A Hybrid Approach for Spatial Web Personalization

Yanwu Yang and Christophe Claramunt

Naval Academy Research Institute BP 600, 29240, Brest Naval, France
{yang,claramunt}@ecole-navale.fr

Abstract. In the context of Web personalization, Markov chains have been recently proposed to model user's navigational trails, in order to infer user preference and predict future visits through computation of transitional probabilities. Based on these principles, the research introduced in this paper develops a hybrid Web personalization approach that applies k-order Markov chains towards an integration of spatial proximity and semantic similarity for the manipulation of geographical data on the Web. This framework personalizes Web navigational experiences over spatial entities embedded in Web documents. A reinforcement process is also introduced to evaluate and adapt interactions between the user and the Web on the basis of user's relevance feedbacks. An illustrative case study applied to spatial information available on the Web exemplifies our approach.

1 Introduction

Web personalization attempts to overcome the information overload, navigation problems and cognition mismatch between Web authors and end users. This implies Web designers to approximate user preferences with elicitation approaches, and to offer appropriate personalisation processes to adapt Web structures and contents according to user profiles and expectations. The problem to deal with is not to develop searching mechanisms such as well known keywords-based facilities, but rather to develop inference mechanisms that monitor user actions on the Web, infer user preference and personalize Web accesses. This requires to consider more deeply the way that spatial information, entities and relationships are embedded within Web pages, and the semantic exhibited by the information content of these Web pages [1], a step towards the semantic Web [2] [3].

In the context of Web personalization, many research proposals have been oriented towards the modeling and prediction of user's behaviors on the Web [4] [5]. In particular, Markov chains have been applied for the predictive modeling of contiguous visit sequences on the Web [6]. Our research is oriented to data personalization in the particular context of spatial information available on the Web. Spatially enhanced information retrieval and search engine on the Web recently attract wide research efforts [7] [8]. Nonetheless, and to the best of our knowledge, a few researches have been oriented to the modelling of user's interests and preference, and inference of personalized services when manipulating spatial information on the Web. Personalization services are expected to reduce information flows, adapt information delivery to user's needs and interests, thus improving his/her satisfaction.

K.-J. Li and C. Vangenot (Eds.): W2GIS 2005, LNCS 3833, pp. 206 – 221, 2005.

In a previous work, we introduced a personalization framework for the manipulation of spatial and semantic information personalization on the Web [9]. The proposed framework monitors user's manipulation of spatial entities and reference locations available on a given Web site, and recommends the most appropriate spatial entities given the user preferences the system is able to infer. User's interests and preference are derived from the assumption that the higher the number of closer spatial entities of similar interest to a given spatial entity of interest, the higher the value is given to this entity. The personalization process is supported by a competitive back propagation neural network that derives class preference patterns, and qualifies spatial entities of interest.

The research presented in this paper extends our previous work by integrating user's navigational trails within the user preference elicitation and personalization processes. We introduce a hybrid personalization approach and reinforcement process that facilitate user's navigations and interactions with spatial entities embedded in Web pages. The approach combines semantic similarity, spatial proximity, and k-order Markov chains to predict the next spatial entity which is likely to be in interaction with a given user. The semantic similarity reflects to which degree a spatial entity is close to another in the semantic domain, while the spatial proximity gives a contextual form of inverse distance between two spatial entities. Markov chains implicitly monitors and records user's trails on the Web, and derives navigational patterns and knowledge in order to predict user's interactions on the Web. A reinforcement process complements the approach by adapting the interactions between the user and the Web, that is, a sequence of iterative negative/positive rewards evaluated on the basis of user's relevance feedbacks to personalized presentations.

The remainder of this paper is organized as follows. Section 2 briefly describes the main principles of current personalization techniques and their applications on the Web. Section 3 introduces the modeling principles behind the presentation of spatial entities over the Web. Section 4 develops the hybrid Web personalization approach. Section 5 develops a reinforcement process to adapt transition probabilities. Section 6 illustrates the framework using an experimental personalization process. Section 7 gives some conclusive remarks and discusses some research perspectives.

2 Web Personalization

Personalization techniques on the Web attempt to overcome the information overload, to remove irrelevant information, and to increase utility and user's satisfaction by providing the user with effective services tailored to his/her specific needs. Among many personalization tools, recommender systems and Web personalization are amongst the most successful systems applied so far to improve information services and searching engines on the Web [10]. A personalization process can be categorised according to three main components, namely a *personalization goal*, a *user preference elicitation process* and a *personalization engine* (Figure 1).

A *personalization goal* is generally considered as a way of positively increasing the utility of the information delivered and user's satisfaction. A *user preference elicitation* over a given domain knowledge requires either observing user's choice behaviors, or directly interacting with the user with predefined questions. The range of techniques used varies from the implicit tracking of user actions to explicit user's feedbacks on the information provided. Evaluating user preference is often derived by explicit information such as direct user feedbacks, keywords based evaluation of user's interests, and implicit user feedbacks such as analysis of reading times, frequency of document downloads and page browsing [11] [12]. A p*ersonalization engine* supports personalizing activities, services and attitudes to a specific individual or group taking into account the personalization goal and user preference. Personalization engines usually correspond to a set of computational components that take user profiles and logs as inputs and as outputs a series of information items (e.g. Web pages or components) that might be of interest to the user.

