Data Mining of Tourists' Spatio-temporal Movement Patterns ---A Case Study on Phillip Island

Jianhong (Cecilia) Xia[1], Vic Ciesielski[2] and Colin Arrowsmith[3]

[1]Department of Geospatial Science
RMIT University
Tel: +61 3 99253277
X02250@ems.rmit.edu.au

[2]Department of Computer Science and Information Technology
RMIT University
Tel: +61 3 9925 2926
vc@cs.rmit.edu.au

[3]Department of Geospatial Science
RMIT University
+ 61 3 9925 2042
colin.arrowsmith@rmit.edu.au

Abstract

In these days, understanding tourists' spatio-temporal movement behaviour is becoming more and more important factor for success of tourist marketing. This paper uses a general purpose data mining tool to identify tourist spatio-temporal movement patterns and across patterns between tourist profiles and their spatio-temporal movement patterns. Major temporal movement sequences and spatial movement sequences are discovered and compared. For example, if tourists only visited one attraction in the evening for their day-trip it is Penguin parade. Clustering method is used to identify tourists' market segment for each attraction. There are various key attributes to group tourists for different attractions. Type of visitors (international or domestic tourists) is the most important attribute to cluster the tourists on the Phillip Island. Differences of tourists for each attraction and movement pattern are also compared using classification algorithms. However these differences are not significant.

1 Introduction

Data mining can be viewed as the process of discovering "golden nuggets" of information in large set of data (Ciesielski & Lalani 2003; Witten & Frank 2000). The "golden nuggets" or patterns are extracted, analysed, interpreted and displayed in a meaningful way that could lead to some economic advantages. Today market competition among company is very fierce. Fast and accurate market positioning and personalised service are the key factors for company success. There are a large number of research have been done to identify the customer's purchase behaviours. Some of research focus on developing new data mining algorithms to extract customer's purchase patterns from data (Min & Han 2005; Yada et al. 2004; Yan & Han 2002). Existing data mining algorithms also were experimented and tailored to a specific application (Ciesielski & Lalani 2003; Hsieh 2004; Witten & Frank 2000; Yada et al. 2004).

This paper belongs to the later category that is to identify the patterns between tourist profiles and their spatio-temporal movement. However, so far, there are only limited research on identification of tourists spatio-temporal movement patterns in tourist destinations at the Microscopic level (Arrowsmith & Chhetri 2003; Bardram 2005). Furthermore, there has not been any research on identify the patterns between tourist profile and their spatio-temporal movement sequential patterns, especially temporal movement sequence. For example, How did tourists organise their time to move around Phillip Island? Did different types of tourists have various temporal movement behaviours? Understanding tourist movement behaviour can help park manager make fast and accurate market positioning strategies and offer personalised service based on the data mining results.

The goals of this paper are: (1) to identify the significant patterns of tourists' spatio-temporal movement; (2) to discover patterns between tourist profiles and their spatio-temporal movement patterns using existing data mining techniques

Tourist profiles are age, education, country of residence (international or domestic), gender, lifecycle, and travel model that refers to transport, how long to stay on the Phillip Island (duration), and travel group (who they travel with).

The proposed methodology consists of three parts: (1) To obtain the tourist spatial-temporal movement patterns by surveys (2) To analyse the tourist spatial-temporal movement patterns using statistics package STATISTICA, EXCEL (3) To identify patterns between the tourist profile and their spatio-temporal movement using data minding algorithms such as clustering (EM), classification (OneR, J48/C4.5 decision tree).

The rest of this paper is organized as follows. Section 2 describes the research background. Section 3 presents the proposed methodology. Section 4 explains the results of the evaluation experiment. Section 5 presents the summary and future research issue.

2 Research Background

2.1 Tourist spatio-temporal movement

Movement is the act or process of moving; *especially* **:** change of place or position or posture (Merriam-Webster 2004). It is a dynamic process that is characterised not only by spatial and attribute components, but also by temporal references (Worboys & Duckham

2004). Time is considered as a linear dimension. Movement is a continuous phenomenon on a timeline. In order to represent this continuum, movement can be discretised on a grid (Raper 2001.).

Sequence of movement can be represented as a collection of timestamped states. At a time point or instant, movement is spatially points represented as (x, y) in a coordinate system. But during a time interval, spatial movement can be viewed as a line or network. Nodes of network are the spatial point that movement stop. The edge of network is the trajectory of a continue moving that connect two stop spatial points. Representation of this network can be a collection of time slices that are integrated together to show the spatial distribution from past to present (Ott & Swiaczny 2001).

