

行政院國家科學委員會專題研究計畫 期中進度報告

數位家庭助理之研究與實作(2/3) 期中進度報告(精簡版)

計畫類別：個別型
計畫編號：NSC 95-2221-E-004-017-
執行期間：95年08月01日至96年07月31日
執行單位：國立政治大學資訊科學系

計畫主持人：陳良弼

處理方式：期中報告不提供公開查詢

中華民國 96年05月31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

數位家庭助理之研究與實作(2/3)

Research and Implementation of Digital Home Assistants

計畫類別： 個別型計畫 整合型計畫
計畫編號：NSC-94-2213-E-004-010
執行期間：94年08月01日至97年07月31日

計畫主持人：陳良弼

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 05 月 31 日

數位家庭助理之研究與實作(2/3)

Research and Implementation of Digital Home Assistants

計畫編號：NSC-94-2213-E-004-010

執行期間：94年08月01日至97年07月31日

計畫主持人：陳良弼

執行單位：國立政治大學資訊科學系

中文摘要

為了達成實作數位家庭助理的目的，本計劃致力增進兩大類的技術。一為多媒體與跨媒體處理技術，另一為股票分析技術。在多媒體與跨媒體處理技術方面，不同於去年我們分別就音樂和影像做處理，今年，我們開始發展在不同媒體環境下都能使用的技術。主要有多媒體高維索引(high dimensional index)技術和多媒體可調整距離函式(adjustable distance function)查詢技術兩方面。由於我們第三年想要完成的系統是橫跨影像與音樂兩大媒體，而這兩項技術都與媒體種類無關，因此都可以被使用在第三年所要做的系統上。另外，在股票分析方面，我們利用不同股票之間的關聯來預測股價未來的走勢，以幫助使用者進行家庭理財。

關鍵詞：多媒體、股票分析、查詢、索引、可調整距離函式

英文摘要

To implement digital home assistants, our project makes efforts in two kinds of techniques. One is the multimedia and cross-media processing technique. The other is the stock analysis technique. For the multimedia and cross-media processing technique, we developed a high dimensional indexing technique and a search technique with an adjustable distance function, which can be used in various media. Since the two techniques are media independence, they can be used for implementing the system which combines two media: image and music in the third year. Moreover, for the stock analysis, we utilized the relationship between different stocks to predict the stock trends.

Keyword: Multimedia, Stock analysis, Querying, Indexing, Adjustable Distance Function

一、前言

本子計劃的數位家庭助理希望能為使用者提供兩方面的幫助，一個是在家庭娛樂方面，一個是在家庭理財方面。在家庭娛樂方面，由於計畫的主要目的是希望能在第三年建立出家庭式 KTV 個人化背景影片播放系統，而這個系統需要整合影像與音樂兩個不同種類的媒體。因此，在今年我們致力完成兩項可使用在各種媒體的技術，一是多媒體高維索引技術，另一是多媒體可調整距離函式查詢技術。由於多媒體資料所對應的特徵空間其維度通常很高，因此一個良好的多媒體高維索引技術能加速對於多媒體資料的處理，以縮短使用者的等待時間。再者，使用者可能會想要下達不同語意(semantic)的查詢，而為了表達不同的語義，每一次查詢所使用的距離函式可能會不同。因此，我們發展了多媒體可調整距離函式查詢技術來滿足使用者的需求。由於這兩項技術都與媒體種類無關，因此都可以被使用在第三年所要做的系統上。而在家庭理財方面，我們發展了股價關聯分析及未來走勢預測的方法。透過分析不同股票之間的關聯，我們找出在時序性上具有共同趨勢的股票來幫助預測股票未來的走勢。這些資訊將提供給使用者作為買賣股票的參考，進而達到幫助家庭理財的目的。

