Journal of Ambient Intelligence and Smart Environments 8 (2016) 697–707
DOI 10.3233/AIS-160404
IOS Press

A measure of semantic similarity between a reference context and a current context

Djamel Guessoum [a,*], Moeiz Miraoui [b], Atef Zaguia [a,c] and Chakib Tadj [a]
[a] *Electrical Engineering Department, École de Technologie Supérieure, Quebec University, Canada*
[b] *Higher Institute of Applied Science and Technology, University of Gafsa, Tunisia*
[c] *Taif University, College of Computers and Information Technology, Saudi Arabia*

Abstract. Context-aware applications are intended to facilitate the adaptation of services in a pervasive computing system. The semantic similarity between contexts and the application of a semantic similarity measure as a mechanism for service adaptation are topics that have yet to be thoroughly explored in the literature. This study measured semantic similarities between quantitative contextual and categorical variables in the field of pervasive computing. The measure was applied to a current context and to several reference contexts, which were predefined based on a contextual data set. Built on the overlap measure because of its simplicity, the proposed weighted method is easy to implement and can be used to evaluate the actual weight of each contextual variable.

Keywords: Pervasive computing, context, concept, similarity, reference, context

1. Introduction

Pervasive computing, also known as ubiquitous computing, is based on an idea introduced by Mark Weiser: "The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it". Through this concept, computing becomes part of a user's general environment, extending beyond personal computing to exploit the full potential of computing and communication technologies, to classify complex human behavior, and to react to it in a context-specific way [2]. Thus pervasive computing refers to the full extent of accessibility supplied by an omnipresent computing power that is continuously available to a user, allowing him or her to access a certain application, service, or document, etc., at any time and at any moment [22].

The ultimate objective of such an environment consists in providing appropriate services to a user in a transparent fashion. This goal is achieved by using a dynamic adaptation process in which services are provided to a user on an individual basis, either reactively in response to a change in the current context or proactively by predicting a change and adapting accordingly [29], or by using explicit and implicit inputs in conjunction with the context in which these inputs are acquired [22]. By using the environment and the user profile as a source of information, a pervasive computing system is able to adapt services dynamically in accordance with a specific purpose [33].

Several types of mechanisms exist: 1) Rule-based adaptation is a dynamic adaptation mechanism for services that consists in writing a set of logical rules that determine the triggering of a service in a given context. 2) Machine learning adaptation is a dynamic adaptation process for services that detects contextual information. The user-driven selection of appropriate services is then based on some type of learning mechanism. 3) Data mining adaptation is a method that consists in exploring current context data to detect known situations or tendencies in order to provide appropriate services to the user. 4) Comparison-based adaptation consists in comparing conditions linked to the implementation of a service with current context data [3].

*Corresponding author.
E-mail: djamel.guessoum.1@ens.etsmtl.ca.

Each mechanism has its own advantages and disadvantages as well as a particular application domain. Based on the available literature, adaptation mechanisms incorporating a comparison of semantic similarities have rarely been tested for the dynamic adaptation of services in a pervasive computing system.

In pervasive computing, in which the notion of context is very important, the similarity measure serves as a tool to evaluate similarities between instances of a context, which makes it possible to provide the user with the most appropriate services. Various types of similarity measure – depending on the context model and on the representation of the objects that describe this context – have been proposed in the literature [16,30]. The similarity measure between quantifiable objects (simple objects, vectors, or probability distributions), such as temperature or geographic coordinates, is evaluated by measuring the distance; a vector space is used. Semantic similarity measures are applied for categorical variables (non-quantifiable), such as user activity or user mood.

Our method is based on the concept of case-based reasoning (CBR) [18,20]. In CBR, past experiences (reference contexts) are stored and subsequently compared with a currently experienced situation (current context) by using semantic similarity measures in order to select the most similar experience (context) and to provide the user with the corresponding services.

The paper is organized as follows. Section 2 presents related researches and highlights the novelty of our work. Section 3 introduces the notion of context in pervasive computing. Section 4 provides the definition of the reference context and discusses the selection criteria. Section 5 introduces the proposed weighting method for categorical variables. Section 6 provides a case study that serves to evaluate the proposed measure. Conclusions are presented in Section 7.

2. Related work

Semantic similarity measures are used in various fields with different types of applications. In pervasive computing, the application of these measures is connected to the concept of "context" and its impact on the adaptation of services provided to the user.

Several studies have applied these measures to service recommendation systems [23] in which context is represented by the user's profile-related preferences. By including context and allocating a weight to the relationship between concepts, Ning and O'Sullivan [26] developed a similarity measure for the ontological concepts described by Ganesan et al. [13]. Similarly, Miraoui et al. [25] measured the similarity between a known context and a current context for the adaptation of mobile phone incoming call indications.