In this paper, Tourist spatio-temporal movement is the act or process of moving from one destination to another one during a certain time interval such as a one-day trip. It is represented by a movement network. This network is constrained by the road network in a physical environment. The nodes of movement network are attractions. The edges are the roads that connect attractions. The assumption to define this movement network is that tourists stop at attractions or nodes and move among the attractions along the roads. Actually the stop that is mentioned above doesn't mean that tourists physically stand on one point without change of position and velocity. It means that tourists move round a small area that we call it attraction without change position in a large distance. Each note or attraction has two temporal attributes: arrival time (time point) and duration (time interval). The sequence of movement, therefore, can be simplified as a sequence of attractions or a sequence of combination of attractions (position) and their arrival time and duration. The edge of network is the route that tourists choose to travel between two attractions. An edge can compose of more than one road. And the edge is assigned a direction and label such as road name or speed limit for traffic. The important temporal attributes for edges is time span. The issue related is, for example, how long to travel form one attraction to another.

2.2 Research area

Phillip Island, located at the mouth of Westernport baylies, is 140 kilometres south-east of Melbourne (see Figure 1). It covers about 10 000 hectares and is 26 km long and 9 km wide. The permanent population on the island is around 7000, but almost 1.5 million visitors travel around the island each year currently (Phillip Island Internet Services 2005).

Figure 1 Location of Phillip Island (Phillip Island Nature Park 2005)

There are a large number of natural resources on the island, such as penguins, koalas, seals, shearwaters, mangroves, wetlands, sandy beaches and rugged rocky cliff faces. The major attractions are Penguin Parade that is the Australia's second most visited natural attraction and designed to protect little penguins, the Koala Conservation Centre is a natural habitat for protecting Koala from cars and dogs and offering better opportunities for tourists viewing Koala, Cowes- the principal town of Phillip Island, Churchill Island-historic farm, other attractions such as Rhyll Inlet, Woolamai and the Nobbies, where visitors can experience wildlife in its natural environment (Phillip Island Nature Park 2005) (see Figure 2).

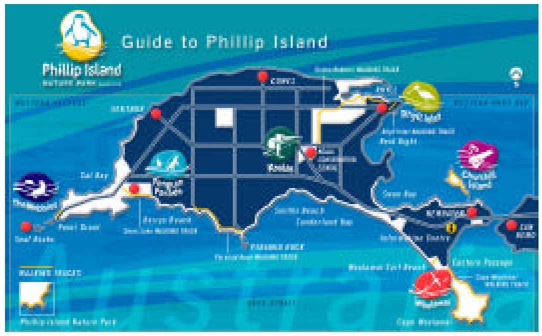

Figure 2 Attractions of Phillip Island (Phillip Island Nature Park 2005)