二、研究目的

以下茲就多媒體高維索引、多媒體可調整距離函式查詢與股價關聯分析及未來走

勢預測等三大主題說明本計畫於今年度之研究目的：

多媒體高維索引技術

當多媒體資料的特徵被擷取出來以後，會形成一個維度極高的特徵空間，在一些應用中，這個空間的維度可能達到上萬維。而每一個多媒體物件都會被對應到此空間中的一個點。當我們在此高維空間中進行查詢時，如何建立一個有效的索引結構來加快查詢速度，並且避免高維空間所帶來的一些問題，一直是很重要的課題。本研究的目的是設計一套在高維空間中有效的索引方法，這個方法可以減少因為高維空間所產生的大量維度運算問題。而在第三年的時候，這套方法將會被應用在家庭式 KTV 個人化背景影片播放系統上。

多媒體可調整距離函式查詢

多媒體資料的數量每天都在增加，因此，如何快速且準確地搜尋使用者所需要的資料是一個很重要的議題。為了搜尋所需要的資料，我們通常需要比對多媒體資料之間的相似度。而多媒體資料之間的相似度，通常都可以使用度量空間(metric space)的距離函式計算。另外，考慮到使用者可能會想要下達不同語意(semantic)的查詢，而為了表達不同的語義，每一次查詢所使用的距離函式可能會不同。因此，我們設計了一套索引結構及演算法，能夠讓使用者在每次執行查詢時，使用不同的距離函式。我們提供的查詢種類包括 k 最近鄰點(k -nearest neighbor)及反向 k 最近鄰點(Reverse k -nearest neighbor)查詢。由於此方法與媒體種類無關，因此，將可以在第三年時，使用到家庭式 KTV 個人化背景影片播放系統上。

股價關聯分析及未來走勢預測

試圖在股票市場中投資獲利者，不可不重視股票價格的變化。然而影響股票價格的因素極為廣泛，對於如此大量且複雜的資訊，實非一般投資人可以輕易掌握的。我們試著尋找股票之間的關聯，尋找具備高度關聯的股票組合，藉由分析股價

的歷史資料來推導出群體股價變化之關聯法則，進而達到預測未來股價走勢的目的。本研究著重在於非即時資料分析，以「日」作為預測時間單位，讓忙碌現代人不需花費太多時間在每日的交易上。

三、文獻探討

以下茲就與多媒體高維索引、多媒體可調整距離函式查詢與股價關聯分析及未來走勢預測等三大主題有關的參考文獻做討論：

多媒體高維索引技術

在索引技術方面，[11]將過去的索引方法分成兩大類。一類為座標導向(coordinate-based)，另一類為距離導向(distance-based)。其中，座標導向的索引方法又可以分成空間基礎(space foundation)和資料基礎(data foundation)的。下面，我們將探討各種索引結構的優缺點。

座標導向索引方法的主要精神，是將空間中的點切割(separate)成數個部份。在此類別下，空間基礎和資料基礎的主要不同在於切割的觀念。空間基礎的索引方法[12]主要是將空間切割成數個大小相等或是所包含的點數相近的子空間。相對的，資料基礎的索引方法[17]主要是用矩形(rectangle)或是圓形(circle)來覆蓋具有聚集性的點，而這些矩形或是圓形通常會形成階層式的結構。但是，座標導向的索引方法在高維空間中是無效率的，主要是因為空間切割的觀念。當一個空間被切割以後，所有的切割資訊將會被保留下來。而使用者的查詢被執行時，這些空間切割資訊將會被拿來做比對，而這些比對的代價，將會隨著維度的提高而增加。

接著，我們來介紹距離導向的索引方式。距離導向的索引方法[2][19]會先挑選一些參考點(reference point)，然後特徵空間中的每個點都會跟參考點計算距離並且將其記錄下來。當查詢處理被執行時，這些所記錄的資訊將會與三角不等式一起被用來刪剪(pruning)不可能的答案。而由於三角不等式的計算只需要一次減

法，因此，距離導向索引方法的計算量是跟維度無關 (dimensional independence) 的。但是，因為距離導向索引方法將距離的資訊過度壓縮，因此維度一高的時候，整個刪剪的效果會很差，甚至有可能無法刪剪掉任何點。

