Acta Technica 62 (2017), No. 6B, 1157–1168 

# Intelligent City Planning and Construction based on Large Data Mining and Knowledge Discovery Technology

Huifang Ji[1]

**Abstract.** The planning and construction method of smart city based on fuzzy TOPSIS decision of dynamic Bayesian network is proposed to increase reasonableness of planning and construction of smart city and improve functional construction of smart city. Firstly, the dynamic Bayesian network is used to divide the factors involved in planning of smart city into three levels according to the analytic hierarchy process to build level model for construction of smart city, where the target decision-making level is the total assessment scores; the criterion level includes four factors and the sub-criterion level totally includes 17 factors. Secondly, the analysis method of fuzzy TOPSIS decision is used to analyze the level model of dynamic Bayesian network for planning and construction of smart city to obtain the model assessment results for use in assessment and guidance of planning and design process. Finally, the effectiveness of algorithm is verified by simulation experiment.

**Key words.** Bayesian network, Smart city, TOPSIS decision, Level model.

## 1. Introduction

In planning of smart city, the planning department and planners have always tried to assess the results by using artificial intelligence and use the quantitative method to analyze and deal with the problems previously resolved by using the qualitative method. However, the smart city is a huge system and there are so many development factors to be assessed. In addition, the difficulties exist in data collection and some development factors are applicable to be represented by numbers. Therefore, these reasons have impacts on the planning department and planners to make decision. The utilization of analytic hierarchy process (AHP) in planning of

[1]Beijing Institute of Control Engineering, Beijing 100190, P.R.China

smart city can support the planning department and planners to make decision and assess the scheme. But the key problem is to build the judgment matrix in analytic hierarchy process.

The previous method is that many experts are invited to give scores and then process these scores. Such method involves in many subjective human factors and opinions on some specific problems are difficult to be obtained from experts. Therefore, such method has limitations. The Fuzzy Dynamic Bayesian Network is used in spatial situation assessment in recent years. When the state information obtained in data collection is in fuzzy state and has dynamic uncertainty in time sequence, the impacts arising from change of information in the whole smart city planning system can be continuously detected and assessed by building FDBN assessment model to provide a more active and accurate quantitative analysis method and decision-making method for understanding, research and judgment of trend in smart city system planning.

The smart city planning and design method is proposed in this paper according to the analysis aspect of smart city system planning and based on fuzzy TOPSIS decision method of dynamic Bayesian network. In addition, the systematic assessment method is established for quantitative components in smart city planning to increase the reasonableness of smart city planning assessment.

## 2. Level model on planning of smart city

### 2.1. Design index

The factors involved in planning of smart city are divided into three levels (as shown in Fig. 1) according to the analytic hierarchy process, where the target decision-making level is the total assessment scores; the criterion level includes four factors and the sub-criterion level totally includes 17 factors. Taking the criterion level as an example, the judgment matrix is built, namely, the score level of importance for each two factors is given.

According to human's thinking ability, it is easier to judge the extreme case than the middle state. Therefore, the experts are easier to reach an agreement in this regard. In consideration of these cases, it is meaningful to build judgment matrix in analytic hierarchy process by using Bayesian network. The Bayesian network consists of three layers, including one input layer (including two nerve cells), one hidden layer (including three nerve cells) and one output layer (including one nerve cell). The extreme case is used as input for Bayesian network training. When the Bayesian network training is completed, the physical case input can obtain factor value of judgment matrix in analytic hierarchy process. For example, assuming that when the value of location condition is 9 (upper limit) but the value of construction condition is 1 (lower limit), the output is 7; when the value of location condition is 1 but the value of construction condition is 9, the output is 4 etc. These cases constitute input and output vectors.