3 Methodology

3.1 Spatio-temporal Movement database design and data collection

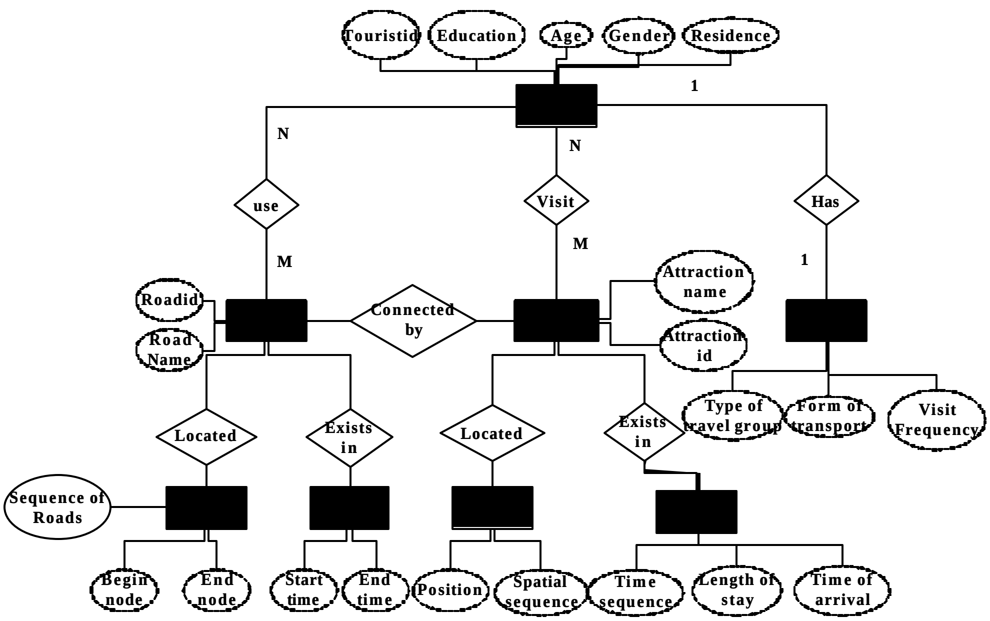

Figure 3 the ER Model for Tourist spatio-temporal movement database

Database is a collection of information that is organised in an efficient way for computer process (Elmasri 2000). Tourist spatio-temporal database deal with information related to tourist and their spatial movement and temporal reference. This database are organised as a relational database. All the information is presented as a collection of relations. Each relation is shown as a table. Columns are attributes and rows ("tuples") are entities (Dilip 2005).

The Entity-Relationship (ER) model above shows that the basic entities in the database are tourists, roads, attractions, travel mode, space, and time. The attributes of tourists include ID, age, gender, residency, and education. Different tourists might have various trave modes, visit different attractions and use different roads. Travel mode refers to form of transport, visit frequency, and type of travel group. Attractions visited by tourists are located in space with attributes as position. A movement is represented spatially a sequence of attractions. Movement in time reference, represented as a sequence of time intervals (or time categories), has attributes of length of time to stay, arrival time at an attraction, start time for tourists to enter road and end time to leave the road. Attractions are connected by roads that are also located in space and stored in the computer as an object with attributes as ID, start nodes and end node.

Data for this database were collected by survey from 6 to 8 March 2004 and 17 to 20 Jan. 2005. Currently, 500 questionnaires that were filled in by visitors on the island have been returned and 466 questionnaires were input into database. The other 34 questionnaires are discarded because of two many data missing.

3.2 Movement pattern extraction

3.2.1 Identification of spatio-temporal movement patterns

Temporal movement sequence means an ordered set of arrival time for a series of spatial positions. For example, a set of arrival time for each attraction during a tourist one-day trip. However, arrival time is time point that scattered along a time line. It is very hard to identify the major time sequence patterns. The solution is to simplify the time points, in other words, to discritise or category the time points into time interval.

The biggest issue for grouping the time points is to decide the level of granularity. Time points could be categorised into one-hour time interval, two-hour time interval, even to six-hour time interval. From domain knowledge, time sequence based on six-hour time interval category can represent general temporal movement behaviours in the case study clearly, because firstly the information that we want to extract from the data is at macro-level. That means the movement is simplified as a sequence of attractions spatially. The important attributes are arrival time and duration for each attraction. The specific time points or position for tourists moving between attractions and staying at an attraction is not significant to affect their general movement behaviour. In addition, the number of attractions that the tourists visited on the Phillip Island is less than 7, mostly less than four attractions according to the survey statistics. And duration for staying at each attraction arrange from half an hour to six hours. So a high level of granularity for the spatio-temporal movement is not necessary.

In the case study, time points were grouped into four categories: morning [6:00, 12:00), afternoon [12:00 18:00), evening [18:00 24:00), and night [24:00 6:00) based on a six-hour time interval. Therefore, morning is time group 1; afternoon is time group 2; evening is time group 3; night is time group 4. If a tourist one-day trip likes (Cowes 16:00), (the Nobbies 18:00) (Penguin Parade 19:00), then time point 16:00 was categorized into time group 2, time point 18:00 into 2, time point 19:00 into 3. After this data discretisation, the arrival time points in one movement itinerary was put into a set of time sequence such as “223”, which means a tourist arrived at the Phillip Island in the afternoon. He or she visited two attractions in the afternoon and one in the evening.

Spatial movement sequence means an ordered set of spatial positions. In other words, during one itinerary of movement such as one-day trip, what is the combination of attractions that tourist visited. So sequence of movement Patterns are simply represented as a series of attractions such as “BADCD” or “DCAB”. Here an alphabet is a symbol of an attraction.

