Suggesting Tourist Destinations by means of Time-Slice Density Estimation

Michelangelo Ceci, Annalisa Appice, Donato Malerba, Nicola Schirone, Nicola Davide Traversa, and Valerio Valrosso

Dipartimento di Informatica, Università degli Studi di Bari
via Orabona, 4 - 70126 Bari - Italy
{ceci,appice,malerba}@di.uniba.it

Abstract. In recent years, a growing interest has been given to trajectory data mining applications that permit to support mobility prediction with the aim of anticipating or pre-fetching possible services. Proposed approaches typically consider only spatio-temporal information provided by collected trajectories. However, in some scenarios, such as that of tourist supporting, semantic information which express needs and interest of the user (tourist) should be taken into account. This semantic information can be extracted from textual documents already consulted by the tourists. In this paper, we present the application of a time-slice density estimation approach that permits to suggest/predict the next destination of the tourist. In particular, time-slice density estimation permits to measure the rate of change of tourist's interests at a given geographical position over a user-defined time horizon. Tourist interests depend both on the geographical position of the tourist with respect to a reference system and on semantic information provided by geo-referenced documents associated to the visited sites.

1 Introduction

Tourism has become, in the second part of last century, one of the most important economic activities in the world. According to World Travel and Tourism Council (WTTC), in 2005, about 11% of World Gross Domestic Product (GDP) was generated by the tourism sector and a consistent part (more than 200 million people) of the global workforce is employed [3]. In addition, it is predicted to be one of a few businesses that will continue to grow at an appreciable rate (around 5% per year) and to generate job opportunities in the future. Tourism is nowadays an important vehicle for regional and national developments and, for many countries, it represents the major contributor to the local economy.

Moreover, recently tourism has become an extremely dynamic system [6] and the intensified (web-based) marketing efforts of all tourism organizations have paved the way for new advances in knowledge based technologies applied to the destination management problem [15]. According to the definition provided in [9], a tourism destination may be intended as a geographical area that offers the tourist the opportunity of exploiting a variety of attractions and services. Recent advances in positioning technology (Global Positioning Systems - GPS) permit to track the tourist position in time and space (trajectories) and, then, to consider the tourist as a moving object into a predefined spatio-temporal space. Knowledge on the past positions of the tourist can be used to suggest/predict the next destination and anticipate or pre-fetch possible services in the next destination. At present, proposals in the literature use either the history of movements of the single object on the basis of which future destination is predicted [10, 18] or the movements of all objects in a certain area to learn a classifier [12]. In both cases, the location predictor capitalizes only on history of movements (trajectories) according to the idea that individuals tend to follow common paths. The innovative aspect of this work is that of considering each moving object (tourist) as an independent individual with a personal movement history, preferences and interests. Additionally, the semantic information provided by the documents geo-referenced at the visited sites is considered in order to maintain an informative profile of the tourist. Starting from an empty profile, the tourist profile is dynamically updated on the basis of both the current spatial position of the tourist and the textual content of the consulted documents.

Our basic assumption is that a tourist moves towards a close destination which is semantically consistent with her/his current profile as more as possible. Destination which minimizes the profile drift can be most likely suggested/predicted as next destination. Profile is computed by resorting to the concept of time-slice density estimation [1]. The profile drift is measured in order to estimate the rate of change of tourist's interests over a user-defined time horizon when the tourist moves towards the new destination.

The paper is organized as follows. In the next Section, some related works are discussed. In Section 3, we present our system ITiS (Intelligent Tourism information System) to suggest/predict next tourist destinations by minimizing profile drift. Finally, In Section 4, some empirical results are presented. Conclusions and future research directions are presented in Section 5.

