International Journal of Innovative Computing, Information and Control
Volume **9**, Number **1**, January **2013**
ICIC International ©2013 ISSN 1349-4198
pp. **47–67**

LEARNING AND REASONING ON BACKGROUND NETS FOR TEXT CATEGORIZATION WITH CHANGING DOMAIN AND PERSONALIZED CRITERIA

Sio-Long Lo[1] and Liya Ding[1,2]

[1]Faculty of Information Technology
Macau University of Science and Technology
Avenida Wai Long, Taipa, Macau SAR, P. R. China
akennetha@gmail.com; ding.liya@nus.edu.sg

[2]Institute of Systems Science
National University of Singapore
25 Heng Mui Keng Terrace, Singapore 119615

Received August 2011; revised January 2012

Abstract. *A term or word in text is not only a symbol but also with rich semantic meanings that are dependent on the domain of content as well as reader's personal experience. Therefore, understanding of the concept of a term forms a key task in text categorization especially when personalization is required. This article proposes a novel approach for personalized article selection as a special case of general text categorization by capturing user's interests with a long-term knowledge background accumulated through incremental learning. The learning is achieved by using the contextual association of terms that appear in sample articles provided. Knowledge background is represented as weighted undirected graph called background net that captures contextual associations among the terms in the recommended articles. With a background net constructed, the understanding of the concept of a term is personalized to a fuzzy set based on contextual associations of the given term to other terms involved in the corresponding background net. The similarity and acceptance measures are defined to evaluate candidate article through the reasoning on background net. One key and also unique feature of our approach is the use of association of terms rather than individual terms. The set of terms used here is naturally expending during the incremental learning rather than a predefined and fixed one as in other methods, which greatly relieves the pressure of feature selection for dimension reduction. The algorithms of reasoning on background net have been proposed based on a "focus spreading" spirit, so computational complexity can be well controlled when the entire background net is expending. This approach applies not only to personalized articles selection but also to text categorization.*

Keywords: Background net, Similarity and acceptance measure, Incremental learning, Personalized article selection, Text categorization

1. **Introduction.** The rapid increasing information distributed throughout the Internet has raised a new issue in finding appropriate information from a huge amount of data [1], which often forces us spending a lot of time for filtering relevant results. Personalized document selection has attracted more attention recently for improving the quality of searching results based on the idea of personalization [2-7]. On the other hand, machine learning approaches have been applied in capturing and learning relatively stable and long-term criteria for information retrieval and classification, such as fuzzy data retrieval [8], and text categorization [9-12].

One common and useful approach applied in classical techniques for searching documents, articles and web pages, is using a provided set of keywords [13,14] under the well-known vector space model (VSM) [15-17]. In VSM, a document or a keyword set is represented as a feature vector in a universal feature space. The task of document retrieval or text categorization is then considered a process of computing the similarity between feature vectors, and the result returned is the most similar documents based on the criteria provided. In this model, the features obtained through statistical methods can be insufficient in representing documents with rich semantic information.

In order to represent a document more accurately matching its content in the form of feature vector, several techniques from information science and machine learning areas have been proposed. Term weighting technique is a statistical method used to evaluate how important a term is to a document in a collection [9,18,19]. Feature selection and extraction is often done for text categorization, which is a process of transforming the original feature space to another feature space through extracting important features [9,16,20]. On the other hand, selecting an appropriate set of features remains a difficult task [21], and either a too small or a too large set of features will often lead to a poor performance.

Article selection based on personal preference can be considered as a special and often more challenging case of text categorization, where multiple categories with possible overlap are indicating different levels of satisfaction of articles according to personal reading preference, and categorization criteria are typically ill-defined and with a lack of explicit description. It is agreed that article selection with personal preference is an open domain problem because new articles can be added into archive and the focus of reading interests can also be updated from time to time. With a fixed set of features determined, those feature vector based techniques are obviously incapable in handling such an open domain.

Some daily life phenomenon is observed that people with similar knowledge backgrounds tend to have a more effective communication. Such an advantage comes from their similar understanding of concepts involved in communication. When focusing our discussion on text media, a concept is usually represented as a symbolic term, sometimes also called *word*. Different from applications in pattern recognition and image processing, though a word appears in an article as a precise symbolic term, its meaning can be fuzzy. So, it is necessary to get semantic information behind the term for the concept understanding.

In short, the new challenge is mainly from two aspects: 1) how to capture the semantics of a concept in an open domain; 2) how to reflect a changing personal reading interest described by terms of concepts. Handling an open domain means it is almost impossible to extract a complete semantics of a term from a linguistic point of view. On the other hand, considering a changing personal reading interest makes it difficult to learn personal background through classical batch learning on a limited set of samples.

In this article, we propose a novel approach to handle text categorization with changing domain and personalized criteria. Being in mind that we are handling an open domain problem and facing a huge amount of data, our intension is to achieve a simple representation, an easy processing and a robust performance. The key ideas that distinguish our approach from other representative methods in text categorization (such as graph-based model [22-25], fuzzy set model [26,27], or fuzzy-rough hybrid approach [28]) are 1) using a freely expending set of words together with the associations between words to represent article, instead of a predetermined fixed set of features, so be able to avoid the tedious and time-consuming processing of feature selection and also better cope with an open and a changing domain of content; 2) accumulating simple appearance of words as an easy indicator for association of words to relieve the demand of large sample data for construction of classifiers; 3) using a graphic representation with an expending set of vertices, called *background net* to represent an accumulation of background information of a domain which can either be a specific content domain or an individual's personal interest; 4) applying different measures in evaluating symmetrical similarity and unsymmetrical acceptance to serve different purposes of article selection, which helps reduce the impact of irrelevant information from background domain and so to increase the robustness; 5) carrying out inference on background net in a "focus spreading" manner rather than having processing on the entire network, so to ensure a reasonable complexity of computation. Combing all the key points above, the proposed approach is novel.

The rest of the article is organized as follows. Section 2 gives an introduction to background net with its representation and learning algorithm. Section 3 discusses the acceptance of an article based on given background information represented by a background net, and defines similarity and acceptance measures for evaluation purpose. Section 4 provides algorithms of applying background nets in both text categorization and personalized article selection, with the discussion of algorithm complexities. Experimental results of our approach with performance analysis and comparison to several representative methods are shown in Section 5. Section 6 highlights and discusses some key issues in text categorization. Finally, conclusions are given in Section 7.

2. **Construction of Background Net.**

2.1. **Representation of background net.**

For a given set of articles from a collection, there should be some useful information for understanding the knowledge background of the collection. *Background net* [29,30] has been proposed to represent contextual association between terms in articles, so to capture the background information not only for a domain but also an individual's personal reading interest. The basic idea of background net is inspired by the understanding of how human accumulates experience: a concept related to a background domain gets stronger association to the particular knowledge background through an accumulation of relevant readings. A general definition of background net is given in Definition 2.1. It applies to a single article as well as a set of articles that contains the former as a special case.

Definition 2.1. *A background net $N =< V, E >$ is a weighted undirected graph, with vertex set V representing a set of terms or association of terms as a vertex:*

$$V = \{v_i | v_i = Symbol(term_i), term_i \in W, i = 1, \dots, q\} \tag{1}$$

where W is a set of q terms obtained from article(s), and $Symbol(term_i)$ is the symbolic representation of $term_i$; edge set E represents the relation indicating contextual association between two terms:

$$E = \{e_{i,j} | e_{i,j} = (v_i, v_j), v_i, v_j \in V, i, j = 1, \dots, q, i \neq j\} \tag{2}$$

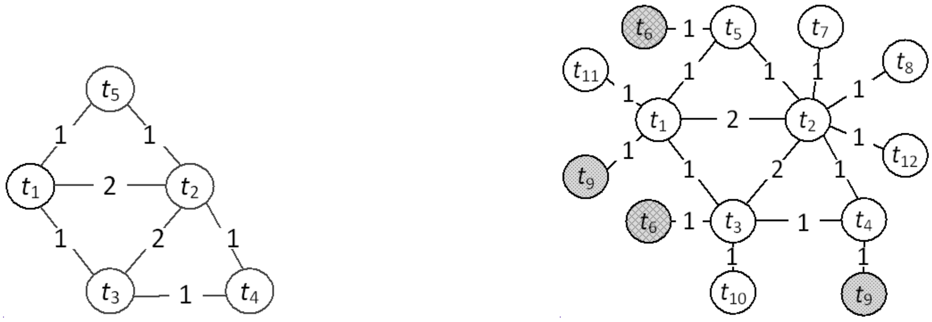
where each edge $e_{i,j}$ is associated with a weight $w(e_{i,j}) = w(e_{j,i}) = Count(v_i, v_j), w_{i,j}$ for short, defined as the number of occurrence of the association of v_i and v_j in basic units of article(s).

It is worthy to mention that Definition 2.1 does not provide a specific guideline for the partition of article for capturing contextual association between v_i and v_j. Following the normal practice in document processing [15,16], sentence is used as a basic unit to partition an article in our current development, but other methods may also be possible.

Example 2.1. *Given an article A_1 that contains three sentences:*

$$A_1 : \quad w_1, w_2, w_3 \,|\, w_2, w_3, w_4 \,|\, w_1, w_2, z_1$$

Corresponding background nets of article A_1 under different considerations are shown in Figure 1: (a) considering the contextual association between two terms; (b) considering



(a) (b)

$t_1 = w_1$, $t_2 = w_2$, $t_3 = w_3$, $t_4 = w_4$, $t_5 = z_1$, $t_6 = w_1 - w_2$, $t_7 = w_1 - w_3$,
$t_8 = w_1 - z_1$, $t_9 = w_2 - w_3$, $t_{10} = w_2 - w_4$, $t_{11} = w_2 - z_1$, $t_{12} = w_3 - w_4$

FIGURE 1. Background nets of Example 2.1

the contextual association between two and three terms, where $w_i - w_j$ denotes the group of two terms w_i and w_j as a unit represented by one vertex in background net.