多媒體可調整距離函式查詢

度量空間中的索引結構已經被廣泛地探討過了，[8]中討論了許多有效的解決方案。這些索引結構只需要知道任兩點之間的距離，而不需要知道距離如何被計算的細節。[6]提出了M-Tree這個索引結構，並且提供了M-Tree上的範圍查詢(range query)及 k 最近鄰點查詢。這個索引結構後來被許多研究所引用。

上述在度量空間的搜尋演算法，在建立索引結構時都必需先指定一個距離函式，而在查詢時必需使用相同的距離函式查詢。[5]首次探討了每次查詢時都可以使用不同距離函式的演算法，其中定義了兩種不同的距離函式。用來建立索引的距離函式，被稱為索引距離函式(index distance)。而使用者查詢時所下達的距離函式，稱為查詢距離函式(query distance)。並基於這樣的條件下發展了範圍查詢與 k 最近鄰點查詢的演算法。[4]延伸了這個方法，在索引結構中記錄了一些額外的資訊以達到加速的目的，但距離函式被線性組合(linear combination)所限制。

近年來，反向最近鄰點查詢是一個相當受關注的議題。大部份的演算法是為了歐吉里德空間(Euclidean distance)上的查詢所設計的[13][14]。然而最近，度量空間上的反向最近鄰點查詢也開始受到重視。[1]提出了基於預先計算的反向最近鄰點查詢演算法。[15]整理了M-Tree的許多特性，定義了一個不需預先計算的反向最近鄰點查詢演算法。

股價關聯分析及未來走勢預測

對於股價變化之關聯法則，在過去的文獻探討中，技術指標常被用來當作預測未來股價變化的工具，方法可分為兩類。

(一) 利用多種技術指標[9][18]，互

相組合來作為買賣策略的依據，組合的指標顯示為多頭時作買進，顯示為空頭時作賣出。

(二) 利用技術指標的數值變化搭配人工智慧，如類神經網路演算法[7][10]，算出最有效的數值變化當作買進與賣出的訊號，也可視為動態的技術指標組合，每日要使用的組合指標要經過人工智慧的方法來求出。

考慮到股票會有群體效應，在[16]的研究中提到，股票之間會有互相影響的關聯，例如聯電上漲時，台積電與矽品都上漲的機率相當高，利用股價的同漲同跌來作為兩兩之間的關聯度分析，並且延伸到群體同漲同跌的機率關係。

四、研究方法

以下茲就多媒體高維索引、多媒體可調整距離函式查詢與股價關聯分析及未來走勢預測等主題說明本計畫於今年度所提出之研究方法：

多媒體高維索引技術

根據文獻探討中的分析，我們選擇以距離導向索引方法為基礎，來設計我們的索引方法，主要是因為距離導向索引方法的計算量是跟維度無關的，因此較能克服高維空間所引發的問題。我們希望設計出來的索引方法能被使用在常用的查詢方式，如： k 最近鄰點查詢或是範圍查詢等。

我們採用[2]所提出的MRP(*multiple reference points*)方法。一開始，先從所有點中選擇幾個點來當成參考點，並且將每個點與這些參考點計算距離。以

Table 1為例，我們選擇 g 和 h 兩個點當作參考點，而所有其他點都與此兩點算距離，並將結果記錄在如Figure 1的 B^+ -tree中，以加快存取的速度。以下，當給定一個查詢時，我們以最近鄰點查詢(nearest neighbor query)為例，來說明我們的索引方法。主要可以分成下列兩步驟：

步驟一：找尋近似最近鄰點

我們使用 B^+ -tree中所記錄的資訊，來快速的求取一個近似的最近鄰點。我們的主要想法是，如果兩個點很近，那麼他們

對於某一個參考點的距離也應該很接近。所以我們從 B^+ -tree中該點的兩邊開始掃描並計算其與查詢真正的距離。以Figure 1為例，假設查詢是 e 點，則我們從 B^+ -tree中 e 兩旁的 a, b, c 和 d 四個點開始計算與 e 的真正距離。透過這種方式，我們可以只計算很少的點就能找出一個近似的最近點。

