行政院國家科學委員會專題研究計畫 成果報告

數位音樂典藏之資料探勘與智慧型檢索技術(Ⅱ)
研究成果報告(完整版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 95-2422-H-004-002-
執 行 期 間 ：95 年 03 月 01 日至 96 年 02 月 28 日
執 行 單 位 ：國立政治大學資訊科學系

計 畫 主 持 人：沈錳坤

計畫參與人員：博士班研究生-兼任助理：邱士銓
碩士班研究生-兼任助理：江孟芬、魏綾音、彭冠誌、董信宗、廖家慧

報 告 附 件 ：出席國際會議研究心得報告及發表論文
國際合作計畫研究心得報告

處 理 方 式 ：本計畫涉及專利或其他智慧財產權，2 年後可公開查詢

中 華 民 國 96 年 07 月 20 日

行政院國家科學委員會補助專題研究計畫 ☑ **成 果 報 告**
□**期中進度報告**

數位音樂典藏之資料探勘與智慧型檢索技術(Ⅱ)

計畫類別：☑ 個別型計畫　　□ 整合型計畫
計畫編號： NSC　95－2422－H－004－002－
執行期間：95 年 3 月 1 日 至 96 年 5 月 31 日

計畫主持人： 沈 錳 坤
共同主持人：
計畫參與人員： 江孟芬、魏綾音、彭冠誌、董信宗、廖家慧、邱士銓

成果報告類型(依經費核定清單規定繳交)：□精簡報告　☑完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
☑涉及專利或其他智慧財產權，□一年☑二年後可公開查詢

執行單位：國立政治大學資訊科學系

中　華　民　國　96　年　5　月　31　日

數位音樂典藏之資料探勘與智慧型檢索技術(Ⅱ)

摘要

本計畫的研究重點即為利用音樂探勘技術，研究數位音樂典藏中的智慧型擷取技術。數位音樂典藏的擷取方式包括以後設資料(metadata)查詢、音樂內容擷取, 音樂曲風查詢, 相關回饋, 音樂瀏覽與個人化音樂推薦等，以有助於使用者方便地擷取典藏的數位音樂。在本計畫中，我們主要目的將研究利用音樂探勘中的使用者概念學習(User's Concept Learning)在相關回饋上，發展以內容主的音樂查詢(Content-based Music Retrieval)技術。

傳統的音樂檢索系統主要在提供使用者特定音樂的查詢(target search)。除此之外，使用者也有類型音樂查詢(category search)的需求。在類型音樂查詢中，該類型的所有音都共同具備使用者所定義的概念(semantic concept)。這個由使用者定義的概念在音樂檢索系統上是主觀的且動態產生的。換句話說，同一使用者在不同情境之下對於同一首音樂可能產生不同的解讀概念。為了動態擷取使用者的概念，讓使用者參與在查詢過程的互動機制是必要的。因此， 我們提出將相關回饋(relevance feedback)的機制運用在以內容為主的音樂查詢系統上，讓系統從使用者的相關回饋中學習使用者的概念，並利用這學習出的概念來幫助音樂查詢。

由於使用者可能從整首音樂或音樂片段兩種角度來判斷該音樂是否具備使用者定義的概念。因此，本論文提出用以片段為主的音樂模型(segment-based modeling approach)將音樂表示成音樂片段的集合。進一步再從整首音樂和片段中擷取特徵。其次，我們針對該問題提出相關演算法來探勘使用者的概念。該演算法先從相關和不相關的音樂資料庫中個別探勘常見樣式，再利用這些樣式建立分類器以區分音樂的相關性。

最後，我們分析各種系統回饋機制對搜尋效果的影響。Most-positive 回傳機制會選擇根據目前系統判斷為最相關的物件。Most-informative 機制則是回傳系統無法判斷其相關性的音樂物件。Most-informative 機制的目的在增加每回合系統從使用者身上得到的資訊量。Hybrid 則是中和前兩種機制的優點。本文中，我們模擬並比較各種回傳機制的效能。實驗結果顯示相關回饋機制確實能提升查詢的效果。

Data Mining and Intelligent Retrieval Techniques for Digital Music Archives（Ⅱ）

Abstract

In this project, we investigated the data mining techniques for intelligent retrieval of digital music archive. The way of the digital music archive retrieval, including metadata search, content-based music retrieval, music style retrieval, music browsing, personalized music recommendation and etc., is helpful for retrieving music archive easily. In this project, we utilize the data mining technique to learn user’s concept of relevance feedback for developing content-based music retrieval technique.

Traditional content-based music retrieval system retrieves a specific music object which is similar to the user’s query. There is also a need, category search, for retrieving a specific category of music objects. In category search, music objects of the same category share a common semantic concept which is defined by the user. The concept for category search in music retrieval is subjective and dynamic. Different users at different time may have different interpretations for the same music object. In the music retrieval system along with relevance feedback mechanism, users are expected to be involved in the concept learning process. Relevance feedback enables the system to learn user’s concept dynamically.

In this project, the relevance feedback mechanism for category search of music retrieval based on the semantic concept learning is investigated. We proposed a segment-based music representation to assist the system in discovering user’s concept in terms of low-level music features. Each music object is modeled as a set of significant motivic patterns (SMP) achieved

by discovering motivic repeating pattern. Both global and local music features are considered in concept learning.

Moreover, to discover user's semantic concept, a two-phase frequent pattern mining algorithm is proposed to discover common properties from relevant and irrelevant objects respectively and based on which a classifier is derived for distinguishing music objects.

Except user's feedback, three strategies of the system's feedback to select objects for user's relevance judgment are investigated. Most-positive strategy returns the most relevant music object to the user while most-informative strategy returns the most uncertain music objects for improving the discrimination power of the next round. Hybrid feedback strategy returns both of them. Comparative experiments are conducted to evaluate effectiveness of the proposed relevance feedback mechanism. Experimental results show that a better precision can be achieved via proposed relevance feedback mechanism.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

The amount of digital multimedia data increases with the advances in multimedia and computer technologies. Digital data enriches our lives and technologies for storing, analyzing and accessing multimedia data are increasingly demanded. One of the technologies is multimedia information retrieval which has been conducted for many years. In the area of music retrieval, a typical music retrieval system can discover what user needs by given keywords such as title, author name etc. Except querying by metadata, a content-based retrieval (CBMR) system is introduced which searches for music object by analyzing music content. Traditional content-based music retrieval system discover a specific music object which is similar to a given music segment. Issues related to CBMR include music representation, similarity metric, indexing and query processing.

Instead of searching for a particular music object ("**target search**"), there is a need for retrieving a specific category of music objects ("**category search**"). These music objects, of the same category, share a common semantic concept which is defined by the user. For example, when a user wishes to search for romantic music, the retrieved music objects should share the common concept of romantic feeling.

Different users at different time may have established different interpretations of concept for the same music object. Therefore, the concept for category search in music retrieval is subjective and dynamic. The taxonomy of categories of music objects with respect to user's concept can't be constructed in advance and in a fixed way.

To attack this problem, an on-line and user-dependent learning process is needed. Users are expected to be involved in the learning process for two reasons. One is the system lacks of prior knowledge with respect to user's concept and thus users are required to provide examples. The other is that owing to the scarcity of training samples, *relevance feedback* provided by the user is needed to improve the retrieval results. In a query session of music retrieval, the process may proceed for several iterations until the user satisfies the result. User's relevance feedback in each round enables the system to learn user's concept dynamically. The system accumulates relevance feedback data throughout the session. The performance is expected to be improved via relevance feedback.

In this report, the relevance feedback mechanism for category search of music retrieval based on the semantic concept learning is investigated. There are four main contributions in our work. The first is the investigation of relevance feedback in content-based music retrieval. Little attention has been paid to the design of relevance feedback approach while in content-based music retrieval most users are frustrated in specification of music query.

The second is the proposed segment-based music modeling approach. In traditional multimedia retrieval, most research on relevance feedback approaches models the multimedia object as a whole. Some approaches in image retrieval extended to deal with local features by decomposing an image into regions. In music retrieval, for the sake that concept may be constituted by entire music object or only parts of it, we propose the segment-based music representation to facilitate the system to capture user's concept in compound granule. In our 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.

The third contribution is the developed algorithm for learning user’s semantic concept. The algorithm is presented to enable a learning process based on the segment-based music representation and to discover the concept which is constituted by different music features. We transform the relevance feedback problem to the binary classification problem where examples from user’s feedback are regarded as either irrelevant or relevant to the user’s concept. An on-line and user-dependent classifier is trained to classify music objects from music archive and return the result to the user.

The last one is a comparative performance is evaluated based on three system feedback strategies for returning results for user’s feedback. The strategy adopted will determine how much discrimination power the system can obtain for the next iteration. **Most-positive strategy** will return the most relevant music object to the user, **most-informative strategy** will return uncertain music objects that provide more discriminative information for systems to learn user’s concept, and hybrid feedback strategies (HB) returns both of them. Traditional method (most positive strategy) always returns the most relevant music objects to the user. Most-informative will select a set of music objects such that user’s feedback will improve the discrimination of uncertain music objects at the next iteration. The strategy highly depends on user’s willingness to interact with the system. For impatient users, the system should applied MP strategy. On the contrary, if users are willing to interact with the system, MI strategy can be applied. The hybrid one is compromise of these two.

Figure 1.1 illustrates the music retrieval along with the relevance feedback mechanism. Music objects in music archive and user’s example query music are preprocessed by the music object modeling module. The semantic concept learning module is designed to discover from user’s relevance feedback the relationships between user’s semantic concept and music features. A classifier with respect to the user’s relevance concept is derived to classify each music object in the music archive either as relevant or irrelevant for the next round. The system then will select a collection of music objects for user's relevance judgment and based on which a further semantic learning process may proceed again once the user isn't satisfied with the retrieval result. Therefore, a query session may involve more than one rounds of learning process. In each round, there are two types of feedback, the feedback (relevance feedback) given from users to systems (**U2S**) and the feedback (search result) returned from systems to users (**S2U**). In the U2S feedback, the user judges each retrieved music object either relevant or irrelevant to the concept. The system proceeds to learn user's semantic concept from U2S feedback and then returns a collection of music objects for user's relevance judgment.

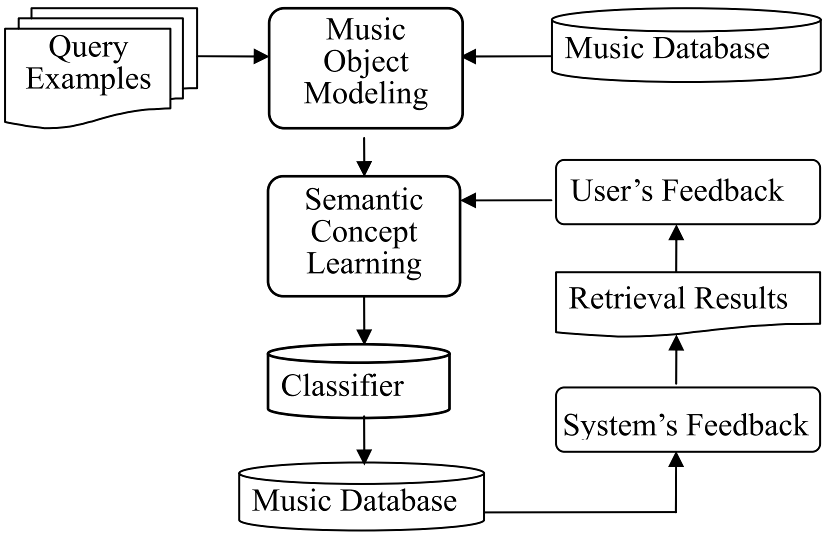

Figure 1.1 Framework of music retrieval along with relevance feedback mechanism.

This report is organized as follows. Related work about relevance feedback and music retrieval is introduced in section 2. The music modeling approach is described in section 3.

Section 4 presents the semantic concept learning task. Experimental results are presented in section 5. We conclude our work in section 6.

CHAPTER 2

Related Work

2.1 Relevance feedback schemes

In the area of modern information retrieval, a variety of approaches were developed for improving query formulation through query expansion and term reweighting [20]. Relevance feedback (RF) in image and video retrieval area has been addressed a lot. Most of them extend RF techniques in text retrieval while some of them adopt machine learning techniques. Note that all the existing work may have different assumption and problem settings. In this section, we will list some major variants in different conceptual dimensions in the aspect of user behavior model and algorithmic assumptions.

Object

According to the target that the user looks for, relevance feedback work can be categorized into two types, category search, and target search. In content-based image retrieval (CBIR), most of the work assumes the user is looking for a specific category of images. On the other hand, Cox et al. [4] assume what user looks for is a particular target object and a Bayesian framework is used to evaluate underlying probabilistic distribution over test data in a database via user's relevance feedback.