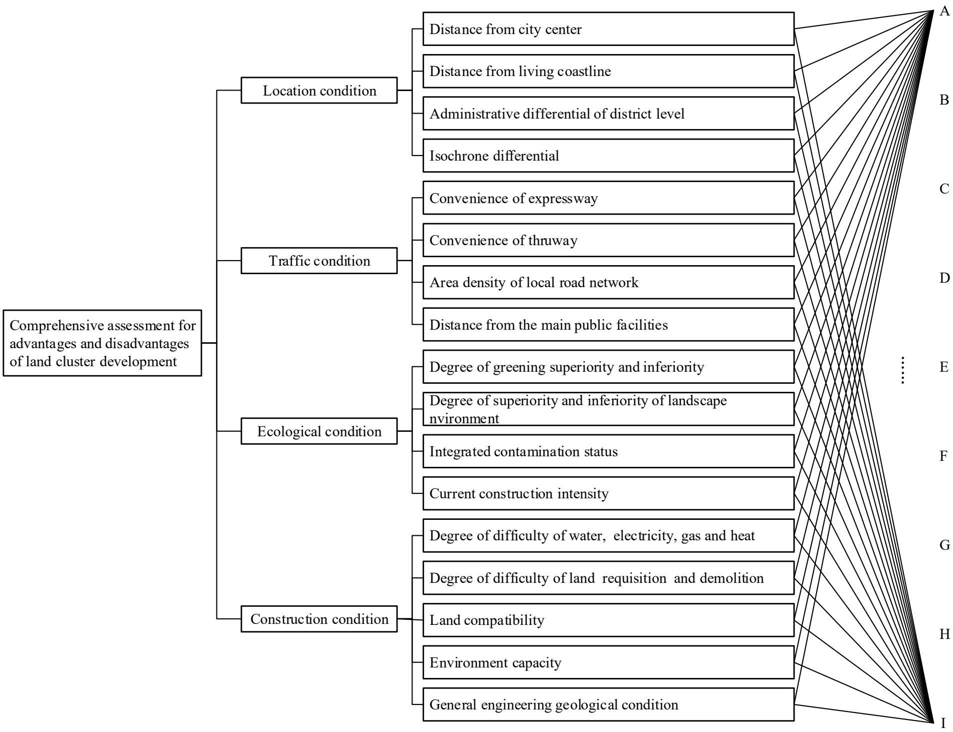


Fig. 1. Hierarchical structure of planning factors

## 2.2. Dynamic Bayesian network model

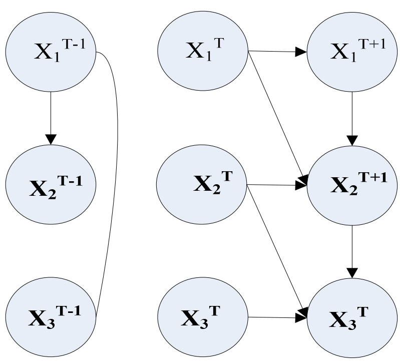


Fig. 2. Dynamic bayesian network structure

The dynamic Bayesian network (DBN) is the time sequence extension of Bayesian network (BN), which has the functional characteristics of static Bayesian network and more accurately reflects the impacts of sample data on network structure in time sequence. Such method is applicable to estimate the impacts of trend factors from smart city planning trend assessment on entire planning system. The time sequence causal relation between adjacent time slices is integrated with the causal relation in the same time slice. The dynamic analysis is conducted through quantitative reasoning. The dynamic Bayesian network can be simply defined as $(B_0,\ B_{\rightarrow})$, where $B_0$ is the BN at $T_0$ (time slice in original state). The prior probability $P\left(X_0\right)$ of hidden node and observation node can be obtained. $B_{\rightarrow}$ is the graph formed by BN in each time slice.

DBN has the following functions, including new knowledge integration, complete expression, deduction and learning. When the modeling analysis is conducted for the uncertain problems with random process characteristics, DBN has favorable effects. The DBN structure is shown in Fig. 2.