We note that an association of terms has richer information than isolated terms. An English word uses multiple characters following certain rules for representing meaning that is often not unique and fuzzy. Multiple words appeared as an expression or elements of sentences or phrases assign constraints to words to provide more specific or complicated meanings. Background net proposed is to represent the contextual association between terms. It should be made clear that background net is different from Semantic net for representing semantic meaning between terms under certain domain knowledge defined by knowledge engineers [31], and is also different from latent semantic indexing (LSI) for the purpose of feature extraction in information retrieval through a terms-documents matrix [9,15,16].

2.2. **Learning on background net.** Given a collection of articles with each article associated with a non-zero degree indicating how much it belongs to the collection, the learning process is carried out in an incremental manner using the sample articles from the collection one by one. It is to merge the information of a new article to the category's background net representing the information of previously learned articles. Initially a category's background net is set to be $N^{(c)} =< \varnothing, \varnothing >$. A single article for learning is first represented as a background net $N^{(a)}$.

Assume category's background net $N^{(c)} =< V^{(c)}, E^{(c)} >$ and the background net representing a newly article $N^{(a)} =< V^{(a)}, E^{(a)} >$ with a preference degree $\lambda^{(a)} \in (0, 1]$. After learning, the updated category's background net is $N^{(new)} =< V^{(new)}, E^{(new)} >$, where $V^{(new)} = V^{(c)} \cup V^{(a)}$, and the weight $w_{s,t}^{(new)}$ of each edge $e_{s,t}^{(new)} \in E^{(new)} = E^{(c)} \cup E^{(a)}$ for $v_s^{(new)}$, $v_t^{(new)} \in V^{(new)}$ can be determined by (3)

$$w_{s,t}^{(new)} = \frac{k \times w_c\left(e_{s,t}^{(new)}\right) + \left(\lambda^{(a)}\right)^{1-\mu\left(e_{s,t}^{(new)}\right)} \times w_a\left(e_{s,t}^{(new)}\right)}{k+1} \tag{3}$$

where

$$w_c\left(e_{s,t}^{(new)}\right) = \begin{cases} w_{s,t}^{(c)} & e_{s,t}^{(new)} \in E^{(c)} \\ 0 & \text{otherwise} \end{cases} \qquad w_a\left(e_{s,t}^{(new)}\right) = \begin{cases} w_{s,t}^{(a)} & e_{s,t}^{(new)} \in E^{(a)} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$\mu\left(e_{s,t}^{(new)}\right) = \begin{cases} w_{s,t}^{(a)}/M^{(c)} & e_{s,t}^{(new)} \in E^{(c)} \\ 0 & \text{otherwise} \end{cases} \qquad M^{(c)} = \max_{i,j: e_{i,j} \in E^{(c)}} \left(w_{i,j}^{(c)}\right) \tag{5}$$

and $k \geq 1$ is the number of articles learned. The function $\mu(\cdot)$ in (5) serves the purpose to make a more significant impact of the preference degree $\lambda^{(a)} \in (0, 1]$ to $N^{(new)}$ when the contextual associations in the recommended article are not forming the major part of $N^{(c)}$.

(3)-(5) give conceptual definitions. For a better efficiency of computation, in actual implementation we have made simplification on calculation and the key idea of such simplification is summarized as below:

a) given a set of training documents, the step of updating weight of the edges by training document k is moved to the end of the learning algorithm based on the idea of accumulation based learning, and so to avoid updating the background net of entire category with each individual training document in the learning phase; according to the definition, the simplified calculation for the weight updating is equivalent to (3)-(5) in terms of final learning result, because multiplying a common factor to the weights of edges does not affect the result of the function $\mu(\cdot)$ in (5);

b) having words sorted in lexicographic order, the processing of searching, adding or removing a word can be done in $O(|V^{(c)}| \cdot \log(|V^{(c)}|))$ time complexity.

It is straightforward to find that Algorithm-1 has a linear complexity to the number of documents and linearithmic or loglinear complexity to the number of terms.

Algorithm-1: *Learning*($\boldsymbol{D}$), the input $\boldsymbol{D} = \{d_i | i = 1, 2, \ldots, k\}$ is a set of training documents, k is the number of documents currently available, and for each document d_i with a preference degree λ_i $(i = 1, 2, \ldots, k)$.
Initially, the category's background net is $N^{(c)} =< \varnothing, \varnothing >$
  For each training document d_i, $i = 1, 2, \ldots, k$
    Construct the article's background net $N^{(a)}$ for d_i by Definition 2.1
    Update category's background net $N^{(c)}$ by:
      $V^{(c)} \leftarrow V^{(c)} \cup V^{(a)}$
      $E^{(c)} \leftarrow E^{(c)} \cup E^{(a)}$
      $w^{(c)}_{s,t} \leftarrow w_c(e^{(c)}_{s,t}) + (\lambda_i)^{1-\mu(e^{(c)}_{s,t})} \times w_a(e^{(c)}_{s,t})$, where $v^{(c)}_s$, $v^{(c)}_t \in V^{(c)}$
Update the weights $w^{(c)}_{s,t} \leftarrow w^{(c)}_{s,t}/k$, where $v^{(c)}_s$, $v^{(c)}_t \in V^{(c)}$
Return $N^{(c)}$

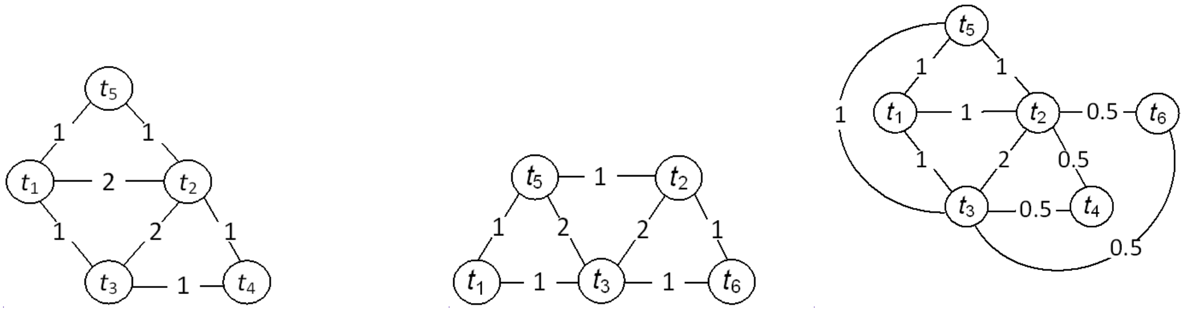
Example 2.2. *Given two background nets N_1 and N_2 representing articles A_1 and A_2, respectively. After N_1 and N_2 were learned with preference degree $\lambda^{(1)} = \lambda^{(2)} = 1$, the updated background net $N^{(new)}$ is shown in Figure 2.*

$$A_1: \quad w_1,\ w_2,\ w_3 \,|\, w_2,\ w_3,\ w_4 \,|\, w_1,\ w_2,\ z_1$$
$$A_2: \quad w_2,\ w_3,\ z_1 \,|\, w_2,\ w_3,\ z_2 \,|\, w_1,\ w_3,\ z_1$$

3. Reasoning on Background Net.

3.1. **Association degree of terms.** A background net $N =< V, E >$ captures the contextual association between terms v_i and v_j, for v_i, $v_j \in V$, and $i \neq j$. The weight $w_{i,j}$ of edges $e_{i,j} \in E$ is the number of appearance of both terms v_i and v_j in the same partition. Based on this number, for a given term v_i, we can determine an association degree to indicate in what level a term v_j $(i \neq j)$ is associated to v_i. The reasoning on background net for text categorization is achieved by similarity comparison of concepts based on the association degree of related terms.

We define $\text{Degree}_{(1)}(v_i, v_j)$ as the degree of direct contextual association from term v_i to v_j in fixed one step, named as 1-step *association degree*.



(a) Background net N_1 (b) Background net N_2 (c) Background net $N^{(new)}$
$t_1 = w_1$, $t_2 = w_2$, $t_3 = w_3$, $t_4 = w_4$, $t_5 = z_1$, $t_6 = z_2$

FIGURE 2. Background nets of Example 2.2

Definition 3.1. *The* association degree *of term v_i to term v_j with a background net provided, is defined as in (6)*

$$\text{Degree}_{(1)}(v_i, v_j) = \begin{cases} 1 & v_i = v_j \\ w_{i,j} \Big/ \sum\limits_{k:k\neq i} w_{i,k} & v_i \neq v_j \end{cases} \tag{6}$$

When v_i and v_j have no direct contextual association in 1-step, there can be two possible situations: 1) the contextual association from v_i to v_j can be made through other term(s), or 2) they do not have an association within finite steps. In the first case, if there are k_1 ($k_1 > 0$) terms $v_{s,1}, v_{s,2}, \ldots, v_{s,k_1}$ making up k_1 2-step associations $v_i - v_{s,q}$ and $v_{s,q} - v_j$, ($q = 1, 2, \ldots, k_1$) then 2-step association degree from v_i to v_j, $\text{Degree}_{(2)}(v_i, v_j)$, is defined as the maximum among k_1 values obtained through multiplications of 1-step association degrees of $v_i - v_{s,q}$ and $v_{s,q} - v_j$ ($q = 1, 2, \ldots, k_1$). On the other hand, if there are no 1-step or 2-step associations from v_i to v_j, then $\text{Degree}_{(1)}(v_i, v_j) = \text{Degree}_{(2)}(v_i, v_j) = 0$. Thus, we can have m-step association degree from term v_i to term v_j for a fixed value m ($m > 1$) defined as

$$\text{Degree}_{(m)}(v_i, v_j) = \max_{\substack{\forall r:r<m \\ \forall (v_{k_1},\ldots,v_{k_r})\in V^r}} (\text{Degree}_{(1)}(v_i, v_{k_1}) \times \ldots \times \text{Degree}_{(1)}(v_{k_r}, v_j)) \tag{7}$$

where V^r is r-ary Cartesian product over vertex set V for r times. $\text{Degree}_{(m)}(v_i, v_j)$ is the maximal among all possible r multiplications ($r = 1, 2, \ldots, m-1$) of 1-step association degrees alone the path from v_i to v_j.