2 Related Works

Roots of this work are in the research field of moving objects prediction. The pioneering work in this area is presented in [11] where moving object behavior is modeled as repetitions of elementary movement patterns (e.g., linear or circular). The next location is predicted by means of a mobile motion prediction algorithm that is highly sensitive to random movements of the object. Subsequently, Markov chain models have been studied in order to estimate the probability of an object's movement from one region or cell to another at the next time period. In particular, Ishikawa et al. [8] propose to derive transition probabilities between cells over the space from indexed trajectories. A special type of histogram, called mobility histogram, is used in order to describe mobility statistics based on the Gauss Markov chain model. Mobility histogram is then used to predict the next cell in which the object would probably move in the future. Other approaches use sequential patterns in order to model trajectories in terms of ordered sequences of time-stamped locations [7, 17]. Most of these approaches try to predict the movement of an object on the basis of the assumption that people typically follow the crowd. Morzy [13] proposes to periodically mine offline historical data of other objects moving on the same area and discover frequent trajectories of objects representing popular movement routes. The unknown location of a moving object can be predicted by ranking on-line trajectories which match past history of the object, according to support and confidence, and using the selected trajectory to predict next destination. More recently, Monreale et al. [12] propose *WhereNext* which extracts trajectory patterns as a concise representation of behavior of moving objects, that is, sequences of regions frequently visited within a travel time. A decision tree is then learned from trajectory patterns that insist a certain area and it is used to predict the next location of a new trajectory by finding the best matching path in the tree.

All methods described above suggest/predict the next destination of a moving object based upon the movement history of either the object itself or the other objects which move in the surrounding area. Although both space and time are considered, none of previous methods takes into account additional semantic information which may be descriptive of the object profile and be bearing of information on the next destination.

The concept of trajectories enriched with semantic information was originally formalized in [2] where authors consider, as semantic information, the name of the geographic layer (e.g., hotels, museums, schools) associated to each site in the trajectories (called stops). In a semantic based moving environment, as that considered in this paper, we do not consider only the information on the geographical layer, but all the semantic information which can be automatically extracted from textual documents possibly geo-referenced with the trajectory sites. This semantic information is intended to express the interests, preferences and needs of the object. As in a stream, each time the object moves towards a new site, semantic information geo-referenced with the site contribute to dynamically construct/update the object profile. Hence, it is reasonable to assume that an object moves to a destination site that slightly changes the object profile or, in other terms, that is semantically close to the object profile. Following this idea, our point of view is that suggesting/predicting a semantic based next destination can be intended as an application of the change diagnosis in evolving data streams. In this area, the seminal work is that of Aggarwal [1], which firstly proposes to capture the change of spatially referenced characteristics over time with the concept of velocity density. The idea of velocity density is that of measuring the rate of change of data concentration at a given spatial location over a user-defined time horizon. Our assumption is that the destination which is spatially close to the current one and which minimizes the rate of change in profile is the most probable next destination.

3 ITiS

The task we address is that of suggesting/predicting the next destination of a tourist which moves on a map given: (1) a spatial referencing system that permits to uniquely define a spatial position on the map (e.g., latitude and longitude); (2) the set of destinations, each of which has a spatial position with respect to the spatial referencing system and geo-references a set of textual documents; (3) the current tourist position with respect to the spatial referencing system; and (4) the trajectory followed by the tourist and the associated profile which is updated according to the semantic information extracted from textual documents consulted by the tourist over a user-defined time horizon.

ITiS addresses this task by automatically updating the profile each time the tourist visits a new site on the map. The profile takes into account the following assumptions:

(1) a tourist consults a document if the document content is interesting for him/her;

(2) content of documents recently consulted is more interesting for the tourist than that of documents consulted in the past.

A time-slice density estimation is then used to suggest/predict the next destination.

Before presenting how the suggestion is performed, we introduce some preliminary definitions, describe how to extract the semantic information from the consulted textual documents in order to update the profile, how to use the time-slice density estimation in order to suggest/predict next destination.

3.1 Preliminary Concepts

Let :

- $P = \{p_i = \langle x_{p_i}, y_{p_i} \rangle | i = 1 \ldots n\}$ be the set of candidate destinations on a map towards a tourist can move, such that x_{p_i} and y_{p_i} represent the spatial coordinates (e.g., latitude and longitude) of p_i with respect to the given spatial referencing system, and n is the cardinality of P (i.e., number of candidate destinations);
- $D = \{d_j | j = 1 \ldots, N\}$ be a set of textual documents.

