A new approach for service discovery and prediction on Pervasive Information System

Salma Najar*, Manuele Kirsch Pinheiro, Carine Souveyet

Centre de Recherche en Informatique / Université Paris 1 – Panthéon Sorbonne, 90 Rue Tolbiac, Paris 75013, France

Abstract

Recent evolution of technology transformed the way we interact with Information Systems (IS), leading to a new generation of IS, the Pervasive Information Systems (PIS). These systems have to face heterogeneous pervasive environments, whose complexity they must hide from end-user. In order to reach transparency and proactivity necessary for successful PIS, new discovery and prediction mechanisms are necessary. In this paper, we propose a new user-centric approach for service discovery and prediction on PIS based on both user’s context and intentions. Intentions allow focusing on goals user wants to satisfy when requesting a service. Those intentions rise in a given context, which may condition the service implementation. We propose then a service discovery mechanism that observes user’s context and intention in order to offer him the service that may best satisfy her/his intention on the current context. We also propose a prediction mechanism that tries to anticipate user’s intentions considering the observed context and user’s history.

Keywords: pervasive information system; service oriented computing; intention; context-aware computing; service discovery; service prediction

1. Introduction

New technologies transform the way we interact with IS and the services they offer, expanding the frontiers of IS outside the companies’ environment. The BYOD (Bring Your Own Device) illustrates this tendency: employees bring their own devices to the office and keep using them to access the IS even when they are on the move. The
consequence of such evolution is that IS now have to cope with a pervasive environment and may integrate services from very different natures. A new generation of IS is then rising, the Pervasive Information Systems (PIS).

Pervasive Information Systems intend to increase user’s productivity by making IS services available anytime and anywhere. These systems changed the interaction paradigm from desktop computing to new technologies. They evolved from a fully controlled environment (the office) to a dynamic pervasive one\(^1\) Contrary to traditional IS, PIS have to support a multitude of heterogeneous device and service types, challenging its design. We argue that PIS should be designed for helping user to better satisfy her/his needs according to her/his environment. PIS must not only consider the goals it must reach as an IS, but also handle pervasive environment heterogeneity. Indeed, it should hide this heterogeneity from the user, allowing her/him to concentrate on her/his needs, instead of on the technology itself. Transparency and proactivity become then key aspects on PIS, which requires offering user appropriate services considering her/his goals and the context in which such goals appear, as well as the capability of anticipating future goals in this context. New service discovery and prediction mechanisms are then necessary.

We propose a new user-centric approach for service discovery and prediction considering PIS. This approach is based on both user’s intentions, representing the goals she/he wants to achieve without saying how to perform it\(^2\), and on the context in which these intentions have been formulated. The notion of context can be seen as any information that can be used to characterize the situation of an entity\(^3\). We consider that context information can influence service execution and, consequently, what service can be chosen to satisfy a given intention. In our opinion, both concepts should be considered during service discovery, since the main purpose is providing user with a service that can fulfill his goals in a fairly understandable and non-intrusive way. Thus, we propose a new service discovery mechanism that intends discovering the available service that can satisfy the immediate user’s intention in the current context. Based on the discovery results, we propose a new prediction mechanism that identifies common situations representing usage patterns, i.e., recurrent context and intentions observed during PIS usage. By analyzing these patterns, the prediction mechanism learns user’s behavior when using a PIS, and therefore anticipates future intentions and the most appropriate services that may satisfy it.

This paper is organized as follows: Section 2 presents related works on service discovery and prediction. Section 3 introduces the proposed service discovery mechanism, while Section 4 presents the service prediction mechanism. Section 5 presents an evaluation of the proposed mechanism, before concluding in Section 6.

2. Related works

During the last decade, a lot of research has been conducted concerning pervasive systems, mainly on context-aware services\(^4,5,6\). Context-awareness becomes a necessary feature for providing adaptable services, for instance when selecting the best-suited service according to the relevant context information or when adapting the service during its execution according to context changes\(^7\). Different service discovery mechanisms have been proposed in the literature\(^5,6,8\). Most of them consider context information as a non-functional aspect of service\(^5,8\), or as a condition for service selection and execution\(^5\). On both cases, a matchmaking, using semantic matching\(^5\) or similarity measures\(^5\), is performed between context information related to the service and the one related to user or execution environment. Nevertheless, only a few research works\(^9,10,11\) consider the notion of intention during the service discovery. Intentions can be associated with service descriptions as a set of capabilities, with their pre- and post-conditions\(^9\). They also can be used as a guide for service discovery, by a refining process in which intentions are decomposed on low-level intentions\(^10\). Unfortunately, the influence of context on the intention satisfaction is merely considered on the literature, context being often seen as a simple input on intention-based mechanisms\(^11\).