Assume M is the maximum number of steps between all v_i and v_j with non-zero association, the *full-step association degree* from term v_i to term v_j is defined as the maximal among M m-step *association degrees* from v_i to v_j, for $m = 1, 2, \ldots, M$.

$$\text{Degree}(v_i, v_j) = \text{Degree}_{(M)}(v_i, v_j) \tag{8}$$

Example 3.1. *Given a background net $A_1 =< V_1,\ E_1 >$ represented as an adjacency matrix M_1 shown below:*

$$M_1 = \begin{bmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 1 & 0 \\ 1 & 0 & 1 & 2 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix} \quad \textit{and } V_1 = \{v_1, v_2, v_3, v_4, v_5\}$$

The 1-step association degree and full-step association degree is shown in Tables 1(a) and 1(b), respectively.

TABLE 1. The 1-step association degree and full-step association degree in Example 3.1

1-step	$\boldsymbol{v_1}$	$\boldsymbol{v_2}$	$\boldsymbol{v_3}$	$\boldsymbol{v_4}$	$\boldsymbol{v_5}$
$\boldsymbol{v_1}$	1.00	0.25	0.25	0.25	0.25
$\boldsymbol{v_2}$	0.50	1.00	0.50	0.00	0.00
$\boldsymbol{v_3}$	0.33	0.33	1.00	0.33	0.00
$\boldsymbol{v_4}$	0.33	0.00	0.33	1.00	0.33
$\boldsymbol{v_5}$	0.50	0.00	0.00	0.50	1.00

(a) 1-step association degree

Degree	$\boldsymbol{v_1}$	$\boldsymbol{v_2}$	$\boldsymbol{v_3}$	$\boldsymbol{v_4}$	$\boldsymbol{v_5}$
$\boldsymbol{v_1}$	1.00	0.25	0.25	0.25	0.25
$\boldsymbol{v_2}$	0.50	1.00	0.50	0.17	0.13
$\boldsymbol{v_3}$	0.33	0.33	1.00	0.33	0.11
$\boldsymbol{v_4}$	0.33	0.11	0.33	1.00	0.33
$\boldsymbol{v_5}$	0.50	0.13	0.17	0.50	1.00

(b) Full-step association degree

3.2. **Comparison between background nets.**

3.2.1. *Concept of term.* With the contextual association between terms captured and represented as a background net $N =< V, E >$, a personalized "understanding" of a single term v can be established, which is much more informative than the symbolic form of the term. We shall use the expression "concept of term" to denote this personalized understanding of v based on the given background net.

Definition 3.2. *The concept of a term v in a given background net $N =< V, E >$ is defined as a fuzzy set [32] Concept(v).*

$$\text{Concept}^{(N)}(v) = \sum_{i:v_i \in V} \rho^{(N)}(v, v_i)/v_i \tag{9}$$

where

$$\rho^{(N)}(v, v_i) = \begin{cases} \text{Degree}^{(N)}(v, v_i) & v \in V \\ 0 & v \notin V \end{cases} \tag{10}$$

$v_i \in V$, and the superscript N indicates the background net N under discussion.

It is important to note that the concept of term v is defined as a fuzzy set and represented through the contextual association degree from the term v to other terms v_i ($v_i \in V$), while the term v itself only serves as a label of the concept.

Example 3.2. *The concepts of v_1 and v_2 in Example 3.1 are*

$$c_1 = \text{Concept}^{(A_1)}(v_1) = 1/v_1 + 0.25/v_2 + 0.25/v_3 + 0.25/v_4 + 0.25/v_5$$

$$c_2 = \text{Concept}^{(A_1)}(v_2) = 0.5/v_1 + 1/v_2 + 0.5/v_3 + 0.17/v_4 + 0.13/v_5$$

3.2.2. *Similarity of concepts of a term.*

Definition 3.3. *The similarity of two concepts c_1 and c_2 of term v in the corresponding nets $N_1 =< V_1, E_1 >$ and $N_2 =< V_2, E_2 >$ is defined as*

$$\text{Similarity}_{\text{Concept}}(c_1, c_2) = \frac{\sum\limits_{i:v_i \in V_1 \cap V_2 - \{v\}} K_1 \cdot \rho^{(N_1)}(v, v_i) \wedge K_2 \cdot \rho^{(N_2)}(v, v_i)}{\sum\limits_{j:v_j \in V_1 \cup V_2 - \{v\}} K_1 \cdot \rho^{(N_1)}(v, v_j) \vee K_2 \cdot \rho^{(N_2)}(v, v_j)} \tag{11}$$

where

$$c_1 = \sum_{i:v_i \in V_1} \rho^{(N_1)}(v, v_i)/v_i \qquad c_2 = \sum_{i:v_i \in V_2} \rho^{(N_2)}(v, v_i)/v_i \tag{12}$$

and K_1 and K_2 are the normalization factors with

$$K_1 = 1 \Big/ \sum_{k:v_k \neq v} \rho^{(N_1)}(v, v_k) \qquad K_2 = 1 \Big/ \sum_{k:v_k \neq v} \rho^{(N_2)}(v, v_k) \tag{13}$$

where the $\wedge$ and $\vee$ are minimum and maximum operations, respectively. It can be understood as an objective measurement of symmetric similarity between two concepts in two corresponding background nets that are considered at the same level of discussion. When only considering the existence of association between two concepts having no concern on the degree of such association, we have the binary *similarity defined in Definition 3.4.*

Definition 3.4. *The binary similarity between two concepts of a same symbolic term is calculated as*

$$\text{Binary\_Similarity}_{\text{Concept}}(c_1, c_2) = \frac{|\text{Supp}(c_1) \cap \text{Supp}(c_2)|}{|\text{Supp}(c_1) \cup \text{Supp}(c_2)|} \tag{14}$$

where $\text{Supp}(c_i)$ $(i = 1, 2)$ *is the support of fuzzy set* c_i*.*

In a different scenario when we need judge how well a concept c_2 in background net N_2 can be accepted by another given background net N_1 containing the concept c_1, which is symbolically equal to c_2, an acceptance measure is useful.

Definition 3.5. *The acceptance of concept* c_2 *in* $N_2 =< V_2, E_2 >$*, based on* $N_1 =< V_1, E_1 >$ *is defined as*

$$\text{Acceptance}^{(N_1)}_{\text{Concept}}(c_1, c_2) = \frac{\sum\limits_{i:v_i \in V_1 \cap V_2 - \{v\}} K'_1 \cdot \rho^{(N_1)}(v, v_i) \wedge K_2 \cdot \rho^{(N_2)}(v, v_i)}{\sum\limits_{j:v_j \in V_2 - \{v\}} K_2 \cdot \rho^{(N_2)}(v, v_j)} \tag{15}$$

where K_2 *is as given in (13), and* K'_1 *is the normalization factors shown below.*

$$K'_1 = 1 \Big/ \sum_{k:v_k \in V_2 - \{v\}} \rho^{(N_1)}(v, v_k) \tag{16}$$

Similarly, we can also define the *binary* acceptance to evaluate the acceptance of an article with its background information represented as a background net.

Definition 3.6. *The binary acceptance of concept* c_2 *in* $N_2 =< V_2, E_2 >$*, based on* $N_1 =< V_1, E_1 >$ *is defined as*

$$\text{Binary\_Acceptance}^{(N_1)}_{\text{Concept}}(c_1, c_2) = \frac{|\text{Supp}(c_1) \cap \text{Supp}(c_2)|}{|\text{Supp}(c_2)|} \tag{17}$$

Algorithm-2: *Similarity*(v, N_1, N_2), the inputs are a term v, and two background nets N_1, N_2.
1: Extract the concept of term v in N_1 by Definition 3.2 as $c_1 = \text{Concept}^{(N_1)}(v)$
2: Extract the concept of term v in N_2 by Definition 3.2 as $c_2 = \text{Concept}^{(N_2)}(v)$
3: Based on different similarity *measure* (11), (14), (15) or (17) to calculate:
4: *degree*$\leftarrow$ *measurement* (c_1, c_2)
5: **Return** *degree*

Algorithm-2 is at a conceptual level, and an appropriate evaluation measure will be selected according to specific application need. So we shell provide more specific algorithms for text categorization and personalized article selection in Section 4 and then discuss their complexity accordingly.