### *2.3. Reasoning algorithm of dynamic Bayesian network*

The reasoning algorithm of dynamic Bayesian network is deduced from formula (1) Bayesian formula:

$$p(x\,|y) = \frac{p(yx)}{p(y)} = \frac{p(yx)}{\sum\limits_{x} p(yx)}\,. \tag{1}$$

Its reasoning process is actually same with the reasoning process of static Bayesian network. For the disperse static Bayesian network with $n$ hidden nodes and $m$ observation nodes, its reasoning theory can be reflected as the mathematics process in formula (2) based on condition independence:

$$\begin{gathered} p(x_1, x_2, ..., x_n\,|y1, y2, ..., y_m) = \\ \frac{\prod\limits_{j} p(y_j\,|p_a(Y_j)) \prod\limits_{i} p(x_i\,|p_a(X_i))}{\sum\limits_{x_1,x_2,\ldots,x_n} \prod\limits_{j} p\left(yj\,|p_a(Yj)\right) \prod\limits_{i} p\left(x_i\,|p_a(X_i)\right)} \\ i \in [1, n]\,, j \in [1, m]\ . \end{gathered} \tag{2}$$

Where, $x_i$ refers to one state value of $X_i$; $p_a(Y_j)$ refers to parent nodes collection of $Y_j$.

When there are fewer hidden nodes and observable nodes in the network, or stronger node coupling, fewer network structure layers and fewer time slices to be considered, each time slice of DBN can be seen as one static Bayesian network. When the node is gradually increased or node coupling is strengthened, DBN consisting of $T$ time slices can be obtained in time domain. Its reasoning process can be reflected as formula (3):

$$p(x_{11},...,x_{1n},...,x_{T1},...,x_{Tn}\,|Y_{11o},Y_{12o},...,Y_{1mo}\ ,...,Y_{T1o},Y_{T2o},Y_{Tmo}) =$$

$$\sum_{y_{11}y_{12}\cdots y_{Tm}} \frac{\prod\limits_{i,j} p(y_{ij}\,|p_a(Y_{ij}))\prod\limits_{i,k} p(x_{ik}\left|p_a(X_{ik}))\prod\limits_{i,j} p(Y_{ijo}=y_{ijo})\right.}{\sum\limits_{x_{11},x_{21},...,x_{T1}...x_{Tn}}\prod\limits_{i,j} p\left(y_{ij}\,|p_a(Y_{ij})\right)\prod\limits_{i,k} p\left(x_{ik}\,|p_a(X_{ik})\right)} \tag{3}$$

$$i\in[1,T]\,,j\in[1,m]\,,K\in[1,n]$$

Where, $x_{ij}$ refers to one state value of $X_{ij}$; $i$ refers to the time slice; $j$ refers to the hidden node; $y_{ij}$ refers to value of observational variable $Y_{ij}$; $p_a(Y_{ij})$ refers to parent nodes collection of $y_{ij}$; $Y_{ijo}$ refers to the observation state of observation node $j$ in $i$ time slice; $p(Y_{ijo}=y_{ijo})$ refers to the membership degree of continuous observation value $Y_{ij}$ belonging to $y_{ij}$.

## 3. Fuzzy TOPSIS decision analysis

### *3.1. Algorithm description*

The TOPSIS weight criterion $W_C$ is generally calculated based on the preference of users or experts. But under actual application environment, such weight value cannot be accurately assigned. The information shortage may exist. In this regard, the fuzzy decision analysis method based on TOPSIS is proposed. The specific calculation process is shown as below:

**Step 1:** assuming that $m$ group of available factors are involved in the smart city planning, the decision criterion is $C_j\ (j=1,2,\cdots,n)$, the decision matrix of smart city planning process is shown as below:

$$X = \begin{array}{c} \\ S_1 \\ S_2 \\ \vdots \\ S_m \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \cdots & C_n \end{array} \\ \left[\begin{array}{cccc} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{array}\right] \end{array}. \tag{4}$$

In formula (2), $x_{ij}$ refers to the quantitative value of smart city planning $S_i$ to evaluation criterion $C_j$.