以 MNIST 手寫數字資料庫為例，此資料庫中有 60000 個資料，並且每個資料會被對應到 784 維的空間中的一個點。當給定一個查詢點時，我們從每個參考點的 B^+ -tree 中該查詢點的位置往兩旁開始掃描，並對掃描到的點計算其與查詢點真正的距離。根據實驗，只需這樣計算一百個不同的點，我們便可以找到平均距離查詢點約第 8.45 近的点。

步驟二：計算距離時，採用維度刪剪

透過第一步，我們可以得到一個近似的最近鄰點。然後利用此近似最近鄰點到查詢點的距離，我們可以加速距離運算。主要的想法是在計算某個點到查詢點的距離時，我們一維一維的將兩個相同維度的值相減取平方後相加。如果在過程中，這個相加的總和已經大過近似最近鄰點到查詢點距離的平方，那麼這個點就不可能為答案了。一樣以 MNIST 資料庫為例，我們對每個點都求取其最近鄰點，利用步驟二，我們可以將兩點間原本需要 784 維的距離運算，降低到平均只有 91.7 維。

我們的方法除了在 MNIST 資料庫中能看出好的效果以外。我們亦有從理論的角度來分析我們的方法能在做距離運算時節省大量的維度運算。

Table 1. 範例資料庫

Point	Coordinates	Class
a	(5,5)	1
b	(7,8)	1
c	(7,5)	1
d	(7,6)	1
e	(6,6)	1
f	(3,7)	2
g	(3,8)	2
h	(11,11)	3

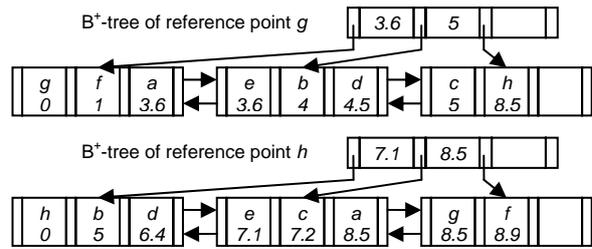


Figure 1. The B^+ -tree set of reference points

多媒體可調整距離函式查詢

[5]是目前在度量空間中，唯一可以在查詢時處理不同距離函式的演算法，其同時能支援範圍查詢及 k 最近鄰點查詢。我們基於[5]做出一些改進，設計一套索引結構，記錄不只一組的索引距離，以加快查詢的速度。另外，我們發展了可以處理不同距離函式的反向 k 最近鄰點查詢，至目前為止這個問題還沒有被其他論文所探討過。

使用者自訂距離函數之 k 最近鄰點查詢

步驟一：建立索引

我們提出的索引結構(MDM-Tree)是基於M-Tree[6]，並且進行一些改進。在建立完M-Tree之後，我們額外使用一些不同的距離函數，計算每個節點的覆蓋範圍，以及每個節點至其父節點的距離，並將這些資訊記錄在M樹之中。換句話說，在MDM-Tree中我們記錄了不止一種的索引距離，查詢時我們可以找出最有用的一組索引距離幫助刪剪。

步驟二： k 最近鄰點查詢

我們的 k 最近鄰點查詢演算法是基於QIC-M-Tree[5]，並且做出一些改善。首先，我們使用了記錄額外資訊的MDM-Tree，而非原本的M-Tree。另外，相較於QIC-M-Tree在走訪樹結構時計算索引距離函式，我們的演算法直接計算查詢距離函式，以達到更好的刪剪效果。

