Relevance feedback for category search in music retrieval based on semantic concept learning

Man-Kwan Shan · Meng-Fen Chiang · Fang-Fei Kuo

Published online: 5 March 2008

Abstract Traditional content-based music retrieval systems retrieve a specific music object which is similar to what a user has requested. However, the need exists for the development of category search for the retrieval of a specific category of music objects which share a common semantic concept. The concept of category search in content-based music retrieval is subjective and dynamic. Therefore, this paper investigates a relevance feedback mechanism for category search of polyphonic symbolic music based on semantic concept learning. For the consideration of both global and local properties of music objects, a segment-based music object modeling approach is presented. Furthermore, in order to discover the user semantic concept in terms of discriminative features of discriminative segments, a concept learning mechanism based on data mining techniques is proposed to find the discriminative characteristics between relevant and irrelevant objects. Moreover, three strategies, the Most-Positive, the Most-Informative, and the Hybrid, to return music objects concerning user relevance judgments are investigated. Finally, comparative experiments are conducted to evaluate the effectiveness of the proposed relevance feedback mechanism. Experimental results show that, for a database of 215 polyphonic music objects, 60% average precision can be achieved through the use of the proposed relevance feedback mechanism.

Keywords Category search · Music retrieval · Relevance feedback · Semantic concept learning

M.-K. Shan (✉) · M.-F. Chiang
Department of Computer Science, National Chengchi University, Taipei, Taiwan
e-mail: mkshan@cs.nccu.edu.tw

M.-F. Chiang
e-mail: g9309@cs.nccu.edu.tw

F.-F. Kuo
Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan
e-mail: ffkuo@csie.nctu.edu.tw

1 Introduction

With the advance of multimedia technologies, the demand for storing, analyzing and accessing multimedia data has increased. One demand is for the content-based music retrieval system which searches for a music object by analyzing its content. Most content-based music retrieval systems attempt to retrieve a specific object which is similar to the one being sought. However, instead of searching for a specific target music object (target search), there is a need for retrieving music objects that conceptually belong to a particular category (category search). Music objects retrieved by category search share a common user-defined concept. For example, when a user wishes to search for romantic music objects, category search will retrieve music objects which are romantic in nature.

However, in category search, the user concept formed is subjective and dynamic. Different users at different times may have different concepts for the same set of music objects. Moreover, the taxonomy of categories of music objects with respect to user conceptualization can not be constructed in advance or in a fixed way. In order to address this problem, an on-line, user-dependent learning process is needed. In such a process, users are expected to be involved in the learning process as relevance feedback provided by them is needed to improve the retrieval results. Therefore, in each query session of category search, the process may require several rounds until the user is satisfied with the results. User relevance feedback in each round enables the system to learn the user concept adaptively. As a consequence, performance improves with relevance feedback. Although much work has been done on content-based music retrieval, little attention has been paid to the design of a relevance feedback mechanism for music retrieval 0. This paper investigates the relevance feedback mechanism for the category search of polyphonic symbolic music objects in content-based music retrieval systems.

Relevance feedback is developed originally for text retrieval. In text retrieval, most research uses the vector space model to represent a document. Vector space model represents each document as a weighted term frequency vector. Relevance feedback is achieved by modification of query vectors such as query expansion or term reweighting. Rochio formulation is one of the most popular query expansion approaches in which the positive and negative feedbacks from users are considered [2].

In the mid-1990, relevance feedback was introduced into image retrieval [26]. In image retrieval, most research on the relevance feedback models the whole image as a feature vector. Semantic concepts are described in terms of discriminative image features. Some approaches in image retrieval have been extended to deal with local features by segmenting an image into regions [11, 12]. Similar segmented regions from all images in image database are grouped into region clusters. An image is therefore represented as a weighted vector with each dimension corresponding to a region cluster. In these approaches, semantic concepts are described in terms of discriminative image regions, rather than discriminative image features.

In music retrieval, the semantic concepts are best described by discriminative music features which are further described by a subset of feature representations. For instance, the concept for the category of inspiring music can be characterized by a high tempo rate and a high average pitch interval. Furthermore, in music retrieval, global features corresponding to the entire piece of music and local features with respect to music segments should be considered. For example, to characterize the concept for a favorite piece, e.g. the four-note opening motif in Beethoven's Symphony No. 5 in C Minor, local features, e.g. the average pitch interval and the density of this four-note music segment, are useful. In order to accommodate a specific concept which may comprise either an entire music object or only

part of it, a segment-based music representation is proposed here. In this approach, a music object is treated as a whole, as well as a set of music segments. These music segments are extracted from a music object based on music theory.

Moreover, while most relevance feedback research in text retrieval and image retrieval adopts the vector space model, query expansion techniques (e.g. Rochio) and classification techniques (e.g. Support Vector Machine) are popular approaches for learning a user concept [26]. However, these techniques are not suitable for the discovery of concepts in terms of discriminative music features of discriminative segments. Therefore, this paper proposes a binary classification algorithm based on the data mining techniques to discover the user concept in terms of the discriminative music features of the discriminative music segments. A binary classifier, with respect to the user concept, is discovered through user feedback and is then employed to classify music objects in the next round.

This paper is organized as follows. Related work about relevance feedback and music retrieval is reviewed in Section 2. The music modeling approach is described in Section 3. Section 4 presents the semantic concept learning algorithm. The relevance feedback methods are listed in Section 5. The experimental results are presented in Section 6. Finally, conclusions are drawn in Section 7.

2 Related work

2.1 Relevance feedback in image retrieval

Relevance feedback has attracted much attention in content-based image retrieval research. Its objective is to establish a link between a high level semantic concept and low level image features based on user feedback [26]. Major approaches on relevance feedback in image retrieval include the query point movement, feature re-weighting and machine learning (classification) algorithms.

The query point movement approach attempts to reformulate a new query point which is closer to the relevant results and farther from the irrelevant ones. In the feature re-weighting approach, each image is modeled as a set of visual features, each of which is associated with a set of representations. Relevance feedback is achieved by dynamically updating the weights of features, as well as the weights of feature representations, in order to accommodate the information needs of the user [4, 21]. Recent work on image retrieval treats the relevance feedback problem as a classification problem [6, 7, 24] in which machine learning techniques, for example, support vector machines, are employed. In such a paradigm, the goal is to discover a decision boundary based on relevant and irrelevant multimedia object information collected from the user.

