International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 4 Issue 7, July 2015

Model-based Methods for Recommender Systems

Yuriy Stekh[1], Mykhaylo Lobur[2], Vitalij Artsibasov[3]

1. CAD Department, Lviv Polytechnic National University, UKRAINE, Lviv, S. Bandery street 12.
2. CAD Department, Lviv Polytechnic National University, UKRAINE, Lviv, S. Bandery street 12.
3. CAD Department, Lviv Polytechnic National University, UKRAINE, Lviv, S. Bandery street 12.

Abstract - **In this paper the analysis methods and tools for building recommender systems performed. In modern recommendation systems are widely used methods and models collaborative filtering. The main problems in collaborative filtering method: a very large dimension, new user and new item, the initial user logon. To solve these problems using an approach based on models. Models are based on data mining methods. We study the following methods of data mining: cluster analysis, association rules, fuzzy set theory. These methods are used in the overall scheme of collaborative filtering. The association rule method is applied to solve the initial user logon problem. C-means fuzzy clustering method is applied in user-based and item-based collaborative filtering methods. K-mode categorical clustering method is applied to solve new user and new item problems.**

Keywords – **Recommender system, collaborative filtering, cluster analysis, association rules, fuzzy set theory.**

I. Introduction

Recommendation system compares the data collected from users and create a list of items recommended to the user. They are an alternative search algorithm because it helps users to find items and information that they would not find themselves.

Since the first works in the mid-1990s, recommendation systems have become the subject of intense scientific attention. During the last decade a lot of work has been done both theoretical and applied, dedicated to the development of recommendation systems. At present the problem of recommendation systems keeps to himself a lot of interest. In this area there remain many unsolved problems. Addressing these challenges promises many opportunities for practical application. This will allow users to cope with the huge amount of information, and provide them the tools making personalized recommendations. An example of a specific application of recommendation systems can serve a system recommendations of books, CD and other items at Amazon.com [1], movies on MovieLens [2], news on VERSIFI Technologies (former AdaptiveInfo.com) [3].

In modern recommendation systems are widely used methods and models collaborative filtering [4, 5, 6, 7, 8]. Model of collaborative filtering is user-item matrix **A** dimension mxn. Rows of the matrix correspond to the vector of users **U**, column correspond to the vector of items **I**. Each element of the matrix a_{ij} contains i_j product rating assigned by user u_i. The rating scale is estimated to be integers $a_{ij} \in \{1,2,3,4,5\}$.

The main problems in collaborative filtering method: a very large dimension and greate sparsity matrix **A**, new user and new item.

To solve these problems using an approach based on models [4, 8]. Originally formed descriptive model of user preferences, items and the relationship between them. Then recommendations are formed on the basis of the obtained model. The advantage of this approach is the availability of the model. It gives more insight generated recommendations and relationships in data availability. The formation of recommendations is divided into two stages: intensive training model in off-line mode and a fairly simple calculation based on the recommendations of the existing models in real time.

II. Recommendations based on association rules for the initial user logon

The input data for the solution of this problem is the matrix of user-items. The essence of the problem lies in the fact that the user selects a specific item (movies, music videos, books, articles, etc.). Recommender system uses the selected item as input and recommends the user a number of items that have chosen other users. This problem can have two interperpretations:1) the problem of items recommendations for the user who is already working with the advisory system;2) the problem of recommendations objects for the new user that implements the first login. In the first problem, the system recommends the user to those items that he has not chosen. In the second problem the system recommends the user a lot of items that have been selected by other users with the selected user data item. In both tasks predict the possibility of user interest.

Solving these problems by using the method of mining association rules. Formal problem mining association rules as follows [9,10].

ISSN: 2278 – 1323

Let a set of items $\mathbf{I}=\{\mathbf{i}_1,\mathbf{i}_2,\ldots,\mathbf{i}_m\}$, a set of transactions $\mathbf{T}=\{\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_k\}$.

