Demonstrator of a tourist recommendation system

Erol Elisabeth, Richard Nock, and Fred Célimène

CEREGMIA - Centre d'Etude et de Recherche en Economie, Gestion, Modélisation et Informatique Appliquée,
Campus de Schoelcher, B.P. 7209 97275 Schoelcher Cédex, Martinique
erol@beepway.com,
Richard.Nock@martinique.univ-ag.fr,
Fred.Celimene@martinique.univ-ag.fr,
http://www.ceregmia.eu

Abstract. This paper proposes a way of using data collected from tracking gps installed in rental tourist cars. Data has been collected during more than one year. The gps positions are lined to the gps positions of the tourist sites (restaurants, beaches, museums ...). [9] These links are presented as a summary of the data. This summary is used to run specific versions of machine learning algorithms because of their geo-graphical dimension. This experiment shows how gps summaries of data can be used to extract relationships between stops of a car and touristic places.

Keywords: gps, association rules, sequential patterns, k-means,Q patterns, Geographical Center of Sequential patterns

1 Introduction

In this paper, we begin with data summaries and we use 5 types of data mining algorithms to process these summaries: association rules, sequential patterns, Q patterns, geographical center of sequential patterns and k-Means.

The aim is to provide the best recommendation for another tourist site for a tourist in his car.

The device used in the car is a tracking gps with a PND (Personal Navigation Device). With this PND it is possible to send in real time a recommendation to the tourist, and if he accepts the recommendation, the system shows him/her the best way to join this place as any gps navigation system.

2 Data Summaries

In our demonstration we have 852 dierent data summaries. These are the activities of 12 cars during more than 14 months in 2008/2009. A path is a succession of stops between the first stop of the first day in the car and the return of the car to the park. The rows are a gps position (date, time, latitude, longitude, instant speed, altitude, cap, status of the car [stopped, running]). These rows are used to produce a table of gps stops. After resuming, the positions collected are organized in sequential rows [date, gps stop position of, duration]. We associate the stop with the gps position of the tourist place. The result is a synthetic table with the succession of the visited tourist sites. Each row is presented as follow :

[sequential number] {date/time, tourist site, duration of visit}

[1] {date/time,1 'Casino Bateliere Plazza', 1:55}

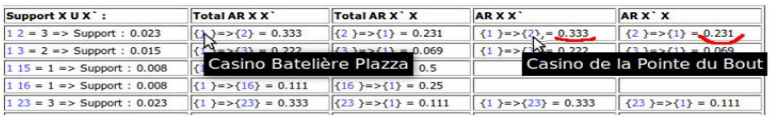
{date/time,2 'Casino Bateliere Pointe du Bout', 3:15}

[2] {date/time,1 'Habitation Clement', 0:53}

{date/time,1 'Casino Bateliere Pointe du Bout', 5:15}

3 Association rules

Association rule mining [2] is defined as : $I = i_1, i_2, ..., i_n$ as a set of $_n$ binary attributes called items. $D = t_1, t_2, ..., t_n$ a set of transactions called the database. Each transaction in D has a unique transaction ID and contains a subset of the items in I. A rule is defined as an implication of the form $X \Rightarrow Y$where $X, Y \subseteq I and X \cap Y = \oslash$. The conviction of a rule [3] is defined $conv(X \Rightarrow Y)$ With association rules algorithm, it is possible to know the relationship between 2 sites. In the screen display below we can see 33,3% of tourists who went to Casino Batelire Plazza went also to Casino de laa Pointe du Bout ; but 23,1% of tourists who went to 2 : Casino de la Pointe du Bout went also to 1 : Casino Batelire Plazza. With the proposed map we can see that the behaviors of tourists from south of the island are dierent from the one from the center.



Support X U X` :	Total AR X X`	Total AR X` X	AR X X`	AR X` X
1 2 = 3 => Support : 0.023	{1 }=>{2} = 0.333	{2 }=>{1} = 0.231	{1 }=>{2} = 0.333	{2 }=>{1} = 0.231
1 3 = 2 => Support : 0.015	{1 }=>{3} = 0.222	{3 }=>{1} = 0.069	{1 }=>{3} = 0.222	{3 }=>{1} = 0.069
1 15 = 1 => Support : 0.008	[illegible]	[illegible] 0.5		[illegible]
1 16 = 1 => Support : 0.008	{1 }=>{16} = 0.111	{16 }=>{1} = 0.25		
1 23 = 3 => Support : 0.023	{1 }=>{23} = 0.333	{23 }=>{1} = 0.111	{1 }=>{23} = 0.333	{23 }=>{1} = 0.111

Fig. 1. Association rule between 2 sites

4 Sequential patterns

Extraction of sequential patterns [1] and [4] make possible the discovery of temporal relations between 2 sites.In this sequence we notice a relationship between

52 : 'ART et Nature' AND 257 : 'Hotel le Panoramique'. We may suppose that before coming back to the 'hotel Le Panoramique' the tourist went to 'Art et Nature'. We propose two types of representations. The geographical map of pattern allows to have a visual representation of behaviors of tourist stops. It is possible to create an oriented diagram that shows the inter site links.