Example 3.3. *Consider the concept of v_3 in background net A_1 in Example 3.1, and that in another background net A^* given below:*

$$M^* = \begin{bmatrix} 4 & 2 & 1 & 1 & 1 \\ 2 & 1 & 5 & 0 & 2 \\ 1 & 5 & 2 & 1 & 0 \\ 1 & 0 & 1 & 3 & 1 \\ 1 & 2 & 0 & 1 & 1 \end{bmatrix} \quad \textit{and } V^* = \{v_1, v_2, v_3, \boldsymbol{z_1}, \boldsymbol{z_2}\}$$

TABLE 2. The full-step association degree of A^* in Example 3.3

Degree	$\boldsymbol{v_1}$	$\boldsymbol{v_2}$	$\boldsymbol{v_3}$	$\boldsymbol{z_1}$	$\boldsymbol{z_2}$
$\boldsymbol{v_1}$	1.00	0.40	0.22	0.20	0.20
$\boldsymbol{v_2}$	0.22	1.00	0.56	0.08	0.22
$\boldsymbol{v_3}$	0.16	0.71	1.00	0.14	0.16
$\boldsymbol{z_1}$	0.33	0.24	0.33	1.00	0.33
$\boldsymbol{z_2}$	0.25	0.50	0.28	0.25	1.00

We can have two corresponding concepts of term v_3:

$$c_3^{(A_1)} = \text{Concept}^{(A_1)}(v_3) = 0.33/v_1 + 0.33/v_2 + 1/v_3 + 0.33/v_4 + 0.11/v_5$$

$$c_3^{(A^*)} = \text{Concept}^{(A^*)}(v_3) = 0.16/v_1 + 0.71/v_2 + 1/v_3 + \mathbf{0.14}/\boldsymbol{z_1} + \mathbf{0.16}/\boldsymbol{z_2}$$

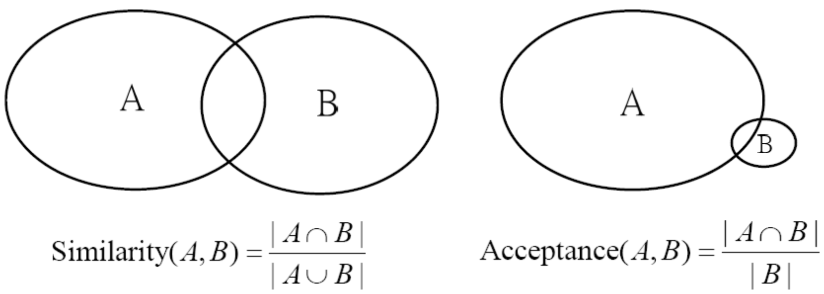
Therefore, we have the similarity obtained as

$$\text{Similarity}_{\text{Concept}}\left(c_3^{(A_1)}, c_3^{(A^*)}\right) = \frac{0.14 + 0.30}{0.30 + 0.60 + 0.30 + 0.10 + 0.12 + 0.14} = 0.282$$

and the acceptance of $c_3^{(A^*)}$ under A_1 is obtained as

$$\text{Acceptance}^{(A_1)}_{\text{Concept}}\left(c_3^{(A_1)}, c_3^{(A^*)}\right) = \frac{0.14 + 0.52}{0.52 + 0.60 + 0.12 + 0.14} = 0.48$$

The similarity measure of two concepts is at a symmetric basis and measuring the closeness of two concepts at a same level of discussion. While the acceptance measure is not symmetric in the sense that it measures how well a concept in the *guest* background net be accepted by the concept involved in the *host* background net, with both having the same symbolic representation. If a concept in the *guest* background net is considered better *fit* in to a *host* background net or a *host* background net better *accept* a concept in the *guest* background, then a higher acceptance is obtained. Figure 3 gives a conceptual illustration.



(a) Similarity measure (b) Acceptance measure

FIGURE 3. The similarity and acceptance measures

3.2.3. *Similarity between background nets.*

Definition 3.7. *Based on the similarity of two concepts on difference background nets* $N_1 =< V_1,\ E_1 >$ *and* $N_2 =< V_2,\ E_2 >$*, the* similarity *of two background nets is defined as*

$$\mathrm{Similarity}_{\mathrm{Net}}(N_1, N_2) = \frac{\sum\limits_{i:v_i \in V_1 \cap V_2} \mathrm{Similarity}_{\mathrm{Concept}}\left(\mathrm{Concept}^{(N_1)}(v_i), \mathrm{Concept}^{(N_2)}(v_i)\right)}{|V_1 \cup V_2|} \tag{18}$$

which serves as an objective measurement of the similarity between two difference background nets.

Considering a different scenario where the main concern is on how much the background net N_2 can be accepted based on background net N_1, an acceptance measure is needed.

Definition 3.8. *The acceptance of* $N_2 =< V_2, E_2 >$ *based on background net* $N_1 =< V_1,$ $E_1 >$*, is defined as*

$$\mathrm{Acceptance}^{(N_1)}_{\mathrm{Net}}(N_2) = \frac{\sum\limits_{i:v_i \in V_2} \mathrm{Acceptance}^{(N_1)}_{\mathrm{Concept}}\left(\mathrm{Concept}^{(N_1)}(v_i), \mathrm{Concept}^{(N_2)}(v_i)\right)}{|V_2|} \tag{19}$$

The acceptance measure is not symmetric in the sense that it measures how well a *guest* background net is accepted by the *host* background net. The two background nets should not be treated as at the same level of discussion when the former is representing a background domain or a user's reading interest and the latter a candidate article, because the former usually captures much more information than the latter does.

Example 3.4. *Considering given nets:* $A_1 =< V_1, E_1 >$ *as in Example 3.1 with* $V_1 = \{v_1, v_2, v_3, v_4, v_5\}$*,* $A_2 =< V_2, E_2 >$ *with* $V_2 = V_1$ *and the corresponding adjacency matrix* $M_2 = 2 \cdot M_1$*,* $A_3 =< V_3, E_3 >$*,* $A_4 =< V_4, E_4 >$*,* $A_5 =< V_5, E_5 >$ *and* $A_6 =< V_6, E_6 >$ *with* $V_3 = \{z_1, z_2, z_3, z_4, z_5\}$*,* $V_4 = V_5 = V_6 = \{v_1, v_2, v_3, v_4, v_5, z_1, z_2, z_3, z_4, z_5\}$ *and their adjacency matrices shown below:*

$$M_3 = \begin{bmatrix} 4 & 2 & 1 & 1 & 1 \\ 2 & 1 & 5 & 0 & 2 \\ 1 & 5 & 2 & 1 & 0 \\ 1 & 0 & 1 & 3 & 1 \\ 1 & 2 & 0 & 1 & 1 \end{bmatrix} \qquad M_4 = \begin{bmatrix} M_1 & 0 \\ 0 & M_3 \end{bmatrix}$$

$$M_5 = \begin{bmatrix} 2 \cdot M_1 & \mathrm{One}_{5\times 5} \\ \mathrm{One}_{5\times 5} & 2 \cdot M_3 \end{bmatrix} \qquad M_6 = \begin{bmatrix} 2 \cdot M_1 & 0 \\ 0 & M_3 \end{bmatrix}$$

The similarity and acceptance measures among these background nets are shown in Tables 3(a) and 3(b), respectively. In this example, A_2 duplicates the associations of terms in A_1 two times, A_1 and A_3 are two background nets from two difference domains, A_4 contains A_1 and A_3, A_5 contains not only A_1 and A_3 but also other associations between A_1 and A_3, and A_6 contains A_3 and two times A_1. Looking at the results, we observed that: a) the duplication of a background net will not affect the similarity and acceptance of two background nets; b) the similarity between two background nets is an objective and a symmetric measure, i.e., $\mathrm{Similarity}(N_1, N_2) = \mathrm{Similarity}(N_2, N_1)$; c) given A_4 as a background net that contains the associations in both A_1 and A_3, the results of the acceptance of A_1, A_2, A_3 based on A_4 are 1, which shows the usefulness of acceptance measure in articles selection; d) for background net A_6 that contains the associations in both A_4 and A_1, the acceptance of A_1, A_2, A_3 and A_4 based on A_6 are as the same as that of A_1, A_2, A_3 and A_4 based on A_4; e) the row of A_5 in Tables 3(a) and 3(b) show

TABLE 3. The similarity degrees and acceptance degrees in Example 3.4

Similarity (N_1, N_2)		N_2					
		A_1	A_2	A_3	A_4	A_5	A_6
N_1	A_1	1.00	1.00	0.00	0.50	0.19	0.50
	A_2	1.00	1.00	0.00	0.50	0.19	0.50
	A_3	0.00	0.00	1.00	0.50	0.27	0.50
	A_4	0.50	0.50	0.50	1.00	0.45	1.00
	A_5	0.19	0.19	0.27	0.45	1.00	0.45
	A_6	0.50	0.50	0.50	1.00	0.45	1.00

(a) Similarity degrees

Accept $^{(N1)}(N_2)$		N_2					
		A_1	A_2	A_3	A_4	A_5	A_6
N_1	A_1	1.00	1.00	0.00	0.50	0.27	0.50
	A_2	1.00	1.00	0.00	0.50	0.27	0.50
	A_3	0.00	0.00	1.00	0.50	0.35	0.50
	A_4	1.00	1.00	1.00	1.00	0.62	1.00
	A_5	0.54	0.54	0.69	0.62	1.00	0.62
	A_6	1.00	1.00	1.00	1.00	0.62	1.00

(b) Acceptance degrees

the change of associations between terms affects both of the similarity and acceptance of background nets, but the affection to acceptance is less significant than that to similarity.