One or more documents in D are geo-referenced to a destination p_i according to the function $\delta : P \rightarrow 2^D$ such that $\delta(p_i) = \{d_j \in D | d_j$ is geo-referenced to $p_i\}$. The function δ is not injective, that is, the same textual document can be geo-referenced to two or more destinations.

Given U be the set of tourists, it is also possible to define the set of visits of the tourist $u_j \in U$, that is, the movement history of the tourist, as:

$$v_{u_j}(t) = \left(\langle p_{j_1}, t'_{j_1}, t''_{j_1}, D_{j_1} \rangle, \ldots, \langle p_{j_s}, t'_{j_s}, t''_{j_s}, D_{j_s} \rangle \right) \quad (1)$$

where $p_{j_k} \in P$ is the k-th ($k = 1, \ldots, s$) destination the tourist u_j visited, $D_{j_k} \subseteq \delta(p_{j_k})$ is the set of consulted documents geo-referenced to p_{j_k} and $[t'_{j_k}, t''_{j_k}]$ is the time interval (starting time and ending time) of the k-th visit such that $t'_{j_k} \leq t''_{j_k}$ and:

$$\begin{cases} t''_{j_k} \leq t'_{j_{k+1}} & iff \quad k \leq s - 1 \\ t''_{j_k} \leq t & iff \quad k = s \end{cases}$$

The set of consulted documents at the time t by the tourist u_j is defined as:

$$d_{consulted}(v_{u_j}(t)) = \bigcup_{k=1 \ldots s} D_{j_k}. \quad (2)$$

Analogously, the set of documents which are still not consulted at the time t is defined as:

$$d_{notConsulted}(v_{u_j}(t)) = D - d_{consulted}(v_{u_j}(t)). \quad (3)$$

The set of visited destinations at the time t by the tourist u_j is defined as:

$$p_{visited}(v_{u_j}(t)) = \bigcup_{k=1 \ldots s} \{p_{j_k}\}. \quad (4)$$

Finally, the set of destinations which are still not visited at the time t is defined as:

$$p_{notVisited}(v_{u_j}(t)) = P - p_{visited}(v_{u_j}(t)). \quad (5)$$

3.2 Document Representation

A document is pre-processing in order to remove *stopwords*, such as articles, adverbs, prepositions and other frequent words and to determine equivalent stems (*stemming*) by means of Porter's algorithm for English texts [14]. Pre-processed documents are subsequently represented by means of a feature set which is determined on the basis of some statistics whose formalization is reported below.

Let C be a set of documents, with $C \subseteq D$, and w be a token of a stemmed (non-stop) word which occurs in a document of D, it is possible to define:

- $TF_d(w)$ as the *relative* frequency of w in a document $d \in D$,
- $TF_C(w) = max_{d \in C} TF_d(w)$ the maximum value of $TF_d(w)$ on all documents $d \in C$,
- $DF_C(w) = \frac{|\{d \in C| \quad w\ occurs\ in\ d\}|}{|C|}$ the percentage of documents in C in which w occurs,
- $CF_{C',C'',\ldots C^{(s)}}(w)$ is the number of sets of documents where the token w occurs. In this formulation, sets of documents are denoted as $C', C'', \ldots C^{(s)}$ with $C^{(i)} \subseteq D$.

Then the following measure, used in text categorization [4], associates a token w_i with its score v_i and selects relevant tokens for the representation of documents in D:

$$v_i = \frac{TF_{d_{consulted}(v_{u_j}(t))}(w_i) \times \left(DF_{d_{consulted}(v_{u_j}(t))}(w_i)\right)^2}{CF_{d_{consulted}(v_{u_j}(t)), d_{notConsulted}(v_{u_j}(t))}(w_i)} \tag{6}$$

Tokens that minimize v_i ($maxTF \times DF^2 \times ICF$ criterion) are penalized since they are commonly used in documents of both $d_{consulted}(v_{u_j}(t))$ and $d_{notConsulted}(v_{u_j}(t))$ and do not permit to discriminate between the two sets. Differently, tokens that maximize v_i can be reasonably used to represent documents in D. In particular, the set of the best n_{dict} tokens forms the *dictionary* $Dict(v_{u_j}(t))$ of the tourist u_j at the time t. $Dict(v_{u_j}(t))$ is then used to index the set of N documents in D according to the normalized $TF \times idf$ measure [16].