In order to satisfy user focus, some research has extended the relevance feedback mechanisms from global image features to region-based ones [6, 11, 12]. In these approaches, each image is segmented into several regions. Then, the similar regions from all images are clustered into region clusters. One approach represents each image as a weighted vector with each dimension corresponding to a region cluster [11, 12]. Traditional relevance feedback methods, e.g. query point movement, can be employed to find the optimal query. The other approach uses SVM to represent an image as a set of regions. This creates a variable length image representation for which standard SVM kernels are not suitable. Therefore, a generalization of the Gaussian kernel with Earth Mover's Distance is introduced to find the optimal query [6].

Both of these region-based approaches capture the user semantic concept by finding the optimal *region importance*. The principles of the region-based approaches for image

retrieval can be applied to the proposed segment-based approach for music retrieval, provided that the music segmentation technique is available. However, it should be noted here that *feature importance* is not considered in the region-based image retrieval. As the user semantic concept for music retrieval is best described by discriminative music features as well as discriminative music segments, the proposed learning algorithm can discover the discriminative feature representations of the discriminative segments that constitute the information needs of a user.

2.2 Relevance feedback in music retrieval

Traditional music retrieval systems focus on target search to locate a specific music object similar to a query specified by a user [15, 20]. Little attention has been paid to category search in music retrieval. Moreover, while much work has been done on music retrieval, relevance feedback has attracted little attention.

As far as the authors are aware, the earliest research on relevance feedback in music retrieval was proposed by Mandl and Womser-Hacker [17, 18]. In this approach, a model for music retrieval which automatically adapts to preference of users is proposed. Learning is achieved by a fusion approach which is a linear combination of the results of different music representation scheme. It is adapted according to the success in previous rounds where the success is measured by the relevance feedback given by users.

In [8, 9], a music retrieval system was proposed for acoustic music search based on user preference. Two types of profiles based on user ratings and genre preferences were constructed to form the prior knowledge. After TreeQ vector quantization of the music [5], vector representation was employed to represent these two types of profiles. To accommodate possible user dissatisfaction with the initial retrieval results, a relevance feedback mechanism based on query point movement was presented to improve the retrieval results. Mandel et al. developed a relevance feedback mechanism in music retrieval using SVM [16]. In that model, a number of audio features based on the Gaussian mixture models of mel frequency cepstral coefficients (MFCCs) [19] are extracted for music representation. Therefore, each piece of music is represented as a fixed-length vector.

While the aforementioned two approaches deal with acoustic music, this paper focuses on symbolic music such as MIDI. Nonetheless, the proposed relevance feedback mechanism, with the well-developed acoustic music segmentation techniques (e.g. [3]), can be extended to included acoustic music feature representation. More importantly, another advantage of the proposed approach is that, unlike the other works which consider global music features only, it takes both global and local features into consideration. The following sections present the proposed approach in more detail.

3 Music object modeling

The modeling process consists of four steps. The first step attempts to extract *potential significant segments* for each music object. Next, for each potential significant segment, the *local feature representations* are extracted. Potential significant segments are regarded as the same if their local feature representations are identical. Then, the *significant segments* are selected from the potential significant segments based on their importance. Finally, for each music object, the *global feature representations* are extracted from the entire music object.

3.1 Potential significant segment extraction

In musicology, a motive is a salient recurring segment of notes that may be used to construct all or some of the melody and themes. The repetition of a motive may have some variations and not necessarily be an exact repetition in the music object [23]. We address a repeating pattern with *motivic variations* as *motivic repeating pattern*. Motivic repeating patterns may be regarded as potential significant segments to characterize the melody feature of a music object for content-based music retrieval.

In our work, six common types of motivic variations are considered: repetition, transposition, sequence, contrary motion, retro-gradation, and augmentation/diminution. Figure 1 lists examples of these motivic variations where each occurrence of the motives is marked with a solid box. These six types of motivic variations are described as follows:

(1) Repetition: The repetition is the exact repeating, where Fig. 1a is an example.
(2) Transposition: In the transposition (Fig. 1b), a motive repeats at another pitch level.
(3) Sequence: The sequence refers to the moving of a motive in pitch in a constant level. It is a specialization of transposition. Figure 1c shows an example where the motive is moved downwards by two semitones. The variation in Fig. 1b doesn't belong to the sequence type. The second occurrence of the motive is moved downwards by two semitones while the third occurrence is moved upwards by two semitones.
(4) Contrary motion: The contrary motion refers to a motive where interval directions have been made to move in the opposite direction of the original motive. For example, in Fig. 3d, the interval sequence, in semitones, of the original motive is "2, 2, 1, -3, 2, -4" while that of the first variation is "-2, -2, -1, 3, -2, 4".
(5) Retro-gradation: In the retro-gradation, the pitches of a motive are repeated in reverse order. For example, in Fig. 1e, the pitch sequence of the original motive is "Fa, Fa, Sol, Sol, Si, Do, Do" while that of the variation is "Do, Do, Si, Sol, Sol, Fa, Fa".
(6) Augmentation/Diminution: In the augmentation/diminution, a motive is repeated while the rhythmic durations are proportionately doubled or halved. In Fig. 1f, the rhythmic durations of the former variation are doubles while those of the latter are halved.

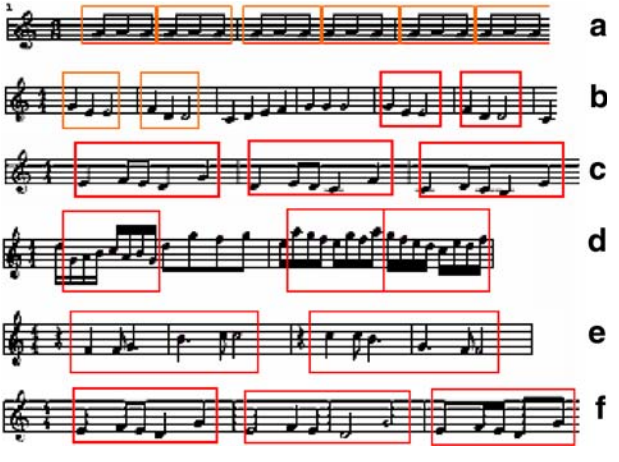


Fig. 1 Examples of six common types of motivic variations. (**a**) Repetition (**b**) Transposition (**c**) Sequence (**d**) Contrary motion (**e**) Retro-gration (**f**) Augmentation/Diminution [23]

To extract the motivic repeating patterns from a polyphonic music object, first of all, a well-known melody extraction method, all-mono, developed by Uitdenbogerd and Zobel [25], is employed to extract the main melody. All-mono combines all the music tracks and, among the simultaneous notes, includes the highest notes as the main melody. The extracted main melody is represented as the note sequence in which each note is expressed by its pitch and duration.