Each transaction is a binary vector, wherein $\mathbf{t}_i(l)=1$, if the object $\mathbf{i}_l$ is in transaction, otherwise $\mathbf{t}_i(l)=0$. Let $\mathbf{X},\mathbf{Y}$ some subset of the set $\mathbf{I}$ of objects. Then associative rule - is the implication $\mathbf{X}\Rightarrow\mathbf{Y}$. In this problem, a set of items is a set of items in domain recommender system. A set of transactions is a set of user profiles. In the user profile vector $\mathbf{u}_i(j)=1$, if the user *i* selects an item *j*, $\mathbf{u}_i(j)=0$ otherwise. In the initial matrix of user-item vectors user profiles contain ratings of the selected items. To represent the interests of users, each rating score is replaced by 1 (Fig.1).

	i_1	i_2	i_3	i_4	i_5
u_1	1	0	0	3	5
u_2	0	4	0	1	0
u_3	5	1	0	4	0
u_4	0	1	2	1	0

$\Longrightarrow$

	i_1	i_2	i_3	i_4	i_5
u_1	1	0	0	1	1
u_2	0	1	0	1	0
u_3	1	1	0	1	0
u_4	0	1	1	1	0

Fig.1 Transformation user-item matrix to user-interest to items matrix

The resulting matrix $\mathbf{U}=\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3,\mathbf{u}_4\}$ is a transaction vectors. If $\mathbf{u}_i(j)=1$ the user i is interested in the item j, if $\mathbf{u}_i(j)=0$ - not interested in the item j. In general statement of the problem for mining association rules $|\mathbf{X}|\geq 1, |\mathbf{Y}|\geq 1$. In this problem $|\mathbf{X}|=1, |\mathbf{Y}|=1$.

Solution to the problem is to find rules:

$$\textbf{if } \{\mathbf{i}_k\} \textbf{ then } \{\mathbf{i}_j,\ j=1(1)m,\ j\neq k\}. \quad (1)$$

The problem is solved in two stages [10]: 1) search all popular sets; 2) search association rules of the form (1) that meet certain criteria. These criteria are the support (2), confidence (3), lift (4), leverage (5). The input values for the algorithm are minimal support - minsupp, minimum confidence - minconf, N-number of top N rules.

$$\text{Sup }(\mathbf{X}\Rightarrow\mathbf{Y})=\frac{|\mathbf{X}\cup\mathbf{Y}|}{|\mathbf{T}|} \quad (2)$$

$$\text{Conf }(\mathbf{X}\Rightarrow\mathbf{Y})=\frac{\text{sup}(\mathbf{X}\Rightarrow\mathbf{Y})}{\text{sup}(\mathbf{X})} \quad (3)$$

$$\text{Lift }(\mathbf{X}\Rightarrow\mathbf{Y})=\frac{\text{conf}(\mathbf{X}\Rightarrow\mathbf{Y})}{\text{sup}(\mathbf{X})} \quad (4)$$

$$\text{Leverage}(\mathbf{X}\Rightarrow\mathbf{Y})=|\ \text{sup}(\mathbf{X}\Rightarrow\mathbf{Y})- -\text{sup}(\mathbf{X})\times\text{sup}(\mathbf{Y})\ | \quad (5)$$

The algorithm for solving the problem involves the following steps.

Step 1: In a variety of popular sets search rules that satisfy the condition

$$\text{Sup}\,(\mathbf{X}\Rightarrow\mathbf{Y})>\text{minsup} \quad (6)$$

Step 2. Exclude association rules that satisfy the following condition

$$\text{conf}\,(\mathbf{X}\Rightarrow\mathbf{Y})\leq\text{minconf} \quad (7)$$

Step 3. In the resulting set reserve rules that satisfy the condition

$$\text{lift}\,(\mathbf{X}\Rightarrow\mathbf{Y})>1 \quad (8)$$

Step 4. Select the top - N rules with a maximum value leveradge$(\mathbf{X}\Rightarrow\mathbf{Y})$ and form a matrix of item-item of associative links between objects.

Hypothetical example of such a matrix is shown in Fig. 2.

	i_1	i_2	i_3	i_4	i_5
i_1	0	1	0	1	0
i_2	0	0	0	1	0
i_3	1	1	0	0	0
i_4	0	0	1	0	0
i_5	0	0	0	1	0

Fig.2 Example of item-item association matrix

Rows of the matrix correspond to items that the user selects. Columns of the matrix correspond to the items will be recommended to the user. The item can not be recommended if the matrix element is 0.

The user logs in and selects an item $\mathbf{i}_j$. Then the system recommends items from i_j= 1.