In the matrix representation we have:

$$\omega(v_{u_j}(t)) = \begin{bmatrix} \omega_{1,1} & \omega_{1,2} & \cdots \\ \vdots & \ddots & \cdots \\ \vdots & \vdots & \omega_{N,n_{dict}} \end{bmatrix} \tag{7}$$

where $\omega_{p,q} = \frac{TF_{d_p}(w_q) \times ln \frac{N}{1+N \times DF_D(w_q)}}{\|\omega(v_{u_j}(t))\|_1}$ and $d_p \in D$ and $w_q \in Dict(v_{u_j}(t))$. It is noteworthy that $\omega_{p,q} \in [0, 1]$.

3.3 Time-Slice Density Based Profile

We define the profile of the tourist u_j at the time t as the triple $\langle x_{u_j}(t), y_{u_j}(t), X(v_{u_j}(t)) \rangle$, where $(x_{u_j}(t), y_{u_j}(t))$ represents the geographical position of the tourist with respect to the spatial referencing system, while $X(v_{u_j}(t))$ represents the semantic position of the tourist over the space $[0,1]^{n_{dict}}$. Since it would be computationally impractical to represent and search this continuous space, ITiS uses a discrete version of the same space. The discrete space is defined by resorting to a discretization of the interval $[0,1]$ according to a supervised discretization function $\psi : [0,1] \to \Phi$, where Φ is a finite set of values whose cardinality β is apriori defined by the user. This way, the continuous space $[0,1]^{n_{dict}}$ is transformed into the discrete space $\Phi^{n_{dict}}$. In ITiS, ψ is based on the equal-width discretization algorithm [5] that associates x with its nearest value in $\Phi = \{0, \frac{1}{\beta}, \frac{2}{\beta}, \ldots, \frac{\beta-1}{\beta}, 1\}$.

The semantic position $X(v_{u_j}(t))$ is computed by a forward time-slice density estimator $F(X, t, h_t, u_j)$ that is obtained by adapting the forward density estimator presented in [1] to our scenario. Formally, $X(v_{u_j}(t)) = \underset{X \in \Phi^{n_{dict}}}{\operatorname{argmax}} F(X, t, h_t, u_j)$, where the density function $F(X, t, h_t, u_j)$, that is measured for all possible semantic positions $X \in \Phi^{n_{dict}}$ of the tourist u_j at the time t, is maximized. The value of density at a given semantic position X is forward estimated on the basis of the sequence S of time-stamped textual documents which belong to $d_{consulted}(v_{u_j}(t))$ and have been consulted during the visits of the tourist in the time slice $[t - h_t, t]$. More precisely, $S = \langle d_1, t_1 \rangle, \ldots, \langle d_{|S|}, t_{|S|} \rangle$, where $\forall \langle d_i, t_i \rangle \in S$, $\exists \langle p_{j_k}, t'_{j_k}, t''_{j_k}, D_{j_k} \rangle \in v_{u_j}(t)$ such that: *i*) $d_i \in D_{j_k}$, *ii*) $t'_{j_k} \leq t_i \leq t''_{j_k}$ and *iii*) $t - h_t \leq t_i \leq t$.

A kernel density estimation is used in order to provide us a continuous estimate of the density $F(X, t, h_t, u_j)$ as sum of smoothed values of kernel functions $K_{h_t, u_j}(X, t)$.

$$F(X, t, h_t, u_j) = C_F \times \sum_{\langle d_i, t_i \rangle \in S} K_{h_t, u_j}(X - \omega_{d_i}, t - t_i) \tag{8}$$

In this equation, $\omega_{d_i} = [\omega_{d_i,1}, \ldots, \omega_{d_i,n_{dict}}]$ is the vector representation of the document $d_i \in D$ (see Equation 7), C_F is a constant value that makes $\sum_{X \in \Phi^{n_{dict}}} F(X, t, h_t, u_j) = 1$ and $K_{h_t, u_j}(X - \omega_{d_i}, t - t_i)$ is a semantic-temporal kernel function that uses a time fading factor to give more importance to recently consulted documents:

$$K_{h_t, u_j}(\Delta X, \Delta t) = \left(1 - \frac{\Delta t}{h_t}\right) K'(\Delta X) \tag{9}$$

Specifically, $K'(\Delta X)$ is the product of n_dict identical Gaussian kernel functions:

$$K'(\Delta X) = \prod_{q=1}^{n_{dict}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{\Delta X_q^2}{2\sigma^2}} \tag{10}$$

where σ is a user defined smoothing parameter.