After that, in order to discover the motivic repeating patterns from the note sequence, the correlative matrix method is modified. The correlative matrix method, proposed by Shih et al., was originally designed to find exact repeating patterns [10]. In this method, a data structure called correlative matrix is constructed by lining up the note sequence the horizontal and vertical dimensions respectively to keep the intermediate information of substring matching. Each cell of this matrix denotes the length of a founded repeating pattern. After the construction of this matrix, the repeating frequencies of all repeating patterns can be derived by computing the non-zero-cells.

To discover the motivic repeating patterns which are not necessarily exact repetition, the conditions of substring matching in each cell of the correlative matrix need to be modified in order to accommodate the motivic variations. For example, to discover the retrogradation, each cell of the correlative matrix should consult the lower-right cell, rather than the upper-left cell in exact repetition finding. More details concerning the motivic repeating finding algorithms can be found in earlier work completed by the authors [22].

3.2 Feature representations

In the current implementation, each music object is modeled as a six-attribute global feature and a set of five-attribute local features where each set corresponds to a significant segment. Table 1 lists the six attributes for global feature representation and the five attributes for local feature representation. In this study, the music features considered include tempo, rhythm, and melody. It should be noted that the proposed relevance feedback mechanism is an open framework to allow other music features or feature representations to be incorporated if needed. Detail descriptions of these features and associated attributes are listed as follows:

(1) Tempo: Tempo denotes the speed of a music object and is defined as the number of beats per minute.
(2) Rhythm: Rhythm features are represented by density attribute. The density of a music piece is defined as the number of notes divided by its total duration.

Table 1 Attributes of global and local feature representations

Global features	Local features
Density (GD)	Density (LD)
Tempo (GT)	Average Pitch (LAP)
Average Pitch (GAP)	Pitch Standard Deviation (LPSD)
Pitch standard deviation (GPSD)	Average Pitch Interval (LAPI)
Highest Pitch (GHP)	Chord Sequence (LCS)
Lowest Pitch (GLP)	

(3) Melody: Melody feature is represented by the following attributes.

(a) Average Pitch: This indicates the average pitch value of notes within a music piece (an entire one or a segment).
(b) Pitch Standard Deviation: This is the deviation of pitch values of notes within a music segment.
(c) Highest and Lowest Pitch: These are global representations for entire music object.
(d) Average Pitch Interval: This indicates the average pitch differences between each pair of two consecutive notes within a music segment. It should be noted that instead of using the highest and the lowest pitch, the average pitch interval is used to represent local melody feature. This is because the average pitch interval captures the pitch variation between consecutive notes and seems more meaningful for a music segment containing a few notes.
(e) Chord Sequence: This is a representation for melody style proposed in an earlier work by the authors [13]. A chord sequence is a sequence of chords within a segment. Two music segments with similar melody lines are not necessarily of the same styles. In harmony, it is the chords with which a melody was accompanied. Therefore, the chord is utilized as one of the melody feature representations. The chord sequence is generated by using a chord assignment algorithm, which is a heuristic method based on harmony theory. The chord assignment algorithm first determines the chord sampling unit (e.g. a measure). Then for each unit, this algorithm chooses 60 common chords as the candidates, and counts the score of each candidate chord according to some heuristic rules from harmony. The chord with the highest score is assigned to the unit. In our implementation, each chord sequence is represented as an integer.

Figure 2 illustrates an example of a music object along with its feature representation. There exist three significant segments in this music object.

3.3 Significant segment selection

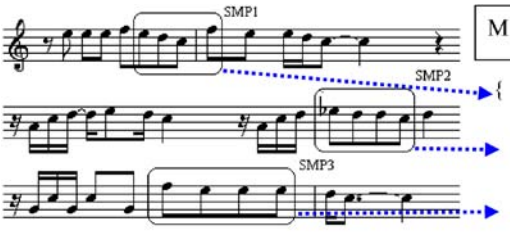
Significant segments are selected from potential significant segments based on their importance (weights). The rationale of segment importance is inspired by the Term Frequency and Inverse Document Frequency weighting scheme widely used in text retrieval [2].

Definition 1 The weight $w(p, m)$ of a potential significant segment p with respect to a music object m is given by:

$$w(p,m) = \frac{f(p,m) \Big/ \max_{p' \in m} f(p,m)}{\sup(p, DB)} \tag{1}$$

where $f(p, m)$ is the frequency (the number of occurrences) of the potential significant segment p in the music object m, $\sup(p, DB)$ is the support of p (the number of music objects in which p appears) in the database DB.

A potential significant segment with a higher occurrence rate in a music object is not necessarily more important than the one with a lower occurrence rate in the other music object. Therefore, the frequency of a potential significant segment p, $f(p, m)$, is normalized



M	GD	GAP	GPSD	GT	GHP	GLP
	1.7	73	3.5	7	81	72

LD	LAP	LPSD	LAPI	LCS
2	75	2	3	0

LD	LAP	LPSD	LAPI	LCS
1.1	74	2	1	7

LD	LAP	LPSD	LAPI	LCS
0.4	76	2	0.3	3

Fig. 2 An example of a music object with three significant segments and its feature representation (excerpt from the song "Don't let the sun go down on me.")

by dividing the frequency of the most frequent potential significant segment p' in music m. Moreover, a potential significant segment is more important if it is more specific in the music database DB. Therefore, the importance of a potential significant segment is divided by the support of p in the database DB. To select significant segments, a *significance threshold* is employed to discard those potential segments with less importance.

4 Semantic concept learning

In order to discover the user concept for the category search, the semantic concept learning process is performed on the accumulated user feedback in each round. Two databases, a positive one (MDB^P) and a negative one (MDB^N), are constructed in each round to store the relevant music objects and the irrelevant objects respectively. The concept learning process should be able to capture the user concept by separating the common characteristics of the positive music objects from those of the negative ones.

The proposed concept learning mechanism is a binary classification algorithm. There are two major steps. The first step finds the *common characteristics* of positive music objects and negative objects respectively. The second step discovers the *discriminative characteristics* to distinguish the common characteristics of positive ones from those of negative ones. The details are provided in the following subsections.

4.1 Common characteristic mining

Before the process of common characteristic mining, the positive database and the negative database are respectively transformed into the music-item forms. A *music-item* is a pair of (A, v) where A is an attribute of music feature representation, v is its corresponding value. In music-item form, the global feature of a music object is represented as an itemset of six music-items, while the local feature of a significant segment is represented as an itemset of five music-items. A music object with m segments is therefore treated as a set of $(m+1)$ itemsets.