Feedback Algorithm

Work on CBIR with RF can be considered as a ranking problem where an ordered object list will be returned to the user based on the system's RF algorithm. RF algorithm is developed based on defined query representation along with distance function. A typical RF technique which takes user's relevance judgment into account to provide an improved retrieval results is the query reformulation approach. In [18], two main query reformulation approaches, single-point and multipoint, in single feature representation feedback are mentioned. The well-known technique of single point approach is query point movement (QPM). The objective of QPM approach is to reformulate a new query point such that this new point is close to relevant results and far from irrelevant results. In QPM, a single query is used and regarded as a point in a multidimensional space along with a single distance function (e.g. Euclidean distance function). Multipoint approach opposite to single query point uses multiple query points and aggregate their individual distances to data points into an overall distance. A well known approach is multipoint query expansion (QE). In such paradigm, the objective is to search for objects similar to more than one example. The distance of an object to the query points in a feature space is measured based on a weighted summation of the distances to each query point. When the new multiple query points are constructed through relevance feedback, the membership of the query points and weight that correspond to each query point may be changed. In other words, work adopts QE approach not only reformulate the query by adding new relevant examples or eliminating irrelevant examples but also adjusting weight of each example in each iteration. On the other hand, instead of representing the query in a single multidimensional space, multifeature representation feedback treats each feature representation individually. The query is represented as a collection of single-feature representation queries and an aggregation function is employed to combine the individual query distances into an overall distance. During a query session, the user can select feature representations that he is interested in along with the RF approach for each single feature representation. For example, users can give a query by selecting color histogram representation using QPM approach for RF and the wavelet feature representation using a multipoint query expansion for RF. A query based on the multifeature query model may be modified by adding or deleting single-feature representation queries and by updating the weights for each single feature query, which is known as query reweighting. Work relative to multifeature representation feedback can refer to [21]. Eventually, an ordered set of multimedia objects will be returned to the user by decreasing relevance by using RF techniques mentioned above.

Research of content-based video retrieval focused on formulating appropriate query via interaction has also been presented in [2] where not only features but also multiple modalities are considered. Information of video objects is implicitly represented by many modalities such as audio, color, and motion vectors. To search a collection of multimedia object is more challenging than searching text documents where natural human language is used to represent a query. Therefore, Amir et. al. [2] probed into a question of what modalities compose of user's information need when retrieving video objects and presented a mechanism to transform an abstract information need into a concrete search query by using mutual relevance feedback. A query expansion task across multiple modalities proceeds along a query session and will formulate a proper query at the end of a query session. Finally, the multimodal search query will be kept as a meta-representation of the corresponding semantic concept and may be applied later for the same concept without going through tiring interactions again.

Recent work on RF treat it as a classification problem [8][10][22][25][28][28][30]. In classification paradigm, user's relevance feedback is regarded either as relevant class or irrelevant class and the objective is to incorporate user's relevance feedback for building classifiers to classify multimedia objects. Most of the previous work utilizes support vector machines (SVMs). Since the SVMs not only set up a decision boundary between relevant and irrelevant images but also provides a mechanism to rank all of them. Moreover, it's efficient in facilitating interactions in an on-line environment although it lacks of incremental learning property. In addition to SVMs, some work adopts Bayesian framework [4] and some may use boosting technique to build a composite classifier to make binary decision on each returned objects. For instance, [30] proposed two online pattern classification methods, called interactive random forest (IRF) and adaptive random forests (ARF) which form a composite classier known as random forest for relevance feedback. Both improve the performance of regular random forests in different aspect. IRF improves the efficiency by using a two-level resampling technique, while ARF improves the effectiveness by using dynamic feature extraction and adaptive sample selection techniques.

In addition to classification approach, Yan [32] also mentioned the issue of selecting negative examples since negative instances are less well-defined as a coherent subset. [32] presented a negative pseudo-relevance feedback mechanism which uses the bottom-ranked examples for negative feedback identified based on a similarity metric. The training examples containing the positive examples (the query images) and the negative examples are then fed back to train a margin-based classifier. An adaptive similarity space will be learned following the pseudo-relevance feedback mechanism.

User's Feedback

There are three main ways for user to provide relevance feedback. The first one is binary feedback where the user judges each returned object either as relevant (or positive) or irrelevant (negative). The second type is degree of (ir)relevance where the user has to score each returned object showing how (ir)relevant an object is with respect to his semantic concept. In this way, it's difficult and burdensome for users to give (ir)relevance degree for each object consistently. The third type of feedback is a comparative judgment where no definite relevant and irrelevant judgment is made. Some feedback algorithms may take both positive and negative samples into account while some only consider positive samples.

System's Feedback

In addition to machine learning technique introduced in RF to increase the discrimination power, an active learning technique related to returning strategy is introduced to improve the discrimination of uncertain object in the next round. It's a strategy of selecting the best set of object at a feedback round to maximize potential information from the user. A standard strategy always returns the most positive objects based on previous training process. On the other hand, the system can actively query the user for labels to achieve the maximal information. The objects whose labels the system most uncertain about are named most informative objects in [33].

Representation

Most work on CBIR tends to model an image object as a vector in a multiple feature space [18][21]. Each feature dimension will be assigned a feature weight to represent the importance of that dimension. Hence, the feature weight of all dimensions forms a feature weight vector. The goal of algorithms using vector representation is to adjust the weight vector dynamically which may captures user's interest more precisely via few rounds of feedback learning.

Instead of describing an object as a whole, some work concerns local properties of an object and model an object as a set of feature vector corresponding to a local part. Owing to different expectations of the user, the user may mark an object relevant either based on global or local properties among the object. In some applications, such as [13][14], an image is considered as a set of regions and the system should be capable of leaning regions which are emphasized by the user if user's feedback is concerned based on region properties.

Long/Short-Term Learning

Much research about CBIR on RF has been addressed by using machine learning techniques [28][31]. Most learning method only concerns about the feedback information during the current query session. In addition to short-term information, [8][10] also took the knowledge from the past user interactions into account. The learning technique which takes information within the current query session into account is named as **short-term learning**, while **long-term learning** will learn knowledge over many query sessions. The function of different learning techniques is slightly different. Short-term learning aims to improve the retrieval result of the current query session and has more flexibility to fit user's need. On the contrary, long-term learning collects knowledge from the past and aims to boost performance of future sessions. However, the user may have different interpretation for the identical object at different circumstances and hence the past information doesn't help all the time.

The framework proposed in [8] demonstrated how long-term learning is incorporated into short-term learning in. He et. al [8] proposed a long-term learning approach for constructing a semantic space from user's past interaction and image content. The high-level semantic space is updated when more and more queries is made. During short-term learning in a query session, the low-level features of the query are extracted to conduct the first round of the retrieval. After that, the system uses the feedback examples to form the semantic representation of the query example. The query is refined and a classifier is generated to differentiate semantically relevant object from irrelevant ones. Then, the user judges the refined retrieval results and the system keeps on next round of short-term learning process if it's necessary.

2.2 Music Information Retrieval

A traditional information retrieval system aims to search for a particular music object which is close to a rough excerpt of the particular music given by the user. It belongs to the target search problem where key issues include music representation, similarity measure, indexing and query processing techniques [12][17][23]. A typical model of MIR system extracts low-level features (rhythm, melody, chords) from a music object and represents each music object as feature strings. After that, exact or approximate matching process is performed between query object and each one in database. A similarity measure will be addressed in a system allowing approximate matching based on music theory or heuristic rules. On the other hand, in order to speed up the searching process, issues related to index music objects in database may also be conducted.

Relevant research related to music information retrieval includes music recommendation and filtering systems [3][5][15][19][24]. Some of them recommend music object via collaborative filtering technique. Lack of objective similarity metric between multimedia objects has complicated many multimedia applications. When there is no nature idea of similarity between music objects, collaborative filtering strategy seems helpful for prediction. However, the similarity recommendations created by analyzing behaviors and ratings of users do not necessarily correspond to actual music similarity. Besides, popular music objects may dominate the recommendation result.

Chen [3] analyzes polyphonic music object and properly group music objects according to extracted music features and content-based, collaborative and statistics-based recommendation methods are proposed based on the favorite degrees of users to the music groups. Ringo [24] is one of the earliest proposed music recommendation systems by collaborative filtering method. Users are grouped by similar preference and music objects are recommended based on ratings derived from the users with similar preference. Ragno [19] presented a graph-based preference modeling framework where preference is derived from learning an expertly authored stream (EAS) from radio stations. EAS provides similarities among music objects judged by experts. Based on EAS, the burden of defining similarity metrics by content is avoided. However, the system may always generate same playlists for the same seed song which lack of desirable variety.

Music recommendation can be solved via categorizing music objects by user's preference. M. Grimaldi et al. [5] used classification approach to predict user's taste. An instance based classifiers based on user profiles is applied to learn music preferences. In his work, a reasonable accuracy can be achieved if user's taste is driven by a certain genre preference. A personalized music recommendation system proposed by Kuo [15] use associative classification methods to learn user's preference based on chords which are extracted from MIDI files. A user's profile was constructed by marking a sub set of music genre that a user is interested in.

Research about RF in music information retrieval has seldom been conducted. The first work on content-based music information retrieval based on RF was presented in [11]. In Hoshi's research, a music retrieval system was proposed for searching music objects based on user's preference. He assumes that the system only has insufficient knowledge about a user's preference in real world and proposed a retrieval method based on vector representation to address this problem. The user may not be satisfied with the retrieval results which are produced based on insufficient learning data. A relevance feedback mechanism is applied to improve retrieval result and experiments were conducted to show its effectiveness. An advantage of this approach is that it enables the user to discover new songs according to user's preference. Hoashi also presented two types of profile constructed from user ratings and from genre preference respectively. Comparative experiments show that precision of user rating based profiles is higher then that of the genre based profiles. When relevance feedback is conducted, genre based method outperforms user rating based method.

CHAPTER 3

Music Object Modeling

A good music representation should be able to assist the system in capturing user's semantic concept in terms of low-level music features. A music object can be characterized by multiple features such as tempo, rhythm, melody etc. Each feature can be represented as a set of representations. For example, average pitch difference and pitch standard deviations can be used for representations of the melody feature. In the representation space, the semantic concept can be characterized as a subset of representations which discriminates the concept from others. For instance, an inspiring music which rise and fall seriously in melody is describable by average pitch difference.

To understand user's concept, **global features** corresponding to an entire objects and **local features** with respect to each music segment should be considered. A music object is composed of a set of music segments. A music object can be globally described by a set of representations in feature space or locally described as multiple sets of representations in feature space where each representation set corresponds to a music segment.

We proposed a segment-based music modeling technique to represent music object in segment level. In our work, the modeling approach consists of three steps. In the first step we represent each music object as a set of segments found by the motivic repeating pattern finding algorithm. Then, multiple feature representations are extracted from each music segment. Moreover, global feature representations are also extracted from an entire music object to represent the music object as a whole. The last step is to filter significant motivic patterns based on frequency of patterns.

The music modeling approach is organized as follows. Section 3.1 describes the technique for finding motivic repeating patterns. Section 3.2 introduces the step of feature extraction. After that, section 3.3 introduces how to filter significant motivic repeating patterns.

3.1 Motivic Repeating Pattern Finding

In music, a motive is a salient recurring fragment of notes that may be used to construct the entirely or parts of complete melodies and themes. Therefore, each music object can be described by a set of motives. The recurrence of a motive may not be an exact repetition in the music object but with some variations. This is called as motivic treatment in musicology [26]. Six common motivic treatments (a)repetition (exact repeat), (b)transpose (interval repeat), (c)sequence, (d)contrary motion, (e)retrograde, and (f)augmentation/ diminution repetition are considered in our work (Figure 3.1).

We first apply the all-mono method to extract main melody. The extracted main melody will be represented as a note sequence where each note is expressed by pitch and its duration. Then, we modified the correlative matrix method [12], originally designed for exact repeating pattern finding, to discover six variations of motivic repetitions [9]. Finally, a minimum constraint on the length of a fragment is used to retain motivic patterns of more than four notes.

The correlative matrix method is utilized for repeating pattern discovery with a given note sequence. It includes the following three steps:

1) Construct Correlative Matrix:

The correlative matrix is the data structure which is initially formed by the given note sequence. Namely, if the length of note sequence is n, the size of the matrix is nxn. The purpose of the first step is to fill the matrix row by row. For the i^{th} note and the j^{th} note in the note sequence, the cell of i^{th} row and the j^{th} column in the matrix will be set as one if they are the same, otherwise it will be empty. In addition to the current matching results, the value of cell in the i^{th} row and the j^{th} column is also decided based on the result of the cell in the $(i\text{-}1)^{th}$ row and $(j\text{-}1)^{th}$ column. Assume the value of the cell in $(i\text{-}1)^{th}$ row and $(j\text{-}1)^{th}$ column is v. The value of the cell in i^{th} row and j^{th} column will be set to v+1 if the i^{th} note is the same as j^{th} note in the sequence. The value in the cell indicates the length of a potential repetition.

After the construction step, the matrix will keep all of the intermediate results of substring matching.