3.4 Next Destination Suggestion/Prediction

In order to suggest the next possible destination, ITiS assumes that a tourist moves towards a site spatially close to her/his current position and is not interested to visit that same site more than once. According to these assumptions, the set of candidate next destinations is defined as:

$$\begin{aligned} P_r(v_{u_j}(t)) = \{\ & p \in p_{notVisited}(v_{u_j}(t)) | \\ & EuclideanDistance(p, (x_{u_j}(t), y_{u_j}(t))) \leq r\} \end{aligned}$$

where r represents the maximum spatial distance that the tourist is willing to cover. Among the candidate destinations in $P_r(v_{u_j}(t))$, ITiS suggests the tourist to move toward the destination which geo-references the set of documents whose consultation will lead to minimize her/his profile drift, that is:

$$p_{next}(v_{u_j}(t)) = \underset{p \in P_r(v_{u_j}(t))}{\operatorname{argmin}} \; drift(X(v_{u_j}(t)) \,,\, \langle p, t, t, \delta(p) \rangle) \tag{11}$$

ITiS suggests the tourist to move toward one of the destination which minimizes the drift of her/his profile. If several destinations minimize the drift measure, then ITiS suggests all of them ordered according to the Euclidean distance from the current geographical position of the tourist. The function $drift(\cdot,\cdot)$ in Equation 11 can be computed in two alternative ways, that is, (1) by computing the simple cosine similarity between the semantic position of the tourist profile at the time t and the set of textual documents $\delta(p)$ which are geo-referenced to the candidate next destination p; or (2) by measuring the variation of the semantic position of the tourist profile due to the simulation of a visit to the candidate next destination. By computing the cosine similarity, $drift(\cdot,\cdot)$ is computed as:

$$drift(X(v_{u_j}(t)) \,,\, \langle p, t, t, \delta(p) \rangle) = \frac{1}{|\delta(p)|} * \sum_{d \in \delta(p)} \frac{X(v_{u_j}(t)) \cdot \omega_d}{\|X(v_{u_j}(t))\| \|\omega_d\|} \tag{12}$$

Alternatively, by measuring the variation of the semantic position of the tourist profile due the visit, $drift(\cdot,\cdot)$ is computed as:

$$drift(X(v_{u_j}(t)) \,,\, \langle p, t, t, \delta(p) \rangle) = \|X(v_{u_j}(t)) - X(v_{u_j}(t),\, \langle p, t, t, \delta(p) \rangle \,)\|_2 \tag{13}$$

4 Experiments

In this Section we present applications of ITiS in suggesting the next destination of a tourist on the basis of the time-slide density estimation of her/his semantic-base profile. We consider two touristic areas, that is, Stockport (United Kingdom) and Paris (France).

Due to difficulty in obtaining real data, we asked sixteen users to perform virtual thematic tours over either Stockport or Paris. The basic hypothesis is that the tourist has a Java enabled mobile device with GPS and remotely access geo-referenced textual documents stored in the server. Documents have been selected by a tourism expert.