A similar situation can be observed when considering service prediction. Even if several works have considered context prediction\(^12,13\) or service recommendation based on context information\(^14,15\), at the best of our knowledge, none has proposed combining intention aspect with context information. In the one hand, several works propose anticipating context information based on historical data\(^12,16\) or pattern matching\(^13\). In the other hand, service prediction works proactively propose services based on user’s historical context information. Most of these works\(^1,4,15,17\) consider the correlation between context information and an item (e.g. a service) using different filtering techniques\(^17\), sometimes correlated with ontology-based matching\(^15\). Unfortunately, the notion of intention, representing concrete user’s goals, remains unexplored by these works.
3. Service Discovery Mechanism

In our user-centric approach, we propose a service discovery mechanism guided by user’s intention and context. Its objective is to hide implementation complexity, and consequently to achieve the transparency promised by PIS. This service discovery selects the most appropriate service for the user, i.e., the service that satisfies his/her immediate intention in a given context. It is based on a semantic service description and on a semantic matching algorithm. The goal of this algorithm is to rank the available services based on their contextual and intentional information. It compares semantically the user’s intention with the intention that the service satisfies and user’s current context with the service’s context conditions. Then, the service having the highest matching score is selected. It represents the most appropriate service that satisfies user’s immediate intention in his current context.

The semantic matching algorithm, as illustrated in Fig. 1, is a two-step process: intention matching and context matching. In the first step, the intention matching is based on the use of ontologies and a semantic matching. Indeed, an intention can be represented as a n-uplet by a verb \( (v_l) \) and a target \( (t_l) \) representing user’s goals. In this step, we propose to compare the user’s required intention \( I_v = <v_l, t_l> \) and the service’s intention \( I_{svi} = <v_{svi}, T_{svi}> \) in two separated matching process. For the verb matching, we use a verb ontology (OntoV), which contains a domain-specific set of verbs, representing significant actions authorized by the PIS, with their different meanings and relations. The degree of similarity is calculated based on the distance between these verbs in the verb ontology: \( 1/L + 1 \), where L represents the number of links between two concepts in the ontology. We define five levels of similarity, inspired from \(^{19}\), as illustrated in Table 1 (more details about the intention matching are presented in \(^{20}\)).

<table>
<thead>
<tr>
<th>Matching Relation</th>
<th>Explanation</th>
<th>Link</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact</td>
<td>Required verb is equivalent to the provided verb</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Synonym</td>
<td>Required verb share a common signification with the provided verb</td>
<td>-</td>
<td>0,9</td>
</tr>
<tr>
<td>Hyponym</td>
<td>Required verb is more specific than the provided one</td>
<td>L</td>
<td>1/(L+1)</td>
</tr>
<tr>
<td>Hypernym</td>
<td>Required verb is more generic than the provided one</td>
<td>L</td>
<td>1/(L+1)</td>
</tr>
<tr>
<td>Fail</td>
<td>No relation between the two verbs</td>
<td>-1</td>
<td>0</td>
</tr>
</tbody>
</table>

Similarly, for the target matching, we use a domain-specific ontology, namely target ontology (OntoT). This ontology represents the possible targets that are made available through the PIS. We compare the required target \( T_v \) and the provided target \( T_{svi} \) using a degree of similarity also based on the distance between these concepts in the ontology. This semantic similarity, based on the works of \(^{19}\), uses four levels: exact, plugin, subsume and fails. The plugin relation is similar to the hyponym relation in the verb matching, while the subsume relation is similar to the hypernym relation.

In the second step, the context matching is based on a context ontology (OntoC) and a set of similarity measures. It matches individually the different context elements constituting the user \((Cx_u)\) and service context descriptions \((Cx_{svi})\). Indeed, context information is often semantically represented using ontologies in which context information is structured (e.g. \(^{5,8,16}\)). In our case, context information is represented as context elements (location, available memory, user’s expertise, etc.) that are observed from a given subject (a user, a device, etc.). Both context elements and subjects are semantically described using ontologies. Thus, the context description for a user \((Cx_u)\) represents a set of context observations \( Cx_u = \{c_x | j > 0 \} \) associated to this user and the context description for a service \((Cx_{svi})\) represents a set of context conditions \( Cx_{svi} = \{c_x | i > 0 \} \) expressed over context elements. The former
corresponds to the current observed user’s context, while the latter represents contextual conditions for which service was designed (i.e. context under which service is able to better satisfy its intention).

