Plan Mining-Based Tracking Information System

Hay Mar Hla Thein
University of Computer Studies (Mandalay)
circle.lay@gmail.com

Abstract

Information is pivotal in today's business environment. Success is dependent on its early and decisive use. A lack of information is a sure for failure. The rapidly changing environment in which business operates demands ever more immediate access to data. Many corporations are actively looking for new technologies that will assist in becoming more profitable and competitive. This system is the prototype of package tracking system. Users can track their packages with an inexpensive way. This system is implemented by using plan mining concepts and shortest path. This prototype system will be based on Air Line Package Delivery Service. The system will provide the user for package detecting services to desired destination. Firstly, the flight plan will be used to draw package travel plan using (shortest path algorithm). Secondly, updating information for package travel will be provided based on plan mining (sequential pattern mining).

1. Introduction

The term "parcel post" refers to the sending of packages through the mail service. In 1878, the Congress of the Universal Postal Union established an international parcel post system. Four years later, the British parliament approved a bill implementing domestic, colonial and foreign parcel post services. Other countries quickly followed service.

In the late 1800's, the National Grange and similar organizations concerned with farmers' welfare lobbied Congress for the free delivery of mail to rural households. Many rural residents had to travel for days to retrieve their mail from distant post offices or pay private express companies for delivery. Finally, in October 1896, Congress approved the establishment of rural free delivery.

This system is implemented by combining shortest path theory, plan mining theory and web protocol. This plan mining-based tracking information system intends to satisfy both the user and the service provider. This system can operate the package tracking service from Myanmar to other limited countries with shortest path. To calculate shortest path, system need airlines to reach destination. There are many different airlines all over the world. The service selected the airline plan from the flight planbase by using plan mining techniques. "Shortest path" is the best path between two countries and any sub path is the shortest path. The mining of gold from rocks or sand is referred to as gold mining. As same as, mining of plan from plan database is termed as "plan mining".

2. Related Works

For shortest path problems in computational geometry, see Euclidean shortest path.

The traveling salesman problem is the problem of finding the shortest path that goes through every vertex exactly once, and returns to the start. Unlike the shortest path problem, which can be solved in polynomial time in graphs without negative cycles, this problem is NP-complete and , as such, is believed not to be efficiently solvable.

The problem of finding the longest path in a graph is also NP-complete [9].

The Canadian traveler problem and the stochastic shortest path problem are generalizations where either the graph isn't completely known to the mover, changes over time, or where actions (traversals) are probabilistic.

The problems of recalculation of shortest paths arises if some graph transformations (e.g., shrinkage of nodes) are made with a graph [7].

Label setting techniques are all based on Dijstra's condition of always scanning the node with the minimum label, which guarantees that each node will be scanned exactly once; while this condition is sufficient it is not necessary. In this paper, we discuss less restrictive conditions that allow the scanning of a node that does not have the minimum label, yet still maintaining sufficiency in scanning each node exactly once; various potential shortest path schemes are discussed, based on these conditions. Two approaches, a label setting and a flexible hybrid one are designed and implemented. The performance of the algorithms is assessed both theoretically and computationally. For comparative analysis purpose, three additional shortest path algorithms – the commonly cited in the literature – are coded and tested. The results indicate that the approaches that rely on the less restrictive optimality conditions perform substantially better for a wide range of network topologies [9].

While heuristics based on geometric constructs of the networks would appear to improve performance of Dijkstra's algorithm, the fallacy of depreciated accuracy has been an obstacle to the wider application of heuristics in the search for shortest paths. The authors presented a shortest path algorithm that employs limited area heuristics guided by spatial arrangement of networks. The algorithm was shown to outperform other theoretically optimal solutions to the shortest path problem and with only little accuracy lost. More importantly, the confidence and accuracy levels were both controllable and predictable [6].

3. Mining Complex Types of Data

Previous studies on data mining techniques have focused on mining relational database, transactional database and data warehouses formed by the transformation and integration of structured data. An increasingly important task in data mining is to examine complex types of data, including complex objects, spatial data, multimedia data, time-series data, text data and the World-Wide-Web.

The storage and access of complex structured data have been studied in object-relational and object-oriented database systems. These systems organize a large set of complex data objects into classes, which are in turn organized into a class/subclass hierarchies. Each object in a class is associated with (1) an object- identifier, (2) a set of attributes that may contain sophisticated data structures, set- or list-valued data, class composition hierarchies, multimedia data, and so on, and (3) a set of methods that specify the computational routines or rules associated with the object class [3].