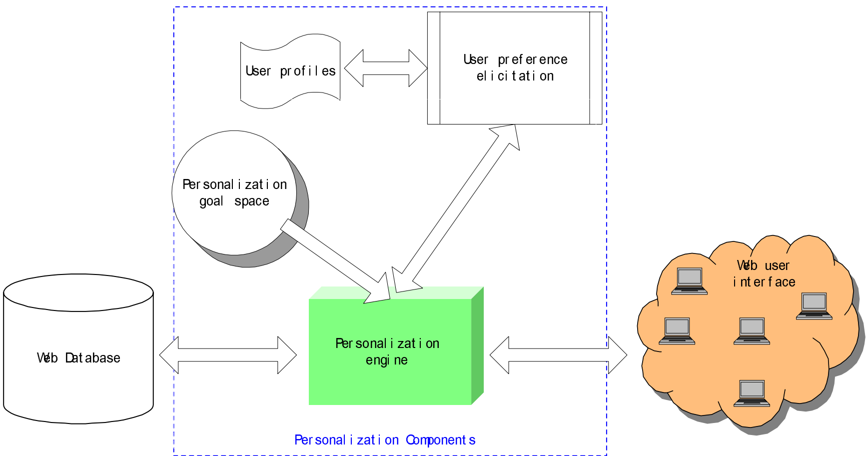


Fig. 1. Personalization components

Web personalization favours the presentation of Web information, content and structure according to user's explicit or implicit preferences. Two orthogonal filtering dimensions are often applied in Web personalization techniques: content-based filtering and collaborative filtering. Content-based filtering is oriented towards the personalization of Web pages and information retrieval processes, based on the analysis of Web document similarities and user's personal profiles. Collaborative filtering takes as inputs user's explicit (or implicit) ratings on items of interest and generates recommendations according to user preference similarities using proximity or correlation measures. However, and despite recent developments of Web personalisation techniques over conventional applications, there is still a need for capturing more complex relationships and patterns [13]. With the emergence of the semantic Web [2] [3], incorporating semantic knowledge and domain ontology into personalization processes should at least influence the next generation of Web personalization systems. Applied to the spatial domain, the semantic Web should include spatial ontologies, qualitative reasoning and representation of spatial knowledge on the Web [14].

3 Spatial Entities on the Web

A reasonable proportion of Web resources can be mapped to some degree to geo-referenced entities associated to a given location. Statistics collected by search engines and systems on the Web found that spatial information is pervasive on the Web, and that many queries contain spatial information [15]. However, current Web pages and interfaces are not completely adapted to the manipulation and personalisation of geo-referenced information. Modelling and identifying semantic and spatial relationships amongst entities embedded in Web documents is still a challenging task.

A spatial entity usually materializes a real world or virtual object. Different classes of spatial entities of interest (e.g. sightseeing places, hotels, universities) are commonly embedded in multi-media Web documents, either in textual or map forms. Maps favour the representation and display of spatially referenced data to the users, whereas other media forms such as text prevalent in Web documents serve as supplementary means to describe some semantic and spatial contents. Spatial entities represented on maps explicitly denote their locations and their geographical distribution. They can be linked to some semantic documents that can describe additional semantic properties. Maps on the Web are so far provided by either graphic files or interactive maps software. The latter provides an effective way to dynamically present spatial information on the Web. Interactive maps allow the user to have an access to various levels of manipulation and interaction, thanks to some customized user interfaces [16] [17]. For instance, the image of a spatial entity can be linked to an interactive map viewer, thus allowing the users to perform some map-oriented operations and hyperlink interactions (e.g. clicking on the map to view additional information on a spatial entity).

Let us represent a given set of spatial entities E explicitly embedded in a Web pages as $E = \{e_1, e_2, \ldots e_p\}$ where p is the cardinality of the set. Spatial entities materialised on the Web and associated by hyperlinks form a graph that relates them in the Web space. We assume that a spatial entity is likely to possess a semantic content and some spatial properties: $e = (spatial(e), semantic(e))$ where $spatial(e)$ denotes the spatial component and $semantic(e)$ the semantic component of the spatial entity e. The spatial component describes the location of a spatial entity as an abstract data type and as, $spatial(e) = (x, y)$ where (x, y) denotes the coordinates of the spatial entity e in a two dimensional space. The semantic component can be considered as an h-dimensional vector that specifies the semantic parameters of a given spatial entity: $semantic(e) = \{w(c_1, e), w(c_2, e), \ldots w(c_h, e)\}$ where h is an integer that denotes the number of semantic parameters; $w(c_i, e)$ gives the relevance of the spatial entity e associated to a semantic class c_i. A semantic class corresponds to a categorization of entities that share some semantic properties. Relevance values are membership values given by the unit interval [0,1], they qualify the membership degree of an entity with respect to a given semantic class C_i.