III. FUZZY MODEL FOR RECOMMENDATION

Perspective is the application of methods and models of clustering. Clustering allows divide the whole set of users and products to compact clusters. Each cluster contains the most similar objects. General clustering model can be represented as follows: **K** - set of items and users; $\mathbf{K}_1$, $\mathbf{K}_2$, ... $\mathbf{K}_c$ - clusters.

For a hard clustering the following conditions are performed [12]:

$$\mathbf{K}_1\cup\mathbf{K}_2\cup\ldots\cup\mathbf{K}_c=\mathbf{K} \quad (9)$$

ISSN: 2278 – 1323

$$\mathbf{K}_i \cap \mathbf{K}_j = \varnothing;\ \mathbf{K}_\mathrm{i}, \mathbf{K}_\mathrm{j} \neq \varnothing \tag{10}$$

$$|\mathbf{K}_1| + |\mathbf{K}_2| + \ldots + |\mathbf{K}_c| = |\mathbf{K}| \tag{11}$$

With a hard clustering each object belongs to one cluster.

For a fuzzy clustering the following conditions are performed:

$$\mathbf{K}_1 \cup \mathbf{K}_\mathbf{2} \cup \ldots \cup \mathbf{K}_\mathbf{c} = \mathbf{K} \tag{12}$$

$$\mathbf{K}_i \cap \mathbf{K}_j \neq \varnothing;\ \mathbf{K}_\mathrm{i}, \mathbf{K}_\mathrm{j} \neq \varnothing\ \forall i, j \quad i \neq j \tag{13}$$

$$|\mathbf{K}_1| + |\mathbf{K}_2| + \ldots + |\mathbf{K}_c| = c|\mathbf{K}| \tag{14}$$

where c is the number of clusters.

Let Y - matrix partition of objects into clusters. For a hard clustering the following conditions are performed:

$$y_i(l) = y_{il} = \begin{cases} 1; l \in \mathbf{K}_i \\ 0; l \notin \mathbf{K}_i \end{cases} \tag{15}$$

$$\sum_{l=1}^{n} y_{il} > 0; \forall i \tag{16}$$

$$\sum_{i=1}^{c} y_{il} = 1; \forall l \tag{17}$$

For a fuzzy clustering the following conditions are performed [11,12,13]:

$$y_{il} \in (0,1) \tag{18}$$

$$0 < \sum_{l=1}^{n} y_{il} < n \tag{19}$$

$$\sum_{i=1}^{c} y_{il} = 1 \tag{20}$$

The classical model of collaborative filtering is to predict the unknown product rating for the active user by user-user or product-product model.

For a method user-user

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u \in \mathbf{U}} (r_{u,i} - \bar{r}_u) w_{a,u}}{\sum_{u \in \mathbf{U}} |w_{a,u}|} \tag{21}$$

where $p_{a,i}$ - rating i-th item for active users;

$\bar{r}_a$ - the average of the rating of the active user;

$\bar{r}_u$ - average rating value of the u-th user;

$r_{u,i}$ - rating of the i item for u user;

$w_{a,u}$ - coefficient of similarity for a rating vector of active user and rating vector of u user.

For the method of the item-item

$$p_{u,i} = \frac{\sum_{u \in \mathbf{U}} r_{u,n} w_{i,n}}{\sum_{u \in \mathbf{U}} |w_{i,n}|} \tag{22}$$

where $w_{i,n}$ - coefficient of similarity for a rating vector of i item and n item.

In fuzzy clustering, each product belongs to each cluster. Degree of membership of each product to fuzzy cluster is determined by the characteristic function

$$y_{il} \in (0,1),\ 0 < y_{il} < 1 \tag{23}$$

$$\sum_{i=1}^{c} y_{il} = 1 \tag{24}$$

Let Tl - vector of values characteristic function for l item. $\mathbf{M}_l = \mathbf{T}_l \times N$ - vector quantity of items with the highest ratings will be used in the forecast calculation formulas. M_{lc} - number of items from top-rated belonging to cluster c.