3.1 Plan Mining

A plan consists of variable sequences of actions. Simply a planbase, is a large collection of plans. Plans Mining is the task of mining significant patterns or knowledge from a planbase. It is different from sequential pattern mining , where a large number of frequently occurring sequences are mined at a very detailed level. It is the extraction of important or significant generalized (sequential) patterns from a planbase.

Plan mining can be used to discover travel patterns of business passengers in an air flight database or to find significant patterns from the sequences of actions in the repair of automobiles. Plan mining is different from sequential pattern mining, where a large number of frequently occurring sequences are mined at a very detailed level. Instead, plan mining is the extraction of important or significant generalized (sequential) patterns from a planbase.

"So, how should we go about mining a planbase?" We would like to find a small number of general (sequential) patterns that cover a substantial portion of the plans, and then we can divided our search efforts based on such mined sequences. The key to mining such patterns is to generalize the plans in the planbase to a sufficiently high level. A multidimensional database model, for the air flight planbase, can be used to facilitate such plan generalization. Since low-level information may never share enough commonality to form succinct plans, we should do the following: (1)) generalize the planbase in different directions using the multidimensional model; (2) observe when the generalized plans share common, interesting, sequential patterns with substantial support; and (3) drive high-level, concise plans.

The plan mining technique can be further developed in several aspects. For instance, a minimum support threshold similar to that in association rule mining can be used to determine the level of generalization and ensure that a pattern covers a sufficient number of cases. Additional operators in plan mining can be explored, such as less-than. Other variations include extracting associations from subsequences, or mining sequence patterns involving multidimensional attributes for example, the patterns involving both airport size and location. Such dimension-combined mining also requires the generalization of each dimension to a high level before examination of the combined sequence patterns.

4. Proposed System

The basic problem: Find the "best" way of getting from s(source) to where s and t (target) are vertices in a graph. It can be measured "best" simply as the sum of edge lengths of a paths. For instance the graph could be a map representing intersections as vertices, road segments as edges: The user want to find either the shortest or fastest route from user's house to other area. Although both of these problems have different solutions, they are both shortest path problems; in one the length of an edge represents the actual mileage of a segment of road, while in the other it represents the time it would take to drive it, but in both cases the important fact is that the total length of a path is measured by adding the lengths of individual edges.

It is going to make a big assumption: that all the edges have lengths that are positive numbers. This is often but not always the case; it makes sense in the examples above, but it is conceivable that an airline could pay people to take certain routes, so that the lengths of those edges in the airport graph might be negative.

Rather than computing one distance d(s,t), we'll compute d(s,x) for all vertices x. This is known as the single source shortest path problem (s is the source). It turns out that computing this extra information makes things easier, because then the user can put together information about paths with fewer edges to get paths with more edges.

4.1 Drawing Graph

Service draw the graph presentation to reach from A to F. A (city**)** can also be defined as Myanmar and F can be termed as Moscow. This digraph shows possible ways to arrive at Moscow from Myanmar. B,C, D, E are another countries along the way and transit such as Bangkok, Brussels, Hongkong and Italy. These data are got from flight planbase [9].

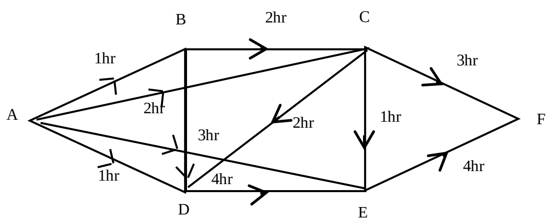

Figure 1. Directed graph from A to F

4.2 Creating Weighted Matrix

The digraph in Figure 1 could be represented by a 6 x 6 matrix to whose entries have one of three values:

$$w(x,y) = \begin{cases} 0 & \text{if } x = y \\ \infty & \text{if } (x,y) \text{ is not an edge} \\ \text{the weight of the edge } (x,y), & \text{otherwise} \end{cases}$$

the symbol ∞ is read "infinity" and represents a value greater than any finite real value. This matrix w is called the weighted digraph. The weight matrix for the graph in Figure 1 is shown below.

The value of elements in this matrix represents time (hour) [9].