4. Applications.

4.1. Text categorization.

4.1.1. *Main algorithms.* Text categorization (TC) is the process of finding the correct topic for each document from a collection of text documents given a set of categories [9-12,15,16]. Considering single label multiple categories text categorization tasks, the training processing is to learn each category c_i in categories set $\boldsymbol{C} = \{c_i|i = 1, 2, \ldots, n\}$ by provided training document $d_{i,j}$ in document set $\boldsymbol{D} = \{d_{i,j}|i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, n_i\}$, where n is the number of categories, and n_i is the number of documents currently available in i-th category c_i. The result of learning of i-th category c_i is represented by a background net called *category background net* of c_i and denoted as CN_i. For each document $d_{i,j}$, a corresponding background net $N_{i,j}$ is first constructed and then be merged with the current category background net CN_i in an incremental basis as introduced in Section 2.2, to represent the background information of the i-th category captured (Figure 4). Finally, we obtain a set of category background nets $\boldsymbol{CN} = \{CN_i|i = 1, 2, \ldots, n\}$, where CN_i is the corresponding category background net for category c_i.

When a new document d^* is given for the categorization task, we first construct its background net N^*, and then evaluate its acceptance with each of the category background net CN_i, for $i = 1, 2, \ldots, n$. The final decision will be the category that gives the maximum acceptance to d^* (Figure 5):

$$c_{d^*} = \underset{c_i \in \boldsymbol{C}}{\arg\max} \left(\text{Acceptance}_{\text{Net}}^{(CN_i)}(N^*) \right) \tag{20}$$

In text processing, the association between terms can be more important than the frequency of terms or that of term associations [33,34]. When the *binary* acceptance

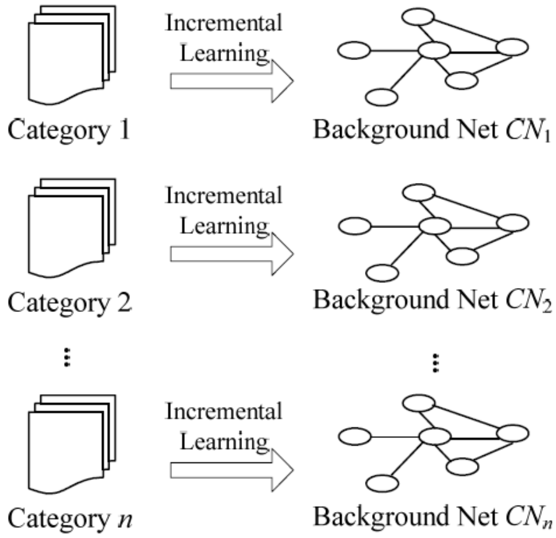


FIGURE 4. Illustration of training background net classifier

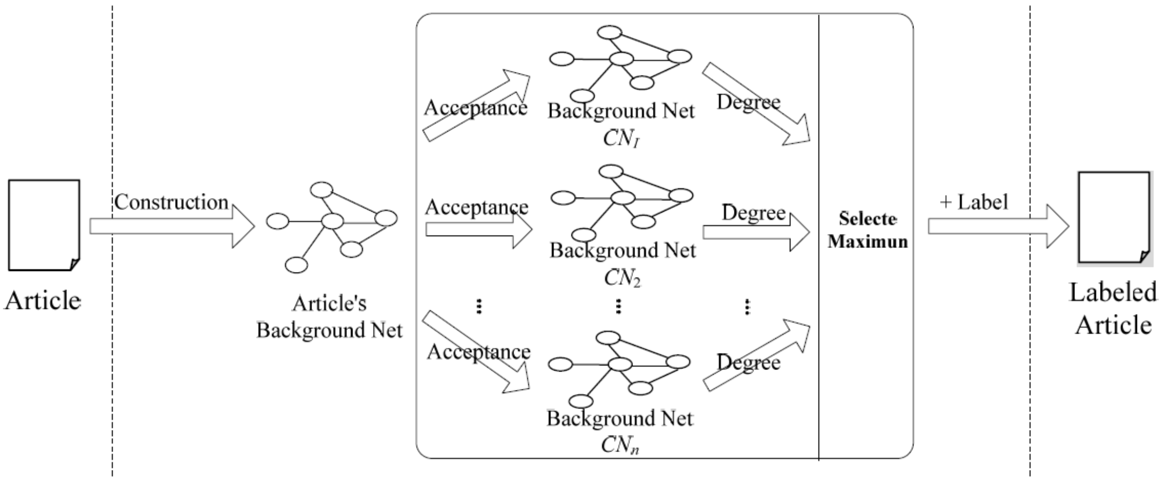


FIGURE 5. Illustration of text categorization by background net classifier

measure is adopted, the information captured in a background net is actually viewed as a set of combination of terms:

$$N = \bigcup\nolimits_{m=1}^{\tau} \mathrm{AssoSet}(m) = \{\{w_1\}, \{w_2\}, \{w_3\}, \ldots, \{w_1, w_2\}, \{w_1, w_3\}, \ldots\}$$

where w_k $(k = 1, 2, \ldots)$ are original single terms in the background net, $\tau \geq 2$, and AssoSet (m) is a set that contains all m-term associations within carefully selected partitions of an article. A set of simple algorithms is developed for the learning and reasoning on background net for the text categorization. This approach simply records the association between terms in the learning phase, and computes the number of elements of the intersection of category background nets and article background net for the categorization task. The algorithms are given in Algorithm-3 and Algorithm-4, where the operations on "background net" only make use of the information of association of terms.

4.1.2. *Algorithm complexity.* The computation complexity of these algorithms is: 1) in learning phase, assuming the average of the length of each partition is L_{avg}, there is $O(L_{\mathrm{avg}}^{\tau})$ time complexity for capturing τ combinations with possible repetition of words in each partition, and assuming the average of the number of partition of each article

Algorithm-3: $\mathit{Training}(\boldsymbol{C}, \boldsymbol{D}, \tau)$, the inputs are $\boldsymbol{C} = \{c_i | i = 1, 2, \ldots, n\}$ a categories set, and $\boldsymbol{D} = \{d_{i,j} | i = 1, 2, \ldots, n, \, j = 1, 2, \ldots, n_i\}$, n is the number of categories, and n_i is the number of documents currently available in i-th category, and τ is a parameter indicates τ combinations with repetition of words for computing.

```
For each i-th category, i = 1, 2, ..., n
  For each training document d_{i,j}, j = 1, 2, ..., n_i
    Partition set: P_{i,j} ← PartitionArticle(d_{i,j})
    For each partition p_l ∈ P_{i,j}, l = 1, 2, ..., |P_{i,j}|
      Ω ← τ-CombinationsWithRepetitionOfWords(p_l)
      Update article background net: N_{i,j} ← N_{i,j} ∪ Ω
    Update category background net: CN_i ← CN_i ∪ N_{i,j}
Return CN = {CN_i | i = 1, 2, ..., n}
```

Algorithm-4: $\mathit{Categorization}(\boldsymbol{C}, \boldsymbol{CN}, d^*, \tau)$, the inputs are $\boldsymbol{C} = \{c_i | i = 1, 2, \ldots, n\}$ a categories set, $\boldsymbol{CN} = \{CN_i | i = 1, 2, \ldots, n\}$ is the set of category background nets, where, n is the number of categories, d^* is a newly document for categorization, τ is a parameter indicates τ combinations with repetition of words for computing.

```
Partition set: P* ← PartitionArticle(d*)
  For each partition p_l ∈ P*, l = 1, 2, ..., |P*|
    Ω ← τ-CombinationsWithRepetitionOfWords(p_l)
    Update article background net: N* ← N* ∪ Ω
Return arg max_{c_i ∈ C}(|CN_i ∩ N*|)
```

is P_{avg}, and there are total $|\boldsymbol{D}|$ documents in the training set, the time complexity in learning step is $O(|\boldsymbol{D}| \cdot P_{\text{avg}} \cdot L^{\tau}_{\text{avg}})$; 2) in categorization task, the acceptance measure is used for the evaluation of a new article under a given category background. Different from using similarity measure, only those term associations appeared in the article background net will be involved in calculation, instead of both of the entire category background net as well as the entire article background net. By sorting words in lexicographic order, the time complexity in categorization step is $O(|\boldsymbol{C}| \cdot P_{\text{avg}} \cdot L^{\tau}_{\text{avg}} \cdot K^{\tau})$, where $|\boldsymbol{C}|$ is the number of categories, $K = \log V$ is the time complexity of searching each word in the category background nets with V words, P_{avg} is the average number of partitions of new article and L_{avg} the average number of words in each partition.

Typically, there are $12 \sim 15$ terms for composing a meaningful sentence, thus, $L_{\text{avg}} = 12 \sim 15$. On the other hand, selecting $\tau = 2$ for the learning and categorization is considered enough, based on the understanding that there are typically no more than three meaningful words (after removal of meaningless words) appeared simultaneously in two different sentences that describing different subjects.

A typical kNN classifier has $O(|\boldsymbol{D}| \cdot P_{\text{avg}} \cdot L_{\text{avg}})$ time complexity in learning step, and $O(|\boldsymbol{D}| \cdot P_{\text{avg}} \cdot L_{\text{avg}})$ time complexity in categorization task that calculates the similarity between all training documents and a new document for finding the k-nearest neighbors. As a comparison, in our approach the time complexity of learning of background nets is slightly more than that of kNN classifier, but the categorization processing is much faster than that of kNN classifier. Furthermore, in order to represent documents well representing their content, techniques of term weights, feature selection and feature extraction are usually applied before the training of kNN classifiers. Thus, the time complexity of learning of kNN classifier is actually more than $O(|\boldsymbol{D}| \cdot P_{\text{avg}} \cdot L_{\text{avg}})$, depending on the representation techniques adopted.

4.2. **Personalized articles selection.**

4.2.1. *Personalized text categorization.* Background net provides a possible solution for the task of personalized article selection that can be considered a special case of text categorization where a person's interest behaves as a special "category" covering multiple domains of content. For personalized articles selection, we construct a single background net for a given set of articles the user provided which captures his/her personal preference of reading by a long-term knowledge background accumulated through incremental learning.