Mention should also be given to applications of similarity measures in other domains that may be of interest to the field of pervasive computing; such applications include data clustering/mining [1,11] as well as research by Slimani et al. [34], who improved on the previous semantic similarity measure of Wu and Palmer [38] by taking into account the context of the measure.

A pervasive computing system is designed to provide services to a user by minimizing the user's direct involvement. The few existing studies that have applied semantic similarity measures have each provided a particular definition of the context and its specific purpose. Examples include Kirsch-Pinheiro et al. [17], who proposed a dynamic adaptation of services to solve the problem of incomplete information occurring for a selection process of adequate services in a particular context. Benazzouz [3] used the same type of similarity measure to cluster data in order to determine the specific situations that trigger a particular service.

A simple semantic similarity measure (overlap measure) was used for the present study. The proposed weighting method for contextual categorical variables aims to provide appropriate services to a user by defining a set of reference contexts (see Section 4) based on the user's location. We introduce the method for weighting contextual categorical variables with this similarity measure, followed by a comparison of our approach with other, better known methods, which we found to be more suitable in our case. The data collected in the current context were compared with data for each reference context, and the semantic similarity measure was then applied to categorical data types. This model allows the comparison of individual concepts (categorical-type variables) unrelated to a taxonomic or ontological structure, which in turn avoids design problems.

However, the model's main difficulty lies in determining the weighting parameters of the concepts to be compared. Although several authors have previously published work on weighting categorical variables – for example, Smirnov [35], Gambaryan [12], and others in [4] – the weighting methods used in these studies fail to take into account the contribution of each contextual categorical variable concerning the calculation of the semantic similarity measure between two data sets (contexts).

To mitigate some of these shortcomings, we propose a weighting method that assumes that the weight of a characteristic is dependent on how frequently this characteristic occurs for the categorical variables that are compared for each reference context. A large number of occurrences of a particular characteristic in different reference contexts indicated that it weakly characterized these reference contexts and was not specific to a single reference context, as shown in Section 5.

3. Context in pervasive computing

The dictionary defines context as “the set of circumstances or facts that surround an event or a particular situation.” It can nevertheless be shown that this seemingly generic definition contains the essence of the definition of context proposed in most professional domains. Before the term context can be defined, its characteristics should be stated precisely.

The general characteristics of “context” described in [10,15,28] can be summarized as follows:

1. There is no context without “context”: the notion of context must be determined according to a purpose. For example, context dynamically adapts to the interactive capacity of a system.
2. A context is an information space used for interpretation: capturing the context is not an end unto itself, but rather the collected data must serve a purpose.
3. A context is an information space shared by several actors: in the proposed case, the user and the system.
4. A context is an infinite and scalable information space: context is not fixed permanently; it evolves over time.
5. It may be difficult to determine the information needed to infer a contextual state. The relevance of any information is highly dependent on the particular situation.
6. Context and activity are separable. The context as a set of features can be encoded and made available to a software system together with an encoding of the activity itself.

Two key features emerge from these common characteristics: 1) the dynamic nature of the context and 2) its purpose.

The following definitions of context focus on these two aspects. Brezillon and Pomerol defined two concepts related to context [5]: First, a set of contextual data (e.g., time, location) can be used in a decision-making problem; knowledge gained in this way is latent and cannot be used without an objective. Second, context is the product of an emerging objective or intention and requires a large amount of contextual knowledge.

Dalmau et al. [7] showed that the objective of defining the context in pervasive computing is to enable a context-aware application to discover and react to situational changes. Schilit et al. [31] considered context to have three important aspects in response to the following questions: Where are you? Who is with you? What resources are available nearby?

The authors thus categorized context according to six factors. The first three factors relate to the human component, namely information concerning the user (e.g., clothing, biophysical conditions), the social environment (e.g., proximity to other people), and the tasks of the user (e.g. smart tasks of the user). The remaining three factors relate to the physical environment, namely location, infrastructure (e.g., resources, communication), and environmental conditions (e.g., noise, light and climatic conditions).

Dey, whose definition is cited most frequently, defines context as “any information that can be used to characterize the situation of an entity (person, object, or physical computing” [9]. This definition clearly resembles that of Schilit et al. because context is considered as a data set collected from a user environment (person), physical environment (physical object), or system environment. The characterization of these environments is the purpose of data collection.

Brezillon and Pomerol subsequently provided the following definition: “Context is what does not intervene directly in the resolution of a problem but compels its resolution” [5]. This rather generalized definition does not specify the nature of what may compel (constrain) the resolution of a problem.

Miraoui and Tadj [24] proposed a service-oriented definition of context according to which relevant information is used to provide appropriate services to a user. This definition, however, includes only the expectations of the user and has many varied applications, because context is defined as any information for which a change in value triggers a service or alters the quality (form) of a service.

Benazzouz [3] categorized the sources of contextual information: user preferences, behavioral history, physical environment (i.e., ambient temperature, geographical location), as well as the system environment (i.e., applications and networks in which the system functions).