	A	B	C	D	E	F
A	0	1	2	1	3	∞
B	∞	0	2	4	∞	∞
C	∞	∞	0	2	1	3
D	∞	∞	∞	0	3	∞
E	∞	∞	∞	∞	1	4
F	∞	∞	∞	∞	∞	0

Figure 2. Create matrix

4.3 Finding Distance from A

Step 0: (Initialization). We set the value of d[x] to 0 when x = a, to ∞ when(A,x) is an edge and to the weight of the edge to (A,x), when (A,x) is an edge. These values are shown in the first row of Figure 2. Since A is closest to itself, we mark it. All other vertices are unmarked at this point.

Step 1: The next to be marked is B because it is the unmarked vertex closest to A. Then we remove B from the list of unmarked vertices. For each of the other unmarked vertices y such that (B,y) is an edge, we check to see if the new path to v through B is shorter than the previously computed path. In this instance, we are able to reach C and E for the first time, so we reduce their distances from ∞ to 3 and from ∞ to 6 respectively.

Step 2: Of the remaining unmarked vertices, D and C are tied for closest. We arbitrarily pick D to mark first. Because E can be reached through D with a path of length 5, we reduce the entry under E from 6 to 5.

Step 3: C is still the closest vertex to A, so we will mark it next. Since F is reachable through C for the first time, we reduce its distance to 8.

Step 4: We mark E next, which causes us to reduce the distance to F from 8 to 6.

Step 5: Then, we mark F, which leaves no more vertices that can be reached. The process must end with F left unmarked [9].

4.4 Resulting Shortest Path

There are many different paths to reach F from A. Figure.1 is shown possible ways from A to F, such as

Path (1) A-B-C-F
Path (2) A-B-D-E-F
Path (3) A-B-C-E-F
Path (4) A-D-E-F
Path (5) A-E-F
Path (6) A-C-F
Path (7) A-C-D-E-F
Path (8) A-C-E-F

After solving section 4.2 and 4.3, the least time to reach F from A is A-C-F (shortest path).

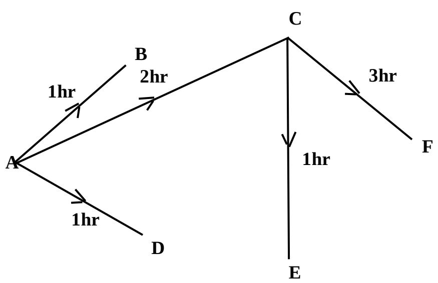

Figure 3. Shortest from A to F.

5. Software Components of System

There are five software components in system implementation. These related functions are represented by the following figure.

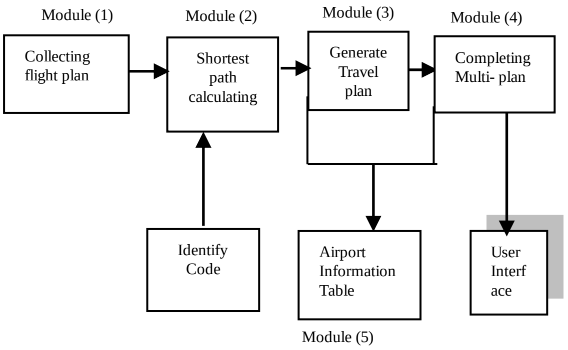

Figure 4. Software components of system

5.1. Flight Plan Collection Module

This module is responsible to collect the plan information of each airlines. The member data containing in this module are departure, departure-time, arrival , arrival time, airline and date. The main function of this module is to collect plan information . The class diagram for this module is described in Figure 5.

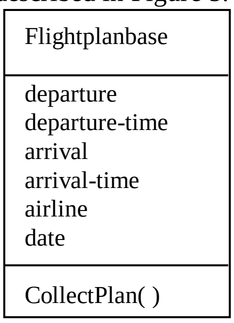

Figure 5. Class diagram of flight planbase

5.2. Shortest Path Calculation Module

According to the user input data(i.e. source and destination) , this module calculates and finds the shortest path among alternative paths. There are 3 main attributes in this module: node, edge, weight. This module firstly draws the graph based on the plan database. Then, distance matrix is calculated by this module as served functions. Finally, this module calculates the shortest path by using Dijkstra's Algorithm. The class diagram for this module is described in Figure 6.