Example 1 Table 2 is an example of positive database MDB^P containing four music objects and Table 3 is an example of negative database MDB^N with three music objects. Table 4 and Table 5 are the music-item forms of Table 2 and Table 3 respectively. To put it simply, a music-item is denoted by concatenating the attribute name with its value. Moreover, in these tables, a music object is associated with a two-attribute global feature (G, H) and a set of three-attribute local features (A, B, C) in which each three-attribute local

Table 2 An example of positive database MDB^P

p_1	G	H	{	A	B	C	,	A	B	C	,	A	B	C	}								
	4	**2**		**1**	5	**1**		2	1	1		1	1	3									
p_2	G	H	{	A	B	C	,	A	B	C	,	A	B	C	}								
	1	**2**		2	2	2		**1**	6	**1**		3	4	2									
p_3	G	H	{	A	B	C	,	A	B	C	}												
	1	**2**		**1**	5	**1**		3	5	3													
p_4	G	H	{	A	B	C	,	A	B	C	}												
	4	**2**		**1**	3	**1**		1	3	2													

feature corresponds to a significant segment. For instance, the object p_3, in Table 2, is associated with a global feature and two local feature representations. This indicates that there are two significant segments extracted from the music object p_3.

Concerning the common characteristics of a database which is a collection of sets of itemsets, the followings are the formal statements.

Definition 2 Let I be the set of all possible music-items, a music object $\mathcal{T} = \{T_1, T_2, \ldots, T_m | \forall i, 1 \le i \le m, T_i \subseteq I\}$ is said to *contain* a *music-pattern* $\mathcal{P} = \{P_1, P_2, \ldots, P_n | \forall j, 1 \le j \le n, P_j \subseteq I\}$ if there is a one-to-one mapping function from P to T such that for each j, there exists an i, where $P_j \subseteq T_i$. $\mathcal{P}$ is called a size-n music-pattern.

Definition 3 Let DB be a collection of music objects, the support of a music-pattern $\mathcal{P}$, $sup(\mathcal{P})$, is the percentage of music objects in DB that contain $\mathcal{P}$. If the support of a music-pattern $\mathcal{P}$, $sup(\mathcal{P})$, is no less than a user-specified minimum support threshold *minsup*, P is called a *frequent music-pattern.*

Definition 4 The frequent music-patterns found in the positive database and the negative database are called the *positive frequent music-patterns* and the *negative frequent music-patterns*, respectively.

Example 2 In Table 5, the music object n_1={{G1, H2}, {A2, B3, C3}, {A1, B4, C1}} in MDB^N is said to contain the music-pattern $\mathcal{P}$ ={{B3), {A1, C1}}, because {B3}⊆{A2, B3, C3} and {A1, C1}⊆{A1, B4, C1}. While both music objects n_1 and n_2 contain the pattern $\mathcal{P}$, the support of $\mathcal{P}$ is 2/3. If the minimum support threshold *minsup* is 2/3, then $\mathcal{P}$ is a frequent music-pattern. Figure 3 lists all the positive frequent music-patterns and all the negative frequent music-patterns.

To discover the frequent music patterns from a collection of music objects, we propose an Apriori-based algorithm. *Apriori* [1] is a well-known data mining approach originally developed for the discovery of frequent itemsets from a database of itemsets. In our work, each music object is a set of itemsets and each frequent music pattern is also a set of itemsets, rather than an itemset. Therefore, a two-phase mining algorithm is proposed here. The first phase finds all the frequent itemsets and the second phase discovers the frequent

Table 3 An example of negative database MDB^N

n_1	G	H	{	A	B	C	,	A	B	C	,	A	B	C	}								
	4	2		2	**3**	1		1	1	3		2	**3**	3									
n_2	G	H	{	A	B	C	,	A	B	C	}												
	1	2		6	**3**	2		3	6	1													
n_3	G	H	{	A	B	C	}																
	4	3		5	5	1																	

Table 4 The music-item form of MDB^P in Table 2

Object ID	Set of Itemsets
p_1	{{G4,H2}, {A1,B5,C1}, {A2,B1,C1}, {A1,B1,C3}}
P_2	{{G1,H2}, {A2,B2,C2}, {A1,B6,C1}, {A3,B4,C2}}
P_3	{{G1,H2}, {A1,B5,C1}, {A3,B5,C3}}
P_4	{{G4,H2}, {A1,B3,C1}, {A1,B3,C2}}

music-patterns constituted by the frequent itemsets found in the first phase. It should be noted that the itemsets found in the first phase correspond to the music segment level, while the music patterns (sets of itemsets) found in the second phase correspond to the music object level.

Phase 1: Mining Frequent Itemsets

The Apriori algorithm is employed to discover all frequent itemsets in which each item must appear in the same itemset [1]. The classic Apriori algorithm for the discovery of frequent itemsets makes multiple passes over the database. In the first pass, the support of each individual item is calculated and those above the *minsup* are kept as a seed set. In the subsequent pass, the seed set is used to generate new potentially frequent itemsets, namely candidate itemsets. Then the support of each candidate itemset is calculated by scanning the database. The candidates with support no less than the *minsup* are the frequent itemsets and are fed into the seed set that will be used for the next pass. The process continues until no new frequent itemsets are found.

In this work, only the support calculation step is different from that of the classic *Apriori* algorithm. This difference is due to the fact that in this work each object is a set of itemsets, rather than just one itemset.

Phase 2: Mining Frequent Sets of Itemsets

The second phase discovers the patterns constituted by frequent itemsets found in the first phase. Similar to the algorithm in the first phase, the algorithm in this phase makes multiple passes over the databases. In the k-th pass, the seed set (the set of size-k candidate patterns) is generated by joining two frequent size-(k-1) patterns found in the previous pass. Then the support of each candidate pattern is calculated by scanning the database. The candidates with a support of no less than *minsup* are considered the frequent patterns and are fed into the seed set that will be used for the next pass. The process continues until no new frequent patterns are found. The only exception is the first pass in which the seeds are the frequent itemsets generated in the first phase.