Compared with the use of only isolated terms, our approach gives a more informative representation of a document and personal interests. Being association-based, and constructed in an accumulation basis through incremental learning, it captures the *contextual association* between terms and represents a document or personal interest as an association graph. The comparison between concepts can be reasonably done based on term associations, based on the belief: for two concepts indicated by two corresponding symbolic terms, the more similar they are, the wider the common set of associations from them could be found. This approach helps minimize the polysemy confusion in concept understanding, and improves robustness in handling changing environment as well as personal interest relating to multiple domains.

4.2.2. *Personalized keywords support.* A keyword is not purely a symbol but a symbolized representation of a concept which is rather personal in terms of its meaning. Considering both of instant requirements of articles selection and long-term personal interest, background net is used to capture personal background of preference and apply the concept of a term as a fuzzy set to express personalized keyword [30] through its contextual association with other terms as introduced in Section 3.2.1. The evaluation of a candidate articles is done through the similarity evaluation between two concepts of terms (Section 3.2.1). Finally, all of selection results are filtered (or re-ranking) using similarity measure based on user's background net with the set of keywords specifically given to better suit personal preference.

5. **Experiments.**

5.1. **Datasets.** We applied our approach for the single label multiple categories text categorization tasks with the Reuters-21578 dataset, which is a widely used benchmarking collection. In our experiments, the ApteMod version [35] of Reuters-21578 has been selected for use. After removing unlabeled documents and documents with multiple class labels, we selected the top 10 largest categories for our experiment. The numbers of training documents and testing documents in the top 10 categories are shown in Table 4. Another benchmarking collection is the 20Newsgroups "bydate" version that consists of 18846 documents uniformly distributed in 20 categories[1].

TABLE 4. The distribution of the top 10 categories in skewed Reuters-21758

Top 10	**Training size**	**Testing size**	**Top 10**	**Training size**	**Testing size**
acq	1596	90	253	2840	191
coffee	696	22	121	1083	81
crude	money-fx	money-supply	ship	sugar	trade
earn	222	123	108	97	250
interest	87	28	36	25	75

[1]http://people.csail.mit.edu/jrennie/20Newsgroups

5.2. **Evaluation measures.** In the TC, the most commonly used performance measures are *recall*, *precision* and their harmonic mean F1 [9,16]. Given a specific category c_i ($i \leq K$) from the set of predefined categories $\boldsymbol{C} = \{c_i | i = 1, 2, \ldots, K\}$ with $K = |\boldsymbol{C}|$, the corresponding *recall*, *precision*, and F1 are defined as follows:

$$recall_i = \frac{TP_i}{TP_i + FN_i} \quad precision_i = \frac{TP_i}{TP_i + FP_i} \quad \mathrm{F1}_i = \frac{2 \times recall_i \times precision_i}{recall_i + precision_i}$$

where TP_i (true positives) is the number of documents *assigned correctly* to category c_i, FP_i (false positives) is the number of documents *assigned wrongly* to category c_i, i.e., those documents do not belong to category c_i but are assigned wrongly to this category, and FN_i (false negatives) is the number of documents *rejected wrongly* from category c_i, i.e., those documents actually belong to category c_i but are not assigned to this category.

The macro-average and mirco-average are performance measures for evaluating the overall performance. Macro-averaged performance scores are determined by first computing the performance measures per category and then averaging these to compute the global means.

$$recall_{macro} = \frac{\sum_{i=1}^{K} recall_i}{K} \qquad precision_{macro} = \frac{\sum_{i=1}^{K} precision_i}{K}$$

$$\mathrm{F1}_{macro} = \frac{2 \times precision_{macro} \times recall_{macro}}{precision_{macro} + recall_{macro}}$$

Micro-averaged performance scores are determined by first computing the subtotals of *TP*, *FN*, *FP* for all categories and then using these subtotals to compute the performance overall measures.

$$recall_{micro} = \frac{\sum_{i=1}^{K} TP_i}{\sum_{i=1}^{K} (TP_i + FN_i)} \qquad precision_{micro} = \frac{\sum_{i=1}^{K} TP_i}{\sum_{i=1}^{K} (TP_i + FP_i)}$$

$$\mathrm{F1}_{micro} = \frac{2 \times precision_{micro} \times recall_{micro}}{precision_{micro} + recall_{micro}}$$

5.3. **Experiment results.** The pre-processings of our experiment includes: removal of numbers and non-alphabetic characters, converting all words into lower case, deleting stop-words using a standard stopwords list[2], and removal of the words with length of less than 2 or greater than 30. We use character '.', '!' and '?' for partition of article into a set of sentences.

The results listed in Table 5 throughout Table 8 show that our approach has successfully achieved a good performance even without complex feature selection/extraction. The columns of tables indicate different settings on algorithms applied to an article: 1) by viewing the information captured on a background net as a set of combination of terms, *binary* acceptance measure is applied ($\tau = 2$ in Algorithm-3 and Algorithm-4); 2) 1-step association degree for a concept on article's background net with acceptance; 3) full-step association degree for a concept on article's background net with acceptance. While for a category's background net, we only apply the 1-step association degree for concept calculation, based on the belief that a category background net captures the accumulation of training documents in the entire category and may contain information irrelevant to the target article evaluation, so m-step or full-step association is considered inappropriate.

The settings of kNN and SVM classifiers used in our experiments are: applied typical *tf-idf* term weighting [9,16] of a term and normalized feature vector by cosine normalization or 2-norm normalization, the best result (Macro F1) of k in the range of 1 to 500 for

[2]http://www.ai.mit.edu/projects/jmlr/papers/volume5/lewis04a/a11-smart-stop-list/english.stop

TABLE 5. Comparisons of Reuters top 10 categories with different settings of algorithm having stop words removed

	Binary Accept. (Algorithm-3 & 4)			**1-Step Acceptance**			**Full-Step Acceptance**			**kNN ($k = 97$)**			**SVM**		
	P.	**R.**	**F1**	**P.**	**R.**	**F1**	**P.**	**R.**	**F1**	**P.**	**R.**	**F1**	**P.**	**R.**	**F1**
acq	0.935	0.957	0.946	0.944	0.962	0.953	0.932	0.961	0.946	0.983	0.769	0.863	0.943	0.978	0.961
coffee	1	0.955	0.977	1	1	1	1	1	1	0.955	0.955	0.955	0.957	1	0.978
crude	0.962	0.843	0.899	0.945	0.860	0.900	0.954	0.851	0.9	0.92	0.95	0.935	0.95	0.934	0.942
earn	0.960	0.985	0.972	0.965	0.978	0.972	0.971	0.973	0.972	0.881	0.994	0.934	0.991	0.985	0.988
interest	0.96	0.593	0.733	0.956	0.531	0.683	0.959	0.580	0.723	0.957	0.827	0.887	0.917	0.815	0.863
mny-fx	0.731	0.908	0.810	0.727	0.920	0.812	0.752	0.943	0.837	0.745	0.874	0.804	0.821	0.793	0.807
mny-sp	0.952	0.714	0.816	0.917	0.786	0.846	0.92	0.821	0.868	0.735	0.893	0.806	0.929	0.929	0.929
ship	1	0.556	0.714	0.905	0.528	0.667	0.909	0.556	0.690	1	0.528	0.691	0.964	0.75	0.844
sugar	1	0.76	0.864	1	0.8	0.889	1	0.84	0.913	0.923	0.96	0.941	1	0.92	0.958
trade	0.831	0.92	0.873	0.833	0.933	0.881	0.833	0.933	0.881	0.811	0.973	0.885	0.923	0.96	0.941
Micro.	0.937	0.936	0.937	0.940	0.936	0.938	0.941	0.937	0.939	0.901	0.901	0.901	0.961	0.961	0.961
Macro.	0.933	0.819	0.872	0.919	0.830	0.872	0.923	0.846	0.883	0.891	0.872	0.882	0.939	0.906	0.922

P. – Precision; R. – Recall; Micro. – Micro-averaged; Marco. – Marco-averaged

the parameter k of kNN, and LibSVM package [36] for linear SVM with default setting. Feature selection/extraction techniques are not applied in both of kNN and SVM.

The typical *tf-idf* term weighting of a term and normalized feature vector is given below.

$$w_{ij} = \frac{\mathrm{tf}_{ij} \times \ \mathrm{idf}_i}{\sqrt{\sum\limits_{k:t_k \in d_j} (\ \mathrm{tf}_{kj} \times \ \mathrm{idf}_k)^2}}; \qquad \mathrm{idf}_i = \log \frac{|\boldsymbol{D}|}{n_i} \tag{21}$$

where w_{ij} is the weight of term t_i in document d_j, tf_{ij} is term frequency that is the number of occurrences of term t_i in document d_j, $|\boldsymbol{D}|$ is the total number of documents in the collection, and n_i is the number of documents where the term t_i occurs in the collection.

It is also worthy to note that our approach works well even without stopword removal. The results are shown in Table 7 and Table 8.

Based on the experiment results, we found that in text processing, the association between terms gives important information, while the frequency of terms or association of terms can be less important in evaluation of acceptance measure on background net for TC application. Table 9 shows this observation. However, it is also interesting to note that the comparison between two background nets with their association degree can be achieved more precisely using similarity measure than using acceptance measure.

Compared with other typical methods in literatures and the results claimed in [18], the experiment results show that our approach has obvious advantage with its characters of simple in implementation, robust with preprocessing and settings, and allowing incremental learning to deal with changing domain. For the TC results of 20Newsgroups, it shows encouraging results achieving the performance level of kNN and SVM classifiers, without the request of heavy processing for feature selection or feature extraction.