3.2.2 Identification of across patterns between tourist segment and their spatio-temporal movement patterns

The across patterns are discovered by Weka (Waikato Environment for Knowledge Analysis) software environment for data mining. Weka, developed at the University of Waikato in New Zealand, is a implementation tool of machine learning algorithms such as data pre-processing, classification, regression, clustering, association rules, and visualization (Witten & Frank 2000).

Data mining can be considered as the task of finding useful patterns in data. In this context the term "pattern" is somewhat vague and depends on the data and the technique used. In some cases some patterns are so unexpected and valuable that they can be called "golden nuggets" in an analogy with mining for gold. There are a number of different approaches that give different kinds of patterns. In this work we use classification and clustering.

In classification we require some kind of rule, or classifier, for distinguishing instances of different classes. For example, if we treat international and domestic visitors to Philip Island as different classes we would like to know what the differences between these two classes are. Ideally, given a set of facts about an unknown visitor we would like to accurately predict whether they are international or domestic. In our situation we have a data file of visitors containing various facts about them and about their visit to Philip Island. It is known whether a visitor is international or domestic. We would use a data mining algorithm, such as a decision tree algorithm or a rule finding algorithm to induce a classifier from the data. The methodology requires us to split the data file into 2 parts, training data and test data. The training data is used in the construction of the classifier. Once the classifier has been constructed it is applied to the test data. A prediction is made for each test case and compared with the known class label. The more accurately the test data can be classified the better the classifier. As a general guideline if the test accuracy is greater than 75% then we assume that a significant difference has been found.

There are a very large number of classification algorithms. We have used the ONER (One Rule) and decision tree algorithms (J48/C4.5) because, unlike many other algorithms, the classification rule can be readily understood and provides insight into the data. The OneR induces a one-rule classifier based on a single attribute. All attributes are tried and one attribute is chosen as the classifier with minimum error in the training dataset. J48 algorithm is an implementation of the C4.5 decision tree learner (Quinlan 1986). The output of algorithm is a decision tree, the non-terminal nodes represent tests on attributes and terminal nodes show decision outcomes (Haglin 2004).

Classification requires each data instance to be labelled with the classes of interest. In contrast, clustering assumes that the instances are unlabelled and that the task is to group together instances that are similar to each other into clusters. The members of each cluster should be very similar to each other and be very different from members of other clusters. Algorithms for clustering require a measure of distance between instances. If the attribute values are numeric Euclidean distance is usually used. There is a considerable number of algorithms for numeric data. If the attribute values are symbolic (for example, "international", and “domestic") other distance measures need to be found. There is only as small number of algorithms that can be used with symbolic attributes. We have used the EM algorithm since it can accept both numeric and symbolic attributes. Given an input file of data the EM algorithm finds the number of clusters in the data, gives a description of each class and assigns each instance to one of the clusters.

In application, data mining can be hypothesis driven or data driven. In hypothesis driven data mining some expectations about the data are generated and a data mining technique used to confirm or reject the hypothesis. For example, if we hypothesise that there is a difference between international and domestic visitors we can build and test a classifier. If its test accuracy is 50% then we conclude that there is no difference. If its test accuracy high, say more than 75% we conclude that there are differences between the two types of visitors.

In data driven mining there are no prior expectations. We apply an algorithm to the data and examine the outputs looking for something unexpected or significant. For example, we could look at the outputs of a clustering algorithm and examine the cluster descriptions to determine whether there are any significant groupings of visitors.

4 Experiment

4.1 Temporal movement sequence

There are 43 temporal movement sequence patterns identified from database. Generally tourists travelled around the Phillip Island in the afternoon and evening. The most popular pattern is "23". Tourists visited one attraction in the afternoon and another one in the evening. One interesting result is that 15% tourists only visited one attraction in the evening during their one-day trip. Seven percent of tourists visited two attractions in the evening. In addition, 105 out of 464 tourists visited three attractions per day, two in the morning one in the evening or two in the evening or one in the afternoon.