4.2 Discriminative characteristic mining

After the two-phase mining process performed on MDB^P and MDB^N, the collections of positive and negative frequent music-patterns are obtained, respectively. This step attempts to find the discriminative characteristics which distinguish the positive frequent patterns

Table 5 The music-item form of MDB^N in Table 3

Object ID	Set of Itemsets
n_1	{{G4,H2}, {A2,B3,C1}, {A1,B1,C3}, {A2,B3,C3}}
n_2	{{G1,H2}, {A6,B3,C2}, {A3,B6,C1}}
n_3	{{G4,H3}, {A5,B5,C1}}

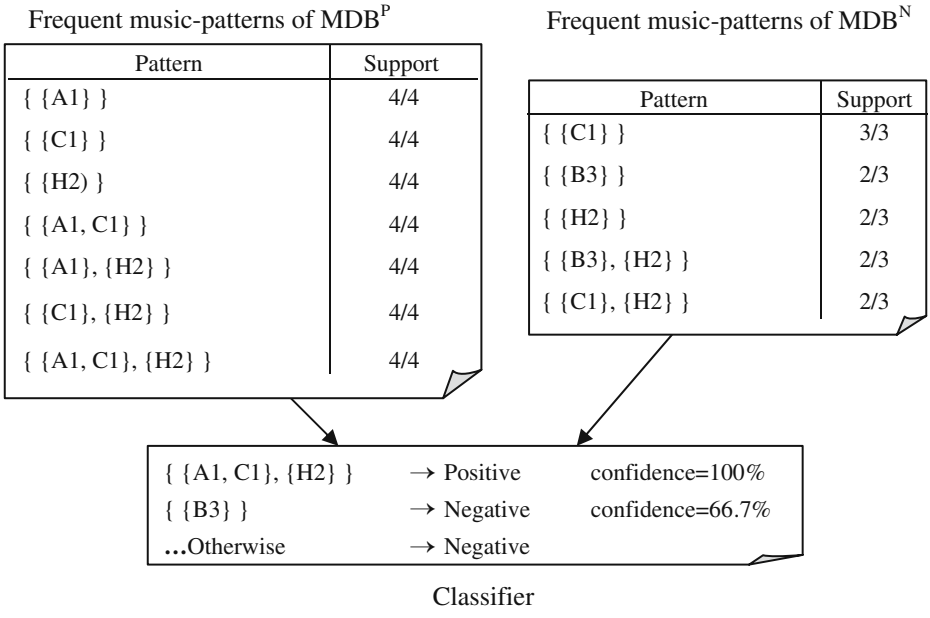


Fig. 3 An example of a classifier learned from Tables 4 and 5

from negative patterns. The result of this step is a binary classifier consisting of classification rules.

Example 3 Figure 3 is an example of a classifier, expressed as an ordered set of classification rules, learned from the feedback information contained in Table 4 and Table 5. One example of a classification rule is "{{A1,C1}, {H2}} → positive" which is derived from the fact that {{A1,C1},{H2}} appears frequently in MDB^P but never appear in MDB^N.

To generate a binary classifier learned from the positive and negative frequent patterns, the *associative classification* algorithm 0, proposed by Liu et al., is employed. This algorithm eventually generates a classifier containing a set of ordered rules. The classifier is of the form $<r_1, r_2,\ldots, r_c, default\_class>$. Each rule r_i is of the form $l \Rightarrow y$, where l is a frequent positive or negative music-patterns and y is a class label which is either positive or negative. The *confidence* of a rule is defined as the percentage of the music objects that contains l belonging to class y.

Example 4 In Fig. 3, the classifier is of the form $<r_1, r_2, negative>$ where the confidence of the second rule r_2: "{{B3}} → negative" is 2/3. The music-pattern {{B3}} is contained in the music objects p_4, n_1 and n_2, where p_4 actually belongs to positive objects.

To generate a classifier, the associative algorithm sorts the rules according to their confidence. Then, the rules that correctly classify at least one music object are selected and are retained as the potential rules in the classifier. A default class referred to as the majority class of the remaining music objects in the database is appended as the last rule of the classifier. Finally, the rules that do not improve the accuracy of the classifier are discarded. The first rule in the classifier that makes the least number of recorded errors is the cut off rule: Subsequent ones are discarded since they only produce more errors. Unlike the

original associative classification, in this study, the frequent pattern on the left-hand side of a rule is a set of itemsets.

5 Relevance feedback method

Once the classifier is constructed by the concept learning process, the system then produces a ranked list of music objects for the next round. Three types of system feedback strategies are investigated here: the Most-Positive, the Most-Informative and the Hybrid. In general, the most-informative music objects will not necessarily be the most-positive music objects. The three strategies along with their corresponding ranking functions are described as follows.

(1) The Most-Positive Strategy (MP)

This strategy displays those music objects which are classified as the most relevant by the system based on previous training. The most-positive music is a list of music objects ordered by the following ranking function:

$$\mathrm{Score}_{\mathrm{MP}}(m) = \frac{\sum_{r \in Rp(m)} \mathrm{conf}(r)}{\sum_{r \in Rp(m) \cup R_n(m)} \mathrm{conf}(r)} \tag{2}$$

where *conf(r)* is the confidence value of the classification rule r, $Rp(m)$ and $Rn(m)$ are the sets of the positive and the negative rules that satisfy the music object m, respectively.

(2) The Most-Informative Strategy (MI)

By sacrificing the performance in the current round and maximizing the information obtained for the next round, a better result can be expected in subsequent rounds. The Most-Informative strategy returns the music objects whose labels the system is most uncertain about. The system which uses the MI strategy displays a collection of the most-informative objects in each round until the user attempts to find out what the system can retrieve at that point. Then, the system switches to the MP strategy and returns the most-positive music objects.

In our work, objects belonging to the default class are selected by the system as the most-informative music objects. This is because that, in the associative classification algorithm, an object which matches no rules belongs to the default class and is likely to be uncertain.

(3) The Hybrid Strategy (HB)

The HB strategy is a compromise between the MI and MP strategies. A system which adopts the hybrid strategy returns both the most-positive and most-informative objects at each round.

6 Experimental analysis

6.1 Experiment setup

The relevance feedback information provided by users is essential to the evaluation of the performance of the relevance feedback mechanism. However, it is difficult to evaluate and

compare the effectiveness of a relevance feedback mechanism with different strategies and different parameters when experimenting with users because they have to undergo many experiments and require too much listening time. Moreover, it is also doubtful that a concept formed by a user is consistent across all experiments. The common alternative for researchers of relevance feedback in image retrieval is to use a collection of images for which the ground truth is available. In other words, a collection in which each image has been annotated with a label which corresponds to a concept.