Table 1
Context types, variables, and attributes

Context type		Variable	Attribute
Spatio-Temporal	Space	Type (TP)	covered, open
		Coordinates (CD)	lt_m, lg_m, al_m (latitude, longitude, altitude)
	Time	Day (D)	weekday, public holiday, weekend, vacation
		Hour (H)	5 p.m. $<$ H $<$ 7 a.m. 7 a.m. $<$ H $<$ 5 p.m.
User		Mobility (MB)	sleeping, walking, running, sitting, standing
Environmental		Temperature (T)	cold, warm, hot
		Surroundings (N)	alone, w/friend(s), w/family, unknown
		Noise level (NS)	silent, quiet, noisy
		Light level (Lm)	dark, dim, bright
System		Internet access (CI)	Cable, Wi-Fi, 3G

Contextual variables are selected according to the purpose of the application. This study therefore used contextual information that could be collected with mobile equipment (smartphones) and that could be used to determine the specific location of a user.

Table 1 provides a general classification of contextual information based on work by Lavirotte et al. [19] and cited in various other sources [21,27,32,37,40]. This classification was adopted throughout the present paper because of its simplicity and comprehensiveness. The selection was restricted to information relevant to the initial objective of providing appropriate services to a user in a pervasive computing system.

4. The reference context

A reference context is a context described by predefined contextual information (environmental, user, or system) including the services to be provided to the user or system. In this study, several reference contexts were defined for a particular device (smartphone) to provide a baseline for the similarity measure. These reference contexts were chosen based on the following criteria:

1. The probability (P) of a particular context occurring for a predetermined duration: the higher P, the more likely the context will be a candidate for a reference context.
2. Reference contexts must be sufficiently dissimilar so that the services provided to the user are of differing natures (it is necessary to determine a minimal threshold of semantic similarity (S)).
3. Reference contexts should not be chosen in relation to the user's emotional or physiological state (e.g., happy, sad, thirsty, hungry) because these conditions cannot be captured accurately.

Studies on contextual models have shown that a user's location, identity, time, and activity are the most important parameters determining the type of service to provide [6,39]. According to this categorization and to meet the criteria mentioned previously, user location-based reference contexts were chosen. A user's location can be determined accurately. Environmental, system, and user information is susceptible to change depending on the user's location; to provide the appropriate service, these location-induced changes must thus be taken into account. Furthermore, the user periodically occupies well-defined locations, such as "at home" (nighttime), "at work", or "at school" (daytime). As a first step, the following three locations were chosen as reference contexts:

1. at home
2. at school
3. in transit

Note that this list is merely preliminary and may be incremented each time a new context meets the predefined selection criteria and is sufficiently dissimilar from other contexts.

4.1. Context variable

The data set that characterizes a context is collected from several sources of information, for example, physical sensors in the environment, intelligent devices, virtual sensors, Internet access, or even telecommunication service providers; this information is thus very heterogeneous.

In accordance with several previous studies that have addressed the classification of contextual infor-

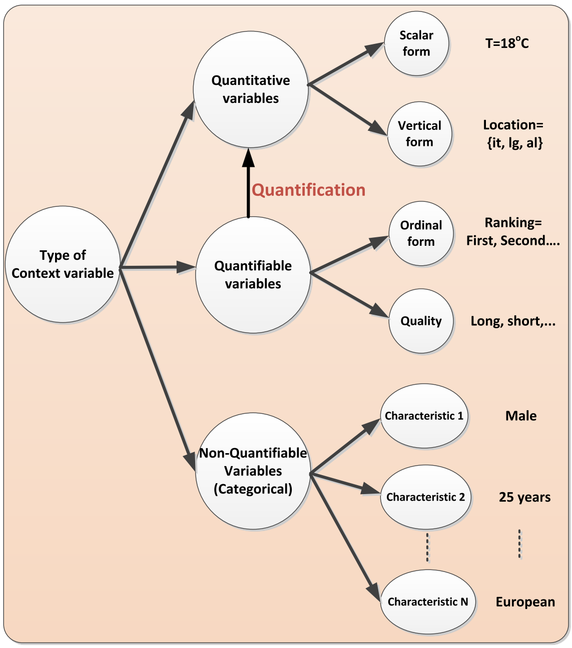

Fig. 1. Classification of context variables.

mation [14,23], the present study adopted the following three categories (Fig. 1):

1. Quantitative variables are expressed in scalar or vector form (i.e., temperature, latitude, longitude, altitude).
2. Quantifiable variables are expressed in qualitative or ordinal form (i.e., large, small, first, second). Quantification is the interpretation of quality in terms of the quantity or projection of an ordinal variable in a metric space; for example, "hot" $\leftrightarrow$ $T > 30\,^{\circ}C$ and "first" are projected on a linear axis, in which case "second" directly follows "first."
3. Categorical variables are not quantifiable. Variables of this type are described as a set of characteristics (e.g., standing, sitting).