Table 1 the number of tourists for each temporal movement sequence

Patterns	23	3	223	233	33	2233	2223	2	22223	22	123	1223	12223	22233
Count	73	71	57	48	34	25	17	10	8	7	7	6	6	6
Percent	16	15	12	10	7	5	4	2	2	2	2	1	1	1

4.2 Spatial movement sequence

There are nine attractions were visited by tourists on the Phillip Island according to the survey. They are symbolised by alphabets from A to J (see Table 2). Penguin parade (G) is the most popular attraction that tourists visited (442 out of 464). One hundred and nine spatial movement sequence patterns extracted from movement database. The most frequent spatial movement sequence is "FG", which means tourists visited Cowes (F) for shopping or dinner first then moved to Penguin Parade (G) (see Table 3). Thirty six percent of tourists travel in this pattern according to the survey. In addition, one third of tourists visited the Nobbies before Penguin Parade (HG). Nearly 20 percent of tourists visited Cowes first then move to Nobbies and Penguin Parade (FHG). There are 17% of tourists who just visited Penguin parade (G) for their trip.

Table 2 Numbers of tourists in each attraction

Attractions	Information Centre (A)	Cape Woolamai (B)	Churchill Island (C)	Koala Conservation Centre (D)	Rhyll Inlet (E)	Cowes (F)	Penguin Parade (G)	The Nobbies /Seal Rock (H)	Ventnor (J)	Total of Sample
Frequency of visit	8	52	45	140	31	286	442	198	14	466

Table 3 Numbers of tourists for major spatial movement sequence patterns

Pattern	G	FG	HG	DG	FHG	DHG	BHG	BFG	DFG	HFG	EFG
Frequency	80	168	154	26	92	15	8	11	34	36	6
Total	466	466	466	466	466	466	466	466	466	466	466
%	17	36	33	6	20	3	2	2	7	8	1

4.3 Comparison between spatial movement sequence and temporal movement sequence

Compared tourist spatial movement patterns with temporal movement patterns using cross tabulation function from STATISTICA (see
Table 4), it is interesting that temporal movement pattern "3" closely associates with spatial movement pattern "G", which means if tourists only visited one attraction in the evening for their day trip it is Penguin parade. It is because little penguins only show up on the beach in the evening around 7 p.m. to 10 p.m.. In addition, 71 out of 464 tourists visit Cowes or the Nobbies in the afternoon then head for Penguin Parade ("23" → "FG", "HG"). If three attractions are visited by tourists such as temporal movement sequence patterns are "223" or "233", related spatial movement patterns mostly are "DFG", "FHG" and "HFG".

Table 4 Comparison between spatial movement sequence and temporal movement sequence

Movement sequence	All Groups	23	3	223	233	33	2233	2223
Totals	443	73	71	57	48	34	25	17
G	78	0	71	0	0	0	0	0
FG	67	46	0	1	0	17	0	0
FHG	45	1	0	13	20	1	0	0
HG	29	15	0	0	0	13	0	0
DFG	28	0	0	12	10	0	0	0
DFHG	19	0	0	0	0	0	10	5
HFG	18	0	0	10	6	0	0	0
DG	13	7	0	0	0	2	0	0
DHG	11	0	0	5	6	0	0	0
BFG	9	0	0	6	1	0	0	0
CDFHG	6	0	0	0	0	0	0	0
DHFG	5	0	0	0	0	0	3	0

4.4 Across Patterns between categories of tourist and spatio -temporal movement patterns

4.4.1 What kind of tourists who visit an attraction (tourist structure)?

Generally, there are two types of tourists visiting the Phillip Island. They are international visitors travelling in an organised group by bus and domestic tourists travelling with friends or relatives or family by car. But for different attractions there are various key attributes to group tourists. For example, there are three types of tourists who visited Ventnor: (1) domestic young couples have visited Ventnor more than three times (2) domestic young singles visit Ventnor at their first time (3) domestic families have visited Ventnor more than three times. So type of visitors (international or domestic tourists) is not an important attribute for Ventnor, because most visitors are domestic. However visit frequency and lifecycle are the key attributes to cluster the tourists in the Ventnor. It is interesting that mainly domestic tourists visited the Cape Woolamia and Rhyll Inlet. But the significant attributes for dividing tourists who visited these two attractions are gender, visit frequency and lifecycle. Possibly male and female tourists behave differently in these two attractions.

Both international and domestic tourists visited Cowes and the Nobbies, but distributions of these two types of tourists are different in the two attractions. More proportion of domestic tourists and less percentage of international tourists visited Cowes than the Nobbies. Therefore subgroups of tourists for both attractions are different. For Cowes the key attributes to divide the subgroups of tourists are lifecycle (young single, young couple without children, family with children) and visit frequency (once or more than three times), but for the Nobbies they are form of transport (car or bus) and travel group (travelling with friends or relatives or travelling in organized group or club).