The experiment in this paper adopts a similar approach. For this study, 215 Western pop songs in polyphonic MIDI format were collected from the Internet. Most files were downloaded from the New Zealand Digital Music Library (http://www.nzdl.org/fast-cgi-bin/music/musiclibrary). Eight subjects were invited to participate in the experiment. Two of the subjects had learned to play the piano for a few years. The others only had the basic music discipline in the school. Each user was asked to create a music concept (e.g. romantic music, or listening preference) in mind and then to label each music object either as relevant or irrelevant according to the created music concept after listening the entire music. The data labeled by a specific user is regarded as the ground truth and the feedback information with respect to the user query session. Table 6 lists these user-defined concepts and the numbers of corresponding music objects.

The retrieval process randomly selects K music objects as the initial query. The user relevance feedbacks on these K music objects were simulated based on the ground truth. The on-line semantic concept learning process generated a binary classifier based on the simulated feedback. The generated classifier evaluated all music objects in the database. According to the specified system feedback strategy, at most K music objects were returned. Note that the learning process was performed on the feedbacks accumulated from the prior rounds. As the number of training objects collected from prior rounds increases, it is expected that the classifier is refined and becomes more precise in its identification of the user concept.

6.2 Effectiveness analysis

Three experiments to compare performances were conducted. Effectiveness was measured by precision, which is defined as the ratio of the number of relevant music objects to the number of total returned music objects. Performance was measured by average precision over all user sessions.

6.2.1 Effectiveness of system feedback strategies

The first experiment observed the effect of the different system feedback strategies. In this experiment, the significance threshold was set to 0.4. The minimum support threshold *minsup* was set to 0.2. The number of music objects returned K was set to 10. For the Most-

Table 6 Description of user-defined concepts and the numbers of corresponding music objects

User ID	Concept	#(Music Objects)	User ID	Concept	#(Music Objects)
1	Rock	72	5	Inspiring music	90
2	Romantic music	38	6	Favorite music	112
3	Spirited music	27	7	Sentimental music	52
4	Soft Music	43	8	Cheerful music	99

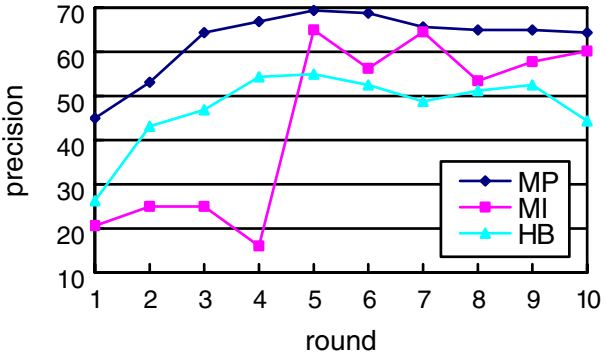


Fig. 4 Performance comparisons of three feedback strategies

Informative strategy, we assumed that the user attempts to find out the positive ones in the fifth round. In other words, the Most-Informative strategy returned the most-informative objects for the first four rounds and then returned the most-positive objects for the rest of rounds. Figure 4 presents a comparison of three system feedback strategies.

As expected, precision increased with an increase in the number of rounds for all the system feedback strategies. The MP strategy ensured a reliable performance in that more than 60% precision was achieved in the third round where 30 feedbacks were provided by the users. After the third round, all strategies maintained a 60% precision level. The MI strategy achieved 65% precision in the fifth round where 50 feedbacks were judged by the user. Subsequent to the fifth round, it maintained a precision level of 50%. The performance of the MI strategy is unexpected in that it does not outperform the MP strategy. It is possible that the training data for the MI strategy is highly unbalanced and as a consequence produced a biased classifier. In the HB feedback strategy, some of the uncertain music objects appeared in the retrieval results and hence the precision of each round was limited since uncertain music objects were not necessarily the positive ones.

From the results, it can be concluded that the MP feedback strategy is a reliable one. More specifically, it is efficient and effective: on average each user can obtain good retrieval results within three rounds. Moreover, gradual improvement via relevance feedback is ensured.

6.2.2 Effectiveness of the number of music objects returned from system

The second experiment evaluated the performance with different numbers (K) of music objects returned from the system. Other parameters were same with the first experiment.

Figure 5 shows the average precision with different K values. From this experiment it can be seen that there was no clear relationship between K and precision. Moreover, the MP strategy was relatively sensitive to parameter K, comparing with the other strategies. In Fig. 5a, the best K value for the MP strategy was 10 and it achieved the highest precision level of 70%. Users obtained 65.6% precision of returned music within three rounds. Figure 5b shows that K did not have a strong impact on the performance of the MI strategy. Since K did not affect performance, users needed only to judge from a total of 25 music objects to obtain 65% accuracy within five rounds. Although it can not achieve a precision level as high as that for the MP strategy (70%), users with limited time can still obtain satisfactory results with the

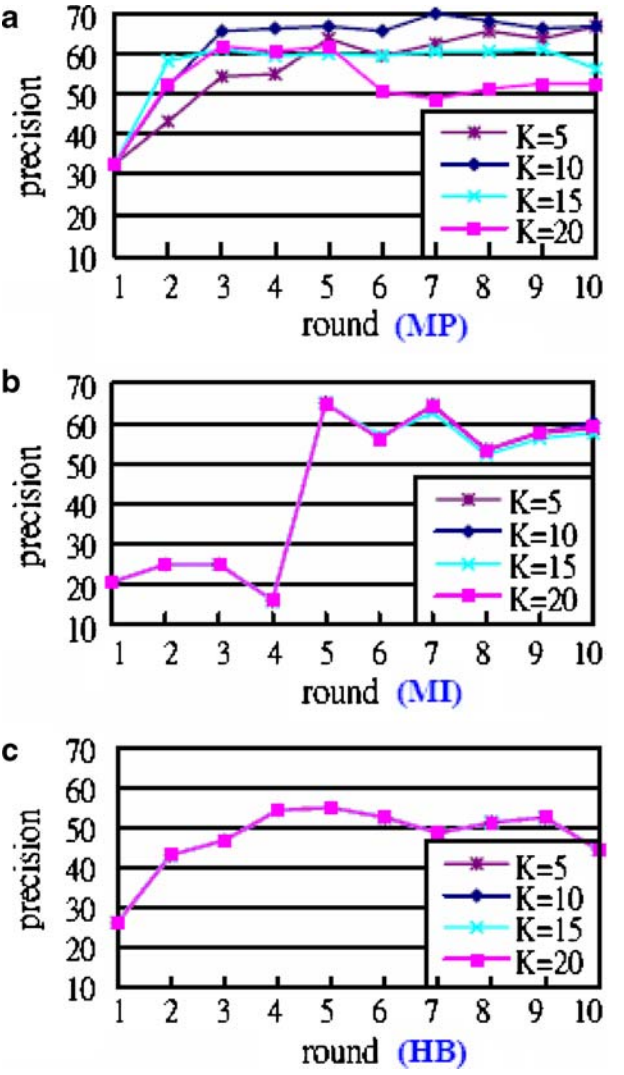


Fig. 5 Performances with different K as a function of the round number for different system feedback strategies, **a** the most positive, **b** the most informative, **c** the hybrid

MI strategy by judging fewer objects. Finally, Fig. 5c shows that the HB strategy was insensitive to parameter K. Since its performance was limited by uncertain objects during the whole session, its growth was also limited. In summary, the MP strategy ensures higher accuracy and steady improvement of retrieval results via relevance feedback, while the MI strategy is another option for achieving satisfactory results with less time.