We do not consider the case, which is far beyond the objective of our research, of spatial entities textually and implicitly embedded on the Web as this leads to explore and develop Web data extraction and classification algorithms. Instead we consider situations where spatial entities are explicitly embedded within a Web page (Figure 3). User's navigational behaviors are recorded using historical Web logs. The basic element of such a log information is a page-view, that is, a "Visual rendering of a Web page in a specific client environment at a specific point in time" as stated by the W3C

Web Characterization Activity (http://www.w3.org/WCA/). At the user level, navigation is triggered by a session. A user session is a sequence of user page-views made during a single visit by a given user. At a finer level of granularity, a transaction denotes a meaningful subset of page-views within a user session. According to [18] [19] these notions are defined as follows:

Definition 1: Spatial Entity-Oriented User Session
A *spatial entity-oriented user session s* is a n-dimensional ordered vector $s = <e_1, e_2, \ldots e_n>$ that materializes a sequence of spatial entities accessed by a given user on the Web.

Definition 2: Spatial Entity-Oriented Transaction
A *spatial entity-oriented transaction t* is a m-dimensional vector $t = <w(e_1, s_1), w(e_2, s_2), \ldots w(e_m, s_m)>$ that materializes a semantically related subset of a user session, and where each spatial entity e_i is associated with a weight s_m, that denotes its semantic and spatial importance with respect to the user's interests.

4 Markov Chains Personalization

Markov chains are used extensively to predict the next state of a system given a sequence of previous states. A Markov chain can be represented by a tuple with three parameters $<S, T, \lambda>$. $S = \{s_1, s_2, \ldots, s_n\}$ corresponds to the state space, namely the set of all possible states for the Markov chain; T is a transition probability matrix, where each entry t_{ij} represents the transition probability from a_i to state a_j; λ corresponds to the initial distribution of the states in S.

The state space of a Markov chain depends on the number of sequences of previous states available to predict the next state. A 0-order Markov chain is an unconditional base-rate probability of x_n denoted as $p(x_n) = Pr(X_n)$. In a 0-order Markov chain, states are independent of each other. A first-order Markov chain only considers one-step transition probabilities $p(x_2|x_1) = Pr(X_2 = x_2|X_1 = x_1)$, that is, the probability of the next state given the immediately previous state. In a first-order Markov chain, each transition corresponds to a state. A k-order Markov chain considers the conditional probability by looking at the last k states to compute the predictions, $p(x_n|x_{n-1}, \ldots, x_{n-k}) = Pr(X_n = x_n|X_{n-1}, \ldots, X_{n-k})$. The state-space of a k-order Markov chain contains all possible sequences of k states.

The dimensionality of a Markov chain has a direct influence on the exhibited properties and performance of a given sequence. Lower-order Markov chains cannot successfully predict the next state of a given sequence because they don't consider far enough past states for a valid discrimination. High-order Markov chains result in high state space and low coverage, and sometimes even worse prediction accuracy due to the high number of sequential states (Deshpande and Karypis 2000). It has been observed, in an empirical analysis of data (Pitkow and Pirolli 1999), that using a 4th order Markov chain is an optimal option upon an assumption that the benefit of making a correct hit equals the cost of marking an incorrect prediction.

Sarukkai [20] introduced Markov chains for link prediction and path analysis to dynamically model URL access patterns, and to predict the next Web page accessed by the user. Padmanabhan and Mogul [4] used n-top Markov models to improve prefetching strategies for Web caches. Pitkow and Pirolli [21] explored the predictive capabilities of user paths and identified user access patterns on the Web. They introduced Longest Repeating Subsequence (LRS) models to predict World Wide Web surfing. LRS models reduced predictive model size and complexity by nearly a third while retaining predictive accuracy. In order to improve prediction accuracy, and at the same time to maintain a low state complexity, Deshpande and Karypis [22] proposed a class of Markov models based on some prediction algorithms called *selective Markov models*. This Markov model is obtained by selectively eliminating a large fraction of the states of the All-Kth-Order Markov model. Empirical results show that the performance of selective Markov models is superior to that obtained by higher-order Markov models to predict Web accesses.