5. Semantic similarity measures

5.1. *A measure of semantic similarity between a current context and a reference context*

Because of the heterogenous character of both the collected information and the context variables, a series of partial semantic similarity measures must be applied; these partial measures are applied to variables of the same type. Therefore, the measure of overall semantic similarity between the current context and the reference contexts is the weighted sum of these partial measures.

Let two contexts – C_r (reference context) and C_c (current context) – be defined by n scalar/vector variables and m categorical variables:

$$\begin{cases} C_r = \{Vrs_1, Vrs_2, \ldots Vrs_n, Vrc_1, Vrc_2, \ldots Vrc_m\} \\ C_c = \{Vs_1, Vs_2, \ldots Vs_n, Vc_1, Vc_2, \ldots Vc_m\} \end{cases} \tag{1}$$

where Vrs_i is the scalar/vector variable and Vrc_i is the categorical variable of the reference context; Vs_i is the scalar/vector variable and Vc_i is the categorical variable of the current context.

Based on the assumption that "two contexts are similar if the context variables of the same type that characterize them are similar," defining the semantic similarity between these two contexts involves determining the similarity between variables of the same type – weighted according to their contribution to the characterization of the reference context – for both contexts.

5.2. *Weighting of contextual variables*

In the literature, the weight of a set of categorical variables is inversely proportional to the number of those variables [4] (Overlap measure, Eskin measure, IOF, OF, etc.). The weighting, which is thus identical for all categorical variables, fails to identify each categorical variable's actual contribution to the semantic similarity measure. Existing measures that account for multiple values of a contextual variable, such as the methods proposed by Smirnov [35] and Gambaryan [12], suffer from the same problem.

Moreover, the total number of attributes for this variable affects these two methods. For example, the attributes "sleeping," "walking," "sitting," and "standing" of the categorical mobility variable (MB) characterize the reference context of "home" ($n = 4$), which can take the following attributes: "sleeping," "walking," "sitting," "standing," and "running" ($nt = 5$).

The weight assigned to each context variable must indicate the contribution of this variable to the characterization of the reference context. Therefore, the more restrictive the value (fewer choices) characterizing the context, the more informative is the variable; it must therefore carry more weight. Consequently, this weighting must not be static and must vary according to context.

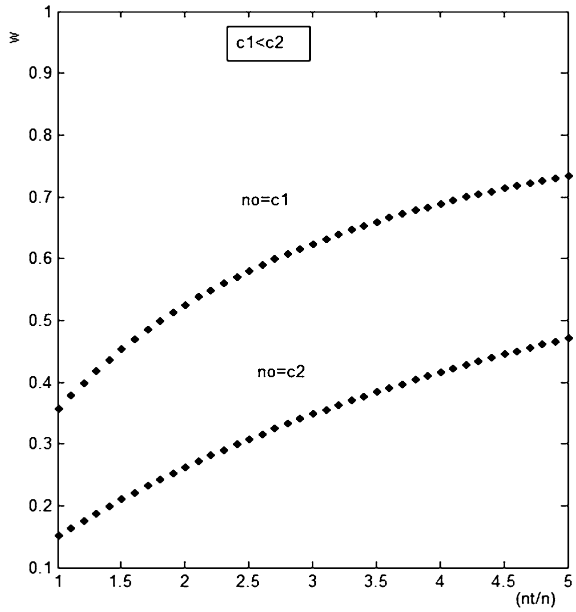

Fig. 2. Weighting of the contextual variables.

For a variable Vrs_i/Vrc_i of the reference context C_r – which can take nt values, but only n ($n \leqslant nt$) of these values characterize the particular reference context – the weight w_i of the context variable is given as follows:

$$w_i = \frac{(\frac{1}{n_o} \cdot \frac{nt}{n})_i}{\sum_{i=1}^{N} \cdot (\frac{1}{n_o} \cdot \frac{nt}{n})_i} \tag{2}$$

$$\text{Note: } \sum_{i=1}^{n} w_i = 1 \tag{3}$$

where N is the total number of contextual variables, and n_o is the number of occurrences of the attribute i of the current context in all reference contexts.

Equation (2) and Fig. 2 shows that the weight of a context variable is proportional to the ratio ($\frac{nt}{n}$) of the total number of attributes allowed for a contextual variable to the number of attributes characterizing a reference context; this weight is inversely proportional to the number of occurrences of the attribute i of the current context in all reference contexts n_o.

5.3. Measures of semantic similarity between quantitative and quantifiable variables

The measure of semantic similarity between two quantities is linked to the measurement of the distance between them in a projected space with identical dimensions. For the distance to have a semantic meaning, the quantities must represent a concept, generally specified as an interval (e.g., nighttime = [7 p.m., 6 a.m.]), low bandwidth = [0 kB/s, 128 kB/s]). Measuring the semantic similarity between a quantity in the current context and a quantity of the same type in the reference context involves measuring its distance to the interval.