(52) ART et NATURE	=>	(257) Hotel Le Panoramique **	9	0.0124826629681
(1) Casino Batelière Plazza	=>	(23) BIBLIOTHEQUE SCHOELCHER	6	0.00832177531207
(1) Casino Batelière Plazza	=>	(3) La Galleria	2	0.00277392510402
(1) Casino Batelière Plazza	=>	(80) HERTZ	4	0.00554785020804
(1) Casino Batelière Plazza	=>	(2) Casino de la Pointe du Bout	2	0.00277392510402
(188) TOTAL	=>	(257) Hotel Le Panoramique **	2	0.00277392510402
(2) Casino de la Pointe du Bout	=>	(80) HERTZ	4	0.00554785020804
(2) Casino de la Pointe du Bout	=>	(134) Anse Noire	2	0.00277392510402
(2) Casino de la Pointe du Bout	=>	(1) Casino Batelière Plazza	2	0.00277392510402
(2) Casino de la Pointe du Bout	=>	(267) Hotel Pagerie ***	2	0.00277392510402
(2) Casino de la Pointe du Bout	=>	(137) Anse Mitan	2	0.00277392510402
(3) La Galleria	=>	(80) HERTZ	7	0.00970873786408
(3) La Galleria	=>	(23) BIBLIOTHEQUE SCHOELCHER	7	0.00970873786408
(3) La Galleria	=>	(280) Hotel Valmeniere***	5	0.00693481276006

Fig. 2. sequential pattern between sites

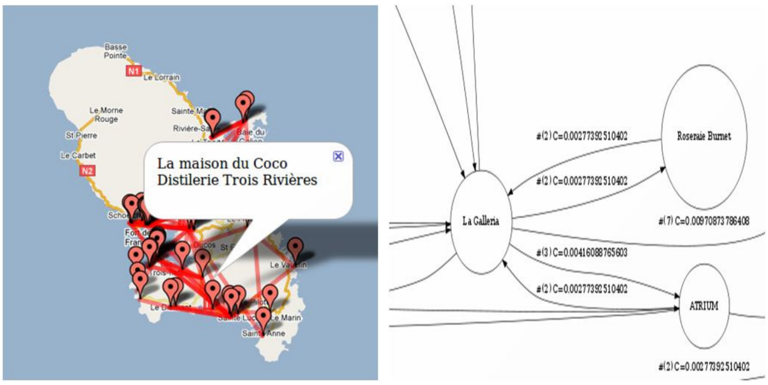


Fig. 3. Map of sequential pattern Diagram inter site links

4.1 Q patterns

The Q patterns are like patterns but the item sets do not have a fixed dimension [5] and [8]. For example item sets are in the database $conv(A, B \Rightarrow C)$ and in the same database we can also have $conv(A, B, D, F \Rightarrow T)$

23 : Bibliothque Schoecher , 217 : La Kasa Saveurs, 298 : Karibea Rsidence La Goelette, 12 : Habitation Clment, 39 : Habitation Depaz, 40 : Distillerie Neisson, 130 : Grande Anse d Arlet

$(23, 217, 298, 12, 298, 39, 40, 298 \Rightarrow 130)$ We can have a representation where each pattern has a specific color on an oriented graph. This sequential pattern shows a succession of stops with the same topic : rum distillery. We can have a representation where each pattern has a specific color on an oriented graph. Each Bubble is a tourist site.

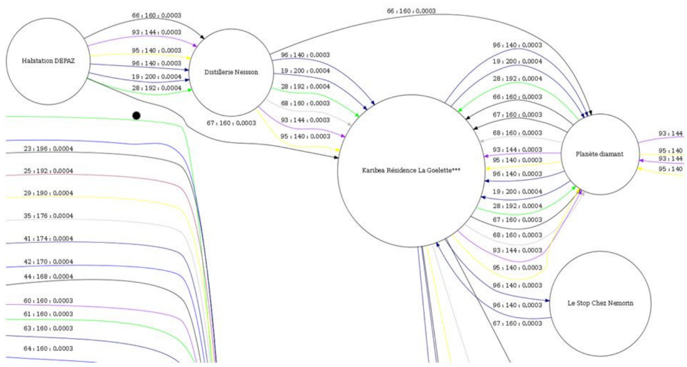


Fig. 4. Part of Q pattern

4.2 Geographical Center of Sequential patterns

We also compute a geographical solution as a recommendation when we already have found the sequential pattern or the cluster in a specific k-means. The objective is to find -in real time- the best tourist site next to a car when we already classify it in a k-means cluster or a sequential pattern [6] and [7]. In this example if the car is IN the cluster (12, 107) or in a sequential pattern where there are item sets (12 AND 107) and if the car is next to the centroid of this data, we can propose a new activity.