由於度量空間一般是難以想像的，我們將查詢之距離函數投射至二維平面上，並舉例說明此演算法的基本精神。在Figure 2中，令索引距離為 d_i ，查詢距離為 d_q 。假設 p 點為 r 點在樹結構中的父節點之中心，則 $d_i(p,r)$ 已經被預先計算過了，而 $d_q(p,q)$ 應在上層走訪時已被計算過。我們可以從索引距離及查詢距離，計算出兩者

之最大差距，而推導出 r 點在查詢距離下的可能位置(即黃色環的範圍)。由此我們可以得知 $d_q(r,q)$ 的最小可能值，若已找到 k 個點比 r 更接近 q ，則 r 可以被刪剪。

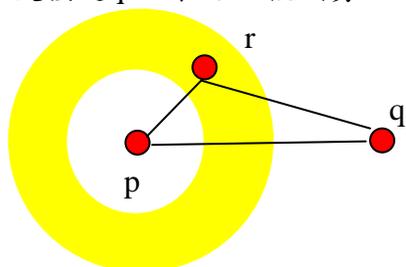


Figure 2. K最近鄰點查詢範例

使用者自訂距離函數之反向 k 最近鄰點查詢

步驟一：建立索引

此部分與 k 最近鄰點查詢的建立索引方式相同，一樣的索引結構可以應用在這兩種不同型態的查詢。

步驟二：反向 k 最近鄰點查詢

目前沒有其他的演算法能夠做到在度量空間下，由使用者自訂距離函數的反向 k 最近鄰點查詢。我們的反向 k 最近鄰點查詢演算法是基於[15]，並且對其六項刪剪規則做出修正，使得我們的演算法可以處理這一種查詢。

我們的方法可以在度量空間中運作，但是，為了舉例的方便，我們以一般的歐吉里德空間為例，如下圖：

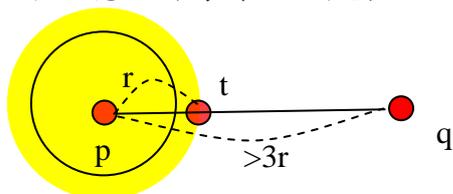


Figure 3. 反向K最近鄰點查詢範例

假設 q 為查詢點， p 為目前試圖刪剪節點之中心。該節點在索引函數下的覆蓋半徑(即實線圓內範圍)已被記錄在索引結構中。根據索引函數及查詢函數間的關係，我們可以算出該節點在查詢函數下最大的覆蓋半徑 r (即黃色圓的範圍)。假設該節點內的資料點數大於 k ，該節點內所有資料點在 $2r$ 的距離都會至少存在 k 個其他資料點。若 $d_q(p, q) > 3r$ ，根據三角不等式，該節點中的任何資料點與 q 的距離皆大於

$2r$ (如距離 q 最近的點 t ，與 q 的距離大於 $2r$)，且已知有其他 k 個資料點比 q 更接近該資料點，故該節點可以成功地被刪剪。

股價關聯分析及未來走勢預測

在文獻[16]中，是以同時間的關聯作研究，而且是以一季作預測時間單位，所以並無法達到預測的效果與買賣的動作。因此，我們考慮分別使用技術指標與股價變化來計算股票之間的關聯度，並且以隔 N 天的股價走勢作為研究目標，利用求出的關聯度來發展出個別股票的買進與賣出推薦值。

我們透過技術指標與股價變化來分別計算股票之間的關聯度，並且考慮了時間延遲問題，分別計算延遲一天、三天和五天的關聯度。例如：A股票顯示為多頭時，B股票在同一天、隔一天、隔三天和隔五天也會顯示為多頭的個別關聯度，希望藉由不同時間的延遲結果與不同類型的關聯度，可以分析出最好的買賣策略。

關聯度的求法是以條件機率的方法來求得。A股票對於B股票的關聯度的計算法是：當A股票在一段時間間隔裡，由技術指標顯示為多頭的總日數中，B股票也顯示為多頭的機率。除此之外，為了判斷某一股票對其他股票的影響力，我們提出了兩種推薦值的計算法，來作為買進與賣出的依據。