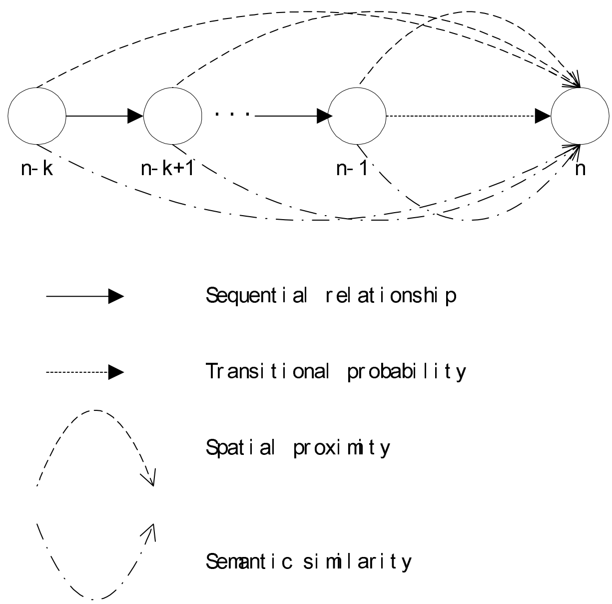


Fig. 2. Spatial entities relationships within k-order Markov chains

Web personalization based on Markov chains predicts the next Web page a given user is most likely to visit by matching the user's current access sequence with historical Web access patterns. The entities extracted and identified from various Web documents constitute the state space. A state is defined as a Web entity, while a transition denotes a hyperlink from one entity to another. Markov chains use a sequence of Web page-views/entities the user accesses as inputs, with the goal of building Markov chains to predict the page-view/entity the user is most likely visit next. The predictive process corresponds to matching user's current navigation trails with historical Web access sequences, and determination of the next visit with transitional probability.

We retained a hybrid Web personalization approach that integrates k-order Markov chains with a combination of semantic similarity and spatial proximity criteria (denoted as *SemSpa similarity*). The intuition behind is as follows

> Given a sequence of k previously visited spatial entities $<x_{n-k},\ldots x_{n-1}>$ on the Web, with consideration of conditional transition probability from $<x_{n-k},\ldots x_{n-1}>$ to x_n, and *SemSpa* similarity between $<x_{n-k},\ldots x_{n-1}>$ and x_n, the Web personalization engine predicts the n_{th} spatial entity which is likely to be visited by the user (Figure 2).

Figure 2 illustrates the principles of a k-order Markov chain that reflects a sequence of spatial entities of interest on the Web. Content-based relationships between the spatial entities are evaluated by a combination of spatial proximities and semantic similarities. This is supported by a stepwise approach that measures semantic and spatial similarities between a sequence of spatial entities on the one hand, and a candidate entity on the other hand. The initialisation of the predictive Markov chain is decomposed into three steps as follows:

1) Determine the *SemSpa* similarity between each spatial entity in a k-sequence of user interactions with spatial entities embedded on the Web, and a candidate spatial entity embedded or connected by a hyperlink to the last page accessed in this k-sequence.
2) Similarly, compute the *kSemSpa* similarity between the spatial entities in a given k-sequence and a candidate spatial entity.
3) Compute the *kSemSpaM* similarity, a combination of *kSemSpa* similarity and transitional probability in a given k-order Markov chain.

1) Semantic and Spatial Similarity Between Two Spatial Entities

The first step of this approach is to determine the *SemSpa* similarity. This is given by a combination of semantic similarity and spatial proximity values between a spatial entity in a k-sequence x_i, and a candidate spatial entity x_j. The *SemSpa* similarity is given as follows

$$SemSpa(x_i, x_j) = \sqrt{(1 + Sem(x_i, x_j)) \times Spa(x_i, x_j)} \tag{1}$$

where $Sem(x_i, x_j)$ denotes the semantic similarity, and $Spa(x_i, x_j)$ the spatial proximity between x_i and x_j.

The most common approach used to compute a degree of similarity between two vectors in Web information retrieval is the standard cosine similarity. Accordingly, and in order to deal with membership degrees of spatial entities, we applied the adjusted cosine similarity introduced by [23] to determine the semantic similarity between two entities x_i and x_j. The semantic function $Sem(x_i, x_j)$ is given as

$$Sem(x_i, x_j) = \frac{\sum_{h=1}^{m}(x_i^h - \overline{x^h}) \times (x_j^h - \overline{x^h})}{\sqrt{\sum_{h=1}^{m}(x_i^h - \overline{x^h})^2 \times \sum_{h=1}^{m}(x_j^h - \overline{x^h})^2}} \tag{2}$$

The domain of the semantic function is given by the unit interval [0, 1]. A semantic parameter x_i^h reflects the membership degree of a spatial entity x_i with respect to a predefined semantic class c_h. The m semantic classes are defined upon the requirement of the application. Membership degree values are given by fuzzy quantifiers bounded by the unit interval [0,1]. $\overline{x^h}$ is the average value of the semantic parameters of the spatial entities with respect to a given semantic class c_h. The value domain of $Sem(x_i, x_j)$ is given by the interval [-1, 1].