6. **Discussion.** In vector space model (VSM), a document is represented as a feature vector. In such a model, statistical methods are usually applied to obtain feature vector for representing a document, e.g., *tf-idf* term weighting as a kind of term weighting method used to evaluate how important a word is to a document in a collection [9,18,19]. After the feature vectors determined, batch learning is applied for training a classifier for TC task, e.g., an SVM classifier is usually used based on its outstanding performance.

TABLE 6. Comparisons of 20Newsgroups categories with different settings of algorithm having stop words removed

	Binary Accept. (Algorithm-3 & 4)			**1-Step Acceptance**			**Full-Step Acceptance**			***k*NN ($k = 483$)**			**SVM**		
	P.	**R.**	**F1**	**P.**	**R.**	**F1**	**P.**	**R.**	**F1**	**P.**	**R.**	**F1**	**P.**	**R.**	**F1**
(1)	0.847	0.799	0.823	0.820	0.8	0.810	0.825	0.812	0.818	0.650	0.646	0.648	0.837	0.743	0.787
(2)	0.667	0.774	0.717	0.648	0.738	0.690	0.665	0.710	0.687	0.805	0.648	0.718	0.68	0.83	0.748
(3)	0.771	0.538	0.634	0.771	0.538	0.634	0.761	0.548	0.637	0.677	0.777	0.723	0.777	0.744	0.76
(4)	0.714	0.75	0.733	0.653	0.719	0.684	0.651	0.719	0.684	0.606	0.727	0.661	0.696	0.793	0.741
(5)	0.824	0.753	0.787	0.816	0.748	0.780	0.781	0.730	0.754	0.755	0.712	0.733	0.832	0.834	0.833
(6)	0.761	0.820	0.789	0.714	0.795	0.752	0.727	0.782	0.754	0.852	0.714	0.777	0.808	0.754	0.78
(7)	0.788	0.821	0.804	0.808	0.854	0.830	0.818	0.844	0.831	0.875	0.679	0.765	0.794	0.879	0.835
(8)	0.884	0.864	0.874	0.889	0.854	0.871	0.882	0.848	0.865	0.85	0.859	0.854	0.906	0.881	0.894
(9)	0.944	0.940	0.942	0.936	0.920	0.928	0.931	0.912	0.921	0.905	0.912	0.909	0.969	0.932	0.95
(10)	0.967	0.884	0.924	0.949	0.884	0.915	0.956	0.879	0.916	0.899	0.877	0.888	0.942	0.942	0.942
(11)	0.919	0.972	0.945	0.903	0.952	0.927	0.905	0.957	0.931	0.817	0.975	0.889	0.979	0.942	0.96
(12)	0.783	0.937	0.853	0.783	0.927	0.849	0.791	0.937	0.858	0.743	0.942	0.831	0.978	0.894	0.934
(13)	0.771	0.659	0.711	0.769	0.677	0.720	0.75	0.672	0.709	0.819	0.506	0.626	0.706	0.814	0.757
(14)	0.863	0.808	0.834	0.860	0.747	0.8	0.851	0.763	0.804	0.910	0.715	0.801	0.873	0.854	0.863
(15)	0.825	0.911	0.866	0.840	0.919	0.878	0.818	0.901	0.857	0.785	0.937	0.854	0.937	0.90	0.918
(16)	0.841	0.945	0.890	0.851	0.930	0.888	0.855	0.920	0.886	0.648	0.935	0.765	0.822	0.902	0.86
(17)	0.764	0.931	0.839	0.791	0.904	0.844	0.805	0.898	0.849	0.657	0.918	0.766	0.752	0.89	0.815
(18)	0.956	0.915	0.935	0.918	0.894	0.906	0.906	0.894	0.9	0.894	0.915	0.904	0.981	0.835	0.902
(19)	0.753	0.648	0.697	0.738	0.671	0.703	0.730	0.697	0.713	0.792	0.552	0.650	0.788	0.6	0.681
(20)	0.733	0.602	0.661	0.740	0.633	0.682	0.691	0.633	0.661	0.809	0.355	0.493	0.703	0.586	0.639
Micro.	0.819	0.819	0.819	0.810	0.810	0.810	0.807	0.807	0.807	0.776	0.776	0.776	0.835	0.835	0.835
Macro.	0.819	0.814	0.816	0.810	0.805	0.807	0.805	0.803	0.804	0.787	0.765	0.776	0.838	0.828	0.833

(1) – alt.atheism, (2) – comp.graphics, (3) – comp.os.ms-windows.misc,
(4) – comp.sys.ibm.pc.hardware, (5) – comp.sys.mac.hardware, (6) – comp.windows.x,
(7) – misc.forsale, (8) – rec.autos, (9) – rec.motorcycles, (10) – rec.sport.baseball,
(11) – rec.sport.hockey, (12) – sci.crypt, (13) – sci.electronics, (14) – sci.med,
(15) – sci.space, (16) – soc.religion.christian, (17) – talk.politics.guns,
(18) – talk.politics.mideast, (19) – talk.politics.misc, (20) – talk.religion.misc

TABLE 7. Comparisons of Reuters top 10 categories without stop word removal

	Binary Accept. (Algorithm-3 & 4)			**1-Step Acceptance**			**Full-Step Acceptance**			***k*NN ($k = 147$)**			**SVM**		
	P.	**R.**	**F1**	**P.**	**R.**	**F1**	**P.**	**R.**	**F1**	**P.**	**R.**	**F1**	**P.**	**R.**	**F1**
Micro.	0.927	0.926	0.927	0.943	0.939	0.941	0.934	0.933	0.935	0.913	0.913	0.913	0.965	0.965	0.965
Macro.	0.927	0.754	0.832	0.921	0.822	0.869	0.915	0.806	0.857	0.894	0.853	0.873	0.942	0.915	0.928

TABLE 8. Comparisons of 20Newsgroups categories without stop word removal

	Binary Accept. (Algorithm-3 & 4)			**1-Step Acceptance**			**Full-Step Acceptance**			***k*NN ($k = 463$)**			**SVM**		
	P.	**R.**	**F1**	**P.**	**R.**	**F1**	**P.**	**R.**	**F1**	**P.**	**R.**	**F1**	**P.**	**R.**	**F1**
Micro.	0.798	0.798	0.798	0.816	0.816	0.816	0.805	0.805	0.805	0.769	0.769	0.769	0.838	0.838	0.838
Macro.	0.811	0.79	0.801	0.817	0.811	0.814	0.806	0.8	0.803	0.789	0.757	0.773	0.841	0.83	0.836

Different from representing a document as a feature vector, background net represents a document as an association graph based on terms' associations, and therefore can keep more information than using vector representation. The experiment results show that

Table 9. 1-step *binary* similarity for TC on two datasets

	Reuter-21578			**20Newsgroups**		
	P.	**R.**	**F1**	**P.**	**R.**	**F1**
Micro.	0.218	0.217	0.217	0.589	0.589	0.589
Macro.	0.389	0.574	0.464	0.640	0.589	0.613

Table 10. $\tau = 1, 2$ in Algorithm-3 and Algorithm-4 for TC on two datasets

	Reuter-21578 ($\tau = 1$)			**Reuter-21578** ($\tau = 2$)			**20Newsgroups** ($\tau = 1$)			**20Newsgroups** ($\tau = 2$)		
	P.	**R.**	**F1**	**P.**	**R.**	**F1**	**P.**	**R.**	**F1**	**P.**	**R.**	**F1**
Micro.	0.761	0.760	0.760	0.937	0.936	0.937	0.750	0.750	0.750	0.819	0.819	0.819
Macro.	0.861	0.502	0.634	0.933	0.819	0.872	0.751	0.745	0.748	0.819	0.814	0.816

term associations in a sentence provide useful information in representing a document. A good performance for TC is obtained when using term associations in a sentence and applying an incremental learning to train a classifier. Table 10 shows that term association is more informative than isolate terms. On the other hand, the appearance of a set of words in a smaller range, such as in a single sentence can be more important and more informative than that in a larger range, such as in an article. Therefore, capturing the terms' associations in a sentence is more meaningful for representing a document.

In Table 10, $\tau = 1$ means using single terms as the features for learning and categorization, which becomes a *binary* feature vector for document representation.

In addition, in VSM a similarity measure is usually used for evaluating the closeness of documents by their feature vectors. With a category represented as a feature vector named as prototype vector of a category, the same manner is also applied to evaluate the closeness of a single article to a category. The rationality of using a category feature set for single document is not self-evident, because it is not always reasonable to evaluate single document possibly with a smaller set of important features through comparing with a category feature vector obtained typically from a large set of samples of the corresponding category and usually with a larger set of features. For this reason, appropriate weights, feature selection and extraction are often done for extracting important features. In a different spirit, an acceptance measure considers the different positions of a category and a document when a comparison is taken, and indicates the level of acceptance of a document having a category as the comparison basis. It is clear that an acceptance measure should be more appropriate for TC, when the representation is based on background net (Table 9) without the processing of feature selection gone through.