Here is an example to identify tourist cluster in the Koala Conservation Centre using EM algorithms. Figure 4 shows output of tourists' categories from EM algorithms in Koala Conservation Centre, there are two clusters are selected. The probability for an instance to be grouped into cluster 0 is 0.6478. The other one is 0.3522. That means that probability for an instance to belong to cluster 0 is 64.78%. Or tourists who visited Koala Conservation Centre belong to cluster 0 is 64.78%. The probability to group an instance in cluster 1 is 35.22%. Or the possibility for tourists belong to cluster 1 is 35.22%. 18 attributes are calculated; four significant attributes are selected (see Figure 4). As we can see from attribute: type of group table, major parameters of type of group are travelling with friends or relatives (44.84/94.01) and travelling with spouse or partner only (27.78/94.01) in cluster 0, and travelling with friends or relatives (24.16/54) and travelling with spouse or partner and children (17.85/54) in cluster 1. It could be explained that main difference between cluster 0 and cluster 1 is that 28 out of 94 tourists travel with spouse or partner only in cluster 0, 18 out of 54 tourists travel with spouse or partner and children in cluster 1. Domestic tourists dominated cluster 1, while international tourists leaded cluster 0. Main lifecycle of tourists for cluster 0 is "Young couple/no children"(27.74/96.01) and "Young singer" (28.97/96.01), while for cluster 1 it is "Middle family (Children 6-15 years)" (22.52/55.99). Major age group for cluster 0 is 18-29 years; Tourists are mainly 40 – 49 years old in cluster 1.

Therefore, it can be concluded that there are two types of tourist visiting the Koala Conservation Centre. One type is the young international tourists travelling with friends or relatives or spouse or partner only. The other one is the middle age domestic tourists travelling with friends or relatives or spouse or partner and children.

Cluster: 0 Prior probability: 0.6478 Cluster: 1 Prior probability: 0.3522

Attribute: Type of group		
	Cluster 0	Cluster 1
Travelling with friends/relatives	44.84	24.16
Travelling with spouse/partner and children	3.15	17.85
Travelling with spouse/partner only	27.78	7.22
Travelling in organised group/club	8.98	1.02
Travelling alone	9.25	3.75
Total	94.01	53.99

Attribute typeofvisitor		
	Cluster 0	Cluster 1
Domestic visitors	20.17	46.83
International visitors	70.84	4.16
Total	91.01	50.99

Attribute Lifecycle		
	Cluster 0	Cluster 1
Young single	28.97	6.03
Mature single	4.38	3.62
Young couple/no children	27.74	3.26
Young family (Youngest child younger than 6 years)	4.15	7.85
Middle family (Children 6-15 years)	3.48	22.52
Mature family (Children older than 15 years)	11.94	8.06
Older couple/no children at home	15.35	4.65
Total	96.01	55.99

Attribute Age		
	Cluster 0	Cluster 1
18-29 YEARS	48.73	8.27
30-39 YEARS	13.89	16.11
40-49 YEARS	19.61	22.39
60 + YEARS	10.78	6.22
Total	93.01	52.99

Figure 4 Tourist market segment for Koala Conservation Centre

4.4.2 Are there differences of tourist distribution among each attraction or each movement pattern? ---- Classification (J48)

- **Across patterns among attraction**

The differences of tourist distribution among each attraction can be identified by classification method based on tourists' profiles such as age, type of visitors, type of group, education, and transport. According to the result of clustering, the most significant attribute is type of visitors (international or domestic). Therefore tourist data are classified based on it. Significant differences between international tourists or domestic tourists are identified from output of J48 decision tree algorithm. The correct classification rate is 76.6234%. The same result is also obtained that international tourists could have more possibilities to visit Koala Conservation Centre than domestic tourists. These international tourists who are young single or old couple without children visit Koala Conservation Centre (16.79/2.0). Domestic tourists who visit Koala Conservation Centre are often with their partner and children (16.79/2.0). It is also interesting to find that both international tourists and domestic tourists use car, but many international tourists also travel by tour bus or coach and campervan (56.37/5.12), (15.1/4.03). Some domestic tourists rode bike when they are travel around Phillip Island (23.15/2.1).