Then the model method the user-user (17) and the item-item (18)

$$p_{a,i} = \bar{r}_a + \frac{\sum_{m=1}^{c} \sum_{s=1}^{M_{as}} w_{a,s} (r_{s,i} - \bar{r}_s)}{\sum_{m=1}^{c} \sum_{s=1}^{M_{as}} |w_{a,s}|} \tag{25}$$

ISSN: 2278 – 1323

$$p_{u,i} = \frac{\sum_{m=1}^{c}\sum_{s=1}^{M_{as}} w_{u,s} r_{u,i}}{\sum_{m=1}^{c}\sum_{s=1}^{M_{as}} |w_{u,s}|} \qquad (26)$$

For clustering applied fuzzy c-means method, which minimizes the following functional

$$J_m(\mathbf{Y}, v) = \sum_{k=1}^{N}\sum_{i=1}^{c} (y_{ik})^m \|\mathbf{u}_k - \mathbf{v}_i\|_A^2 \qquad (27)$$

where $\mathbf{U} = \{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u}_N\}$-set of information objects; *c*-number of clusters; *m*- fuzziness factor; $\mathbf{Y}$ -fuzzy *c*-partition matrix; $\mathbf{v} = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_c}\}$-set of vectors of the centers of fuzzy clusters; $\mathbf{v_i} = (v_{i1}, v_{i2}, \ldots, v_{in})$ -vector coordinates i – th of cluster centers.

C-means fuzzy clustering algorithm consists of the following steps [13].

1.Choose a random number of initial cluster centers and placement

$$2 < c < N \qquad (28)$$

2.Choose option stops δ .
3.Choose a fuzziness factor $m \in (1, \infty)$.
4.Calculate the initial values of the matrix elements of fuzzy partition $\mathbf{Y^{(0)}}$.
5.Calculate the cluster centers

$$\mathbf{v_i}^{(l-1)} = \frac{\sum_{k=1}^{N} (y_{ik}^{(l-1)})^m \mathbf{u}_k}{\sum_{k=1}^{N} (y_{ik}^{(l-1)})^m},\ 1 \le i \le c \qquad (29)$$

6.Calculate the metric distance

$$d_{ik}^{(l-1)} = (\mathbf{u}_k - \mathbf{v}_i^{(l-1)})^T (\mathbf{u}_k - \mathbf{v}_i^{(l-1)}) \qquad (30)$$

7.Calculate the matrix elements of fuzzy partition

$$y_{ik}^{(l-1)} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}^{(l-1)}}{d_{jk}^{(l-1)}}\right)^{\frac{2}{(m-1)}}} \qquad (31)$$

8.Check the stop condition

$$\left\|\mathbf{Y}^{(l)} - \mathbf{Y}^{(l-1)}\right\| < \delta \qquad (32)$$

If the condition is met, then stop. Otherwise $\mathbf{Y}^{(l)} = \mathbf{Y}^{(l-1)}$, go to the step5.

Application of fuzzy clustering can improve the accuracy of forecasting rankings by taking into account degree of membership of each product to each cluster.

IV. SOLVING NEW USER AND NEW ITEM PROBLEMS

The main problems of the method of collaborative filtering: large dimension and large sparse user-item matrix, problem a new user and a new product.

Recommender system provides a forecast of a new item for the active user based on the known ratings of items that have already been given a new user to this. New user has no ratings. For a new user who is registered for the first time in the system, known only demographic information about it, such as: age, sex, occupation, country of residence, and education.

For a new item known only qualitative characteristics, such as for the film: the genre, the main actor, director, studio. Qualitative characteristics of objects and demographic characteristics of users are nominal categorical data. These data are qualitatively characterize the object, and do not have a numerical evaluation. They belong to the string data type. Only comparison operations is used when processing of such data "equal" and "not equal". Thus vectors qualitative characteristics of items and demographic characteristics of users in the future can be considered as a vector of categorical data elements. Let $\mathbf{D} = \{\mathbf{O}_1, \mathbf{O}_2, \ldots, \mathbf{O}_n\}$ - the set of categorical objects, $\mathbf{O}_i = \{x_{i_1}, x_{i_2}, \ldots, x_{i_m}\}$- i - th categorical object, x_{i_j} - categorical meaning A_{i_j} - th attribute object $\mathbf{O}_i$, $x_{i_j} \in dom(A_{i_j})$, $dom(A_{i_j})$ - the domain of attribute values A_{i_j} .

Categorical object $\mathbf{O}_i$ can be represented as a conjunction of attribute values

$$\mathbf{O}_i = [A_{i_1} = x_{i_1}] \wedge [A_{i_2} = x_{i_2}] \wedge \ldots \wedge [A_{i_m} = x_{i_m}] . \qquad (33)$$

Moreover, $\mathbf{O}_i = \mathbf{O}_j$ if $x_{i_j} = x_{k_j} \forall i = 1(1)m, \forall k = 1(1)m$.