6.2.3 Effectiveness of significance selection threshold

While the significance selection threshold is used to select significant segments, an experiment with different significance selection thresholds was conducted for each feedback strategy.

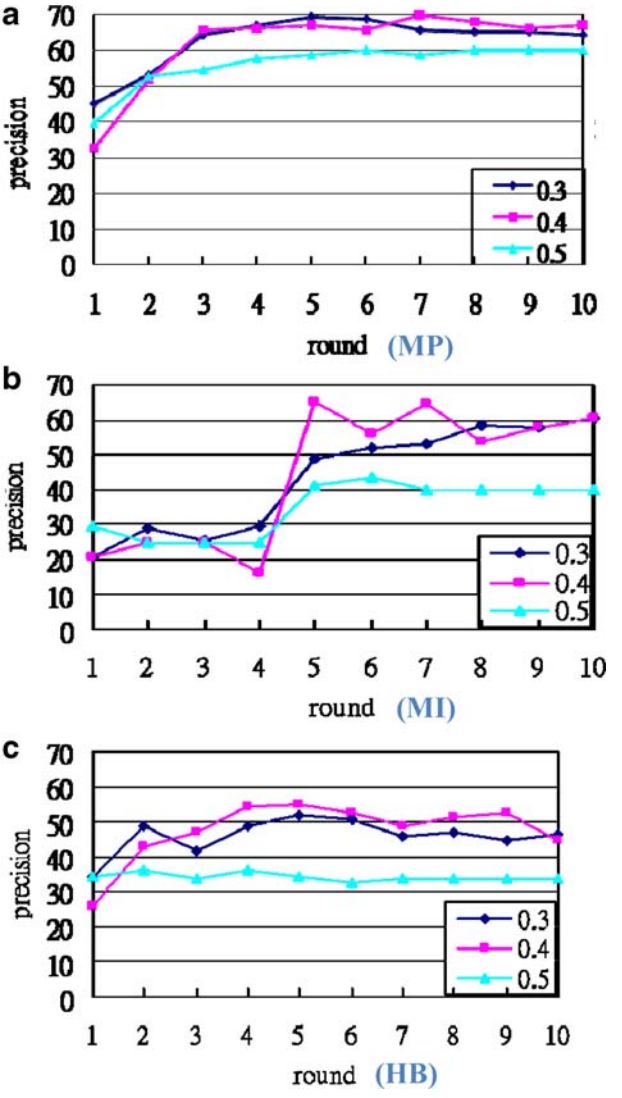


Fig. 6 Performances with different significance selection thresholds as a function of the round numbers for different system feedback strategies, **a** the most positive, **b** the most informative **c** the hybrid

It is anticipated that the system may work worse as less significant segments are kept in modeling the music object. With a higher motive threshold, the less significant segments are retained. If a user concept is highly correlated with the eliminated significant motivic patterns, the system will fail to detect some potential patterns relating to the user concept. Consequently, a derived classifier which is based on an insufficient pattern set will not identify potential discrimination rules and thereby have a reduced level of precision.

The impact of the motive selection threshold ranging from 0.3 to 0.5 was observed. Figure 6a,b, and c illustrate the performances of different system feedback strategies, respectively. The system with a lower significance selection threshold (e.g. 0.3, or 0.4) had

Table 7 Description of music genres and the numbers of corresponding music objects

Genre ID	Genre	#(Music objects)	Genre ID	Genre	#(Music objects)
1	J. S. Bach	48	8	Japanese animation film music	73
2	Blue Note	44	9	Western animation film music	141
3	Chinese Folk	55	10	Popular music	219
4	Dance Music	33	11	RWC classical music	29
5	Jazz	78	12	Scotland folk	96
6	John Williams	52	13	American folk	62
7	March	23	14	Christmas music	55

a better performance than that with a higher one (e.g. 0.5). The experimental results coincide with our expectations. In summary, with appropriable parameters, the MP and MI strategies are effective choices for all users.

6.2.4 Scalability analysis

It is a laboring work for the subjects to listen to thousands of music objects. Neither to say that listening to thousands of 3-min pop music will take 50 h and it seems difficult for the subject to construct the ground truth based on the music concept in a consistent way. An alternative way is to use the music genre database where each genre is regarded as a special kind of music concept.

We have collected 1008 polyphonic MIDI files from internet. These music objects belong to 14 categories. The genres and the amount of the corresponding music objects are shown in Table 7. An experiment was conducted on this music genre database to measure the scalability of the proposed relevance feedback mechanism.

Figure 7 shows the average precision over the 14 genres with the MP strategy. The significance threshold, the minimum support threshold *minsup*, and the number of music objects K were set to 0.4, 0.2 and 10 respectively. In the first six rounds, the average precisions are trending upward. It achieved 55% precision in the sixth round where 60 feedbacks were judged. After the sixth rounds, the average precision remains unchanged and descends slightly.

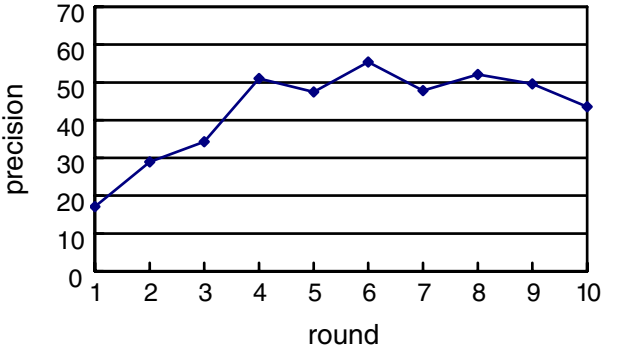


Fig. 7 Performances as a function of the round numbers for Most-Positive feedback strategies on music genre database

7 Conclusions

Relevance feedback is an attractive music retrieval alternative for users who are frustrated by the current requirement for specification in music query. In this paper, a relevance feedback mechanism for category search in content-based music retrieval is proposed. The main idea of this approach is to discover the relationship between the semantic concept behind the cognition of a music category and low level music features. A segment-based music modeling approach is presented which takes both global and local features into consideration. After each music object is properly modeled by the proposed representation, the system learns a user semantic concept from user relevance feedback via a two-phase frequent pattern mining algorithm and a modified associative classification algorithm. Three system feedback strategies have been investigated for music retrieval. The Most-Positive strategy returns the most relevant music. The Most-Informative one returns the most uncertain results based on user feedback in order to acquire more knowledge about the user concept. Finally, the Hybrid strategy is a compromise between these two strategies. Comparative experiments were conducted to evaluate the effectiveness of the proposed refinement mechanism. In short, the experimental results show that 60% average precision can be achieved for a database of 215 polyphonic music objects through the use of the proposed relevance feedback mechanism.