Let Vs and Vrs – two quantities of the same type (except for their location coordinates) – belong to the current context and the reference context as defined in Eq. (1).

If the interval limited by (Vrs_{min}, Vrs_{max}) represents a concept ("warm", "hot", etc.) in the reference context, then the distance to the current context variables Vs and Vrs of the same type is as follows:

$$\begin{cases} d(Vs, Vrs) = 0 \\ \quad \text{if } Vrs_{min} \leqslant Vs \leqslant Vrs_{max} \\ d(Vs, Vrs) = (Vrs_{min} - Vs) \\ \quad \text{if } Vs < Vrs_{min} \\ d(Vs, Vrs) = (Vs - Vrs_{max}) \\ \quad \text{if } Vs > Vrs_{max} \end{cases} \tag{4}$$

Furthermore, the semantic similarity between Vs and Vrs is the following:

$$\begin{cases} Sim(Vs, Vrs) = 1 \\ \quad \text{if } Vrs_{min} \leqslant Vs \leqslant Vrs_{max} \\ Sim(Vs, Vrs) = \frac{1}{1+d(Vs, Vrs)} \\ \quad \text{otherwise} \end{cases} \tag{5}$$

Example. If a "warm" temperature is defined as $T \in [20\,°C, 30\,°C]$, then $T_1 = 20\,°C$ is more similar to $T = 25\,°C$ than to $T_2 = 18\,°C$, even though the distance between them would indicate the inverse (Fig. 3).

Thus, from Eq. (4),

$$T_{min} = 20\,°C < Vs_1 = T_1 = 25\,°C < T_{max} = 30\,°C$$

$$Vrs_{min} \leqslant Vs_1 \leqslant Vrs_{max} \Rightarrow \begin{cases} d(Vs, Vrs) = 0 \\ Sim(Vs, Vrs) = 1 \end{cases}$$

$$Vs_2 = T_2 = 18\,°C < T_{min} = 20\,°C$$

$$Vs_2 < Vrs_{min}$$

$$\Rightarrow \begin{cases} d(Vs, Vrs) = (Vrs_{min} - Vs) = 20 - 18 = 2 \\ Sim(Vs, Vrs) = \frac{1}{1+d(Vs, Vrs)} = \frac{1}{1+2} = \frac{1}{3} \end{cases}$$

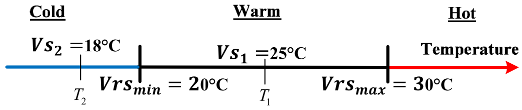

Fig. 3. Scalar variable for temperature.

If the variable Vrs of the reference context relates to geographical coordinates limited by ($\mathrm{Vrs_{min}}$, $\mathrm{Vrs_{max}}$) in three dimensions, then the semantic similarity between Vs and Vrs is

$$\begin{cases} \mathrm{Sim(Vs, Vrs)} = 1 & \text{if } \mathrm{Vrs_{min}} \leqslant \mathrm{Vs} \leqslant \mathrm{Vrs_{max}} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

For all quantitative variables, the overall similarity Sim(Vs, Vrs) between the current context and the reference context is thus

$$\mathrm{Sim(Vs, Vrs)} = \sum_{1}^{N} \mathrm{w_i Sim(Vs_i, Vrs_i)} \tag{7}$$

where N is the total number of quantitative variables.

5.4. *Measures of semantic similarity between categorical variables*

The measure of semantic similarity between categorical variables aims to quantify the intrinsic characteristics shared by these variables. A characteristic is intrinsic to an object when it defines the nature of the object and cannot be separated from it.

Each reference context $\mathrm{C_r}$ is characterized by a set of categorical variables.

Let

$$\mathrm{Vc_i} = \{\mathrm{Car}\} = \{\mathrm{Car_1}, \mathrm{Car_2} \ldots \mathrm{Car_n}\}$$

be one such variable.

To determine the semantic similarity between the categorical variable of the current context $\mathrm{Vc_i}$ and the categorical variable of the reference context $\mathrm{Vrc_i}$ of the same type, the overlap measure [36] was chosen because it is simple to implement for equipment with limited resources such as a smartphone, see Eq. (8).

$$\mathrm{Sim(Vc_i, Vrc_i)} = \begin{cases} 1 & k \neq 0 \\ 0 & k = 0 \end{cases} \tag{8}$$

where k is the number of common attributes shared by $\mathrm{Vc_i}$ and $\mathrm{Vrc_i}$.

For all categorical variables, the overall similarity Sim(Vc, Vrc) between the current context and the reference context is thus

$$\mathrm{Sim(Vc, Vrc)} = \sum_{1}^{N} \mathrm{w_i Sim(Vc_i, Vrc_i)} \tag{9}$$

where N is the total number of categorical variables.