The context matching score \( C_{\text{score}} \) is calculated as the sum of the scores of each context condition, as follows:

\[
C_{\text{score}} = \sum_{i=1}^{n} (w \times \text{ContextConditionMatching}(cx_i, cx_j))
\]

Thus, in order to define the relation ContextMatching, we consider the relation ContextConditionMatching that matches individually the user’s context observations \( (Cx_U) \) and the service’s context conditions \( (Cx_R) \). The context matching proceeds as follows: for each condition \( cx_i \) and observation \( cx_j \), we first match the corresponding subjects, using the context ontology. This match is calculated, like the verb and target matching, based on the distance between both concepts in the ontology. If the matching score between them is higher than a given threshold then we match their context elements. This last matching takes also into account the weight assigned to it. Then, if the matching score between them is higher than a given threshold, only at this moment we evaluate the satisfaction of the context condition regarding to the user’s context observations value one by one. More details about the context matching are presented in 20.

The observable context elements can be divided into several types. Context information values are distinguished between numerical or non-numerical types. In order to take into account this diversity, the relation ContextConditionMatching identifies the nature of the context element value and accordingly triggers the suitable measure to compare them. This relation evaluates if the user’s context element satisfies the service’s context element conditions, based on a specific operator (equal, not-equal, between, higher-than, etc.). For example, we have as a service’s context condition that the device bandwidth must be higher than 12500. From the user’s current context, if the captured value of the user’s device bandwidth is really higher than 12500, then we return an exact match.

Besides, the weight \( (w) \) that the user allocates to each context attribute and whose value is between 0 and 1, represents the importance of an attribute to a given entity. The purpose here is to highlight the real importance of a context attribute according to user’s preferences, and the importance of the attribute is proportional to its weight.

4. Service Prediction Mechanism

We propose also in this approach to predict the user’s future intention. This approach recommends proactively a service that can fulfill user’s future needs. It is based on the assumption that common situations (\( S \)) can be detected, even in a dynamic PIS. We define the notion of situation (\( S \)) as the user’s intention (\( I_U \)), in a given context (\( Cx_U \)), satisfied by a specific service (\( S_v \)) resulting from a previous discovery process: \( S = < I_U, Cx_U, S_v > \). These situations are time-stamped and stored in a database after each service discovery process (history). Let the user’s history \( H \) be defined as a set of all the observed situations \( S \) ordered according to their time of occurrence. Thus, by analyzing the history (\( H \)), the prediction mechanism can learn the user’s behavior model (\( Mc \)) in a dynamic environment, and thus deduce its coming immediately intention.

Two main processes compose this service prediction mechanism: the learning process and the prediction process, as illustrated in Fig. 2. Thus, to realize anticipatory and proactive behavior of PIS, we need first to dynamically learn about the user and his behavior in a frequently changing environment. This represents the learning process where similar situations (\( S \)) are grouped into clusters, during the phase of clustering. In the next
step, these clusters are interpreted as states of a state machine. This phase is called classification. It aims to represent, from the recognized clusters, the user’s behavior model (Mc) based on his situations (S). By interpreting situation changes as a trajectory of states, we can anticipate his future needs. Therefore, the prediction process is based on the user’s behavior model (Mc), on the current user’s intention (Ic) and on the current user’s context (Cc).

More specifically, the main task of the clustering phase, as illustrated in Fig. 3.a, is to detect, for a given situation, the closest set of situations corresponding to highly similar intentions in quite similar context. This provides us a powerful mechanism to evaluate the user’s intention. Indeed, a user can express the same intention in a slightly different way by using verbs and targets that are semantically similar enough. Based on verb and target ontologies, we perform a semantic matching between two intentions in order to determine their degree of similarity.

The same applies to context information, since an intention may rise on similar contexts. More precisely, the input of this clustering phase represents vectors representing user’s situations stored in the history. The main role of clustering is to detect recurrent situations among those previously observed and grouped in a cluster. A cluster consists of a centroid and a set of situations. The centroid represents the identifier of the cluster, which symbolizes the situation the most similar to all the situations grouped in this cluster. It is defined by the triplet < Ic; Cc; S0 >. After a clustering phase, the corresponding cluster identifier is attached to each new situations stored in the history.