The spatial function $Spa(x_i, x_j)$ is specified by the contextual proximity $p(x_i,x_j)$ introduced in [9]. The contextual proximity denotes the closeness of two spatial entities taking into account the overall distribution and context of the underlying spatial structure. The higher $p(x_i,x_j)$ the closer x_i to x_j, the lower $p(x_i,x_j)$ the distant x_i to x_j, and *vice versa*. The spatial function is given as

$$Spa(x_i, x_j) = P(x_i, x_j) \tag{3}$$

The contextual proximity is an asymmetric value, it is bounded by the unit interval [0,1], and given as

$$P(x_i,x_j) = \frac{1}{1 + RD(x_i,x_j)^2} \tag{4}$$

where x_i is the entity under consideration, $RD(x_i,x_j)$ is the relative distance from x_i to x_j.

The relative distance is given as

$$RD(x_i,x_m) = \frac{d(x_i,x_m)}{\frac{1}{p-1}\sum_{m=1,m\neq i}^{p} d(x_i,x_m)} \tag{5}$$

where $d(x_i,x_j)$ stands for the Euclidean distance between x_i and x_j, p is the number of entities in the set $X=\{x_1, x_2, \dots x_p\}$.

2) Semantic and Spatial Similarity Between Spatial Entities in k-Sequence and a Candidate Spatial Entity

The second step of the hybrid approach is to determine the *SemSpa* similarity between a sequence of k spatial entities $<x_{n-k},\dots x_{n-1}>$, and a candidate spatial entity x_n. As Web surfers may have two or more kinds of interests in mind, their interests may change from time to time when they are browsing on the Web. The latter is so-called “concept drift” [24], an issue beyond the scope of traditional Web personalization applications. In order to address the “concept drift” issue, we introduce a discount rate γ to adapt semantic similarity and spatial proximity values, on the basis of an assumption that following the user's navigational trails, "nearer" entities are more related than distant ones. The *kSemSpa* similarity is computed using a discount rate γ that refines the *SemSpa* similarity value between each spatial entity embedded in a sequence $<x_{n-k},\ldots x_{n-1}>$ and x_n. The *kSemSpa* similarity is given as follows

$$kSemSpa(< x_{n-k},\ldots,x_{n-1}>,x_n) = \sum_{k=1} \gamma^{k-1} SemSpa(x_{n-k},x_n) \tag{6}$$

where γ is a discount rate parameter, $0 \leq \gamma \leq 1$, that recursively decreases the *SemSpa* similarity between each spatial entities in a historical sequence $<x_{n-k},\ldots x_{n-1}>$ and x_n with the historical length. The historical length denotes the number of steps from a given spatial entity x_i in a sequence $<x_{n-k},\ldots x_{n-1}>$ to a spatial entity x_n.

3) Combination of k_SemSpa Similarity and Transitional Probability

The third step of the hybrid approach is to combine the *kSemSpa* similarity with the transitional probability exhibited by the k-order Markov chain $<x_{n-k},\ldots x_{n-1}>$ in order to predict the n_{th} candidate spatial entity. The *kSemSpaM* value is given as

$$\begin{aligned} &kSemSpaM(<x_{n-k},\ldots,x_{n-1}>,x_n) = \sqrt{Pr \times kSemSpa} \\ &\text{with } Pr = p(x_n \mid x_{n-1},\ldots,x_{n-k}), \\ &kSempa = kSemSpa(<x_{n-k},\ldots,x_{n-1}>,x_n) \end{aligned} \tag{7}$$

where $p(x_n \mid x_{n-1},\ldots,x_{n-k})$ is the transitional probability of the k-order Markov chain, statistically collected from the Web logs that record user's previous behaviors on the Web.

5 Reinforcement Process

Interactions between the user and the personalization engine form an iterative process when he/she is surfing on the Web. These interactions consist of two kinds of process. First, the Web system provides personalized information services according to his/her interests. Secondly, the user gives some relevance feedbacks through various behaviors. User's relevance feedbacks reflect his/her attitudes to the personalized results. We take into account this component to adjust the transitional probability of the k-order Markov chains $p(x_n \mid x_{n-1},\ldots,x_{n-k})$, using a reinforcement process. The reinforcement process is a learning process based on the observation of user's feedbacks to the predictive result provided (i.e., the n_{th} spatial entity). Possible forms of user's feedbacks to the n_{th} state is valued by two alternative Boolean values *satisfied* and *unsatisfied*. In the former situation, the user is likely to visit the recommended n_{th} spatial entity; while in the latter, the user will follow some other hyperlinks. The reinforcement process gives either positive or negative rewards. The reinforcement process is given as

$$\mathrm{Pr} \leftarrow \mathrm{Pr} \pm \eta \times r \tag{8}$$

where η is the learning rate, r the reinforcement reward.