ITiS is run with $n_{dict} = 5$, $\sigma = 0.5$, $\beta = 20$. Additionally, h_t is appropriately set in order to temporally consider, for each tourist, the entire set of stored visits. r is set to 3 Kms in Stockport experiment and it is set to 32 Kms in Paris experiment. This choice of r permits to consider all sites in the corresponding maps as candidate destinations to be suggested. To evaluate how much a suggested destination p matches the interest of the tourist u, we assign $score(u,p) = \begin{cases} 1 \text{ if } u \text{ accepts to move toward } p \\ 0 \text{ otherwise.} \end{cases}$. By considering that $p_{next}(v_{u_j}(t))$ (see Equation 11) may suggest a set of (equivalent) destinations, denoted as P_{next}, then:

$$score(u, P_{next}) = \frac{\sum_{p_i \in P_{next}} score(u, p_i)}{|P_{next}|} \tag{14}$$

4.1 Stockport

Stockport is a large town in Greater Manchester and, in this study, we consider eight tourists who visited Stockport by moving from one site to another. We consider thirty candidate destination sites which, for descriptive purposes, are classified into several categories as reported in Table 1. Each site has a geographic position (latitude and longitude) over the map of Stockport and it geo-references a set of textual documents (e.g., Wikipedia pages) which are descriptive of the site attractiveness. In all, we consider a total of sixty-four textual documents. Additionally, we have tracked the moving trajectory and the consulted documents of the tourists. A brief description of tourists' trajectories is reported in Table 2. The destinations suggested by ITiS for each tourist with both the cosine similarity measure and the semantic variation measure are reported in Table 3. The score shows that the cosine similarity measure generally outperforms the performance of the semantic variation measure. Indeed, the semantic variation measure often suggests additional non-interesting destinations which result in decreasing the score value. This is mainly due to the fact that semantic variation measure is affected by the discretization function more than the cosine similarity measure.

4.2 Paris Dataset

In this experiment, we consider fifty-one candidate destinations located over the map of Paris. Destinations are classified into seven categories as reported in Table 4. In all, we consider ninety-two textual documents and results are collected on eight tourists. The destinations suggested by ITiS for each tourist with both the cosine similarity measure and the semantic variation measure are reported in Table 6. By analyzing the average score, we observe a significant improvement with respect to the results obtained with Stockport data. This is motivated by the fact that in this experiment, tourists followed trajectories where it is possible to recognize well defined thematic interests of the tourist (e.g., a thematic interest for the impressionist art). This depends on the fact that Paris offers a wide spectrum of touristic attractions which may match distinct thematic interests of a possible tourist. By comparing the score obtained with the cosine similarity measure and the semantic variation measure, we observe that results confirm the main considerations drawn from the analysis of Stockport data (i.e., cosine similarity measure outperforms semantic variation measure).

5 Conclusion

In this paper, we have presented a version of the forward time-slice density estimation that is tailored for suggesting/predicting the next destination where a tourist reasonably would move towards. Results on two distinct real-world datasets show both effectiveness and accuracy of the proposed approach. As future work, we intend to take into account constraints in the suggestion/prediction step. This way, for example, the system can avoid to suggest specific destinations during closing times. Additionally, we intend to use the forward time-slice density estimation in order to suggest complete itineraries.

Site	Category
Hazel Grove Rail Station, Bramhall Rail Station, Rose Hill Marple Rail Station, Stockport Rail Station, Bredbury Rail Station	TRANSPORT NET
The Co-Op Bank Pyramid, Wellington Mill, Stockport Viaduct, Staircase House, Stockport Town Hall, Air Raid Shelters Museum	MONUMENT AND MUSEUM
Vernon Park	PARK
The Bowling Green, Duke Of York, The Hare & Hounds, The Romper Inn, Bredbury Hall Hotel	RESTAURANT AND HOTEL
Stockport College (Town Centre), Stockport College (Heaton Moore)	SCHOOL AND UNIVERSITY
TK Maxx, Merseyway, Debenhams, John Lewis Cheadle	SHOPPING
Stockport Plaza, Bramhall Park Golf Club	SPORT AND ENTERTAINMENT
Unitarian Church, Salvation Army Church, Salvation Army Church (Chesire), St. Barnabas Parish Church, St. Elizabeth Church	CHURCH

Table 1. Candidate destinations description for Stockport data.

Acknowledgment

This work is partial fulfillment of the research objective of ATENEO-2009 project "Estrazione, Rappresentazione e Analisi di Dati Complessi".