**Step 2:** the entropy value is used for weight assignment of decision criterion in smart city planning process and to calculate projection $P_{ij}$ of decision matrix to criterion $C_j\ (j=1,2,\cdots,n)$:

$$P_{ij} = x_{ij}\Big/\sum\nolimits_{i=1}^{m} x_{ij}\,. \tag{5}$$

Where, the entropy value calculation method is shown as below:

$$e_j = -\left(\ln m\right)^{-1} \cdot \sum_{j=1}^{n} p_{ij} \ln p_{ij} \,. \tag{6}$$

The criterion weight in smart city planning is as below:

$$W_{C_j} = \left(1 - e_j\right) \Big/ \sum\nolimits_{k=1}^{n} \left(1 - e_k\right). \tag{7}$$

**Step 3:** the fuzzy TOPSIS decision-making process for smart city planning is as below:

$$\tilde{R} = \left[\tilde{r}_{ij}\right]_{m \times n} \tag{8}$$

In the process of smart city planning, the planning process influence factors should be divided when the decision is made due to great changes of planning process, including functional correlation ($F$) and price correlation ($C$). the $(a_{ij}, b_{ij}, c_{ij})$ triangle fuzzy form is used for fuzzy number:

$$\begin{cases} \tilde{r}_{ij} = \left(\dfrac{a_{ij}}{c_j^+}, \dfrac{b_{ij}}{c_j^+}, \dfrac{c_{ij}}{c_j^+}\right), & if \ j \in F \\ \tilde{r}_{ij} = \left(\dfrac{a_j^-}{c_{ij}}, \dfrac{a_j^-}{b_{ij}}, \dfrac{a_j^-}{a_{ij}},\right), & if \ j \in C \end{cases} \tag{9}$$

In formula (7), the relevant parameters involved include the following:

$$\begin{cases} c_j^+ = \max c_{ij} \,, & if \ j \in F \\ a_j^- = \min a_{ij}, & if \ j \in C \end{cases} \tag{10}$$

**Step 4:** the fuzzy decision weight referred in step 3 is calculated based on step 2 to obtain decision judgment matrix:

$$\tilde{V} = \begin{bmatrix} \tilde{v}_{11} & \tilde{v}_{12} & \cdots & \tilde{v}_{1n} \\ \tilde{v}_{21} & \tilde{v}_{22} & \cdots & \tilde{v}_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{v}_{n1} & \tilde{v}_{n2} & \cdots & \tilde{v}_{nn} \end{bmatrix} = \begin{bmatrix} \tilde{r}_{11} & \tilde{r}_{12} & \cdots & \tilde{r}_{1n} \\ \tilde{r}_{21} & \tilde{r}_{22} & \cdots & \tilde{r}_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{r}_{m1} & \tilde{r}_{m2} & \cdots & \tilde{r}_{mn} \end{bmatrix} \cdot diag\left\{W_{C_1}, \cdots W_{C_n}\right\}. \tag{11}$$

**Step 5:** based on the assessment matrix obtained in the last step, the ideal solution $A^+$ and $A^-$ can be obtained:

$$\begin{cases} A^+ = \left(\tilde{v}_1^+, \tilde{v}_2^+, \cdots, \tilde{v}_n^+\right) \,. \\ A^- = \left(\tilde{v}_1^-, \tilde{v}_2^-, \cdots, \tilde{v}_n^-\right) \,. \end{cases} \tag{12}$$