References

1. Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In Proc International Conference on Very Large Data Bases, pp 487–499
2. Baeza-Yates R, Riberio-Neto B (1999) Modern information retrieval. Addison Wesley, Reading, MA, USA
3. Chai W (2006) Semantic segmentation and summarization of music: methods based tonality and recurrent structure. IEEE Signal Process Mag 23(2):124–132
4. Doulamis N, Doulamis A, Varvarigou T (2003) Adaptive algorithms for interactive multimedia. IEEE Multimed 10(4):2–11
5. Foote JT (1997) Content-based retrieval of music and audio. Proc SPIE Multimedia Storage Archiving Syst II 3229:138–147
6. Gondra, Heisterkamp DR (2004) Learning in region-based image retrieval with generalized support vector machines. In Proc IEEE Computer Vision and Pattern Recognition Workshop
7. He XF, King O, Ma WY, Li MJ, Jiang HJ (2003) Learning a semantic space from user's relevance feedback for image retrieval. IEEE Trans Circuits Syst Video Technol 13:39–48
8. Hoashi K, Matsumoto K, Inoue N (2003) Personalization of user profiles for content-based music retrieval based on relevance feedback. In Proc ACM International Conference on Multimedia pp 110–119, Nov
9. Hoashi K, Zeitler E, Inoue N (2002) Implementation of relevance feedback for content-based music retrieval based on user preferences. In Proc ACM International Conference on Research and Development in Information Retrieval, pp 385–386, Aug
10. Hsu JL, Liu CC, Chen ALP (2001) Discovering non-trivial repeating patterns in music data. IEEE Trans Multimedia 3(3):311–325
11. Jing F, Li MJ, Zhang HJ, Zhang B (2004) Relevance feedback in region-based image retrieval. IEEE Trans Circuits Syst Video Technol 14(5):672–681
12. Jing F, Li MJ, Zhang HJ, Zhang B (2004) An efficient and effective region-based image retrieval framework. IEEE Trans Image Process 13(5):699–709

13. Kuo FF, Shan MK (2002) Music style mining and classification by melody. In Proc IEEE International Conference on Multimedia and Expo, Aug
14. Liu B, Hsu W, Ma Y (1998) Integrating classification and association rule mining. In Proc ACM International Conference on Knowledge Discovery and Data Mining, pp 80–86, Aug
15. Liu CC, Hsu JL, Chen ALP (1999) An approximating string matching algorithm for content-based music data retrieval. In Proc IEEE International Conference on Multimedia Computing and Systems, pp 451–456, Jun
16. Mandel MI, Poliner GE, Ellis DPW (2006) Support vector machine active learning for music retrieval. Multimedia Syst 12(1):3–13
17. Mandl T, Womser-Hacker C (2002) Learning to cope with diversity in music retrieval. In Proc International Conference on Music Information Retrieval
18. Mandl T, Womser-Hacker C (2003) Learning to cope with diversity in music retrieval. J New Music Res 32(2):133–141
19. Mermelstein D (1980) Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech and Signal Processing 28(4)
20. Pardo B (2006) Music information retrieval. Commun ACM 49(8):29–31
21. Rui Y, Huang TS, Ortega M, Mehrotra S (1998) Relevance feedback: a power tool for interactive content-based image retrieval. IEEE Trans Circuits Syst Video Technol 8(5):644–655
22. Shan MK, Ho MC (2007) Theme-based music structural analysis. Submitted to IEEE Multimedia
23. Stein L (1979) Structure and style. Summy-Birchard, New York, NY, USA
24. Tong S, Chang E (2001) Support vector machine active learning for image retrieval. In Proc ACM International Conference on Multimedia, pp 107–118, Sep
25. Uitdenbgerd, Zobel J (1999) Melodic matching techniques for large music databases. In Proc ACM International Conference on Multimedia, pp 57–66, Oct
26. Zhou XS, Huang TS (2003) Relevance feedback in image retrieval: a comprehensive review. Multimedia Syst 8(6):536–544

Man-Kwan Shan received the B.S. degree in Computer Engineering and the M.S. degree in Computer and Information Science both from National Chiao Tung University, Taiwan, in 1986 and 1988, respectively. From 1988 to 1990, he served as a lecture in the Army Communications and Electronics School. Then, he worked as a lecture at the Computer Center of National Chiao Tung University, where he supervised the Research and Development Division. He received the Ph.D. degree in Computer Science and Information Engineering from National Chiao Tung University in 1998. Then he joined the Department of Computer Science at National Chengchi University as an assistant professor. He became an associated professor in 2003. His current research interests include data mining, multimedia systems, and multimedia mining. He has supervised students who were the winner of 2003 National Science Council Excellent M.S. Thesis Award, the winner of 2003 Acer Long Term Award for Excellent Thesis, the winner of 2000, 2003 National Science Council Excellent Undergraduate Research Award.

Meng-Fen Chiang received the Bachelor's and Master's degrees in computer science from the National Chengchi University, Taipei, ROC, in 2004 and 2006 respectively. She is currently a Ph.D. candidate in the department of computer science at National Chiao Tung University, Hsinchu, ROC. Her research interest includes data mining and machine learning, with emphasis on modeling and mining real-world networks.

Fang-Fei Kuo is currently a Ph.D. candidate at Chiao Tung University, Hsinchu, Taiwan. She received her bachelor degree in Management from Tsing Hua University, Taiwan, in 2001 and master degree in Computer Science from Chengchi University, Taiwan, in 2003. She also received the 2003 National Science Council Excellent Master's Thesis Award and 2003 Acer Long Term Award for Excellent Thesis. Her research interest includes data mining, multimedia information retrieval and emotion-based multimedia mining and retrieval.