Two categorical objects are equal only when they are equal categorical components.

The distance between two categorical *x* and *y* variables is defined as follows [14]

$$\delta(x, y) = \begin{cases} 1\, if\ x = y \\ 0\, if\ x \ne y \end{cases} \qquad (34)$$

ISSN: 2278 – 1323

Let $\mathbf{X}=(x_1, x_2, \ldots, x_m)$ and $\mathbf{Y}=(y_1, y_2, \ldots, y_m)$ are two categorical objects with m categorical attributes. Similarity measure between objects X and Y is defined as the number of matching categorical attribute values of two objects

$$d_{sim}(\mathbf{X},\mathbf{Y})=\sum_{i=1}^{m}\delta(x_i, y_i) \quad (35)$$

The contribution of each component $\delta(x_i, y_i)$ in $d_{sim}(\mathbf{X},\mathbf{Y})$ dependents on the frequency which meet the values of categorical variables x_i, y_i in the set of categorical objects. In the formula (36) account weights [15]

$$w_i=\frac{n_{x_i}+n_{y_i}}{n_{x_i}n_{y_i}}, \quad (36)$$

where n_{x_i}, n_{y_i} - the number of objects with categorical variables x_i, y_i in the set of categorical objects.

Categorical similarity measure objects with the weight coefficients

$$d_{sim}(\mathbf{X},\mathbf{Y})=\sum_{i=1}^{m}w_i\delta(x_i, y_i) \quad (37)$$

Mode of the sets of the categorical objects **D** is the vector $\mathbf{Q}=(q_1, q_2, \ldots, q_m)$ that minimizes the function

$$d(\mathbf{D},\mathbf{Q})=\sum_{i=1}^{n}d_{sim}(\mathbf{O}_i,\mathbf{Q}) \quad (38)$$

where $d_{sim}(\mathbf{O}_i,\mathbf{Q})$ is calculated by the formula (35) or (37).

Let the frequencies vector of categorical mode components is $\mathbf{F}_Q=(n_{q_1}, n_{q_2}, \ldots, n_{q_n})$ and frequencies vector of categorical object components is $\mathbf{F}_{O_i}=(n_{O_{i_1}}, n_{O_{i_2}}, \ldots, n_{O_{i_n}})$, where n_{q_j} - the frequency value categorical attribute of an mode vector, $n_{O_{i_j}}$ - the frequency value of the categorical attribute of an object $\mathbf{O}_i$ in a set of categorical objects **D**. To achieve the minimum functions is necessary condition $n_{q_j} > n_{O_{i_j}}$ at each step of the algorithm k-modes.

Let the system contains information about *n* users $\mathbf{U}=\{\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_n\}$ and *l* items $\mathbf{I}=\{\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_l\}$.

The task of the new user and a new item is solved in several stages:

1.Clustering of categorical objects of user or item. Suppose we have a *p* clusters of users or *q* clusters of objects

$$\mathbf{U}=\{\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_n\} \Rightarrow \mathbf{C_u}=\{\mathbf{C}_{\mathbf{u}_1}, \mathbf{C}_{\mathbf{u}_2}, \ldots, \mathbf{C}_{\mathbf{u}_p}\} \quad (39)$$

$$\mathbf{I}=\{\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_l\} \Rightarrow \mathbf{C_I}=\{\mathbf{C}_{\mathbf{I}_1}, \mathbf{C}_{\mathbf{I}_2}, \ldots, \mathbf{C}_{\mathbf{I}_q}\} \quad (40)$$

2.The new user or a new item allocate to the cluster to which distance to mode is maximal.

3.The components of the new user profile or a new object profile is prediction on the basis of user or object profiles corresponding cluster.

For clustering categorical objects used method of k-modes [15-18]:

Step 1. Select the k-modes for k clusters.

Step 2: Calculate the value of the similarity of each object and each categorical mod.

Step 3. Determine the accessory of each object to the cluster on the criterion of maximum similarity.

Step 4. Determine the new modes vector for each cluster.

Step 5. If change belonging the categorical objects to clusters, then go to step 2, otherwise - STOP.