**Step 6:** the distance is calculated for fuzzy number $A_1 = (a_1, b_1, c_1)$ and $A_2 =$

$(a_2, b_2, c_2)$:

$$d(A_1, A_2) = \sqrt{\frac{1}{3}\left[(a_1 - a_2)^2 + (b_1 - b_2)^2 + (c_1 - c_2)^2\right]}\,.$$

$$\begin{cases} d_i^+ = \sum_{j=1}^{k} d\left(\tilde{v}_{ij}, \tilde{v}_j^+\right), i = 1, 2, \cdots m\,, \\ d_i^- = \sum_{j=1}^{k} d\left(\tilde{v}_{ij}, \tilde{v}_j^-\right), i = 1, 2, \cdots m\,. \end{cases} \tag{13}$$

**Step 7:** the characteristic vector of judgment matrix can be obtained in planning process:

$$\omega_i = \frac{d_i^-}{d_i^+ + d_i^-}\,. \tag{14}$$

### *3.2. Comprehensive analysis*

The relevant characteristic vector $\omega$ and maximum characteristic root $\gamma$ in judgment matrix $R$ are solved. Then the characteristic vector $\omega = (\omega_1, \omega_2, \cdots, \omega_n)^T$ solved is uniformly operated, namely, $\bar{\omega} = \omega_i / \sum_{j=1}^{n} \omega_j$. The characteristic vector of secondary index referred in Table 2 is used for operation and calculation with primary index.

The verification and planning forecasting is conducted based on consistent verification index $CI = (\lambda_{\max} - n)/(n - 1)$, consistent random verification index $RI$ and consistent verification ratio index $CR = CI/RI$. If the result shows $CR < 0.1$, it indicates that the judgment matrix is in line with consistency index; if the result shows $CR \geq 0.1$, the different judgment matrixes are required to be compared and adjusted. The index value standardization calculation is made for mart city planning model. The standardization calculation is firstly made for direct ratio and inverse ratio index.

Direct ratio index:

$$y_{ij} = \frac{x_{ij} - \min x_{ij}}{\max x_{ij} - \min x_{ij}}\,. \tag{15}$$

Inverse ratio index:

$$y_{ij} = \frac{\max x_{ij} - x_{ij}}{\max x_{ij} - \min x_{ij}}\,. \tag{16}$$

In formula 13-14, $y_{ij}$ refers to the index value after standardization; $x_{ij}$ refers to the initial value of index $i$ for component $j$. $\max x_{ij}$ and $\min x_{ij}$ respectively refer to minimum and maximum planning process value $(1 \leq i \leq 7, 1 \leq j \leq 8)$ of index $x$ in sequence. The standardization calculation is made for each planning influence factor to obtain the influence factor of component to the construction planning.

$$D_j = \sum_{i=1}^{12} p_i y_{ij} \,. \tag{17}$$

$$Z = \sum_{i=1}^{8} D_i \sum_{j=1}^{8} \sum_{i=1}^{12} p_i y_{ij} \,. \tag{18}$$

### 3.3. Algorithm steps

The steps of fuzzy TOPSIS decision analysis algorithm adopted in planning process assessment are shown as below:

**Step 1:** the layer-based model is used to build process planning model under different scenarios. The planning characteristic under certain scenario is selected to classify the characteristics according to time to obtain the planning characteristic in independent time period. Then the judgment criterion $C_j$ $(j = 1, 2, \cdots, n)$ is selected for analysis based on fuzzy TOPSIS decision;

**Step 2:** the weight of entropy value is assigned for the decision criterion in smart city planning process based on the experts' judgment to obtain the judgment criterion $C_j$ $(j = 1, 2, \cdots, n)$. The specific steps are detailed in step 1-2 of Section 4.1;

**Step 3:** the decision matrix is built for the planning process characteristics in different time periods. The relatively reasonable planning characteristic vector is obtained based on fuzzy TOPSIS decision. As the planning scenario satisfies independence, the fuzzyTOPSIS analysis process can be calculated in machine in parallel manner;

**Step 4:** the characteristic vector obtained can be normalized. Then consistency verification is made to obtain its direct ratio and inverse ratio indexes and the influence factors of components on construction planning.