2) Find Candidate Set:

For each non-empty cell, the corresponding pattern is regarded as a candidate, a potential repeating pattern. The associated information is computed as we find each candidate. The information includes, *pattern*, *rep_count*, and *sub_count*. *Pattern* indicates the repeating pattern, *rep_count* represents the count of matching for the repeating pattern, and *sub_count* means the number of other repeating patterns which contains this pattern. To calculate the *rep_count* and *sub_count* for the ith row and jth column (M_{ij}), conditions of $M_{i-1,j-1}$ and $M_{i+1,\ j+1}$ has be taken into account. After computation for each non-empty cell, patterns with their corresponding repetition count and substring count will be used to calculate pattern frequency in the next step.

3) Discover Non-trivial Repeating Patterns:

The purpose of this step is to discover all non-trivial repeating patterns and calculate the actual frequency of each legal pattern. A pattern is trivial if its *rep_count* equals *sub_count*. The trivial case indicates that there exists a superstring *S'* containing the pattern *S* and *S* appears along with *S'*. In such case, the superstring *S'* is considered more representative and hence the trivial pattern *S* will be removed. After removal, the frequency *f* of each pattern *p* in a music object *m* is calculated by the formula:

$$f(p,m) = (1+\sqrt{1+8\times rep\_count})/2 \tag{1}$$

Table 3.1 shows an example. Given a note sequence of "CAACCAACD", the correlative matrix is constructed by substring matching row by row. For the 1st note "C", it repeats in 4th, 5th, and 8th position of the sequence. For the cell M_{26}, because "A" in the 2nd row matches the "A" in the 6th column and M_{15} is 1, the value of M_{26} is set to 2. The value 2 indicates the pattern "CA" with length 2. To find all candidates, all non-empty cells is scanned and associated information of each candidate is computed. Take M_{37} as an example. The corresponding pattern of M_{37} is "CAA", whose count of match so far is one and is a substring of the pattern "CAAC" since M_{48} isn't empty. Hence, the associated information of "CAA" is ("CAA", 1, 1). Since the *rep_count* of "CAA" equals sub_count, "CAA" is a trivial pattern and will be removed. The pattern "C" is an example of non-trivial patterns.

Table 3.1 An example of correlative matrix.

	C	A	A	C	C	A	A	C	D
C	–			1	1			1	
A		–	1			2	1		
A			–			1	3		
C				–	1			4	
C					–			1	
A						–	1		
A							–		
C								–	
D									–

The method described is the standard version for discovering exact repeating patterns and can’t be applied for other repeating variants shows in Figure 3.1 without modification. For exact repetition (Figure 3.1 (a)), we can utilize the method directly.

For transpose (interval repeat) (Figure 3.2 (b)), we have to transform the pitch sequence into pitch interval sequence(Figure 3.1). After that, the correlative matrix method is applied on the pitch interval sequence.

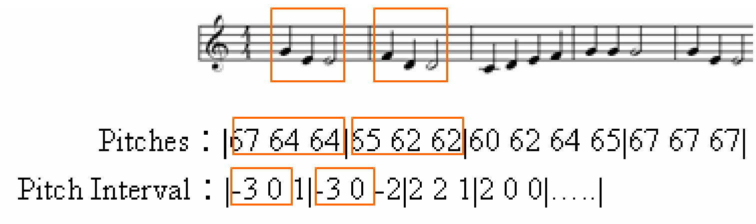

Figure 3.1 Transformation from pitch sequence into pitch interval sequence.

Sequence is a type of motive treatment which contains more than three consecutive motive transpositions (Figure 3.2 (c)). Beside, the direction of the transposition has to be the same, namely ascending or descending. In Figure 3.2 (c), the first rectangle indicates the original motive. The second and third are the transposition of the original motive. To discover sequence, the method is the same as the case of transpose except that we have to check whether the discovered pattern is repeated consecutively.

Contrary motion (Figure 3.2 (d)) is a motive treatment where pitch interval sequence is inversely repeated while the rhythm keeps the same. Namely, the contrary motion of the original motive can be obtained by assigning opposite sign for each pitch interval. To discover contrary motion, the correlative matrix is constructed by two different sequences. One is the original pitch interval sequence and the other is the one with opposite sign. Others remain the same.

Retrograde is a repetition where pitch contour is inversely repeated while rhythm keeps the same. Figure 3.2 (e) gives an example. The second motive <72, 72, 71, 67, 65, 65> is the retrograde of the first one <65, 65, 67, 71, 72, 72>. To discover retrograde in the sequence, The conditions to decide the value for each cell is changed. To assign the value of M_{ij}, the original method will take $M_{i-1,j-1}$ into account while $M_{i-1,j+1}$ will be considered in the retrograde case. Others remain the same.

Augmentation (diminution) repetition is repetition where pitch sequence remains the same while rhythm becomes faster (slower) with a ratio. In Figure 3.2 (f), the second motive is the augmentation repetition of the first one and the third motive is the diminution of the first one. To discover augmentation (diminution) repetition, the process for discovering repeating patterns remains the same while an additional check on the results is needed to ensure the rhythm of repetitions with regarded to one pattern is changed in a ratio.

After six discovery processes perform on each music object, we only keep original motive to represent the structure feature of that music. Next step describes the process of feature extraction.

3.2 Feature Extraction

We extract six kinds of global feature representation and five kinds of local feature representation shown in Table 3.2. In other words, each music object is modeled as a six-attribute global feature and a set of five-attribute local features. Music features considered in this report are melody, rhythm and tempo. Representations for melody features include average pitch, pitch standard deviation, highest/lowest pitch, chord sequence and average pitch difference. Rhythm feature is represented as density while tempo is represented as the tempo value only.

Average pitch is the average pitch values of notes within a music piece (an entire one or a segment). Pitch standard deviation is the standard deviation of pitch values of notes within a music piece. Highest and lowest pitch value is extracted from a music object and average pitch difference indicates the average of difference in pitch between two consecutive notes within a music segment. Chord sequence is a sequence of chord within a segment calculated by chord assignment algorithm which is a heuristic method based on harmony and music theory. Details on chord sequence can be seen in [14]. Density of a music piece is defined as number of notes dividing by the total duration of a music piece. Tempo denotes the speed of a music object and is defined as number of beats per minute.

Table 3.2 Global and local features considered in our work.

Global feature	Local feature
Density (*gd*)	Density (*ld*)
Average Pitch (*gap*)	Average Pitch (*lap*)
Pitch standard deviation (*gsd*)	Pitch standard deviation (*lsd*)
Tempo (*gt*)	Chord sequence (*lcs*)
Highest Pitch (*ghp*)	Average Pitch difference (*lapd*)
Lowest Pitch (*gl*p)	

In some cases, two different music segments may be approximately sounds the same. In order to consider the fault-tolerant cases, we intent to quantize the feature values in each segment. In the aspect of global feature, density and pitch standard deviation are quantized by the range of 0.5. More precisely, the quantized value will equal the quotient obtained by dividing the raw value by 0.5. For instance, two densities of 1.7 and 1.9 are quantized as 3. In the same way, the average pitch, highest pitch and lowest pitch are divided by 5. In the part of local feature, density, pitch standard deviation and average pitch difference are divided by 0.5.

The average pitch value is quantized as it does in the global feature, while the chord sequence remains the original value.

In order to observe the impact of quantization on performance, we keep two copies of features, the raw one and the quantized one. These two copies will be processed in the next step and the sequential learning process respectively. We will compare the performance of the two different representations in the chapter 5.

3.1 Significant Motive Selection

We aim to filter **significant motivic patterns** (**SMPs**) in this step. We measure the significance of each motivic repeating pattern and retain those significant one with regard to a music segment. A motivic repeating pattern with high frequency in the music object isn't necessarily more important than the one with low frequency in the other music object. Therefore, the frequency of a motive, *f(p,m)*, is normalized by dividing the maximal frequency of the motivic pattern *p'* in music *m*. A motive is more important with respect to one music object if the motive is more specific in the music database (*DB*) and thus the importance of a motive with respect to one music object is defined as follows:

$$W(p,m) = \frac{f(p,m) \Big/ \max_{p' \in m} f(p',m)}{\sup(p,DB)} \tag{2}$$

where sup*(p,DB)* stands for the support of *p* in the *DB*.

Table 3.3 shows the representations of the song "don't let the sun go down on me" which contains eight SMPs with importance higher than 0.5. Only the representations of global feature and local features of three SMPs are shown in Table 3.2. Figure 3.3 illustrate the corresponding score for these SMPs.

Figure 3.2 Examples of six motivic treatments.

Table 3.3 Representation of the global feature and three SMPs of the music object M.

M	gd	gap	gsd	gt	ghp	glp
	1.7	73	3.5	7	81	72

{

ld	lap	lsd	lcs	lapd
2	75	2	0	3

,

ld	lap	lsd	lcs	lapd
1.1	74	2	7	1

,

ld	lap	lsd	lcs	lapd
0.4	76	2	3	0.3

}

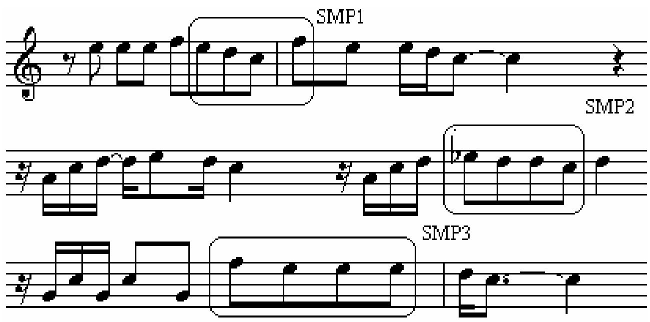

Figure 3.3 Corresponding score of SMPs in Table 3.3.

CHAPTER 4

Semantic Learning from User's Relevance Feedback

The semantic learning process at each round is performed on the accumulated training data. The training data is composed of two databases, the relevant one (MDB^P) and irrelevant one (MDB^N). Each database contains relevant/irrelevant music objects accumulated from previous rounds. The amount of samples in each MDB^P/ MDB^N increases during the session. The concept can be learned by mining common properties of MDB^P and MDB^N respectively first and then discovering discrimination between these properties. Table 4.1 is an example of MDB^P containing four music objects while Table 4.2 is an example of MDB^N with three music objects. For convenience of explanations, in these examples, a music object is modeled as a two-attribute global feature (G, H), and a set of three-attribute local features (A, B, C) where each three-attribute local feature corresponding to a SMP. One example of common properties of MDB^P is (A=1,C=1) and one example of discriminative properties of MDB^P and MDB^N is (B=2) & (H=2) & (A=1, C=1) which appears frequently in MDB^P but seldom or never appears in MDB^N.

The semantic learning process for capturing user's concept proceeds first by frequent pattern mining algorithm followed by associated classification algorithm. Details are shown in the following.

4.1 Frequent Pattern Mining

Relevance/irrelevance is usually defined by a characteristic that is shard by relevant/irrelevant music objects. To capture the characteristic sharing by a class of music objects, we employ the data mining techniques. Before the mining process, each (attribute, value) pair of the global and local feature is transformed to an item. For example, an (attribute, value) pair of (B,2) is transformed to an item "B2". Therefore, the global feature is represented as an itemset of six items while the local feature of an SMP is represented as an itemset of five items. A music object is therefore treated as a set of itemsets. Before presenting the algorithm, we introduce some formal definitions in the following.

Definition 1:

Let I be the set of possible items and $Y = \{X \mid X \subseteq I$, X is the itemset corresponding to a local feature or a global feature$\}$. Let MDB^P/MDB^N be a music database, where each **object** T is a set of itemset such that $T=\{T_1, T_2, \ldots, T_x \mid T_i \in Y\}$, namely $T \subseteq Y$.

Example 1:

Take M1 in table 4.1 as an example. The object M1 is represented as $\{\{G4, H2\}, \{A1, B2, C1\}, \{A2, B2, C1\}, \{A1, B1, C2\}\}$.

The common property found in this mining stage is called a *frequent pattern,*.

Definition 2:

Let X be the itemset corresponding to a local feature or a global feature. The common property, **pattern**, found in the mining stage is a set of itmest, $P = \{P_1, P_2, \ldots, P_v \mid P_j \subseteq X\}$.

Example 2:

An example of pattern in table 3 is $\{\{H2\}, \{A1,C1\}\}$ where itemset $\{H2\}$ and $\{A1, C1\}$ are the subset of a local feature or a global feature.

Definition 3:

We say that an object T **contains** the pattern P if there is a one-to-one mapping function from P to T such that for each P_i, there exists a T_i, $T_i \in T \ni P_i \subseteq T_i$.

Example 3:

Take the pattern {{*A2*}, {*C2*}} as an example. If an object contains {{*A2*}, {*C2*}}, there must exist two distinct itemset containing {*A2*} and {*C2*} respectively. For instance, in table 4.1 M1 and M2 contains {{*A2*}, {*C2*}}, while M3 doesn't contain {{*A2*}, {*C2*}}.