As previously mentioned, associations between terms can be more important than isolated terms for understanding concepts. Now there are more points for discussion. The first question is about the right range to capture terms' association. Using entire article as block without partition will obviously cause too many associations between terms and make less meaningful for a specific association in article selection. On the extreme, however, directly using single terms without association will go back to the original single-term feature based methods. While a simplest way is to use natural sentence in article for partition, other methods are also possible to partition articles for capturing terms association. The second question is the right granularity of term association. How many terms should be considered for terms association? Too many terms involved in association will cause too much overlap between granules and thus less significant of each association in

TABLE 11. Micro- and macro- averaged F1 on Reuters dataset

		Partition Methods			
		". ! ?"	**30**	**60**	**90**
τ	**1**	0.763	0.763	0.763	0.763
	2	0.936	0.942	0.939	0.933
	3	0.947	0.949	0.952	0.952
	4	0.945	0.949	**0.954**	0.950

(a) Micro-averaged F1

		Partition Methods			
		". ! ?"	**30**	**60**	**90**
τ	**1**	0.641	0.641	0.641	0.641
	2	0.871	0.879	0.860	0.847
	3	0.892	0.894	0.890	0.886
	4	0.889	0.894	**0.900**	0.881

(b) Macro-averaged F1

TABLE 12. Micro- and macro- averaged F1 on 20Newsgroups dataset

		Partition Methods			
		". ! ?"	**30**	**60**	**90**
τ	**1**	0.778	0.778	0.778	0.778
	2	0.822	0.829	**0.839**	0.837
	3	0.775	0.822	0.822	0.836

(a) Micro-averaged F1

		Partition Methods			
		". ! ?"	**30**	**60**	**90**
τ	**1**	0.775	0.775	0.775	0.775
	2	0.819	0.828	**0.837**	0.834
	3	0.773	0.819	0.819	0.833

(b) Macro-averaged F1

understanding concept of term, while the other extreme is taking only one term without association which simply goes back to the original single-term feature based methods. We have mainly used 2-term association in our experiments, but also carried out comparisons with different settings. The experiments have shown that m-term associations ($m > 2$) contributed very little in article selection.

Further experiment of Algorithm-3 and Algorithm-4 introduced in Section 4 is carried out with different methods of partition in capturing combinations of words. The results are summarized in Table 11 and Table 12, where, the partition methods used are 1) using the characters '.', '!' and '?' for segmenting an article into a set of sentences; 2) counting n space in article to find n-sequence-words as a block. The results show that an appropriate partition to capture terms' association contributes to the overall performance.

7. **Conclusions.** We have proposed a novel approach for personalized article selection as a special case of text categorization with changing domain and personalized criteria. We use background net built up through learning to capture background domain based on a collection of articles. The learning is achieved by using the contextual association of terms appeared in sample articles provided and carried out in an incremental manner to expend term set as well as term associations. Algorithms are provided for learning and text categorization, and their complexities are briefly analyzed. The reasoning on background net is based on "focus spreading" spirit, so computational complexity can be controlled when the entire background net is expending. Experiment results show that our approach works well (in some cases, better) compared with representative methods in text categorization. An important and probably a unique feature of our approach is that it does not require the processing of feature selection, so that relieves the requirement of relatively large sample data set to support the training of classifiers. As a natural consequence of this feature, our approach has more potential in handling text categorization with open domain and changing criteria. Our future work will be focused on several aspects: (a) to explore a more effective use of the quantity measure of associations for further improving the performance of background net in application of personalized article selection as well as TC; (b) to further study how to utilize the statistical information in a collection of

articles to improve performance of categorization with categories that are fuzzy granules of content domain.

Acknowledgment. This work was supported in part by the Macao Science and Technology Development Fund under grant 040/2010/A. The authors express heartfelt gratitude to anonymous reviewers whose comments and suggestions had greatly helped improving the article.

REFERENCES

[1] S. Brin, L. Page, R. Motwani and T. Winograd, The pagerank citation ranking: Bringing order to the web, *Technical Report*, Stanford Digital Library, 1999.

[2] J. Chen and Z. Yang, A learning multi-agent system for personalized information filtering, *Proc. of the 2003 Joint Conference of the 4th International Conference on Information, Communications and Signal Processing*, vol.3, pp.1864-1868, 2003.

[3] C. Makris, Y. Panagis, E. Sakkopoulos and A. Tsakalidis, Category ranking for personalized search, *Data & Knowledge Engineering*, vol.60, no.1, pp.109-125, 2007.

[4] A. Micarelli, F. Gasparetti, F. Sciarrone and S. Gauch, Personalized search on the world wide web, in *The Adaptive Web: Methods and Strategies of Web Personalization*, P. Brusilovsky, A. Kobsa and W. Nejdl (eds.), Springer, 2007.

[5] M. Speretta and S. Gauch, Personalized search based on user search histories, *IEEE/WIC/ACM International Conference on Web Intelligence*, pp.622-628, 2005.

[6] J. Teevan, S. T. Dumais and E. Horvitz, Personalizing search via automated analysis of interests and activities, *Proc. of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.449-456, 2005.

[7] Z. Zhao, C. Li and S. Feng, Mining users' interest model based on behavior logs, *Bulletin of Advanced Technology Research*, vol.5, no.1, 2011.

[8] T. Takagi and K. Kawase, A trial for data retrieval using conceptual fuzzy sets, *IEEE Trans. on Fuzzy Systems*, vol.9, no.4, pp.497-505, 2001.

[9] K. Aas and L. Eikvil, Text categorization: A survey, *Technical Report*, Norwegian Computing Center, 1999.

[10] L. Lv and Y. Liu, Research of English text classification methods based on semantic meaning, *Information and Communications Technology*, pp.689-700, 2005.

[11] Q. Wu, E. Fuller and C. Zhang, Text document classification and pattern recognition, *Proc. of International Conference on Advances in Social Network Analysis and Mining*, pp.405-410, 2009.

[12] X.-Q. Yang, N. Sun, T.-L. Sun, X.-Y. Cao and X.-J. Zheng, The application of latent semantic indexing and ontology in text classification, *International Journal of Innovative Computing, Information and Control*, vol.5, no.12(A), pp.4491-4499, 2009.

[13] D. Bracewell, J. Yan and F. Ren, Single document keyword extraction for Internet news articles, *International Journal of Innovative Computing, Information and Control*, vol.4, no.4, pp.905-913, 2008.

[14] A. Leubner and W. Kießling, Personalized keyword search with partial-order preferences, *Brazilian Symp. on Database Systems*, pp.181-193, 2002.

[15] R. Feldman and J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press, New York, 2007.

[16] C. D. Manning, P. Raghavan and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.

[17] G. Salton, A. Wong and C. S. Yang, A vector space model for automatic indexing, *Communications of the ACM*, vol.18, no.11, pp.613-620, 1975.

[18] M. Lan, C. L. Tan, J. Su and Y. Lu, Supervised and traditional term weighting methods for automatic text categorization, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.31, no.4, pp.721-735, 2009.

[19] V. Lertnattee and T. Theeramunkong, Effect of term distributions on centroid-based text categorization, *Information Sciences*, vol.158, no.1, pp.89-115, 2004.

[20] S. Li, R. Xia, C. Zong and C. R. Huang, A framework of feature selection methods for text categorization, *Proc. of Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pp.692-700, 2009.

[21] R. Jensen and Q. Shen, Are more features better? A response to attributes reduction using fuzzy rough sets, *IEEE Trans. on Fuzzy Systems*, vol.17, no.6, pp.1456-1458, 2009.
[22] M. Montes-y-Gomez, A. Lopez-Lopez and A. F. Gelbukh, Information retrieval with conceptual graph matching, *Proc. of the 11th Int. Conf. Database Expert Syst. Appl.*, pp.312-321, 2000.
[23] W. Jin and R. K. Srihari, Graph-based text representation and knowledge discovery, *Proc. of the 22nd ACM Symp. Appl. Comput.*, New York, pp.807-811, 2007.
[24] A. Schenker, M. Last, H. Bunke and A. Kandel, Classification of web documents using a graph model, *Proc. of the 7th International Conference on Document Analysis and Recognition*, pp.240-244, 2003.
[25] J. Y. Choi, Clustering analysis of collaborative tagging systems by using the graph model, *Technical Report*, Indiana University Bloomington, 2008.
[26] D. A. Buell, An analysis of some fuzzy subset applications to information retrieval systems, *Fuzzy Sets System*, vol.7, no.1, pp.35-42, 1982.
[27] Y. Ogawa, T. Morita and K. Kobayashi, A fuzzy document retrieval system using the keyword connection matrix and a learning method, *Fuzzy Sets System*, vol.39, no.2, pp.163-179, 1991.
[28] H. H. Huang and Y. H. Kuo, Cross-lingual document representation and semantic similarity measure: A fuzzy set and rough set based approach, *IEEE Trans. on Fuzzy Systems*, vol.18, no.6, pp.1098-1111, 2010.
[29] S. L. Lo, L. Ding and Y. Chen, Incremental learning on background net to capture changing personal reading preference, *Proc. of the 2011 International Conference on Machine Learning and Cybernetics*, pp.1503-1508, 2011.
[30] Y. Chen, L. Ding, S.-L. Lo and W. Sun, Background net for personalized keywords on article selection, *ICIC Express Letters*, vol.5, no.4(B), pp.1339-1346, 2010.
[31] S. Kendal and M. Creen, *An Introduction to Knowledge Engineering*, Springer-Verlag, 2007.
[32] L. A. Zadeh, Fuzzy sets, *Information and Control*, vol.8, no.3, pp.338-353, 1965.
[33] S. L. Lo and L. Ding, Learning and reasoning on background net – Its application to text categorization, *ICIC Express Letters*, vol.6, no.3, pp.625-631, 2012.
[34] S.-L. Lo and L. Ding, Incremental learning of concept cluster on background net for fuzzy granulation of document representation, *ICIC Express Letters*, vol.6, no.3, pp.699-704, 2012.
[35] Y. Yang and X. Liu, A re-examination of text categorization methods, *Proc. of Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, NY, USA, pp.42-49, 1999.
[36] C. C. Chang and C. J. Lin, *LIBSVM: A Library for Support Vector Machines*, http://www.csie.ntu.edu.tw/cjlin/libsvm, 2001.