## 4. Experimental analysis

### 4.1. Model specification

The assessment factors are divided into three levels based on the analytic hierarchy process. The planning scheme is given on the criterion level. The judgment matrix consisting of these four factors must pass the consistency check. Such matrix is shown in Table 1:

The factors in Table 1 are 1-9 scales. The value obtained by Bayesian network is not the whole number, but the significance of their whole numbers is shown in Table 2: the final assessment results for this level can be obtained by the operation process of analytic hierarchy process when the judgment matrix is built. The sub-criterion level has 17 factors, where four factors determine the location condition, four factors determine traffic condition, four factors determine the environmental condition and the last five factors determine the construction condition. Thus four judgment matrixes are required to be built to respectively correspond to four factors in criterion level. Similarly, such complicated multiple-target system is decomposed into comparison problems between each two factors due to errors and mutual effect in planning process assessment.

Table 1. Condition setting

| Condition | Location condition | Traffic condition | Environmental condition | Construction condition |
|---|---|---|---|---|
| Location condition | 1 | Model specification | Model specification | Model specification |
| Traffic condition | Model specification | 1 | Model specification | Model specification |
| Environmental condition | Model specification | Model specification | 1 | Model specification |
| Construction condition | Model specification | Model specification | Model specification | 1 |

Table 2. Condition setting

| Scale | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|
| Meaning | Extremely good | Very good | Pretty good | Good | Same | Bad | Pretty bad | Very bad | Extremely bad |

### *4.2. Result analysis*

Yantai is located in Shandong. Its planning area is 40 square kilometers. The whole region is divided into 9 parts. These 9 parts are respectively assessed.

Table 3 is the judgment matrix of criterion level obtained from Bayesian network. Table 4 is one of 4 judgment matrixes of sub-criterion level obtained from Bayesian network, where the last column indicates weight value. Table 5 shows the final results, where the last line indicates scores of 9 parcels.

Table 3. Model calculation result

| Condition | Location condition | Traffic condition | Environmental condition | Construction condition |
|---|---|---|---|---|
| Location condition | 1 | 6.33 | 7.32 | 4.66 |
| Traffic condition | 0.16 | 1 | 5.69 | 0.25 |
| Environmental condition | 0.14 | 0.18 | 1 | 0.36 |
| Construction condition | 0.21 | 4.01 | 2.78 | 1 |

Table 4. Sub-criterion weight value calculation result

| Environmental condition | Greening | Landscape | Pollution | Construction density | Weight value |
|---|---|---|---|---|---|
| Greening | – | 0.5 | 0.498 | 0.367 | 0.226 |
| Landscape | 0.5 | – | 0.465 | 0.444 | 0.234 |
| Pollution | 0.502 | 0.536 | – | 0.5 | 0.258 |
| Construction density | 0.633 | 0.557 | 0.5 | – | 0.284 |

Table 5.

| Parcel | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Location condition | 8.75 | 7.05 | 7.20 | 5.52 | 6.06 | 6.25 | 4.72 | 4.71 | 3.18 |
| Traffic condition | 8.24 | 5.52 | 7.51 | 7.23 | 8.01 | 6.24 | 5.21 | 6.23 | 5.48 |
| Environmental condition | 6.87 | 5.24 | 7.43 | 6.48 | 8.25 | 7.24 | 8.76 | 6.01 | 6.32 |
| Construction condition | 3.98 | 5.42 | 8.14 | 6.35 | 7.82 | 7.41 | 7.26 | 6.54 | 7.57 |
| Score | 7.48 | 6.38 | 7.46 | 5.94 | 6.81 | 6.58 | 5.64 | 5.37 | 4.64 |

We hope to establish an expert assessment system for the smart city planning by using the current and collected data and train such Bayesian network by using the data subsequently collected. This expert assessment system will give great assistance for planning of smart city.