5.5. *Overall semantic similarity*

The overall semantic similarity between the current context $\mathrm{C_c}$ and the reference context $\mathrm{C_r}$ is as follows:

$$\begin{aligned} &\mathrm{Sim(C_c, C_r)} \\ &\quad = (\mathrm{Sim(Vs_i, Vrs_i)} + \mathrm{Sim(Vc_i, Vrc_i)})/2 \end{aligned} \tag{10}$$

where $\mathrm{Sim(Vs_i, Vrs_i)}$ is the semantic similarity between the ith variable (quantitative or quantifiable) in the reference context $\mathrm{C_r}$ and the ith variable of the same type in the current context $\mathrm{C_c}$, and $\mathrm{Sim(Vc_i, Vrc_i)}$ is the semantic similarity between the ith variable (categorical) in the reference context $\mathrm{C_r}$ and the ith variable of the same type in the current context $\mathrm{C_c}$.

$$\mathrm{Sim(C_c, C_r)} \in [0, 1]$$

6. Case study

To provide an example of an application and to illustrate the weighting of the contextual variables, the measure of semantic similarity between a current context $\mathrm{C_c}$ and reference context $\mathrm{C_r}$ was applied to the data set.[1]

This data set consists of feature files for 43 different recording sessions. In each recording session the same user – carrying a mobile phone, sensor box, and laptop computer – was commuting from home to the workplace or vice-versa. During the commute the user walked, used some kind of public transportation (bus or Metro), and sometimes drove a car. Occasionally, the user took either slightly different or considerably different routes/modes of transport. During the session, triaxial acceleration sensors recorded atmospheric pressure, temperature, humidity, etc. Ambient audio levels were recorded on a laptop com-

[1]The data set NokiaContextData, obtained from the Institut für Pervasive Computing (Johannes Kepler University Linz, Austria), is available at http://www.pervasive.jku.at/Research/Context_Database/.

Table 2
Attributes of contextual variables by location

Contextual variable	Current context	Reference context by location Number of attributes characterizing the reference context (n)				Total number of attributes (nt)
		Location 0	Location 1	Location 2	Location 3	
Day Name	2	(1.2.3.4)	(1.2)	(1.2.3.4)	(1.2)	7
Day Date	4	(3.4.5.6.10)	(3.4.10)	(4.5.6.10)	(4.10)	31
Day Period	4	(2.4)	(2.4)	(2)	(2)	4
Temperature	9	(4.5.6.7.8.9)	(4.5.6.7.8.6)	(4.5.6.7.8.9)	(8.9)	10
Relative Humidity	5	(1.2.3.4.5.6.7)	(1.2.3.4.5.6.7.8)	(1.2.3.4.5.6.7.8.9)	(6.9)	10
Pressure	10	(1.2.3.5.6.7.8.9.10)	(1.2.3.8.9.10)	(1.2.3.4.5.7.8.9.10)	(9.10)	10
User Activity	6	(1.3.4.5.6)	(1.2.3.4.5.6)	(1.2.3.4.5.6)	(3.4)	6
Average Audio Level	1	(1.2.3.4)	(1.2.3.4)	(1.2.3.4)	(3.4)	5

Note: Contextual variable names were taken from and are identical to the original data set.

puter equipped with a microphone and a sound card. On the mobile phone, changes in the user's location were recorded in the form of a Cell ID and a location area code obtained through the GSM network.

Five sessions were randomly selected, with each session characterized by the following contextual variables: categorical variables (Day Name (Saturday, Sunday…(1–7)), Day Period (night, morning, afternoon, evening (1–4)), User Activity (1–6)), quantitative variable (Day Date (1–31), Temperature (1–10), Relative Humidity (1–10), Pressure (1–10), Average Audio Level (1–5)).

These five selected sessions included seven different locations; several hundred values of contextual variables were recorded at each location.

To show the relevance of the weighted calculation, a current context was selected based on the recordings characterizing "Location 0." Results from four locations (Loc. 0, Loc. 1, Loc. 2, and Loc. 3) were compared (Table 2). Table 2, far right column, shows the total number of values that a context variable can take (nt); the "Reference context by location" columns show the contextual variable values (n) characterizing each location (reference context). The column labeled "Current context" contains the current value of the contextual variable.

The values of the categorical context variables were then modified (Day Name, Day Period, and User Activity) to show the effect of weighting these variables (Table 4).

The following semantic similarity algorithm was applied:

Step 1. Determine the weight of the contextual variables (w_{DN}, w_{DD}, w_T, w_H, w_A, w_P, ws_{DP}, ws_{Act}), see Table 3.