Definition 4:

Given a pattern *P*, the **support count** of *P,* supCount(P)*,* is the number of objects in MDB^P / MDB^N that contain *P* and it's **support** *sup*(*P*) in an object database is (*supCount*(*P*))*100%. We called *P* a **frequent pattern** if *sup*(*P*) is no less than a given minimum support threshold *,minsup.*

Example 4:

An example of frequent patterns with support 100% in MDB^P is {{*B2*}, {*H2*}, {*A1,C1*}} which is contained in all objects in MDB^P.

Table 4.1 An example of MDB^P.

M1	*G*	*H*	{	*A*	*B*	*C*	,	*A*	*B*	*C*	,	*A*	*B*	*C*	}
	4	*2*		*1*	*2*	*1*		*2*	*2*	*1*		*1*	*1*	*2*	
M2	*G*	*H*	{	*A*	*B*	*C*	,	*A*	*B*	*C*	,	*A*	*B*	*C*	}
	1	*2*		*2*	*2*	*2*		*1*	*2*	*1*		*2*	*4*	*2*	
M3	*G*	*H*	{	*A*	*B*	*C*	,	*A*	*B*	*C*	}				
	1	*2*		*1*	*3*	*1*		*2*	*2*	*2*					
M4	*G*	*H*	{	*A*	*B*	*C*	,	*A*	*B*	*C*	}				
	4	*2*		*1*	*1*	*1*		*1*	*2*	*2*					

Table 4.2 An example of MDB^N.

M1	*G*	*H*	{	*A*	*B*	*C*	,	*A*	*B*	*C*	,	*A*	*B*	*C*	}
	4	*1*		*1*	*2*	*1*		*1*	*1*	*3*		*2*	*3*	*3*	
M2	*G*	*H*	{	*A*	*B*	*C*	,	*A*	*B*	*C*	}				
	1	*2*		*2*	*3*	*3*		*1*	*4*	*1*					
M3	*G*	*H*	{	*A*	*B*	*C*	}								
	4	*3*		*1*	*4*	*1*									

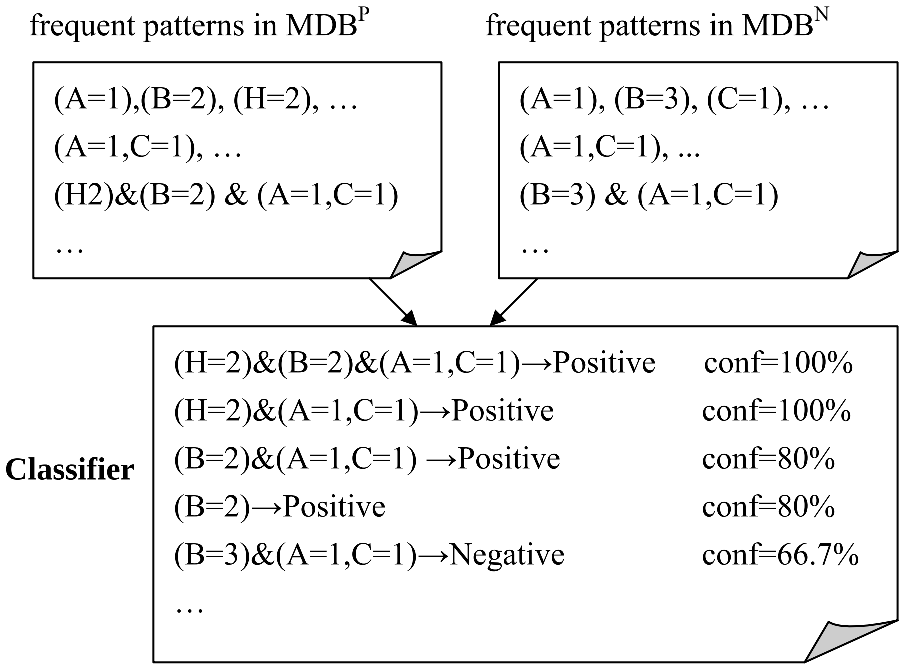

Figure 4.1 An example of classifier.

The task of frequent pattern mining is to find all frequent patterns with support no less than the minimum support threshold minsup. The frequent pattern found in MDB^P, MDB^N are called *positive frequent pattern* and *negative frequent pattern* respectively. Both of them are is the form of set of itemsets.

A well-known approach for mining frequent pattern is *Apriori* algorithm [1]. Apriori is a data mining technique originally developed to discover frequent itemsets from database of itemsets. However, in our work, MDB^P/MDB^N.is a database of sets of itemsets and the frequent pattern is also a set of itemsets. Therefore, we proposed a two-phase mining algorithm modified from Apriori to discover the frequent patterns. The first phase will find the frequent itemsets and the second phase will discover the frequent patterns constituted by the frequent itemset found in the first phase. Note that the itemset found in the first phase corresponds to the music segment (SMP) level while the pattern (set of itemsets) found in the second phase corresponds to the music object level. The mining process will proceed on both MDB^P and MDB^N respectively.

1st phase : mining frequent itemsets

We employ Apriori algorithm to discover all frequent itemset in which each item must appear in the same itemset. The classic Apriori algorithm for discovering frequent itemset makes multiple passes over the database. In the first pass, support of each individual item is calculated and those above the *minsup* will be kept as a seed set. In the subsequent pass, the seed set is used to generate new potentially frequent itemsets, candidate itemsets. Then the support of each candidate itemset is calculated by scanning the database. Those candidates with support no less than *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 our work, only the step of support calculation is different from classic Apriori algorithm, since in our work each object is a set of itemset, rather than an itemset. For the example of Table 3, the support count of the frequent itemset {*A2, C2*} is two. {*A2, C2*} appears in M2, M3, but not in M1, M4.

2nd phase : mining in segment level

The second phase will discover the patterns constituted by frequent itemsets found in the first phase. Similar to the algorithm in the first phase, the algorithm makes multiple passes over the database MDB^P/MDB^N. In the k-th pass, the seed set (the set of candidate patterns of k itemsets) is generated by joining two frequent patterns of k itemsets found in the previous pass. Then the support of each candidate pattern is calculated by scanning the database. Those candidates with support no less than minsup are 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.

In order to improve the efficiency of the above mining process, we introduce the **pattern canonical form** in the following to makes the **candidate generation** more efficient**.** Following shows the definitions and examples.

Definition 5:

Let P_k be a pattern containing k itemsets. Assume that items in an itemset is ordered by lexicographic ordering $\prec_l$. The pattern canonical form of P_k is defined as the set of itemsets in which itemset is rendered based on the ordering $\prec_{pcf}$. We define the ordering $\prec_{pcf}$ as follows. If $\alpha = \{s_1, s_2, \ldots, s_m\}$ and $\beta = \{t_1, t_2, \ldots, t_n\}$ are two itemsets in P_k, then $\alpha \prec_{pcf} \beta$ iff one of the following is true.

(i) $m < n$ or

(ii) $m = n$ and $\exists i, j, 1 < i < j, \ni s_{\alpha k} = t_{\beta k}$ for $1 < k \leq i$ and $s_{\alpha j} \prec_l t_{\beta j}$.

Example 5:

For example**,** the canonical form of $\{\{A1,C1\}, \{A1\}, \{B2,C1\}\}$ is $\{\{A1\},\{A1,C1\},\{B2,C1\}\}$.

Definition 6:

Given two frequent k-patterns in pattern canonical form, $\{P_1, P_2,\ldots,P_k\}$ and $\{Q_1, Q_2,\ldots,Q_k$, they are **joinable** if $P_2= Q_1$ and $P_3= Q_2$, …, and $P_k = Q_{(k-1)}$. A ($k+1$)-candidate pattern $\{P_1, P_2,\ldots, P_k, Q_k\}$ will be generated in canonical form as well.

Example 6:

Given two frequent 2-patterns in canonical form, $\{\{A1\},\{B2\}\}$ and $\{\{B2\},\{A1,C1\}\}$ will generate the 3-candidate pattern $\{\{A1\},\{B2\},\{A1,C1\}\}$.

Moreover, in order to count the support of patterns more efficiently, we maintain a table, **occurrence table**, for each pattern.

Definition 7:

Given a frequent k-patterns P in pattern canonical form, $\{P_1, P_2,\ldots,P_k\}$, the **occurrence** of P in object T is $\{L_{P1}, L_{P2}, \ldots, L_{Pk}\}$ where L_{Pi}, $1 \leqq i \leqq k$ indicates the location of itemset P_i in T. There may be more than one occurrence in an object. The table records all occurrences in each object in the database for P is called the **occurrence table**.

Example 7:

The occurrence tables for the patterns $\{\{B2\},\{H2\}\}$ and $\{\{H2\},\{A1,C1\}\}$ are shown in Table 5(a) and (b) respectively. In Table 5(a), the first occurrence of the pattern $\{\{B2\},\{H2\}\}$ in music object M1 is (2,1) where $\{B2\}$ appears in the 2nd itemset and $\{H2\}$ appears in the 1st itemset of music object M1.

We use the data structure, occurrence table, to store positions where the pattern appears in. Each pattern is associated with an occurrence table. Moreover, we also derive the occurrence table for each candidate during the process candidate generation.

Table 4.3 Examples of occurrence tables.

(a) $\{\{B2\},\{H2\}\}$

M1	(2,1), (3,1)
M2	(2,1) (3,1)
M3	(3,1)
M4	(3,1)

(b) $\{\{H2\},\{A1,C1\}\}$

M1	(1,2)
M2	(1,3)
M3	(1,2)
M4	(1,2)

(c) $\{\{B2\},\{H2\},\{A1,C1\}$

M1	(3,1,2)
M2	(2,1,3)
M3	(3,1,2)
M4	(3,1,2)

Definition 8:

Given two joinable k-patterns along with their occurrence tables, suppose that the occurrence of the first pattern in a specific music object is $(u_1, u_2, ..., u_k)$ while that of the second pattern in the same music object is $(v_1, v_2, ..., v_k)$, an occurrence $(u_1, u_2, ..., u_k, v_k)$ for this object will be generated if $u_2 = v_1$, $u_3 = v_2, \ldots, u_k = v_{(k-1)}$ and $u_1 \neq v_k$.

Example 8:

Given two occurrence tables in Table 5(a) and 5(b), the occurrence table for the candidate $\{\{B2\}, \{H2\}, \{A1,C1\}\}$ is presented in Table 5(c). In Table 5(c), the occurrence in object M1, (3,1,2) is generated by (3,1) in Table 5(a) and (1,2) in Table 5(b) of M1.

By utilizing the occurrence table, it is efficient to check the support count of each candidate pattern without scanning the music database.

4.2 Associative Classification

After a two-level mining process performs on MDB^P and MDB^N, we obtain a collection of positive and negative frequent patterns with respect to the common properties of music objects relevant and irrelevant to the concept respectively. In order to discriminate the concept of relevant music from that of irrelevant music, this step tries to find the discrimination between characteristics of MDB^P and MDB^N. The result of this step is a classifier consisting of rules. Figure 4.1 is an example of classifier learned from common properties discovered from Table 3 and 4. One rule is "(B=2) & (A=1,C=1)→ Positive", which classifies a music object containing attributes of (B=2) &(A=1,C=1) as positive class. This rule comes from the fact that (B=2) & (A=1,C=1) appears frequently in MDB^P but seldom appears in MDB^N.

We employ associative classification algorithm [16] to generate a binary classifier learned from the *positive* and *negative frequent patterns*. The algorithm eventually will generate a classifier containing a set of ranked rules. The classifier is of the form <r_1, r_1, ..., r_1, *default_class*}. Each rule r_i is of the form $l \Rightarrow y$, where $l \in F$, F is the collection of *positive* and *negative frequent patterns* and y is a class label. The confidence of a rule is defined as the percentage of the training music that contains l belonging to class y.

A naïve version of the algorithm will first sort the set of rules according to a defined precedence order. And then select rules following the sorted sequence that correctly classify at least one music object and will be a potential rule in our classifier. Different from the original rule type, the *frequent pattern* on the left hand side of one rule in our work is a set of itemset. We say that a music object is covered by a rule if it contains the *frequent pattern* of the rule. Take the rule {{*B2*},{*A1,C1*}}→ positive in Figure 4.1 as an example, if a music object has two itemsets containing *{B2}* and *{A1,C1}* respectively, then the music objects is covered by rule {{*B2*},{*A1,C1*}}→Positive. If the music object is in MDB^P, we say that it's correctly classified by the rule. A default class referred to the majority class of the remaining music object in database is determined. Finally, it will discard those rules that do not improve the accuracy of the classifier. The first rule in classifier that made the least error recorded in classifier is the cut off rule where rules after the cut off rule will be discarded since they only produce more errors.

4.3 S2U Feedback Strategy

Once the classifier is constructed, the system then produces a ranked list of music objects. As we have mentioned in section 1, what the system return to the user will determine the potential information granted from the user. We present three types of feedback strategies, most-positive, most-informative and hybrid strategies. In general, the most-informative music objects will not coincide with the most-positive music objects. Different strategies along with corresponding scoring function are described as follows.