References

1. C. C. Aggarwal. On change diagnosis in evolving data streams. *IEEE Trans. Knowl. Data Eng.*, 17(5):587–600, 2005.
2. L. O. Alvares, V. Bogorny, B. Kuijpers, J. A. F. de Macêdo, B. Moelans, and A. A. Vaisman. A model for enriching trajectories with semantic geographical information. In H. Samet, C. Shahabi, and M. Schneider, editors, *GIS*, page 22. ACM, 2007.
3. R. Baggio and L. Caporarello. Decision support systems in a tourism destination: literature survey and model building . In *itAIS 2005 - 2nd Conference of the Italian Chapter of AIS: Organising Information Systems: Evolution of Studies in the Field*, 2005.
4. M. Ceci and D. Malerba. Classifying web documents in a hierarchy of categories: a comprehensive study. *J. Intell. Inf. Syst.*, 28(1):37–78, 2007.
5. J. Dougherty, R. Kohavi, and M. Sahami. Supervised and unsupervised discretization of continuous features. In *ICML*, pages 194–202, 1995.
6. B. H. Farrell and L. Twining-Ward. Reconceptualizing tourism. *Annals of Tourism Research*, 31(2):274–295, 2004.
7. F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi. Trajectory pattern mining. In P. Berkhin, R. Caruana, and X. Wu, editors, *KDD*, pages 330–339. ACM, 2007.
8. Y. Ishikawa, Y. Tsukamoto, and H. Kitagawa. Extracting mobility statistics from indexed spatio-temporal datasets. In J. Sander and M. A. Nascimento, editors, *STDBM*, pages 9–16, 2004.
9. J. Jafari. *Encyclopedia of Tourism*. Routledge, London, 2000.

Tourist	Visited sites	No. of consulted docs
T1	Stockport Rail Station	4
T2	Salvation Army Church	1
T3	Salvation Army Church	1
	Salvation Army Chesire	1
	St. Elizabeth Church	3
	Unitarian Church	2
T4	The Bowling Green	1
	Duke Of York	2
	The Hare & Hounds	2
T5	The Co-Op Bank Pyramid	1
	Stockport Viaduct	2
	The Hare & Hounds	1
	Vernon Park	1
	Stockport Plaza	3
T6	Wellington Mill	2
	Staircase House	3
T7	TK Maxx	2
	Merseyway	2
	Debenhams	1
T8	Stockport Plaza	2
	Stockport College (Town Centre)	1
	John Lewis Cheadle	1

Table 2. Tourist trajectories and corresponding number of consulted documents in Stockport data.

10. H. Jeung, Q. Liu, H. T. Shen, and X. Zhou. A hybrid prediction model for moving objects. In *ICDE*, pages 70–79. IEEE, 2008.
11. G. Y. Liu and G. Q. M. Jr. A predictive mobility management algorithm for wireless mobile computing and communications. *Proceedings of the Fourth IEEE International Conference on Universal Personal Communications*, page 268272, 1995.
12. A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti. Wherenext: a location predictor on trajectory pattern mining. In J. F. E. IV, F. Fogelman-Soulié, P. A. Flach, and M. J. Zaki, editors, *KDD*, pages 637–646. ACM, 2009.
13. M. Morzy. Mining frequent trajectories of moving objects for location prediction. In P. Perner, editor, *MLDM*, volume 4571 of *Lecture Notes in Computer Science*, pages 667–680. Springer, 2007.
14. M. F. Porter. An algorithm for suffix stripping. *Readings in information retrieval*, pages 313–316, 1997.
15. J. R. B. Ritchie and G. I. Crouch. *The competitive destination : a sustainable tourism perspective*. CABI, Oxon, UK, 2003.
16. F. Sebastiani. Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1):1–47, 2002.
17. J. Yang and M. Hu. Trajpattern: Mining sequential patterns from imprecise trajectories of mobile objects. In Y. I. et al., editor, *EDBT*, volume 3896 of *Lecture Notes in Computer Science*, pages 664–681. Springer, 2006.
18. G. Yavas, D. Katsaros, Ö. Ulusoy, and Y. Manolopoulos. A data mining approach for location prediction in mobile environments. *Data Knowl. Eng.*, 54(2):121–146, 2005.