方法一:機率模式

以條件機率的方式，考慮所有兩兩股票之間的關聯度分布，訂出高度關聯的threshold值。依歷史資料裡，所有對A股票具有高度多頭關聯性的股票總數作為條件，以每日加總已成為多頭的高關聯度股票總數，因而可以計算出A股票的當日買進推薦值與每日賣出推薦值。例如對於A股票具高度關聯的總股票數有50檔，而當天這50檔股票有30檔股票顯示為多頭時，則A的買進推薦值為 $30/50=0.6$ 。

方法二:PageRank 模式

採取了google[3]在判斷網頁重要性的方法而應用在股票之間推薦值的計算

中。如A股票的買進推薦值，是由其他股票(B)對於A具有高度相關性的買進推薦值，依比率所累加出來的，比率是依照其他股票(B)對於全部股票市場具有高度關聯股票的總數作為依據。

以Figure 4為例，對於A股票具備高度相關的有B、C、D三檔股票，所以A股票的買進推薦值是由B、C、D個別的影響值所累加出來。而B股票對A股票的影響值是由B股票的買進推薦值所分配出來，分配的原則是考慮B股票總共影響了幾檔股票總數。下圖中，B的買進推薦值為80且總共對A、E、F具備高度相關，所以A、E、F個別可以從B中獲取 $80/3$ 的影響值。

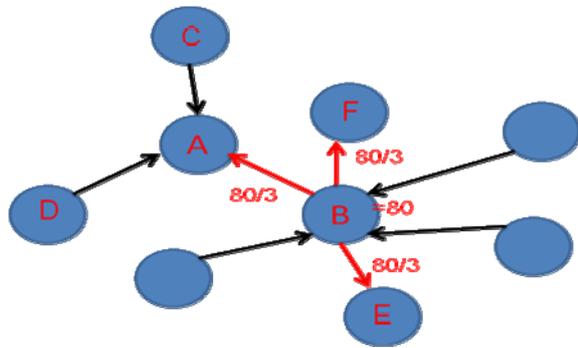


Figure 4. PageRank 示意圖

五、未來工作

在計畫第三年，我們要完成家庭式KTV個人化背景影片播放系統，我們計劃發展跨媒體整合技術來達成這個目的。而在跨媒體整合技術中很重要的一部份就是對多媒體的查詢與處理，因此可以利用今年所發展的兩個與媒體種類無關的多媒體技術來達成。

六、計劃成果自評

在過去的一年裡，我們發展了三個重要的技術。其中，多媒體高維索引和多媒體可調整距離函式查詢這兩個技術都是針對多媒體資料作處理。通常為了能更加精確的描述多媒體資料，我們會使用很多特徵，在這種情況之下所產生的特徵空間其維度通常很高，而多媒體高維索引技術，可以加速在這種環境下對多媒體的處理速度。另外，為了滿足使用者想要表達不

同語意的查詢，因此，我們發展了多媒體可調整距離函式查詢技術，以便讓使用者根據不同的語義來給定不同的距離函式。此外，利用機率模型來分析不同股票之間的關聯，可以幫助我們找出具有共同趨勢的股票，這些資訊將提供給使用者作為買賣股票的參考，進而達到幫助家庭理財的目的。我們完成相關論文三篇，其中兩篇為本年度碩士生之畢業論文，一篇正在準備投稿中。

碩士生之畢業論文

- [1] Y. C. Ling. Nearest and Reverse Nearest Neighbor Search in Metric Spaces with Adjustable Distance Function, National Tsing Hua University, 2007.
- [2] C. W. Chien. Developing a stock recommendation system by stock prices correlation, National Chengchi University, 2007.

準備投稿之論文

- [1] D.Y Chiu and Arbee L.P. Chen, "An Online Classifier Method for Enhancing the Accuracy of Multimedia Data Retrieval," *submitted for publication*.