(1) Most-Positive strategy (MP)

If the user is impatient, the system should present the most positive (i.e. those marked as relevant by the system) music objects learned so far. The most positive music is a list of music object *m* ordered by the score function which is related to the confidence of matched rules.

$$Score_{MP}(m) = \frac{\sum_{r \in Rp} conf(r)}{\sum_{r \in Rp \cup Rn} conf(r)} \tag{3}$$

where *Rp*/*Rn* stands for rules belong to positive/negative class that satisfies each music object m.

(2) Most-Informative strategy (MI)

If we sacrifice the performance at this round for maximizing information obtained for the next round, a better result can be expected in the future process. Most-Informative strategy will select a set of music objects such that their judgment by the user will provide more information for labeling uncertain music objects. The uncertain music objects are those whose class labels the system is uncertain about. These objects are most-informative objects. In other word, the system using MI strategy will display a collection of most informative objects at each round until the user attempts to find out what the system can retrieve in hand. Then, the system will adjust itself to MP strategy and return the most positive music objects.

In the associative classification algorithm, object which matches no rules in hand belongs to the default class. We define those belong to default class as most informative music objects. If the user is willing to interact with the system, our system will display a number of most informative music objects for user's feedback.

(3) Hybrid strategy (HB)

HB is a compromised between MP and MI strategies. The system applied HB strategy will equally return both most positive and most informative objects each round. The score of each music object *m* is defined as follows:

$$score_{HB}(m) = \begin{cases} 0.5, & m \in default\_class \\ score_{MP}(m), & otherwise \end{cases} \tag{4}$$

CHAPTER 5

Experimental Result and Analysis

5.1 Dataset

The dataset contains 215 MIDI music objects collected from the internet. Each music object belongs to western pop music including rock, jazz, and country genres. Subjects involved in the experiment are unfamiliar with some of the music objects. In this case, noisy and inconsistent judgment caused by the user because of familiarity may be avoided.

Automatic melody extraction process is performed on each MIDI file by all-mono algorithm. The raw feature representation and quantized version will be fed to the system separately for performance evaluation.

5.2 Experiment Setup

In order to evaluate the segment-based relevance feedback algorithm, we design an on-line CBMR system with relevance feedback mechanism. The relevance feedback information of users is essential for system evaluation. We invite eight subjects to investigate our system for creating relevance feedback data.

The retrieval process proceeds by randomly selecting 20 music objects for user's labeling. An on-line training process will derive a classifier based on initial U2S feedback. The classifier labels all music objects in database along with a scoring function which defines relevance degree of each one. According to specified S2U feedback strategy, at most 20 music objects will be returned and judged by the user. Once the user isn't satisfied with the current retrieval result, next round proceeds again. The training samples are accumulated from each relevance feedback round. The classifier is expected to be refined based on the accumulated training samples via the relevance feedback mechanism.

In order to compare with performances for experiments with different strategies and parameter settings, the user has to go through many experiments and provide relevance feedback for each one of them in reality. It wastes user's time and somewhat a tiring job. To reduce user's burden, we attempt to collect user's relevance feedback data in advance. Once the user determines the concept in mind, the user labels each music object in the database either as relevant or irrelevant. The relevance feedback data made by the user will be regarded as the groundtruth. After that, a series of experiments for each user will be conducted. Each experiment corresponding to a query session contains many rounds will be simulated and each returned music object will be automatically label based on user's groudtruth.

5.3 Effectiveness Analysis

In order to evaluate the results of our experiments, the performance measure employed is based on the average precision, which is defined as the ratio of the number of relevant music objects of the returned music objects over the number of total returned music objects n for all users.

Note that each experiment intends to evaluate effectiveness of the refinement framework on music retrieval system. Music objects that system has returned in previous round will not be removed from the music database.

We conducted four sets of experiments for performance comparison. The first one is to evaluate the different feedback strategies of the system. The second experiment is to measure the effectiveness of the number of music objects accumulated from user's feedback (top *K*). Subsequently, the experimental result for evaluating effectiveness of the number of rounds (*N*) applied most-informative (MI) will be discussed. Finally we show the effect of motive threshold on performance.

5.3.1 Effectiveness of System Feedback Strategy

We perform three different experiments to compare the effectiveness of system feedback strategy applied for each round during a query session. As mentioned in section 4.3, the system can employ MP, MI or HB strategy at each round. Three different experiments are described as follows:

<u>S2U feedback strategy (MP):</u> the system applies MP feedback strategy each round and only uses the top *K* music objects returned for further refinement.

<u>S2U feedback strategy (MI):</u> the system applies MI feedback strategy for consecutive *N* rounds and then evaluates the final result by applying MP feedback strategy for the rest of rounds. By examining precision at the *N*+1 round among three S2U feedback strategies, how well does MI strategy work can be evaluated.

<u>S2U feedback strategy (HB):</u> the system applies HB feedback strategy each round during the query session.

Figure 5.1 illustrates the performance comparison of three S2U feedback strategies performed on raw feature presentation. *Motive threshold* is set to 0.4 for three different strategies. *Minsup* is set to 0.2. The factor of *N* is set to 4, i.e., the first four rounds are in most informative strategy and the rest of rounds adopt most positive strategy. The number of music objects used for relevance feedback, *K*, is set to 10. As the number of rounds increases, precision grows for all S2U feedback strategies as we have expected. The randomly selected initial training samples may limit initial knowledge learned from the first training round and thus we fix the initial query examples as the seed set for each user to ensure fair comparison among experiments under different parameter settings.

Figure 5.2 shows the performance comparison of three different strategies performed on quantized feature representation. The parameter setting is the same as the raw one. In HB feedback strategy part of uncertain music objects were contained in the retrieval results. The precision of each round is bounded since the uncertain music objects aren't necessarily the positive objects. On the other hand, uncertain objects are counted on to improve the precision. However, there's no clear benefit gained by HB strategy shown in Figure 5.1 and 5.2. In the first *N* round, HB and MP feedback strategies have better performance than MI feedback strategy. With the help of active learning, MI feedback strategy gradually improves the system and outperforms HB strategy after N^{th} round. The performance of HB feedback strategy is a moderate one which is better than MI and less remarkable than MP in the first few rounds. On the contrary, it's worse than MI during the later half session.

The MP strategy ensures reliable performance. The accuracy above 60% can be achieved at round three where the user provides feedback on only 20 music objects. Performance after round three grows initially and then drops slightly. Although the curve doesn't show consistent growth, all cases after round three still keep 60% accuracy.

The system adopted MI strategy in quantized one achieves 65% accuracy at round 5 while the user totally judges 40 music objects. Cases after rounds five keeps accuracy over 50%. According to our expectation, the function of MI strategy is expected to learn more knowledge from user in comparison with MP strategy. Although the system adopted MI strategy has good performance, more efforts on interacting with the system doesn't generate more benefit compared with MP strategy.

Compared with raw feature representation, the quantized one definitely outperforms in all strategies. Beside, quantized one appears more consistent improvement via user's relevance feedback during a session. We refer this phenomenon to the successful approximation of quantization.

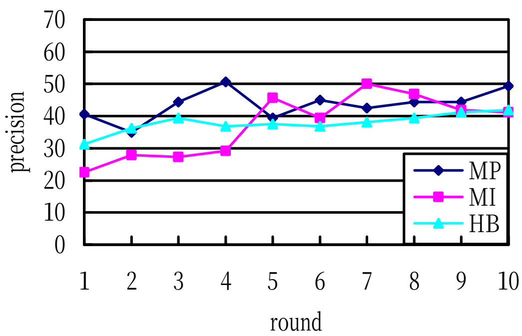

Figure 5.1 Performance of three feedback strategies on raw feature representation.

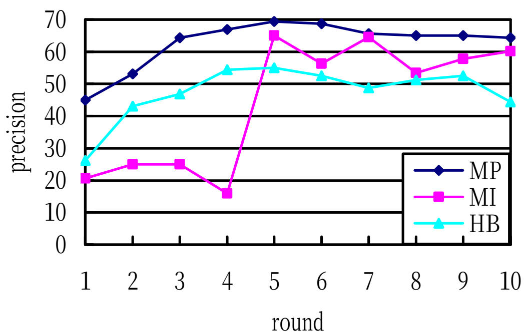

Figure 5.2 Performance of three feedback strategies on quantized feature representation.

We conclude that MP feedback strategy is a reliable strategy. It's efficient and effective since user can obtain good retrieval results in few rounds. It's also suitable for impatient users because user can be satisfied with the retrieval results in the first few rounds without judging many music objects. Besides, it ensures gradual improvement of retrieval results via relevance feedback. Patient users who are willing to feedback more music objects thus can also be satisfied with MP strategy. On the contrary, HB and MI are less effective.

5.3.2 Effectiveness of Number of Music Object Accumulated from User's Feedback

To simulate the learning process based on relevance feedback, the top *K* songs of the ranked retrieval results were collected for next rounds of training. We conduct experiments for each system feedback strategy to illustrate the correlation between *K* and the precision.

The system with respect to one feedback strategy returns a fixed amount of music objects to the user at each round. After binary judgment of each retrieval music object is made, the top *K* music object joined with previous training samples are used to refine the current classifier.

The system returns at most 20 music objects and the top 5, 10, 15, and 20 of them will be accumulated into the training dataset. The motive threshold for raw and quantized feature representation is set to 0.4. The minsup is set to 0.2. The results of each feedback strategy applied with different *K* are illustrated in the following.

Figure 5.3, 5.5, and 5.7 illustrate the result for raw feature representation. In the experiment, we find out there's no regular positive correlation between *K* and the precision. When *K* is small (e.g. *K* =5), it's reasonable that data mining and classification technique can't bring much benefit for refinement due to scarcity of training samples. However, we are confused by the fact that accuracy drops when *K* is larger (e.g. *K*=20). On the other hand, we found out that MP strategy is sensitive to parameter *K* while MI and HB have stable performance under different *K*.

Figure 5.4, 5.6, and 5.8 show the performance of quantized one. In this experiment, there is no clear positive correlation between *K* and the precision as well. Beside, MP strategy is sensitive to parameter *K* while the others aren't.

In the case of MP strategy, the best performance during the whole session in quantized one occurs when *K* is 10. When a user is impatient to interact with the system, it's a good choice to utilize MP strategy by judging 10 examples each round. In such case, the user can obtains about 65.6% accuracy at the third round. In comparison with the raw feature representation, over 60% accuracy can be expected after the third round in quantized version in most cases. The highest accuracy 70% occurs at round seven in the case of *K*=10. On the contrary, raw one doesn't show satisfactory results which keeps accuracy under 60%. In short, MP strategy in quantized one is effective and it is a good choice for all users.

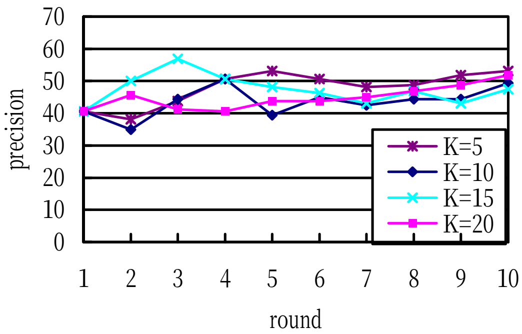

Figure 5.3 Performance of MP strategy with different *K* on raw feature representation.

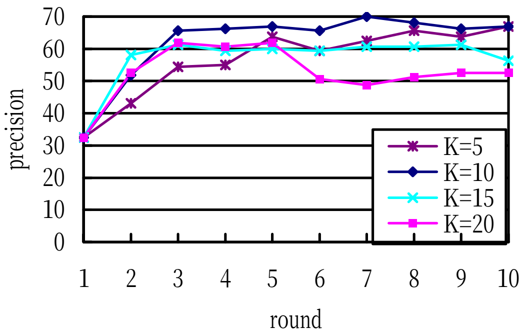

Figure 5.4 Performance of MP strategy with different K on quantized feature representation

In the case of MI strategy, *K* doesn't have notable impact on the performance in raw and quantized feature representation. The difference caused by different *K* only appears during the later half session in Figure 5.6 where positive correlation between *K* and the precision is relatively clear. In quantized version (Figure 5.6), the accuracy of 65% can be expected in all cases at round five. This fact indicates that we can only consider the case of *K*=5 where all users can obtain satisfactory results by only judging 20 examples at most. One may say that MI strategy can't bring astonished improvement or behave worse than MP if we compare MI and MP strategy in terms of the number of rounds. However, if we evaluate MI strategy in terms of efforts that user has to make, we are not disappointed with MI strategy. According to the observation that the system can achieve the same accuracy about 65% in both MI and MP strategy by considering the same amount of relevance feedback, MI strategy is effective as well. The only difference is that MI has to perform four rounds of semantic learning while only two rounds of learning process are needed for MP strategy for the same performance. In this way, MI is less efficient. Moreover, the improvement via relevance feedback isn't consistent enough where the curves shown in Figure 5.6 rise and fall frequently compared to MP strategy in quantized representation.