## 5. Conclusion

The planning and construction method of smart city based on fuzzy TOPSIS decision of dynamic Bayesian network is proposed for assessment and guidance of smart city panning and construction process. The dynamic Bayesian network is used to divide the factors involved in planning of smart city into three levels to build level model for construction of smart city; the analysis method of fuzzy TOPSIS decision is used to analyze the level model of dynamic Bayesian network for planning and construction of smart city to obtain the model assessment results. The experimental result shows that the proposed method can effectively classify and assess the model and has certain guiding significance. In the next step, the algorithm application system will be designed and developed and the actual planning design process will be selected for further verification of algorithm.

### References

[1] Heckerman D: (1997) *Bayesian networks for data mining. Data mining and Knowledge Discovery*[J]. Data Mining & Knowledge Discovery, 1(1):79-119.

[2] Fayyad U M: (1996) *Data Mining and Knowledge Discovery: Making Sense Out of Data*[J]. IEEE Expert Intelligent Systems & Their Applications, 11(5):20-25.

[3] Goebel M, Le G: (1999) *A survey of data mining and knowledge discovery software tools*[M]. ACM, 1999.

[4] Cannataro M, Congiusta A, Mastroianni C, et al.: (2004) *Grid-Based Data Mining and Knowledge Discovery*[J]. Intelligent Technologies for Information Analysis, 8190(1):19-45.

[5] Pahlavan K, Krishnamurthy P and Geng Y: (2015) *Localization challenges for the emergence of the smart workd.* IEEE Access, 3(1), pp. 3058-3067

[6] Lv, Z., Tek, A., Da Silva, F., Empereur-Mot, C., Chavent, M., & Baaden, M. (2013).: *Game on, science-how video game technology may help biologists tackle visualization challenges.* PloS one, 8(3), e57990.

[7] Weisen Pan, Shizhan Chen, Zhiyong Feng: (2013) *Automatic Clustering of Social Tag using Community Detection.* Applied Mathematics & Information Sciences, 7(2): 675-681.

[8] Yingyue Zhang, Jennifer W. Chan, Alysha Moretti, and Kathryn E. Uhrich: (2015) *Designing Polymers with Sugar-based Advantages for Bioactive Delivery Applications*, Journal of Controlled Release, 219, 355-368.

[9] Yingyue Zhang, Qi Li, William J. Welsh, Prabhas V. Moghe, and Kathryn E. Uhrich: (2016) *Micellar and Structural Stability of Nanoscale Amphiphilic Polymers: Implications for Anti-atherosclerotic Bioactivity*, Biomaterials, 84, 230-240.

[10] Jennifer W. Chan, Yingyue Zhang, and Kathryn E. Uhrich: (2015) *Amphiphilic Macromolecule Self-Assembled Monolayers Suppress Smooth Muscle Cell Proliferation*, Bioconjugate Chemistry, 26(7), 1359-1369.

[11] Preuss M, Dehmer M, Pickl S, et al.: (2014) *On Terrain Coverage Optimization by Using a Network Approach for Universal Graph-Based Data Mining and Knowledge Discovery*[M]// Brain Informatics and Health. Springer International Publishing, 2014:564-573.

[12] Chi Y, Wang H, Yu P S: (2009) *System and method for load shedding in data mining and knowledge discovery from stream data: US*, US7493346[P].

[13] Manesis S, Deligiannis V, Koutri M: (2008) *Data Mining and Knowledge Discovery for Monitoring and Intelligent Control of a Wastewater Treatment Plant*[C]// Icinco 2008, Proceedings of the Fifth International Conference on Informatics in Control, Automation and Robotics, Intelligent Control Systems and Optimization, Funchal, Madeira, Portugal, May. DBLP, 2008:86-93.

[14] Li Y, Thomas M A, Osei-Bryson K M: (2016) *Ontology-based data mining model management for self-service knowledge discovery*[J]. Information Systems Frontiers, 2016:1-19.

Received May 7, 2017