- **Across patterns among major movement patterns**

Characteristics or patterns of tourists who used the major movement patterns can also be discovered from the decision tree outputted from J48. Figure 5 displays a partial tree that shows what kinds of tourists use a certain movement pattern. The rules from this decision tree are:

If howlongPI = 1 week AND Type of group = Travelling with friends/relatives then Movement pattern = G (9.71/1.71). That means 10 tourists travelling with friends and relatives in the training set spent 1 week on the Phillip Island had been Penguin Parade. Here 9.71 indicate the numbers of tourists in the training set are correctly classified into this movement pattern (G), while 1.71 shows how many tourists incorrectly classified by the node.

If howlongPI = 1-3 days AND Lifecycle = Mature family (Children older than 15 years) AND Gender = Female: Movement = FG (6.21/1.0), which represents 6 female tourists who are from a mature family and spent 1-3 days on the Phillip Island went to Cowes first then head for Penguin Parade.

If howlongPI = half a day AND Education = Tertiary AND Transport = Tour bus/coach AND type of visitor = International tourists: HG (9.02/3.0), which means 9 international tourists with tertiary degree who spent half a day on the Phillip island and travelled by tour bus or coach went to the Nobbies then Penguin Parade

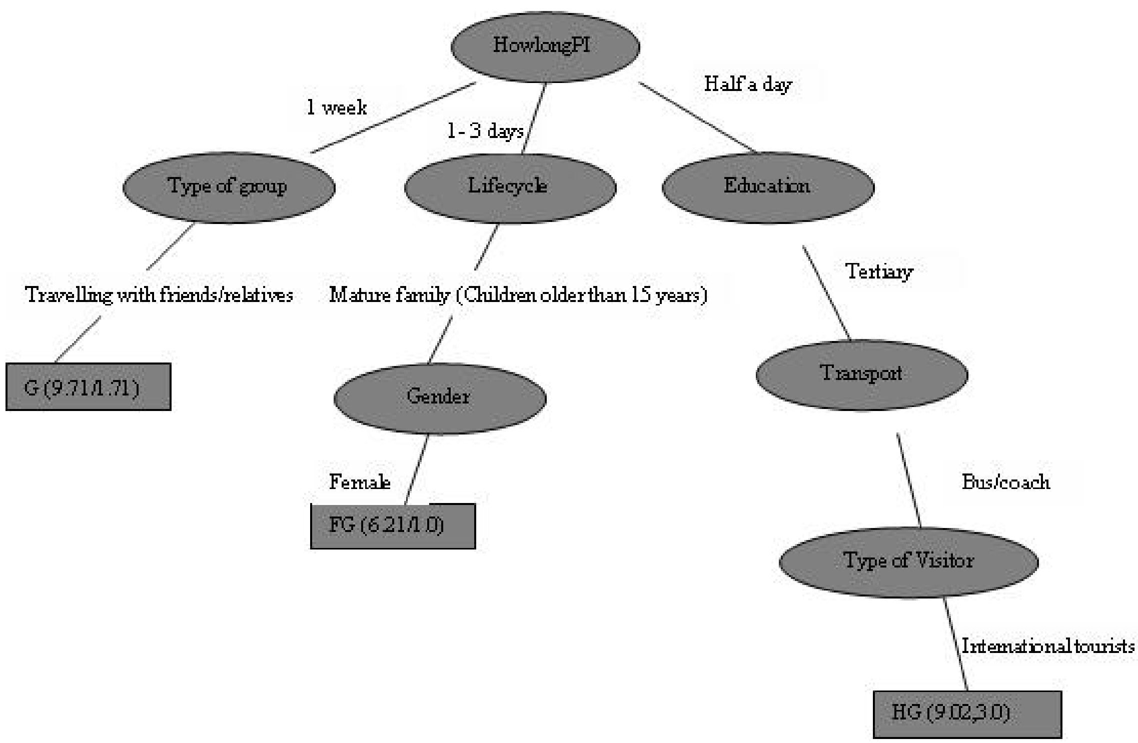

Figure 5 A partial decision tree for across patterns among major movement patterns

There are also other similar rules can be extracted from the decision tree. Here just shows an example. However, Even though this decision tree can show who use movement patterns such as “FG”, “G”, 60% of correct classification rate and small number of terminal nodes outputted from the decision tree prove that it could not be precisely used for predicting movement patterns just based on tourists profiles. Maybe other attributes should also be included or peer-to-peer comparison between two movement patterns is better than this general movement pattern comparison. Another disadvantage of decision tree algorithms is unstable. Slight change of data can significantly affect output of decision tree.