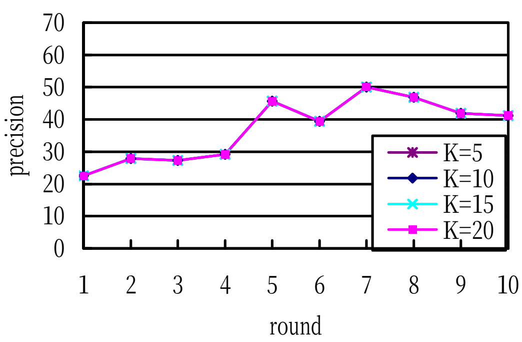

Figure 5.5 Performance of MI strategy with different *K* on raw feature representation.

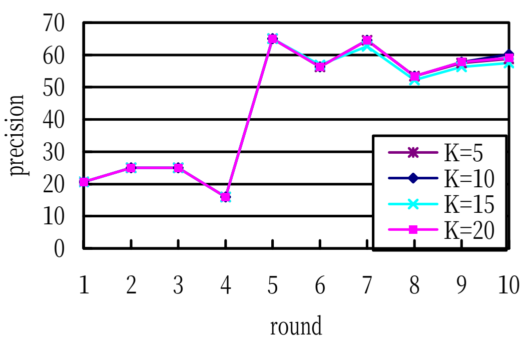

Figure 5.6 Performance of MI strategy with different *K* on quantized feature representation.

In comparison with raw feature representation (Figure 5.5), quantized one is more effective where accuracy is always beyond 50%. Raw version is insensitive to parameter K and doesn't have steady growth as well.

In HB feedback strategy, the performance is insensitive to parameter *K*. In Figure 5.7 and 5.8, all curves with different *K* have the same performance. Accuracy in raw feature representation (Figure 5.7) is about 40% at each round. Quantized one (Figure 5.8) has better accuracy about 50%. The curves in Figure 5.7 grow slightly as more rounds of learning process are performed. Quantized representation shows more improvement via relevance feedback. Hence curves in Figure 5.8 rise faster than raw ones.

In comparison with other strategies, HB strategy has worse performance. Hence, it's not a good choice for the system.

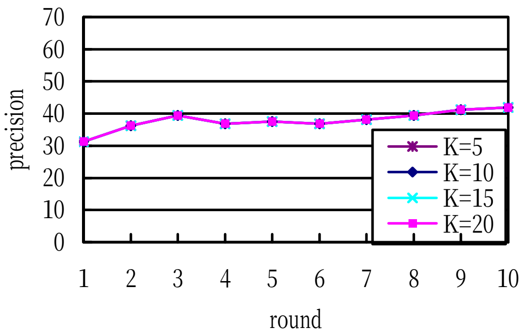

Figure 5.7 Performance of HB strategy with different *K* on raw feature representation.

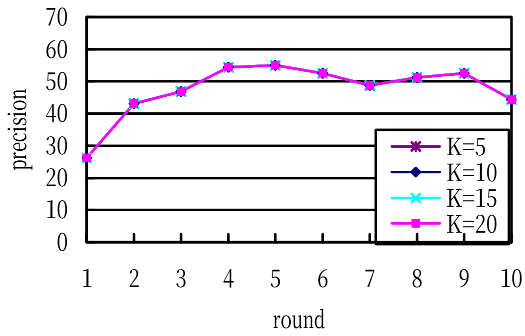

Figure 5.8 Performance of HB strategy with different *K* on quantized feature representation.

In summary, the performance of all strategies in quantized representation is better than that in raw one. MP strategy with 10 examples accumulated each round can achieve accuracy over 60% rapidly. Accuracy about 70% can be expected by judging about 60 music objects. Moreover, MP strategy ensures steady improvement of retrieval results via relevance feedback under all different *K*. On the other hand, MI strategy with K = 5 can be regarded as another option for users since it can achieve 65% accuracy as well by only judging 20 music objects over four rounds. Hence, the above two cases are good choices in terms of effectiveness.

5.3.3 Effectiveness of Number of Rounds Applied MI Strategy

We denote the number of rounds applied MI strategy in a session as *N*. In order to observe the effect of parameter *N*, four experiments with different *N* are conducted. The motive threshold is set to 0.4, and *K* is set to 10. In the first *N* rounds, at most 20 music objects belonging to default class are returned for relevance judgment. In other words, music objects returned in first *N* round may be less then 20 objects and the size is not fixed at each round. Hence, *N* isn't necessarily correlated with the amount of information gained from relevance judgment on uncertain music objects. Instead, what affects is the total amount of uncertain object collected in the first *N* rounds.

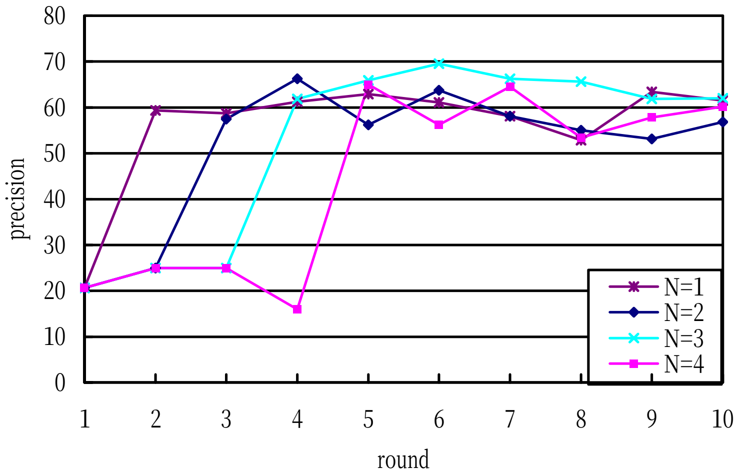

Figure 5.9 Performance of MI strategy with different *N* on quantized feature representation.

Figure 5.9 illustrates the results of each experiment. Intuitively, we expect accuracy of

N+1 round grows as *N* increases. For instance, the accuracy at the fifth round in the case of *N*=4 is higher than that at the fourth round when *N*=3. However, the improvement isn't obvious. Hence, the performance of MI strategy is insensitive to the parameter *N*.

5.3.4 Effectiveness of Support and Motive Threshold

We attempt to investigate how motive threshold affects the performance in this section. Motive threshold for SMP ranges from 0.3 to 0.5. Minsup for mining frequent pattern is fixed to 0.2. *K* is set to 10. Experiments with different motive threshold on raw and quantized feature representation are conducted respectively for each feedback strategy, MP strategy (Figure 5.10, 5.11), MI strategy with *N*=4 (Figure 5.12, 5.13), and HB strategy (Figure 5.14, 5.15).

Intuitively, the system may wok well as more SMPs are kept in one music object. This is because that the higher motive threshold the less SMPs are retained as local features. If user's concept is highly correlated with those pruned SMPs, less knowledge will be obtained concerning the user's concept. Consequently, higher motive threshold may cause the system to miss potential patterns. The classifier derived based on a less tiny pattern set will contain less discrimination rules (even no rules) and may reduce the classification accuracy. The performance of experiments in raw one (Figure 5.10, 5.12, 5.14) roughly shows the correlation between motive threshold and accuracy.

In quantized representation (Figure 5.11, 5.13, 5.15), performance with motive threshold 0.4 is better than those with 0.3 or 0.5. It's reasonable that music objects with motive threshold 0.5 keeps less information. Hence, experiment with motive threshold 0.5 has the worst performance in most cases in quantized one. Music objects with motive threshold 0.3

have the most amount of information while it doesn't show the best performance in all experiments with different motive threshold. Although we are confused with the results, it isn't a bad news for users and the system. In terms of efficiency, the system with lower motive threshold consumes less time and storage.

In the case of MP strategy, performance of raw one clearly demonstrates the impact of motive threshold (Figure 5.10). On the contrary, the impact of motive threshold doesn't affect the performance so much (Figure 5.11). One of the evidence occurs at the first round where the accuracy is descended in the order of motive threshold 0.3, 0.5 and 0.4. However the one with motive threshold 0.4 achieve the highest accuracy in most rounds during the session. Such conflicting evidence indicates the weak impact of motive threshold in quantized case and appears in other feedback strategies as well (Figure 5.13, 5.15).

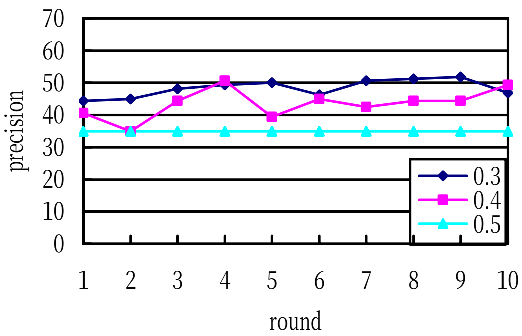

Figure 5.10 Performance of MP strategy with different motive threshold on raw feature representation.

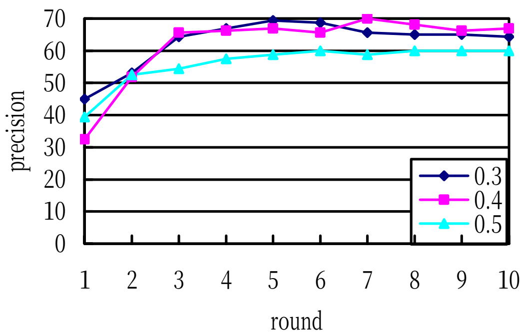

Figure 5.11 Performance of MP strategy with different motive threshold on quantized feature representation.

The significance of motive is proportional to its frequency in one music object and inversely related to its support in the music database. Higher motive threshold may cause mining task more difficult since less SMP information is retained. In such case, lower minsup is preferred and higher precision is achieved. However, the lowest minsup the system can accept is 0.2 (20%). When minsup is lower than 20%, a huge amount of frequent patterns will be generated and will increase useless computation. Considering the case when minsup is 10% in a database containing 10 music objects, each item will form a frequent pattern which is meaningless and unnecessary. Therefore we only consider the best case when minsup is 0.2.

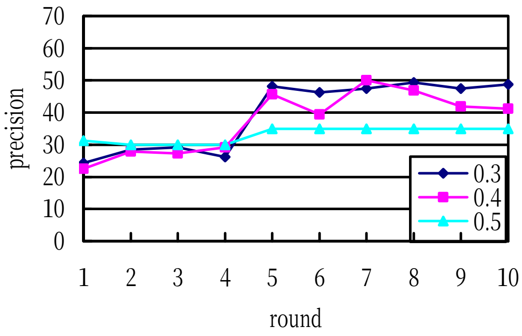

Figure 5.12 Performance of MI strategy with different motive threshold on raw feature representation.

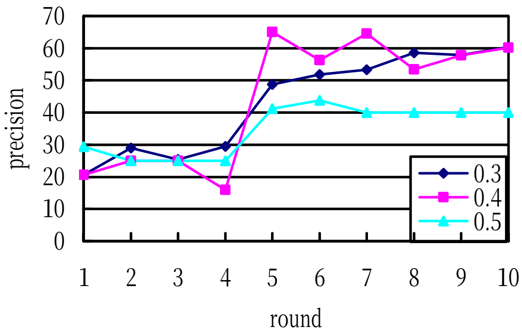

Figure 5.13 Performance of MI strategy with different motive threshold on quantized feature representation.

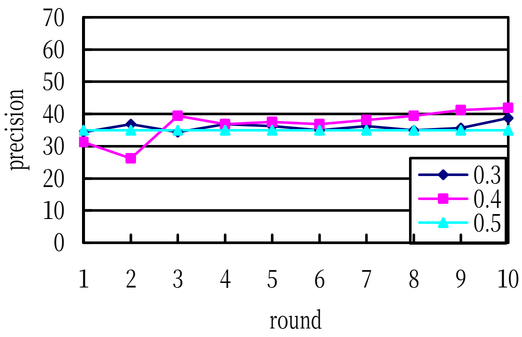

Figure 5.14 Performance of HB strategy with different motive threshold on raw feature representation.

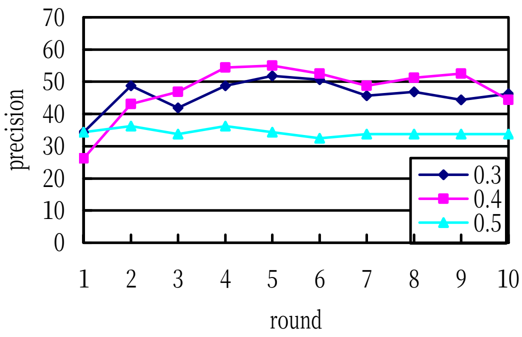

Figure 5.15 Performance of HB strategy with different motive threshold on quantized feature representation.

In summary, the performance in quantized representation is more efficient and effective than that in raw one. With suitable parameter setting, MP and MI strategies in quantized one are good choices for all users.