Let the new object $\mathbf{X}=(x_1, x_2, \ldots, x_m)$ assigned to the cluster **C**. Then the components of the profile of the object is predicted by the following expression

$$x_i=\frac{\sum_{j=1}^{n}y_{i_j}}{s} \quad (41)$$

where x_i - *i*-th component of the profile vector of the new object; y_{i_j} - *j*-th component of the profile vector $\mathbf{Y}_i$ of the object in cluster **C**; *s* - the number of non-zero components in the vector $\mathbf{Y}_i$.

III. Conclusion

The paper analyzes the current state of development and application of recommendation systems, models and methods of construction of recommendation systems. It is shown that the most widely used method came into

ISSN: 2278 – 1323

collaborative filtering. Apply data mining techniques to predict the recommendations. The method of fuzzy clustering is developed, which improves the accuracy of predicting ratings of products. The method of association rules mining is developed, which solving the problem of search associative items for recommendations. The method of categorical clastering is developed, which solving the problem of new user and new item.

REFERENCES

[1] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," *IEEE Internet Computing*, pp.76-80.Jan./Feb., 2003.

[2] B.N. Miller, I. Albert, S.K. Lam, J.A. Konstan, and J. Riedl, "MovieLens Unplugged: Experiences with an Occasionally Connected Recommender System," *Proc. Intrn. Conf. Intelligent User Interfaces*, 2003. pp.263-266.

[3] D. Billsus, C.A. Brunk, C. Evans, B. Gladish, and M. Pazzani, "Adaptive Interfaces for Ubiquitous Web Access," *Comm. ACM*, vol. 45, no. 5, pp. 34-38, 2002.

[4] G. Adomavicius, A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowledge and Data Engineering*, vol. 17, pp.734-749, 2005.

[5] J.L.Herlocker, J.A.Konstan, A.L.Borchers, J.T.Riedl, "An algorithmic framework for performing collaborative filtering", in: *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*,1999,pp.230–237

[6] J.L.Herlocker, J.A.Konstan, J.T.Riedl, L.G.Terveen, "Evaluating collaborative filtering recommender systems", *ACM Transactions on Information Systems*, vol.22, no.1, pp. 5–53, 2004.

[7] L. Candillier, F. Meyer, M. Boullé, "Comparing state-of-the-art collaborative filtering systems", Lecture Notes in Computer Sciece, vol.4571, pp. 548–562, 2007.

[8] X. Su, T.M. Khoshgoftaar, Asurvey of collaborative filtering techniques, *Advance in Artificial Intelligence* , pp. 1–19, 2009.

[9] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases", in *Proc. 1993 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'93)*, Washington, DC, 1993, pp. 207–216.

[10] A. Gkoulalas-Divanis, V.S. Verykios *Association rule hiding for data mining.* New York: Springer, 2010.

[11] I. Gath, A.B. Geva, "Unsupervised optimal fuzzy clustering", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, pp.773-781, 1989.

[12] B.S. Everitt, S. Landau, M. Leese, D. Stahl *Cluster analysis.* London: John Wiley & Sosn, 2011.

[13] C. Bezdek, R. Ehrlich, W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers and Geoscience*, vol. 10, pp.191-203, 1984.

[14] M. K. Ng, M. J. Li, J. Z. Huang, and Z. He, "On the impact of dissimilarity measure in k-modes clustering algorithm", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, pp. 503-507, 2007.

[15] Z. Huang, M. K. Ng, "A note on k-modes clustering." *Journal of Classification*, vol.20, pp. 257-261, 2003.

[16] A. Chaturvedi, P. E. Green, and J. D. Caroll. "K-modes clustering", *Journal of Classification*, vol. 18, pp. 257-261, 2001.

[17] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values." *Data mining and knowledge discovery*, vol. 2, pp. 283-304, 1998.

[18] Z. Huang, "A fast clustering algorithm to cluster very large categorical data sets in data mining", in *Proceedings of the SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*, Canada, 1997, pp. 1–8.

Dr. Stekh Yuriy, PhD, Assistant Professor of CAD department of Lviv Polytechnic National University. Research interests: databases, recommender systems, artificial intelligence methods, data mining.

Prof. Mykhaylo Lobur, DSc, PhD, Academitian of Academy of Science in applied radioelectronics, Head of CAD department of Lviv Polytechnic National University. Research interests: design of CAD tools for microelectronic devices design.

Vitalij Artsibasov, Master of Computer Science. Research interests: recommender systems, artificial intelligence methods, data mining.

ISSN: 2278 – 1323