4.3 k-Means

We can have a k-means representation using 2 clusters. The cluster A (in red) is around the Center of the island and the south, the second one B (in yellow) is in the south.

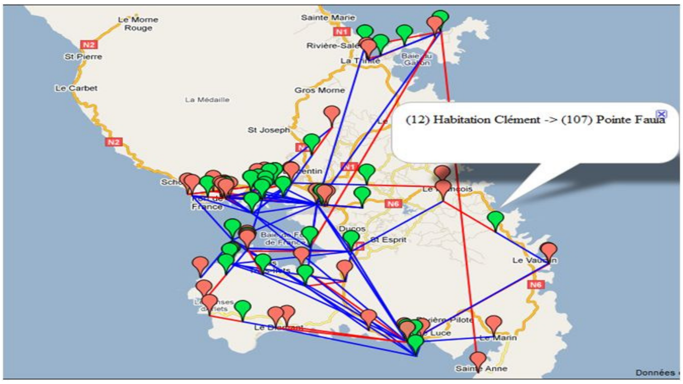


Fig. 5. Map of centroid of sequential patterns

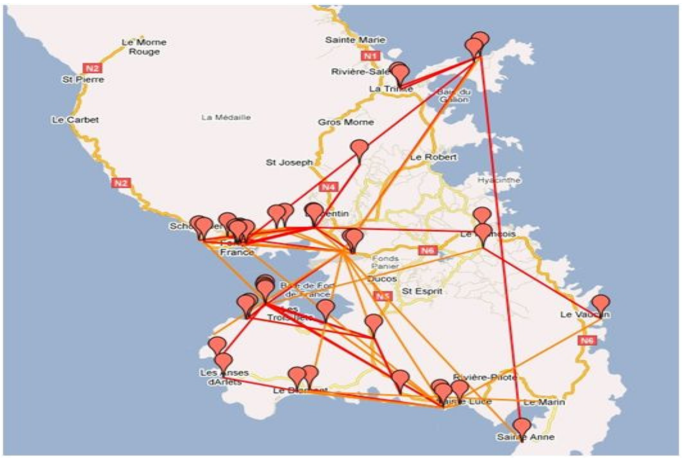


Fig. 6. K means (3 clusters)

5 Conclusion

In this paper, we have applied a set of data mining algorithms using data collected from tracking gps installed in rental tourist cars. Tourist organizations and agencies could look into these applications to find the best way to extract knowledge from their own database systems. GPS tracking companies can also find ideas to improve the uses of their collected data.

References

1. R. Agrawal and R. Srikant. Mining sequential patterns. In Philip S. Yu and Arbee L. P. Chen : Proceedings of the Eleventh International Conference on Data Engineering, March 6-10 (1995)
2. Heikki Mannila and Hannu Toivonen : Multiple uses of frequent sets and condensed representations (extended abstract) (1996)
3. Stephane Lallich et Olivier Teytaud : Evaluation et validation de l'interet des regles d'association (2000)
4. Pasquier Nicolas et Lakhal Lotfi :Data mining : Algorithmes d'extraction et de rduction des regles d'association dans les bases de donnes (2000)
5. Mohammed Javeed Zaki. : Spade : An efficient algorithm for mining frequent sequences. :Machine Learning, 42(1/2) : 3160 (2001)
6. F. Masseglia et M. Teisseire et P. Poncelet , Extraction de motifs sequentiels, Problemes et methodes (2004)
7. J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu. : Mining sequential patterns by pattern-growth : The prefixspan approach. IEEE Transactions on Knowledge and Data Engineering (2004)
8. Nistor Grozavu, Youns Bennani : Classification collaborative non supervisee LPN UMP CNRS, 249-264 CAP 2010
9. Nicolas Bchet, Marie-Aude Aufaure, Yves Lechevallier Construction et peuplement de structures hirarchiques de concepts dans le domaine du e-tourisme : INRIA - 475-506, IFIA 2011