Table 3
Overall semantic similarity

	Overall Similarity $w = 1/d$	Overall Similarity $w = \frac{1}{\sum_{k=1}^{d} n_k}$	Overall Similarity $w = \frac{(\frac{1}{n_o}\cdot\frac{nt}{n})}{\sum_{i=1}^{N}\cdot(\frac{1}{n_o}\cdot\frac{nt}{n})}$
New location-Location 0	1	0.268	1
New location-Location 1	0.9	0.175	0.935
New location-Location 2	0.833	0.203	0.646
New location-Location 3	0.466	0.450	0.4

$$w_i = \frac{(\frac{1}{n_o}\cdot\frac{nt}{n})_i}{\sum_{i=1}^{N}\cdot(\frac{1}{n_o}\cdot\frac{nt}{n})_i}$$

Step 2. For every reference context, determine the semantic similarity with the current context between the context variables of the same type.

For the reference context "Location 0," measure the semantic similarity between quantitative or quantifiable variables (Day Date, Temperature, Relative Humidity, Pressure, and Average Audio Level), see Eq. (7).

Similarly, measure the semantic similarity between categorical variables (Day Name, Day Period, User Activity), see Eq. (9).

Because the data set does not contain discrete values, the similarity between the quantitative and categorical variables is calculated by measuring the overlap [36], see Eq. (8).

Step 3. Calculate the overall semantic similarity, Eq. (10).

Table 3 shows that good agreement can be achieved with the proposed weighted method by using $w = \frac{1}{d}$

Table 4
Weight of contextual variables by location

Context	Categorical variable	$w = \frac{1}{d}$ (Overlap, Eskin, IOF, OF, Burnaby, Goodall1, 2, 3, 4) [4]	$w = \frac{1}{\sum_{k=1}^{d} n_k}$ (Smirnov, Gambaryan) [4]	$w = \frac{(\frac{1}{n_o} \cdot \frac{nt}{n})}{\sum_{i=1}^{N} \cdot (\frac{1}{n_o} \cdot \frac{nt}{n})}$ (Proposed method)
Loc. 0	Day Name	0.333	0.125	0.714
	Day Period	0.333	0.125	0.178
	User Activity	0.333	0.125	0.107
Loc. 1	Day Name	0.333	0.111	0.727
	Day Period	0.333	0.111	0.181
	User Activity	0.333	0.111	0.090
Loc. 2	Day Name	0.333	1	0.727
	Day Period	0.333	0.5	0.181
	User Activity	0.333	0.6	0.090
Loc. 3	Day Name	0.333	0.125	0.714
	Day Period	0.333	0.125	0.178
	User Activity	0.333	0.125	0.107

The number of contextual variables is denoted by d, and n_k is the number of attributes allowed for each contextual variable.

(Overlap, Eskin, IOF, OF, Burnaby, Goodall1, 2, 3, 4) [8]) to measure the semantic similarity. To show the effect of the proposed weighting method, we modified the number of occurrences no and the ratio ($\frac{nt}{n}$) of the categorical context variables (Day Name, Day Period, and User Activity) as follows:

Day Name (no $= 1$, $\frac{nt}{n} = 2$)
Day Period (no $= 4$, $\frac{nt}{n} = 2$)
User Activity (no $= 4$, $\frac{nt}{n} = 1$ and $\frac{nt}{n} = 1.2$)

In accordance with these values, The “Day Name” variable must have a more significant weight than either the “Day Period” or the “User Activity” variable.

Table 4 shows that the most significant contextual variable (Day Name) has a weight that corresponds to its significance and that the other contextual variables are weighted lower.

These results suggest that the proposed approach is applicable to measuring the importance of contextual variables and to providing a true assessment of the contribution of each individual variable while remaining simple to implement (thanks to the Overlap measure) with resource-limited equipment such as smartphones.

Figure 4 shows the dynamic character of the proposed approach to calculating the weight w in comparison with most approaches advocated in the literature ($w = 1/d = 1/3$ et $w = \frac{1}{\sum_{k=1}^{d} n_k} = 1/8$). The proposed approach relies on the number of occurrences of the contextual variable in all reference contexts as well as on the ratio ($\frac{nt}{n}$). This ratio increases with a decreasing number (n) of attributes characterizing a categorical variable.

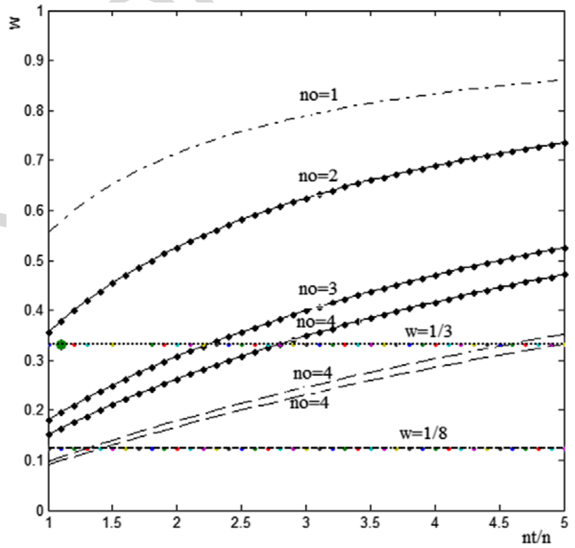

Fig. 4. Weight changes according to no and nt/n.