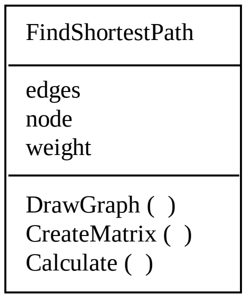

Figure 6 . Class diagram of findshortestpath

5.3. Travel Planbase Generating Module

After calculating the shortest path from Myanmar to Moscow, the package is sent to destination along the transit. Every package's track are shown in this planbase. The function of travel planbase is to track package. The class diagram of this module is shown in Figure 7.

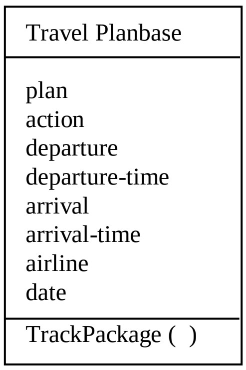

Figure 7. Class diagram of travel planbase

5.4. Multidimensional Planbase Building Module

Package's track is stored in travel planbase. According to the up-to-date information of transit, the data are added in this planbase.The member data are plan, location-sequence, state-sequence and region sequence. The responsibility of this module is to show the user the track of package. All of the packages' track are filling in this planbase by using iteration. The class diagram of multidimensional planbase is shown in Figure 8.

.

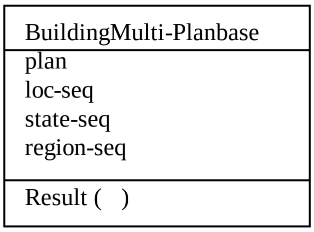

Figure 8. Class diagram of multidimensional planbase

6. Sequence Diagram of Tracking System

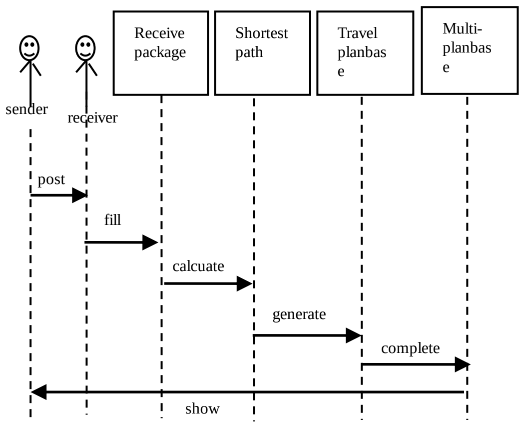

Figure 9. Sequence diagram of tracking system

7. Output Result

345 5569 Mrs. Hay Mar Hla Thein

Oct 30 (10:00 AM)	Bangkok
Oct 31 (02:06 PM)	Brussels
Nov 1 (04:10 PM)	Italy
Nov 1 (10:43 PM)	Moscow
Nov 1 (12:48 PM)	Services in Moscow
Nov 2 (4:30 PM)	Clearence processing complete at Moscow, Russian Fedaration

8. Conclusion

The goal of this system is to track the parcel in every where and every time in the internet. Users can easily get about their package's up-to-date information. From the service provider point of view, it is very benefit and time- consuming. When user post the package, the service stored the detailed information in the user database and gave the identity number to user. User can track their parcel's status online by entering their parcel's identity number. A click of the mouse will bring all information about the parcel's delivery status.

9. Limitation of the System

There are some limitations in this system because it can allow to track the limited countries and limited airline. This system, specially serviced to Russia. So, the countries that consist of along the way of Russia can be tracked. This system uses simple Dijkstra's Shortest Path Algorithm.

REFERENCES

[1] Bontempo,J. Charles
"Copying with complex queries in a data warehouse"

[2] Carickhoff, Rich
"A New face for OLAP"

[3] Chaudhuri, Surajit
"An Overview of Data warehousing and OLAP technology"
Microsoft Research

[4] Debevoise, Nelison Thomas
"The Data ware house Method"
ISBN 0-13-0812306-0
1999 by Prentice Hall PTR

[5] Elmasri, Ramez
"Fundamentals of database systems
ISBN 981-405330-9
Third Edition

[6] Farguson, Mike
"Evaluating and Selecting Data mining tools"

[7] Han, Jiawei
"Data Mining Concepts and Technique

[8] Teorey, J. Toby
"Database Modeling & Design"
ISBN 1-55860-500-2
1999, Third Edition
Morgan Kaufmann Publishers, Inc.

[9] www.cs.usask/content/resources/.../dijkstra/
Dijkstra.html