The value of η may be slightly smaller than 1 when learning begins, and then slowly decreased to 0 as learning progresses. A simple approach is to make use of τ, the number of user accesses, to adjust these values dynamically. Then $\eta = \frac{\alpha}{\tau}$, α is the reinforcement factors with 0< α <1, and is initialised once. The reinforcement reward r is given as

$$\begin{aligned} &\textit{(for positive reinforcement)} \\ &r = (1 - Pr) \times (1 - kSemSpaM) \\ &\textit{(for negative reinforcement)} \\ &r = (1 - Pr) \times kSemSpaM \end{aligned} \tag{9}$$

Then the reinforcement process is given as

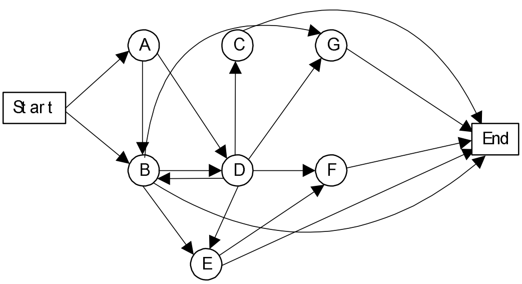
$$\begin{aligned} &\textit{(for positive reinforcement)} \\ &Pr \leftarrow Pr + \frac{\alpha}{\tau}(1 - Pr) \times (1 - kSemSpaM) \\ &\textit{(for negative reinforcement)} \\ &Pr \leftarrow Pr - \frac{\alpha}{\tau}(1 - Pr) \times kSemSpaM \end{aligned} \tag{10}$$

The transactional probability from a Web entity/page-view to the next in sequential Web data mining is statistically calculated from Web logs that record user's historical navigation trails. The reinforcement process introduced is dynamic and provides a mechanism for conditional transition probability according to user's feedbacks to the personalized results. Probabilities of transitions (*Pr*) with a few hits converge toward a null value, while those with many hits converge toward the unit value.

6 Case Study

In a previous work, we introduce a prototype experiment applied to the city of Kyoto [9]. In order to apply the concepts introduced in this paper, we extend this prototype towards the personalization of navigational trails over spatial entities embedded on the Web. Historical and sightseeing places are modeled as spatial entities. The Web system encodes the sightseeing places as spatial entities to support Web-based travel planning for the user. The main interface is enriched with image schemata and affordance concepts (Figure 3), based on the assumption that the user have little knowledge on the given city [9]. These spatial entities contain semantic attributes, and are distributed in space (Figure 8). Each spatial entity embedded on the Web map interface is presented by a symbol or an image, and associated to additional textual information that allows the user to actively interact with the Web interface. The personalization results are presented to the user in various formats and an interactive map with hyperlinks to Web resources of interest.

Fig. 3. The main interface



Transactions	Frequency
A, B, D, C	10
B, D, B, E	10
A, B, D, F	8
A, B, D, B	4
A, D, B	8
B, D, B, C	8
B, D, E, F	4
A, B, D, B, G	2

Fig. 4. Navigation trails and transactions

Without loss of generality, let us consider a set of Web transactions as presented in Figure 4 (right). The set of Web transactions records user's navigational trails involving seven spatial entities that represent some historical and cultural interests. These entities include A (Kitano Temmangu), B (Nijo), C (Higashi Honnganji), D (Yuzen Textile), E (Arashiyama), F (Costume Museum) and G (Nishijin), represented as a labeled directed graph (the left of figure 4). We use a third-order Markov chain to model these transactions. Through some appropriate Web usage mining processes, the set of user's transactions is transformed to a set of transactions represented as a 3-order Markov chains (Figure 5). The left part of these transactions forms the state space, while the transitional probabilities constitute the transitional probability matrix.

Transactions	Frenquency	Transitional probability
A, B, D --> C	10	5/12
A, B, D --> F	8	1/3
A, B, D --> B	6	1/4
B, D, B --> E	10	5/12
B, D, B --> C	8	1/3
B, D, B --> G	2	1/12
B, D, B --> END	4	1/6
D, E, F --> END	8	1.0
B, D, E --> F	4	1.0
D, B, C --> END	8	1.0
A, D, B --> END	8	1.0
B, D, F --> END	8	1.0
D, B, E --> END	10	1.0
B, D, C --> END	10	1.0
D, B, G --> END	2	1.0

Fig. 5. Transaction, frequency and transitional probability for 3-order Markov chains

Suppose a specific user is browsing entity D after visiting A and B successively. The Web personalization component takes the transaction $A \rightarrow B \rightarrow D$, and the semantic and spatial criteria into account to predict users next visits. Possible candidate entities are C, F, B, and transitional probabilities from $A \rightarrow B \rightarrow D$ are 5/12, 1/3, 1/4 respectively. Computed results of *kSemSpa* are presented in the results exhibited by Figure 6 with γ =2:

Fig. 6. Computed results of *kSemSpa*

Computed results of *kSemSpaM* are (Figure 7)

Fig. 7. Computed results of *kSemSpaM*

The Markov chain evaluation recommends the spatial entity which is most likely to be "visited", the entity F in the example above (Costume Museum). Personalization results are illustrated in Figure 8. The recommendation reflects the fact that most previous users with similar trails visit entity F, and F has relative strong relationship