CHAPTER 6

Conclusions

Relevance feedback is helpful for music retrieval. Especially, most users are frustrated by the specification of music query. In this report, we proposed the relevance feedback mechanism for category search in content-based music retrieval. The main idea of our approach is to discover the relationship between the semantic concept behind the cognition of music category and the low level music features. A segment-based music modeling approach is presented which takes both global and local features into consideration. To discover user's semantic concept, a two-phase frequent pattern mining algorithm is presented to discover common properties from relevant and irrelevant music samples respectively. Then the classification algorithm modified from association classification is employed for discrimination of relevant and irrelevant concepts.

Relevance feedback is a process of a series of interaction between the user and system. In the user's feedback, the user judges each returned music object either as relevant or irrelevant to the concept. The system then learns the user's semantic concept and returns a collection of music objects for user's relevance judgment. Three system feedback strategies are investigated for music retrieval. Most-positive returns the most relevance music and thus provides less new information for the system to return results in the next round. Most-informative returns the most uncertain results useful for user's discrimination and therefore is helpful for the next round. Hybrid strategy is the compromise between these two strategies. Comparative experiments are conducted to evaluate effectiveness of the proposed refinement mechanism. Experimental results show that a better retrieval result can be achieved via refinement mechanism.

REFERENCES

[1] Agrawal, R. and Srikant, R. Fast algorithms for mining association rules. *In Proc. of Intl. Conference on Very Large databases (VLDB ‘94)*, (Chile, September 12-15, 1994).

[2] Amir, A. Berg, M. and Permuter. H. Mutual relevance feedback for multimodal query formulation in video retrieval. *In Proc. of ACM SIGMM Intl. Workshop on Multimedia Information Retrieval (MIR ‘05)* (Singapore, November 10-11, 2005).

[3] Chen, H. and Chen Arbee L.P. A music recommendation system based on music data grouping and user interests. *In Proc. of the ACM CIKM Intl. Conference on Information and Knowledge Management (CIKM ‘01)* (Atlanta, Georgia, USA, November 5-10, 2001). ACM press 2001, 231-238.

[4] Cox, I.J. Miller, M. Minka, T.P. Papathomas, T. and Yianilos P. The baysian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments. *IEEE Trans. Im*age *Processing, 9,* 1, (Jan. 2000), 20-37.

[5] Grimaldi, M. and Cunningham. P. Experimenting with music taste prediction by user profiling. *In Proc. of ACM SIGMM Intl. Workshop on Multimedia Information Retrieval (MIR ‘04)* (New York, NY, USA, October 15-16, 2004). ACM press 2004, 173-180.

[6] Haas, M. Lamel, L. Thomee, B. and Lew, M.S. Relevance feedback: perceptual learning and retrieval in bio-computing, photos, and video. *In Proc. of the ACM SIGMM Intl. Workshop on Multimedia Information Retrieval (MIR ‘04)* (New York, NY, USA, October 15-16, 2004). ACM press 2004, 151-156.

[7] Han, E.H. and Karpis, G. Feature-based Recommendation System. *In Proc. of the ACM CIKM Intl. Conference on Information and Knowledge Management (CIKM '05)* (Bremen, Germany, October 31-November 5, 2005). ACM press 2005, 446-452.

[8] He, X.F. King, W.Y. Ma, W.Y. Li, M.J. and Jiang, H.J. Learning a semantic space from user's relevance feedback for image retrieval. *IEEE Trans. Circuits and Systems and Video Technology, 13, 1* (Jan 2003), 39-48.

[9] Ho, M.C. *Theme-Based Music Structural Analysis.* Master Thesis, University of Chen Chi, Taipei, Taiwan, 2004.

[10] Hoi, C.H. and Lyu, M.R. A novel log-based relevance feedback technique in content-based image retrieval. *In Proc. of the ACM Intl. Conference on Multimedia (MM '04)* (New York, NY, USA, October 10-16, 2004). ACM press 2004, 24-31.

[11] Hoashi, K. Matsumoto, K. and Inoue, N. Personalization of user profiles for content-based music retrieval based on relevance feedback. *In Proc. of the ACM Intl. Conference on Multimedia (MM '03)* (Berkeley, CA, USA, November 2-8, 2003). ACM press 2003, 110-119.

[12] Hsu, J.L. Liu, C.C. and Chen, Arbee L.P. Discovering non-trivial repeating patterns in music data. *In Proc. of Intl. Conference on Data Engineering (ICDE '99)* (Sydney, Australia, March 23-36, 1999). IEEE Computer Society Press, 1999, 14-21.

[13] Jing, F. Li, M. Zhang, L. Zhang, H.J. and Zhang, B. Relevance feedback in region-based image retrieval. *IEEE Trans. Circuits and Systems and Video Technology, 14,* 5 (May 2004), 672-681.

[14] Jing, F. Li, M. Zhang, H.J. Zhang, B. An effective region-based image retrieval framework. *In Proc. of the ACM Intl. Conference on Multimedia (MM '02)* (Juan les Pins, France, December 1-6, 2002). ACM press 2002, 456-465.

[15] Kuo, F.F. and Shan, M.K. A personalized music filtering system based on melody style classification. In Proc. of the IEEE Intl. Conference on Data Mining (ICDM ‘02) (Maebashi, Japan, December 9-12, 2002). IEEE Computer Society Press 2002, 649-652.

[16] Liu, B. Hsu, W. and Ma, Y. Integrating Classification and Association Rule Mining. *In Proc. of the Intl. Conference on Knowledge Discovery and Data Mining (KDD'98)* (New York, USA, August 27-31, 1998). AAAI Press, 1998, 80-86.

[17] Liu, C.C. Hsu, J.L. and Chen, A.L.P. An approximating string matching algorithm for content-based music data retrieval. *In Proc. of IEEE Intl. Conference on Multimedia Computing and Systems (ICMCS ‘99)* (Florence, Italy, June 7-11, 1999). IEEE Computer Society, Press 1999, 451-456.

[18] Ortega-Binderberger, and M. Mehrotra, S. Relevance feedback techniques in the MARS image retrieval system. *Multimedia System, 9, 6* (June. 2004)535-547.

[19] Ragno, R. Burges, C.J.C. and Herley, C. Inferring similarity between music objects with application to playlist generation. *In Proc. of ACM SIGMM Intl. Workshop on Multimedia Information Retrieval (MIR ‘05)* (Singapore, November 10-11, 2005).

[20] Rocchio, J.J. and G. Salton. *Relevance feedback in information retrieval.* Prentice Hall. Inc., Englewood Cliffs, New Jersey, 1971.

[21] Rui, Y. Huang, Thomas S. Ortega, M. and Mehrotra, S. Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Trans. Circuits and Systems for Video Technology, 8, 5* (Sep. 1998), 644-655.

[22] Schohn, G. and Cohn, D. Less is more: active learning with support vector machines *In Proc. of the Intl. Conference on Machine Learning (ICML ‘00)* (Stanford, CA, USA, June 29-July 2, 2000). Morgan Kaufmann 2000, 839-846.

[23] Selfridge-Field. E. Conceptual and representational issues in melodic comparison. In Hewlett, W.B. & Selfridge-Field E. Melodic similarity: Concepts, procedures, and applications (Computing in Musicology: 11), The MIT Press.

[24] Shardanand, U. and Maes, P. Social information filtering: algorithms for automating "Word of Mouth" *In Proc. of the Conference on Human Factors in Computing Systems (CHI '95)* (Denver, Colorado, USA, May 7-11, 1995). ACM/Addison-Wesley Press 1995, 210-217.

[25] Shen, X. and Zhai, C. Active feedback in Ad Hoc information retrieval. *In Proc. of ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05)* (Salvador, Bahia, Brazil, August 15-19, 2005). ACM press 2005, 59-66.

[26] Stein, L. *Structure & Style*. Summy-Birchard, 1979.

[27] Tong, S. and Koller, D. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, *2* (Nov. 2001), 45-66.

[28] Tong, S. and Chang, E. Support vector machine active learning for image retrieval. *In Proc. of ACM Intl. conference on Multimedia (MM '01)* (Ottawa, Ontario, Canada, September 30- October 5, 2001). ACM press 2001, 107-118.

[29] Uitdenbgerd, A. and Zobel, J. Melodic matching techniques for large music databases. *In Proc. of ACM Intl. conference on Multimedia (MM '99)* (Orlando, Florida, USA, October 30-Novermber 5, 1999). ACM press 1999, 57-66.

[30] Wu, Y. and Zhang, A. Interactive pattern analysis for relevance feedback in multimedia information retrieval. *Multimedia System, 10, 1* (June. 2004)41-55.

[31] Wu, Y. Tian, Q. and Huang, T.S. Discriminant-EM algorithm with application to image retrieval. *In Proc. of IEEE Intl. Conference on Computer Vision and Pattern*

Recognition (CVPR '00) (Hilton Head, SC, USA, June 13-15, 2000). IEEE computer society 2000, 1222-1227.

[32] Yan, R. Hauptmann. A. and Jin, R. Negative pseudo-relevance feedback in content-based video retrieval. *In Proc. of the ACM Intl. Conference on Multimedia (MM '03)* (Berkeley, CA, USA, November 2-8, 2003). ACM press 2003, 343-346.

[33] Zhou, X.S. and Huang, Thomas S. Relevance feedback in image retrieval: a comprehensive review. *Multimedia System, 8, 6* (Apr. 2003)536-544.

計畫成果自評

■ 就研究內容與原計畫相符程度、達成預期目標情況

本研究計畫已經完成音樂上的相關回饋技術，以提供符合使用者概念的音樂查詢之技術研究。本計劃從使用者的相關回饋中學習使用者的概念，並利用這學習出的概念來幫助音樂查詢。由於使用者可能從整首音樂或音樂片段兩種角度來判斷該音樂是否具備使用者定義的概念。因此，本論文提出用以片段為主的音樂模型將音樂表示成音樂片段的集合。進一步再從整首音樂和片段中擷取特徵。其次，我們針對該問題提出相關演算法來探勘使用者的概念。最後，我們分析各種系統回饋機制對搜尋效果的影響。本文中，我們模擬並比較各種回傳機制的效能。實驗結果顯示相關回饋機制確實能提升查詢的效果。

■ 研究成果之學術或應用價值

本計畫的研究成果，在學術價值方面，我們提出新的音樂表示法。我們也提出演算法來探勘使用者的概念。本研究有助於音樂搜尋與音樂特徵擷取的相關研究。
在應用價值方面，本計畫所開發的技術可利用於文化、娛樂與音樂教育產業、智慧型音樂檢索系統中音樂查詢的功能。也可應用在音樂學習軟體，以幫助使用者對本身喜好音樂的在音樂特性上的了解。

■ 是否適合在學術期刊發表或申請專利

本計畫的研究成果除了三篇碩士論文之外，部分研究成果已經發表在國際學術會議IEEE Syetems, Man and Cybernetics (IEEE SMC'06)、6th IEEE International Conference on Advanced Learning Technologies(ICALT'06)、11th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'07)。此外，我們也將部分研究成果整理後，投稿至國際學術期刊 Multimedia Tools and Applications special issue on Semantic Multimedia。

■ 主要發現或其他有關價值

我們分析各種系統回饋機制對搜尋效果的影響。Most-positive 回傳機制會選擇根據目前系統判斷為最相關的物件。Most-informative 機制則是回傳系統無法判斷其相關性的音樂物件。Most-informative 機制的目的在增加每回合系統從使用者身上得到的資訊量。Hybrid 則是中和前兩種機制的優點。本文中，我們模擬並比較各種回傳機制的效能。實驗結果顯示相關回饋機制確實能提升查詢的效果。

參與計畫的學生，有兩名同學繼續就讀於國立交通大學資訊工程研究所博士班，繼續進行資訊擷取與資料探勘的相關研究。另有一位同學於華寶通訊工作，負責手機軟體的開發。兩名同學即將完成碩士論文。

附件二

可供推廣之研發成果資料表

☐ 可申請專利　　☑ 可技術移轉　　　　　　　　　　　　　　日期：96 年 5 月 31 日

國科會補助計畫	計畫名稱：數位音樂典藏之資料探勘與智慧型檢索技術(Ⅱ) 計畫主持人：沈錳坤 計畫編號：NSC 95-2422-H-004-002　學門領域：數典國家型計畫
技術/創作名稱	根據概念學習發展以內容為主的音樂查詢之相關回饋機制
發明人/創作人	沈錳坤
技術說明	中文： 我們提出將相關回饋(relevance feedback)的機制運用在以內容為主的音樂查詢系統上，讓系統從使用者的相關回饋中學習使用者的概念，並利用這學習出的概念來幫助音樂查詢。 在這個系統技術上，主要可分為三個部分。 第一：我們提出以片段為主的音樂模型(segment-based modeling approach)將音樂表示成音樂片段的集合。進一步再從整首音樂和片段中擷取特徵。 第二：我們針對該問題提出相關演算法來探勘使用者的概念。該演算法先從相關和不相關的音樂資料庫中個別探勘常見樣式，再利用這些樣式建立分類器以區分音樂的相關性。 第三：分析各種系統回饋機制對搜尋效果的影響。
	英文： The relevance feedback mechanism for category search of music retrieval based on the semantic concept learning is proposed. This technique consists of three main parts. 1. We proposed a segment-based music representation to assist the system in discovering user's concept in terms of low-level music features. 2. To discover user's semantic concept, a two-phase frequent pattern mining algorithm is proposed to discover common properties from relevant and irrelevant objects respectively and based on which a classifier is derived for distinguishing music objects. 3. Except user's feedback, three strategies of the system's feedback to select objects for user's relevance judgment are investigated.
可利用之產業及可開發之產品	1. 可利用於文化、娛樂與音樂教育產業 2. 智慧型音樂檢索系統，在網站中提供更精準的音樂搜尋 3. 音樂分析軟體，幫助使用者對自己本身喜好的音樂，其在音樂上特性的了解