In order to represent a situation, we attach a particular service to the couple <Intention, Context>. We believe that this represents a strong constraint (the concept situation is necessarily coupled to a particular service), but it opens a significant performance advantage, since it is not required to launch the service discovery mechanism during the prediction process. Thus, it is important to regularly update the clusters in order to have the service that best meets the couple intention and context of the situation.

Then, from these recognized clusters and the user’s history, the classification phase determines and maintains a user’s behavior model, as illustrated in Fig. 3.b. This model represents the user’s behavior as a set of states with a transition probability. Each state is represented by the centroid of the recognized cluster. Each probability is calculated based on the history and determines the probability of moving from one state to another.

Several classification techniques exist. Among these techniques, the Markov chain represents one of the well-known classification algorithms that can be used in a PIS. It represents a method for representing a stochastic process in discrete time with discrete state space. We represent the Markov chains model (Mc) as the doublet: \(Mc = (S_t, p)\), with \(S_t\) representing the different states and \(p \in [0,1]\) the probability of transition from one state to another.

At a given time \(t\), the user is in a state \(S_t\). In PIS, the user’s intention and/or his context may change. Therefore, the user moves from the state \(S_t\) to \(S_{t+1}\). The state \(S_{t+1}\) represents the successor of \(S_t\) with a certain probability \(p\). This transition probability represents the ratio of the transition from \(S_t\) to \(S_{t+1}\) divided by the number of all the possible transitions from \(S_t\). This probability is represented as follow:

\[
p_{S_{t}S_{t+1}} = P(X_{t+1} = S_{t+1} | X_t = S_t) = \frac{N_{S_{t}S_{t+1}}}{N_{S_{t}S_{t+1}}}
\]

For each new situations stored in the history, we proceed by selecting for each situation \(S_t\) (identified by his cluster), his successor \(S_{t+1}\). The successor \(S_{t+1}\) represents the situation that is directly stored after the situation \(S_t\). Then, we calculate the number of the existing transitions from \(S_t\) to \(S_{t+1}\) (\(N_{S_{t}S_{t+1}}\)). Next, for each situation \(S_t\) we determine the entire possible next situation \(S_{t+1}\) (\(N_{S_{t}S_{t+1}}\)). We note that, in the history database, the former number of transitions from \(S_t\) to \(S_{t+1}\) (\(N_{S_{t}S_{t+1}}\)) and the number of all the possible transitions from \(S_t\) (\(N_{S_{t}S_{t+1}}\)) are already stored. This information is updated, and all the states and transition probability are represented and calculated accordingly.
The prediction process is mainly based on the results of the classification phase in order to predict the next user’s intention and service. This prediction process is based on the semantic matching between the user’s immediate intention and context and those of the user’s behavior model states. Similar to the discovery mechanism, the semantic matching is based on ontologies in order to calculate the intentional and contextual matching scores. The final matching score represents the sum of the intention matching score and the context matching score. This information is stored with the state identifier. And going through all the states of the model, we can determine the state the most similar to the current user’s situation. Subsequently, if a state is identified, then the next state is selected based on the transition probabilities. This transition probability must exceed a certain threshold. If several successor states are retrieved, then the one having the highest transition probability is chosen. By this choice, we can anticipate the user’s future needs by offering him the most appropriate service that can interest him.

5. Evaluation

In order to evaluate the proposed service discovery and prediction mechanisms, we use the test collection OWLS-TC22 that we extend in order to include intentional and contextual information. Besides, we create a database containing user’s traces, recognized clusters and the user’s behavior model. We mixed a set of arbitrary traces with others following a scenario representing a well-defined user behavior. Thereafter, we launch the clustering algorithm on the set of traces in order to determine all the recognized clusters. Then, we execute the classification algorithm on all traces and clusters in order to update the user’s behavior model stored in the database.

The service discovery and prediction mechanisms were implemented using Java language. Moreover, they are based on our OWL-SIC API (OWL-S Intentional & Contextual)18, Jena23 and the reasoner Pellet24.

As part of our experiments, we deployed our algorithms on a machine Intel Core i5 1.3 GHz with 4 GB memory. The purpose of these experiments is to evaluate the validity of our algorithms and their feasibility. Thus, we formulate a set of user’s requests relatives to the travel domain. These request are represented by the user’s intention and his current context. These requests are formalized according to three different distributions. The first distribution considers requests that are very similar to the service (service discovery) or clusters centroid (service prediction). Then, the second distribution illustrates situations where the elements describing the intention and/or the context are not described in ontologies while there are services or clusters that are similar to this request. Finally, the third distribution shows the influence of the threshold by presenting in this distribution requests that are within the limits of the threshold and others that are beyond the threshold.