5 Conclusion

In this paper we describe design and implementation of method to identify the spatio-temporal movement patterns and across patterns between categorise of tourist and their spatio-temporal movement patterns. The frequent spatio-temporal movement sequences in the case study are extracted from the database. The major findings are as follows: if tourists only visited one attraction in the evening for their day trip it is Penguin parade. 71 out of 464 tourists visit Cowes or the Nobbies in the afternoon then head for Penguin Parade. If three attractions are visited by tourists such as temporal movement sequence patterns are “223” or “233”, related spatial movement patterns mostly are “DFG”, “FHG” and “HFG”.

Data mining method such as clustering and classification are used to isolate tourists characteristics associated with their movement patterns. We discovered several “nuggets”, the most significant one is that there are different spatio-temporal movement behaviours between international and domestic tourists. In addition, there are various key attributes to group tourists for different attractions. Visit frequency and lifecycle are important for Ventnor and Cowes. The significant attributes for dividing tourists who visited Cape Woolamia and Rhyll Inlet are gender, visiting frequency and lifecycle. For the Nobbies they are form of transport (car or bus) and travel group (travelling with friends or relatives or travelling in organized group or club).

Unfortunately the output from classification can’t predict tourists’ spatio-temporal movement pattens precisely even it can be used to identify across patterns between categories of tourists and their spatio-temporal movement.

Therefore, the future work is to identify the reasons why this decision tree method can not predict tourists’ spatio-temporal movement. Are there any other methods available to achieve this task? In addition, in this paper EM algorithms are used to identify characteristics of tourists who visited an attraction. However in order to predict which kind of tourists definitely went to an attraction, it needs to clarify whether tourist who did not went to the same attraction have the same characteristics. So further work need to be done to find out the differences between tourists who visit an attraction and who don’t visit an attraction.

Reference:

Arrowsmith, C & Chhetri, P 2003, 'Port Campbell National Park: Patterns of Use'.

Bardram, JE 2005, *Data Management Support for Location Based Services*, http://www.pervasive.dk/projects/LBS/LBS_background.htm#top.

Ciesielski, V & Lalani, A 2003, 'Data Mining of Web Access Logs From an Academic Web Site', paper presented to Proceedings of the Third International Conference on Hybrid Intelligent Systems (HIS'03):Design and Application of Hybrid Intelligent System.

Dilip 2005, *Dilip's Brief Introduction to Relational Databases*, http://www.cs.unc.edu/Courses/wwwp-s98/members/barman/databaseLesson/, 11/05/05.

Elmasri, R 2000, *Fundamentals of database systems*, Addison-Wesley, Reading, Mass ; Menlo Park, Calif.

Haglin, DJ 2004, *Decision Trees for Supervised Learning*, http://grb.mnsu.edu/grbts/doc/manual/J48_Decision_Trees.html, viewed 15/05/05.

Hsieh, N-C 2004, 'An integrated data mining and behavioral scoring model for analyzing bank customers', *Expert Systems with Applications*, vol. 27, no. 4, pp. 623-33.

Min, S-H & Han, I 2005, 'Detection of the customer time-variant pattern for improving recommender systems', *Expert Systems with Applications*, vol. 28, no. 2, pp. 189-99.

Phillip Island Internet Services 2005, *Welcome to Phillip Island*, http://www.phillipisland.net.au/trip/firehouse/firehouse.html, viewed 09/05/05.

Phillip Island Nature Park 2005, *How to get to Phillip Island*, www.Penguins.org.au, 01/05/05.

Quinlan, JR 1986, 'Induction of decision trees.' *Machine Learning*,, no. 1, p. 81–106.

Witten, IH & Frank, E 2000, *Data mining : practical machine learning tools and techniques with Java implementations*, Morgan Kaufmann, San Francisco, Calif.

Yada, K, Motoda, H, Washio, T & Miyawaki, A 2004, *Consumer Behavior Analysis by Graph Mining Technique*, 3214 edn, Lecture Notes in Computer Science.

Yan, X & Han, J 2002, 'gSpan: Graph-Based Substructure Pattern Mining', paper presented to Proceedings of the International Con-ference on Data Mining, Maebashi City, Japan.