技術特點	1. 提出相關回饋的機制運用在以內容為主的音樂查詢系統上 2. 運用資料探勘的技術，從使用者的相關回饋中學習使用者的概念，並利用這學習出的概念來幫助音樂查詢。
推廣及運用的價值	大量音樂典藏或資料庫中，相關回饋技術將可提供使用者更精準搜尋音樂的功能。

出席國際學術會議心得報告

計畫編號	NSC　95－2422－H－004－002
計畫名稱	數位音樂典藏之資料探勘與智慧型檢索技術(Ⅱ)
出國人員姓名 服務機關及職稱	魏綾音(國立政治大學資訊科學系碩士班研究生)
會議時間地點	2006 年 7 月 5 日~2006 年 7 月 7 日, Kerkrade, The Netherlands
會議名稱	IEEE International Conference on Advanced Learning Technologies, ICALT'06
發表論文題目	Mining Spatial Co-orientation Patterns for Analyzing Portfolios of Spatial Cognitive Development

一、參加會議經過

先進學習科技國際學術研討會(International Conference on Advanced Learning Technologies, ICALT)於今年七月五日至七月七日在荷蘭南部的一個小鎮凱爾克拉德(Kerkrade)舉行三天。先進學習科技國際學術研討會是由國際知名的 IEEE 國際電子電機工程學會之電腦與學習科技協會(Technical Committee on Learning Technology)所主辦，是教育方面具有權威性的重要會議。

今年 ICALT 會議共有 492 篇論文投稿，其中 136 篇論文被接受為 Full Paper(論文接受度為 27.64%)，122 篇論文被接受為 Short Paper，42 篇論文被接受為 Poster，且今年有來自 56 個國家的學者們參與此會議。今年的會議主題是 Advanced Technologies for Life-Long Learning，終生學習（Life-long Learning），當行動通訊(Mobile Communication)與網路的基礎建設日漸普及，這使得終生學習(Life-Long Learning)更可容易達成。今年的會議區分了以下 24 個議題，包括：

1. Adaptivity in Learning Systems
2. Advanced Uses of Multimedia and Hypermedia
3. Architecture of Context Aware Learning Technology Systems
4. Artificial Intelligence Tools for Contextual Learning
5. Building Learning Communities
6. Concretizing Technologies in Learning
7. Educational Modelling Languages / Learning Design
8. Educational Paradigms
9. Evolution in Teacher Roles through Technology
10. Information Retrieval and Visualization Methods for Learning
11. Instructional Design Theories
12. Interactive Learning Systems
13. Learning Objects for Personalized Learning
14. Media for Learning in Multicultural Settings
15. Metadata for Learning Resources

16. Mobile Learning Applications
17. Participatory Simulations
18. Pedagogical and Organizational Frameworks
19. Peer-to-Peer Learning Applications
20. Social Software for Collaborative Learning at a Distance
21. Socially Intelligent Agents
22. Technology-Facilitated Learning in Complex Domains
23. Technologies for advanced Teacher Training
24. Virtual Spaces for Learning Communities

我的論文是屬於前面所提到的 24 個議題中的第 22 個議題(Technology-Facilitated Learning in Complex Domains)，而我的論文題目是”應用在空間認知發展的學習歷程分析之空間探勘演算法。此篇論文主要將空間資料探勘(Spatial Data Mining)應用在地理教育方面，透過電腦的技術分析學生空間認知發展的學習歷程檔案。在荷蘭當地時間七月七日上午，我進行了論文發表，這是我第一次於國際會議上發表論文，對我來說是一種新的體驗與歷練，過去以英文報告的機會不多，不過還是帶著緊張的心情，順利完成。

同一個議題的 Section 裡，也有來自清華大學的學生報告”Implementing e-Learning for the Media Industry: A Case Study a Small-sized Advertising Company in Taiwan”，以及來自希臘 National Technical University of Athens 的學生報告”A Model for Interoperability in Computer Supported Collaborative Learning”。

這次會議的主辦學校是荷蘭的 Fontys University，這所大學是荷蘭最大的職業教育大學，共有 33000 位學生、3500 個職員，主要是由 Fontys University of Applies Sciences 所承辦，它包括了六個學習領域，如 Economics and Business、Health Care、Arts、Social Students、Education、Technology。Fontys University of Applied Sciences 是荷蘭在教育方面的知名學校，也是荷蘭在應用科技方面第二大的學校，除此，此學校的其他領域在世界上也都具有相當知名度。

這次會議舉行的地點是在 Kerkrade 的 Hotel Rolduc(見圖三)，Rolduc 原本是修道院興建於十二世紀，是一個寧靜且歷史悠久的地方。這次的會議裡，除了各個學校的學者們參與之外，也有產業界的開發人員參與此會議，他們開發了數位學習方面的相關應用軟體，且藉由這樣難得的機會宣傳至國際。

二、與會心得

很感謝國科會國家型數位典藏計畫補助會議註冊費及生活費約三萬元，讓我有機會參與這次的國際學術會議，與其他國外學者進行交流，聽到來自不同國家學者的看法與建議。在這次的會議裡，認識了來自台灣清華大學、成功大學、中山大學、中央大學與元智大學的教授、博士班研究生或碩士班研究生，也認識了來自其他國家的博士班研究生。很高興有機會和不同國家相關領域的學者們進行交流與學習，分享彼此的經驗，讓我收穫良多。對於我將來在學術生涯的規劃，有很大的幫助。

荷蘭是一個充滿藝術氣息的國家，在當地火車的每節車廂裡，都會有各式各樣特別的畫作在牆壁上。荷蘭也是一個旅遊觀光勝地，荷蘭的面積約是台灣的 1.1 倍，鐵路交通網密集使得交通即為便利，荷蘭也注重觀光產業的發展，火車站的諮詢台服務人員對

於各種交通工具(例如火車、地鐵、公車)等相關資訊都有相當程度的瞭解，這解決了來自其他國家旅客的不少問題，而且服務人員、售票員與查票員的熱心，讓旅客們感到窩心。不僅如此，火車站前一定都會有當地的街道圖，且重要路口也都會有地圖，這使得國外旅客不容易迷路。在荷蘭境內有統一的車票可以搭乘各個城市的公車、捷運或有軌電車，這也方便了想到荷蘭各個城市旅遊的旅客。

出席國際學術會議心得報告

計畫編號	NSC　95－2422－H－004－002
計畫名稱	數位音樂典藏之資料探勘與智慧型檢索技術(Ⅱ)
出國人員姓名 服務機關及職稱	魏綾音(國立政治大學資訊科學系碩士班研究生)
會議時間地點	2006 年 7 月 5 日~2006 年 7 月 7 日, Kerkrade, The Netherlands
會議名稱	IEEE International Conference on Advanced Learning Technologies, ICALT'06
發表論文題目	Mining Spatial Co-orientation Patterns for Analyzing Portfolios of Spatial Cognitive Development

一、參加會議經過

先進學習科技國際學術研討會(International Conference on Advanced Learning Technologies, ICALT)於今年七月五日至七月七日在荷蘭南部的一個小鎮凱爾克拉德(Kerkrade)舉行三天。先進學習科技國際學術研討會是由國際知名的 IEEE 國際電子電機工程學會之電腦與學習科技協會(Technical Committee on Learning Technology)所主辦，是教育方面具有權威性的重要會議。

今年 ICALT 會議共有 492 篇論文投稿，其中 136 篇論文被接受為 Full Paper(論文接受度為 27.64%)，122 篇論文被接受為 Short Paper，42 篇論文被接受為 Poster，且今年有來自 56 個國家的學者們參與此會議。今年的會議主題是 Advanced Technologies for Life-Long Learning，終生學習（Life-long Learning），當行動通訊(Mobile Communication)與網路的基礎建設日漸普及，這使得終生學習(Life-Long Learning)更可容易達成。今年的會議區分了以下 24 個議題，包括：

1. Adaptivity in Learning Systems
2. Advanced Uses of Multimedia and Hypermedia
3. Architecture of Context Aware Learning Technology Systems
4. Artificial Intelligence Tools for Contextual Learning
5. Building Learning Communities
6. Concretizing Technologies in Learning
7. Educational Modelling Languages / Learning Design
8. Educational Paradigms
9. Evolution in Teacher Roles through Technology
10. Information Retrieval and Visualization Methods for Learning
11. Instructional Design Theories
12. Interactive Learning Systems
13. Learning Objects for Personalized Learning
14. Media for Learning in Multicultural Settings
15. Metadata for Learning Resources

16. Mobile Learning Applications
17. Participatory Simulations
18. Pedagogical and Organizational Frameworks
19. Peer-to-Peer Learning Applications
20. Social Software for Collaborative Learning at a Distance
21. Socially Intelligent Agents
22. Technology-Facilitated Learning in Complex Domains
23. Technologies for advanced Teacher Training
24. Virtual Spaces for Learning Communities

我的論文是屬於前面所提到的 24 個議題中的第 22 個議題(Technology-Facilitated Learning in Complex Domains)，而我的論文題目是”應用在空間認知發展的學習歷程分析之空間探勘演算法。此篇論文主要將空間資料探勘(Spatial Data Mining)應用在地理教育方面，透過電腦的技術分析學生空間認知發展的學習歷程檔案。在荷蘭當地時間七月七日上午，我進行了論文發表，這是我第一次於國際會議上發表論文，對我來說是一種新的體驗與歷練，過去以英文報告的機會不多，不過還是帶著緊張的心情，順利完成。

同一個議題的 Section 裡，也有來自清華大學的學生報告”Implementing e-Learning for the Media Industry: A Case Study a Small-sized Advertising Company in Taiwan”，以及來自希臘 National Technical University of Athens 的學生報告”A Model for Interoperability in Computer Supported Collaborative Learning”。

這次會議的主辦學校是荷蘭的 Fontys University，這所大學是荷蘭最大的職業教育大學，共有 33000 位學生、3500 個職員，主要是由 Fontys University of Applies Sciences 所承辦，它包括了六個學習領域，如 Economics and Business、Health Care、Arts、Social Students、Education、Technology。Fontys University of Applied Sciences 是荷蘭在教育方面的知名學校，也是荷蘭在應用科技方面第二大的學校，除此，此學校的其他領域在世界上也都具有相當知名度。

這次會議舉行的地點是在 Kerkrade 的 Hotel Rolduc(見圖三)，Rolduc 原本是修道院興建於十二世紀，是一個寧靜且歷史悠久的地方。這次的會議裡，除了各個學校的學者們參與之外，也有產業界的開發人員參與此會議，他們開發了數位學習方面的相關應用軟體，且藉由這樣難得的機會宣傳至國際。

二、與會心得

很感謝國科會國家型數位典藏計畫補助會議註冊費及生活費約三萬元，讓我有機會參與這次的國際學術會議，與其他國外學者進行交流，聽到來自不同國家學者的看法與建議。在這次的會議裡，認識了來自台灣清華大學、成功大學、中山大學、中央大學與元智大學的教授、博士班研究生或碩士班研究生，也認識了來自其他國家的博士班研究生。很高興有機會和不同國家相關領域的學者們進行交流與學習，分享彼此的經驗，讓我收穫良多。對於我將來在學術生涯的規劃，有很大的幫助。

荷蘭是一個充滿藝術氣息的國家，在當地火車的每節車廂裡，都會有各式各樣特別的畫作在牆壁上。荷蘭也是一個旅遊觀光勝地，荷蘭的面積約是台灣的 1.1 倍，鐵路交通網密集使得交通即為便利，荷蘭也注重觀光產業的發展，火車站的諮詢台服務人員對

於各種交通工具(例如火車、地鐵、公車)等相關資訊都有相當程度的瞭解，這解決了來自其他國家旅客的不少問題，而且服務人員、售票員與查票員的熱心，讓旅客們感到窩心。不僅如此，火車站前一定都會有當地的街道圖，且重要路口也都會有地圖，這使得國外旅客不容易迷路。在荷蘭境內有統一的車票可以搭乘各個城市的公車、捷運或有軌電車，這也方便了想到荷蘭各個城市旅遊的旅客。