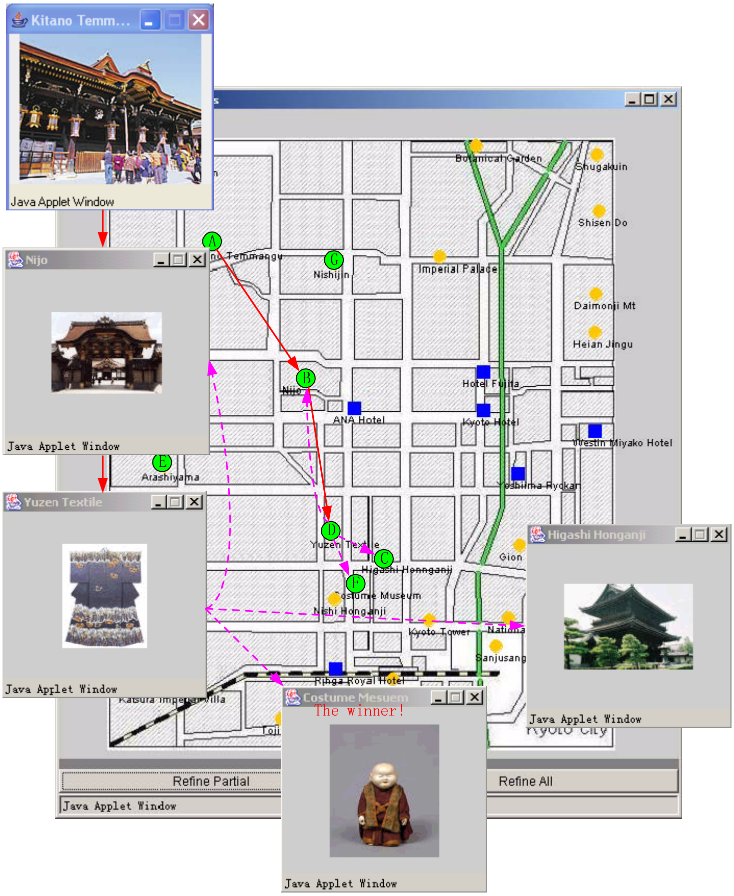


Fig. 8. Personalization results visualization

with previous three entities accessed by the user. Consequently the transitional probability from the sequence <A, B, D> to F, that is, $p(F \mid A, B, D)$ is positively reinforced if the user follows the personalized result, which shows user's feedback as *satisfied*; otherwise, as an example, if the user goes to visit the entity C, that is, in the case of *unsatisfied* feedback, then $p(F \mid A, B, D)$ is negatively reinforced and $p(C \mid A, B, D)$ positively.

The hybrid Web personalization approach uses navigational knowledge and profiles extracted from the users' historical trails on the Web, and then ameliorates them with spatial proximity and semantic similarity responsible for content-based filtering of spatial information entities. This avoids the defects of using each of the two individually. The reinforcement process is used to update the navigational knowledge through unobtrusively observing user's implicit feedbacks.

Whether a user is physically located and acts or not in the city represented on the Web has an influence on the levels of perception and interaction in the environment. The former preferably refers to mobile environments where the user interacts with the environments through portable and embeddable devices. The latter denotes conventional interactions between the user and a Web information space using a client "table" computer. Navigations on the Web constitute sequences of Web pages that can be recorded by Web logs. The main application scenarios we consider at the moment to apply the framework include Web-based travel planning and Location-aware mobile services. Our Current developments are oriented towards the first scenario, that is, an attempt to provide a prototype framework for Web-based travel planning in a "Web urban space". Based on this framework, inference rules can be modelled to identify user's interests and preference, and then personalize user's travel and experience when interacting with spatial entities embedded on the Web. Without loss of generality, the framework can be also applied to location-aware mobile environments with some minor adaptations.

7 Conclusion

The research presented in this paper introduces a hybrid Web personalization approach that combines Markov chains with spatial and semantic similarity, and a reinforcement process in order to model and predict user navigational trails when interacting with spatial information materialized on the Web. The personalization process is based on an integration of two orthogonal dimensions that facilitate the approximation of user preferences, that is, semantic similarities and spatial proximities between spatial entities embedded in Web pages. Markov chains integrated with these relationships allow the system to recommend spatial entities that are of interest for the user. A reinforcement process, based on the monitoring of user actions and relevance feedbacks, complements the personalization mechanisms by employing a learning component. The potential of the approach is illustrated by an exemplified case study. The approach and the underlying modelling concepts can have a several direct or indirect impacts on the benefits of a given Web site. Direct as deriving preference patterns might help Web designers to refine the design of their Web interfaces and tailor the content of Web pages, indirect as categorising users might also have several "marketing" implications for the Web page owners. Further work concerns experimental validation of the personalisation model and reinforcement process, and development of knowledge-based mechanisms that implicitly derive class preferences with flexible learning process, and closer integration of the Markov model and the reinforcement learning process.

Acknowledgement

We would like to thank the three anonymous reviewers for their valuable comments.