參考文獻

- [1] E. Aichert, C. Böhm, P. Kröger, P. Kunath, A. Pryakhin, and M. Renz, "Efficient reverse k-nearest neighbor search in arbitrary metric spaces," *ACM SIGMOD Conference*, pp. 10-18, 2006.
- [2] J. Barros, J. French, W. Martin, P. Kelly, and M. Cannon, "Using the Triangle Inequality to Reduce the Number of Comparisons Required for Similarity-based Retrieval," *International Conference on Storage and Retrieval for Image and Video Databases*, pp. 392-403, 1996.
- [3] S. Brin and L. Page, "The Anatomy of a Large-scale Hypertextual Web Search Engine," *Proceedings of the seventh international conference on World Wide Web*, 1998.
- [4] B. Bustos, and T. Skopal, "Dynamic

- similarity search in multi-metric spaces,” *Proceedings of the 8th ACM international Workshop on Multimedia information Retrieval*, 2006.
- [5] P. Ciaccia and M. Patella, “Searching in Metric Spaces with User-Defined and Approximate Distances,” *ACM Transactions on Database Systems*, 27(4):398-437, 2002.
- [6] P. Ciaccia, M. Patella, and P. Zezula, “M-tree: An efficient access method for similarity search in metric spaces,” *Proc. of International Conf. on Very Large Data Bases*, pp. 426-435, 1997.
- [7] G. Gray and O. Larry, “Forecasting S&P and Gold Futures Prices : An Application of Neural Networks,” *The Journal of Futures Markets*, 1993.
- [8] G.R. Hjaltason and H. Samet, “Index-Driven Similarity Search in Metric Spaces,” *ACM Trans. Database Systems*, 28(4):517-580, 2003.
- [9] Y.K. Kwon, S.S. Choi, and B.R. Moon, “Real World Applications: Stock Prediction Based on Financial Correlation,” *Proceedings of the conference on Genetic and evolutionary computation GECCO*, 2005.
- [10] H. Mizuno, M. Kosaka, H. Yajima, and N. Komada, “Application of Neural Networks To Technical Analysis of Stock Market Prediction,” *Studies in Informatic and Control*, 7(3):111-120, 1998.
- [11] N. Panda and E.Y. Chang, “KDX: An Indexer for Support Vector Machines,” *IEEE Transactions on Knowledge and Data Engineering*, 18(6): 748-763, 2006.
- [12] O. Procopiuc, P.K. Agarwal, L. Arge, and J.S. Vitter, “Bkd-tree: A Dynamic Scalable Kd-tree,” *International Symposium on Spatial and Temporal Databases*, pp. 46-65, 2003.
- [13] A. Singh, H. Ferhatosmanoglu, and A.S. Tosun, “High Dimensional Reverse Nearest Neighbor Queries,” *Proc. Conf. Information and Knowledge Management (CIKM)*, pp. 91-98, 2003.
- [14] Y. Tao, D. Papadias, and X. Lian, “Reverse kNN Search in Arbitrary Dimensionality,” *Proc. of International Conf. on Very Large Data Bases*, pp. 744-755, 2004.
- [15] Y. Tao, M. Yiu, N. Mamoulis, “Reverse Nearest Neighbor Search in Metric Spaces,” *IEEE Transactions on Knowledge and Data Engineering*, 18(9):1239-1252, 2006.
- [16] Y.C. Tsai. “Constructing the Index System of Association to Stock Investment Decision,” 2005.
- [17] D.A. White and R. Jain, “Similarity indexing with the SS-tree,” *IEEE Conference on Data Engineering*, pp. 516-523, 1996.
- [18] S.I. Wu and R.P. Lu, “Combining Artificial Neural Networks and Statistics for Stock-market Forecasting,” *Proceedings of the ACM conference on Computer science*, 1993.
- [19] C. Yu, B.C. Ooi, K.L. Tan, and H.V. Jagadish, “Indexing the Distance: An Efficient Method to KNN Processing,” *Proc. of International Conf. on Very Large Data Bases*, pp. 421-430, 2001.