0</td>
<td>2.0</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>User 5</td>
<td>3.0</td>
<td>1.0</td>
<td>2.0</td>
<td>5.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>User 6</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>5.0</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>User 7</td>
<td>2.0</td>
<td>1.0</td>
<td>2.0</td>
<td>5.0</td>
<td>2.0</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>User 8</td>
<td>2.0</td>
<td>1.0</td>
<td>2.0</td>
<td>5.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>User 9</td>
<td>1.0</td>
<td>3.0</td>
<td>3.0</td>
<td>4.0</td>
<td>3.0</td>
<td>5.0</td>
<td>1.0</td>
</tr>
<tr>
<td>User 10</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td>5.0</td>
<td>4.0</td>
<td>5.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Nevertheless, according figure 5, we have used the Interests and Preferences Matrix to Forming Groups or Communities. The table 5 which contains the users’ preferences about the resources and the table 3 (users’ interests) are used together. Besides that, the Tutor insert the weight of each matrix in calculation and we have a new matrix IP(UxR) (pondered sum using matrix I and P).

---

1 www.cs.waikato.ac.nz/ml/weka
Therefore, we calculate a new matrix, showed in table 6, which is a pondered matrix using users’ Interest and Preferences. In this matrix we applied the k-means algorithm and get a new communities forming. The preference weight was 6 and the interest weight was 4.

Therefore, we calculate a new matrix, showed in table 6, which is a pondered matrix using users’ Interest and Preferences. In this matrix we applied the k-means algorithm and get a new communities forming. The preference weight was 6 and the interest weight was 4.

### Table 6: Interest and Preferences Correlation Matrix Pondered by a Tutor – IP(UXU)

<table>
<thead>
<tr>
<th>Users</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0</td>
<td>0.2</td>
<td>0.8</td>
<td>0.1</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>-0.6</td>
<td>-0.6</td>
<td>-0.4</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>1.0</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.7</td>
<td>0.4</td>
<td>0.1</td>
<td>-0.7</td>
<td>-0.7</td>
<td>-0.6</td>
</tr>
<tr>
<td>3</td>
<td>0.8</td>
<td>-0.1</td>
<td>1.0</td>
<td>-0.1</td>
<td>0.7</td>
<td>0.4</td>
<td>0.1</td>
<td>-0.7</td>
<td>-0.7</td>
<td>-0.6</td>
</tr>
<tr>
<td>4</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>1.0</td>
<td>0.4</td>
<td>0.7</td>
<td>0.9</td>
<td>0.7</td>
<td>0.4</td>
<td>0.7</td>
</tr>
<tr>
<td>5</td>
<td>0.6</td>
<td>-0.5</td>
<td>0.7</td>
<td>0.4</td>
<td>1.0</td>
<td>0.8</td>
<td>0.6</td>
<td>-0.2</td>
<td>-0.6</td>
<td>-0.3</td>
</tr>
<tr>
<td>6</td>
<td>0.4</td>
<td>0.0</td>
<td>0.4</td>
<td>0.7</td>
<td>0.8</td>
<td>1.0</td>
<td>0.9</td>
<td>0.2</td>
<td>-0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>7</td>
<td>0.2</td>
<td>0.0</td>
<td>0.1</td>
<td>0.9</td>
<td>0.6</td>
<td>0.9</td>
<td>1.0</td>
<td>0.5</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>8</td>
<td>-0.6</td>
<td>0.2</td>
<td>-0.7</td>
<td>0.7</td>
<td>-0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>1.0</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>9</td>
<td>-0.6</td>
<td>0.4</td>
<td>-0.7</td>
<td>0.4</td>
<td>-0.6</td>
<td>0.2</td>
<td>0.8</td>
<td>1.0</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>10</td>
<td>-0.4</td>
<td>0.3</td>
<td>-0.6</td>
<td>0.7</td>
<td>-0.3</td>
<td>0.0</td>
<td>0.4</td>
<td>0.8</td>
<td>0.9</td>
<td>1.0</td>
</tr>
</tbody>
</table>

The results of k-means on matrix IP (table 6) is showed in figure 5. As you can see, different groups are formed, because we had interference of preferences values in interest values and it is an important result which shows the discrepancy that can occurs in a WbE System measuring users’ behavior.

### 8 - Conclusion

This paper described an Approach to Forming Communities or Groups of Students who have interests and preferences in common. We have presented a mechanism to measure the interests and preferences values which represents the users’ behavior Web-based Educational Systems. This Approach associates implicit and explicit information about the users to represent their behavior.

The case study has shown that these two types of information can influence the results of approach to forming communities.

Future works may include a more detailed analysis of other algorithms to forming communities in comparison with our Approach. Also, we are finishing a tool to Help Tutors to use the results of Groups/Communities Approach in their class. This tool will be used to instruct how the class can be divided in groups with similar interests and preferences.

### Acknowledgment

We would like to thank the following Brazilian agencies: CAPES for supporting Reginaldo Gotardo in his Master Program; FAPESP, which provides funding for the TIDIA-AE project, process 05/60653-1; and COC Learning System.

### References


### Author Information

Reginaldo A. Gotardo, Professor of COC Colleague and Msc Student of Federal University of São Carlos.

Cesar A. C. Teixeira, PhD Professor of Federal University of São Carlos.

Sergio Donizetti Zorzo, PhD Professor of Federal University of São Carlos.

---

978-1-4244-1970-8/08/$25.00 ©2008 IEEE  
October 22 – 25, 2008, Saratoga Springs, NY  
38th ASEE/IEEE Frontiers in Education Conference  
F4E-20