References

1. Dai, H. and Mobasher, B. Using ontologies to discover domain-level Web usage profiles. In 2nd Semantic Web Mining Workshop at ECML/PKDD02 (2002)
2. Semantic Web Activity of the World Wide Web Consortium. http://www.w3.org/2001/sw

3. Berners-Lee, J., Hendler, J. and Lassila, O. The Semantic Web, Scientific American, vol. 184, no. 5, (2001) 34-43,
4. Padmanabhan, V. and Mogul, J. Using Predictive Prefetching to Improve World Wide Web Latency. In Proc. of the ACM SIGCOMM '96 Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communication, (1996) 22-36
5. Mobasher, B., Cooley R. and Srivastava, J. Automatic personalization based on Web usage mining, Communications of the ACM 43(8), (2000) 142–151
6. Pirolli, P. L. and Pitkow J. E. Distribution of surfers' paths through the World Wide Web: empirical characterization. World Wide Web. 2 (1-2), (1999) 29-45
7. Jones, C. B., Purves, R., Ruas, A., Sanderson, M., Sester, van Kreveld, M. and Weibel, R. Spatial information retrieval and geographical ontologies an overview of the SPIRIT project. SIGIR 2002, 387-388
8. Larson R.R. and Frontiera, P. Ranking and Representation for Geographic Information Retrieval. Presented at SIGIR 2004 Workshop on Geographic Information Retrieval. Sheffield, UK, July 29, (2004)
9. Yang, Y. and Claramunt, C. A Flexible Competitive Neural Network for Eliciting User's Preferences in Web Urban Spaces, in Fisher P (eds.), Developments in Spatial Data Handling, Proceedings of the 11th International Spatial Data Handling Conference, Springer-Verlag, 23-25 August, University of Leicester, (2004) 41-57
10. Shahabi, C. and Chen, Y. Web Information Personalization: Challenges and Approaches, The 3nd International Workshop on Databases in Networked Information Systems (DNIS 2003), Aizu-Wakamatsu, Japan, September (2003) 5-15
11. Oard, D.W. and Kim, J. Modelling Information Content Using Observable Behaviour. In Proceedings of the 64 Annual Meeting of the American Society for Information Science and Technology, USA, (2001) 38-45.
12. Kelly, D. and Teevan, J. Implicit Feedback for Inferring User Preference: a Bibliography. SIGIR Forum 37(2), (2003) 18-28
13. Berendt, B., Hotho, A. and Stumme, G. Towards Semantic Web Mining, In I. Horrocks & J. Hendler (eds.), The Semantic Web - ISWC 2002 (Proceedings of the 1st International Semantic Web Conference, June 9-12th, 2002, Sardinia, Italy). LNCS 2342, Heidelberg, Germany: Springer (2002) 264-278
14. Egenhofer, M.J. Toward the Semantic Geospatial Web, In Proceedings of the Tenth ACM International Symposium on Advances in Geographic Information Systems, McLean, Virginia (2002)
15. Silva, M. J. Martins, B., Chaves, M. and Cardoso N. Adding Geographic Scopes to Web Resources, Ana Paula Afonso ACM SIGIR 2004 Workshop on Geographic Information Retrieval, Sheffield, UK, June (2004)
16. Open GIS Consortium, Inc. (OGC). Open GIS Web Map Server Interface Implementation Specification (Revision 1.3.0). Wayland, Massachusetts: Open GIS Consortium, Inc. http://www.opengeospatial.org/specs/ (2004)
17. Open GIS Consortium, Inc. (OGC). Open GIS Web Feature Service Implementation Specification (Revision 1.1.0). Wayland, Massachusetts: Open GIS Consortium, Inc. http://www.opengeospatial.org/specs/ (2005)
18. Cooley, R., Mobasher, B. and Srivastava, J. Data Preparation for Mining World Wide Web Browsing Patterns, Knowledge and Information Systems, Vol. 1, No. 1, (1999) 5-32
19. Mobasher, B. Web usage mining and personalization. Chapter in Practical Handbook of Internet Computing, Munindar P. Singh (ed.), CRC Press (2004)
20. Sarukkai, R. Link Prediction and Path Analysis Using Markov Chains, Computer Networks, vol. 33, n. 1-6, (2000) 377-386

21. Pitkow J. and Pirolli, P. Mining longest repeating subsequence to predict wolrd wide Web surfing. In Second USENIX Symposium on Internet Technologies and Systems, Boulder, C0 (1999)
22. Deshpande, M. and Karypis G. Selective Markov models for predicting Web page accesses, ACM Trans. Internet Techn. 4(2), (2004) 163-184
23. Sarwar, B., Karypis, G., Konstan, J. and Riedl, J. Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th International WWW Conference, Hong Kong (2001)
24. Webb, G., Pazzani, M.J. and Billsus, D. Machine Learning for User Modelling. User Modelling and User–Adapted Interaction, (11), (2001) 19–29