Our first experiments concern the evaluation of our service discovery mechanism. We measure the performance of the discovery algorithm by varying the number of services between 100 and 400. As illustrated in Fig. 4.a, the execution time follows a polynomial trend of degree three varying from 2.79 s for 10 services to 11.82 s for 400 services. However, even if this time still higher, we can observe that despite the fact that we have increased the number of services over forty times, the response time has only increased by for times.

Besides, in order to measure the quality of the result, we cover the two most useful quality metrics: precision and recall15. Through the experiments, we observe that the precision and recall are interesting factors when considering the intention and context in the service discovery. The result presented in Fig. 4.b shows that we obtained a higher precision percentage, about 80%. This indicates that our service discovery algorithm has a greater chance to retrieve
the most appropriate service according to user’s intention and context. However, the good results of precision are accompanied by less interesting results concerning the recall, as illustrated in Fig. 4.b. We can observe that the average recall approximates the 67%. These results can be explained by the evaluation of some situations that can harm the results quality. For example, we have described some user’s request where the elements of the intention are not described in ontologies, while it exists a set of services able to satisfy this intention in the current user’s context.

Our second experiments concern the evaluation of our service prediction mechanism. We measure the performance of our algorithms with respect to the number of clusters, situations and states, by measuring the average processing time. For example, the execution time of the prediction algorithm is measured by varying the number of states in the user’s behavior model, between 8 and 168 states. As illustrated in Fig. 5.b, the execution time follows a polynomial trend of degree three, like the service discovery algorithm, from 1.63 s for 8 states to 4.16 s for 168 states. We increased the number of states over twenty five times, while the execution time has only increased by two and half times. This allows us to validate the feasibility of our prediction algorithm.

Besides, in order to measure the result quality, we use a quality metric inspired from the precision used to evaluate the service discovery. This metric is used to check whether the predicted service is the one that is expected or not. We determined previously the service that the prediction algorithm must predict for a given user’s request, based on the user’s behavior model. Then, we compared this service with the service returned by the algorithm.

We illustrate in Fig. 5.c the quality percentage achieved by the algorithm by varying the number of states. This percentage represents the average quality obtained for all the user’s requests. The results presented in Fig. 5.c indicate that the prediction algorithm has a good quality that is around 60%. These results can be explained by the evaluation of certain situations that significantly degrade the results quality obtained. For example, in the case where situations are described by intentions where the verb and/or the target are fairly generic or specific, we obtain a quality in some cases below 45%. Thus, when the system designer sets very high threshold settings in the prediction algorithm, some clusters or states that can meet the immediate user’s intention in his current context will not be selected, and this contributes to degradation of the results quality.

The analysis of these results shows the importance of the service discovery and prediction mechanisms. We believe that the proposed mechanisms allow really the selection of the service that fulfills the user’s immediate needs and the anticipation of his future need. This is thanks to both its intentional approach, which is more transparent to the user, and its contextual approach that restricts services to those that are valid. However, it is important to note that we cannot get that good result if the system designer does not establishes from the beginning a rich description of the available services and the different ontologies and the most appropriate threshold setting.

6. Conclusion

In this paper, we have proposed a user-centric approach for service discovery and prediction considering PIS. This approach is needed to hide the complexity of these systems and to achieve the transparency required by their users. It enhances PIS transparency and proactivity throw service discovery and prediction mechanisms, defined considering the user’s point of view. These allow us not only offering user the most suitable services given his current intention and context, but also to anticipate the future user’s needs in a fairly understandable way.
By this approach, we believe contributing to the improvement of PIS transparency and proactivity through a user-centric perspective focusing on the intentions that services satisfy in a given context. Moreover, evaluating the user acceptation of the proposal requires applying it in a real case study. Such evaluation should consider the final user’s point of view. It should consider the user acceptance, mainly considering the prediction mechanism, as well as the level of transparency perceived by these users. As a future work, we expect to evaluate our approach in a large-scale in order to validate its usefulness and compare it with the existing techniques.

References

10. Olsson, T., Bjurling, B., Chong, M., Ohlman, B. Goal Refinement for Automated Service Discovery. 3rd Int. Conf. on Advanced Service Computing, 2011, 46–51
22. OWLS-TC: http://www.semwebcentral.org/projects/owl-tc/