Tourist	Next Destination Suggestion			
	cosine	score	semanticVariation	score
T1	Rose Hill Marple Station	1	Rose Hill Marple Station	1
T2	Salvation Army Church (Chesire)	1	Salvation Army Church (Chesire)	1
T3	St. Barnabas Parish Church	1	*more than one*	1/3
T4	The Romper Inn	1	*more than one*	3/5
T5	Stockport College (Town Centre)	0	*more than one*	1/7
T6	Vernon Park	1	*more than one*	1/4
T7	John Lewis Cheadle	1	John Lewis Cheadle	1
T8	Stockport Town Hall	0	Salvation Army Church	0
Avg.		**0.75**		**0.54**

Table 3. Score computed over the destinations suggested by ITiS for the Stockport data.

Site	Category
Gare de Lyon, Gare de l'Est	TRANSPORT NET
Palais-Royal, Place Vendôme, Conciergerie, Place des Victoires, Place des Vosges, Luxembourg Palace, Tour Eiffel, Place de la Concorde, Arc de Triomphe de l'Étoile, Église de la Madeleine, l'Opéra National de Paris, Place de la Bastille, Palais de l'Elysée, la défense, Montmartre, Avenue des Champs-Élysées'	MONUMENT
Musée Louvre, Musée National Picasso, Centre Pompidou, Musée Cluny, Musee Hôtel National des Invalides, Musée Orsay, Musée Rodin, Musée de Orangerie, Musée Jacquemart-André, Grévin, Musée des Gobelins, Musée de La Poste, Musée dArt Moderne de la Ville de Paris, Musée Marmottan Monet, Fragonard Musee du Parfum, Musée de Les Égouts de Paris	MUSEUM
Bois de Boulogne, Parc des Buttes Chaumont, La Villette, Canal Saint-Martin	PARK
Sorbonne	SCHOOL AND UNIVERSITY
Aquarium du trocadero, EuroDisney, Park Asterix, Moulin Rouge, Hard Rock café, Folies Bergère	SPORT ENTERTAINMENT
Basilique du Sacre Coeur, Sainte Chapelle, Saint Eustache, Notre Dame, Saint Marrie, Le Panthéon, Saint Étienne du Mont	CHURCH

Table 4. Description of the candidate destinations in the Paris data.

Tourist	Itinerary	Visited sites	No. consulted docs
T1	Traditional	Tour Eiffel - Champs Elisée - Arc de Triomphe de l'Étoile - Place de la concorde - Louvre - Notre dame	9
T2	Museum	Musée du Louvre - Musée Orsay - Musée National Picasso - Musée Orangerie - Musée Jacquemart-André - Centre Pompidou	8
T3	Church	Notre Dame - Sacre Cour - Pantheon - Madleine	8
T4	Impressionism	Monmatre - Musée Orsay - Musée Orangerie - Musée Monet	6
T5	Historical	Arc de Triomphe de l'Étoile - Place de la Bastille - Place de la Concorde - Concergerie - Place vendome - Place de josges - Egouts de Paris	16
T6	Historical-political	Palais Royal - Luxembourg Palace - Palais de l'Elysée - La défense - Sorbonné - Musée National des Invalides - Place Vendome	14
T7	Entertainment	Champs élisée - Acquarium du Trocadero - Musée du Parfum - Folies Bergére - Disneyland - Moulin Rouge - Park Asterix - Hard Rock Café	12
T8	Nature	Canal SaintMartine - Parc des Buttes Chaumont - Bois de Boulogne - LaVillette	8

Table 5. Tourist trajectories and corresponding number of consulted documents in Paris data.

Tourist	Next Destination Suggestion			
	cosine	**score**	**semanticVariation**	**score**
T1	Musée Orsay	1	*more than one*	2/9
T2	Musée de La Poste	1	*more than one*	3/3
T3	Monmatre	1	*more than one*	3/11
T4	Musée Rodin	1	*more than one*	2/3
T5	Champs Elisée	1	*more than one*	2/3
T6	Park Asterix	0	*more than one*	9/13
T7	Musée de La Poste	1	*more than one*	4/4
T8	Park Asterix	1	EuroDisney	1
Avg.		**0.88**		**0.69**

Table 6. Score computed over the destinations suggested by ITiS for the Paris data.