7. Conclusion

Based on a simple measure of semantic similarity, the proposed weighting method was used as a mechanism for service adaptation in a pervasive computing system by defining a type of context known as a “reference context” to serve as a basis of comparison with the current context.

The proposed methodology – despite its simplicity – is very intuitive and provides a realistic model for weighting the contextual variables of the semantic similarity measure in a pervasive computing system.

However, the criteria for selecting the reference contexts may be improved to achieve more flexible and more dynamic guidelines and to include several types of context. It would therefore be useful to formulate the definition of partial reference contexts within the global reference context, (e.g., "school" as the global reference context and "library," "classroom," etc., as partial reference contexts).

The selection of variables that define a context in general should be revised, as should the selection of variables and attributes characterizing a reference context.

The proposed weighting method is applicable to our case only. Further research is needed to arrive at conclusions that generalize to other applications in other areas.

References

[1] P. Andritsos, P. Tsaparas, R.J. Miller and K.C. Sevcik, Limbo: Scalable clustering of categorical data, in: *Advances in Database Technology-EDBT 2004*, Springer, Berlin, 2004, pp. 123–146. doi:10.1007/978-3-540-24741-8_9.

[2] J.C. Augusto, H. Nakashima and H. Aghajan, Ambient intelligence and smart environments: A state of the art, in: *Handbook of Ambient Intelligence and Smart Environments*, Springer, US, 2010, pp. 3–31. doi:10.1007/978-0-387-93808-0_1.

[3] Y. Benazzouz, Découverte de contexte pour une adaptation automatique de services en intelligence ambiante, Ph.D. dissertation, École Nationale Supérieure des Mines de Saint-Étienne, France, 2011.

[4] S. Boriah, V. Chandola and V. Kumar, Similarity measures for categorical data: A comparative evaluation, *RED* **30**(2) (2008), 3.

[5] P. Brezillon and J.-Ch. Pomerol, Contextual knowledge and proceduralized context, *AAAI Technical Report* **14** (1999), 16–20.

[6] G. Chen and D. Kotz, A survey of context-aware mobile computing research, Technical Report TR2000-381, 1(2.1), Dept. of Computer Science, Dartmouth College, 2000, 1–16.

[7] M. Dalmau, P. Roose and S. Laplace, Context aware adaptable applications: A global approach, *IJCSI* **1** (2009), 1–13.

[8] A. Desai, H. Singh and V. Pudi, Disc: Data-intensive similarity measure for categorical data, in: *Advances in Knowledge Discovery and Data Mining*, Springer, Berlin, 2011, pp. 469–481. doi:10.1007/978-3-642-20847-8_39.

[9] A.K. Dey, Understanding and using context, *Personal and Ubiquitous Computing* **5**(1) (2001), 4–7. doi:10.1007/s007790170019.

[10] P. Dourish, What we talk about when we talk about context, *Personal and Ubiquitous Computing* **8**(1) (2004), 19–30. doi:10.1007/s00779-003-0253-8.

[11] H. El Sayed, H. Hacid and D. Zighed, *Using Semantic Distance in a Content-Based Heterogeneous Information Retrieval System*, Springer-Verlag, 2008, pp. 224–237.

[12] P. Gambaryan, A mathematical model of taxonomy, *Izvest. Akad. Nauk Armen. SSR* **17**(12) (1964), 47.

[13] P. Ganesan, H. Garcia-Molina and J. Widom, Exploiting hierarchical domain structure to compute similarity, *ACM Trans. Inf. Syst.* **21** (2003), 64–93. doi:10.1145/635484.635487.

[14] K.C. Gowda and E. Diday, Symbolic clustering using a new similarity measure: Systems, man and cybernetics, *IEEE Transactions* **22**(2) (1992), 368–378.

[15] S. Greenberg, Context as a dynamic construct, *Human-Computer Interaction* **16**(2) (2001), 257–268. doi:10.1207/S15327051HCI16234_09.

[16] D. Guessoum, M. Miraoui and C. Tadj, Survey of semantic similarity measures in pervasive computing, *International Journal on Smart Sensing and Intelligent Systems* **8**(1) (2015), 125–158.

[17] M. Kirsch-Pinheiro, Y. Vanrompay and Y. Berbers, Context-aware service selection using graph matching, in: *2nd Non-functional Properties and Service Level Agreements in Service Oriented Computing Workshop (NFPSLA-SOC'08), ECOWS*, CEUR Workshop Proceedings, Vol. 411, 2008.

[18] J. Kolodner, *Case-Based Reasoning*, Morgan Kaufmann, 2014.

[19] S. Lavirotte, D. Lingrand and J.-Y. Tigli, *Définition du Contexte. Fonctions de Cout et Methodes de Selection*, ACM, 2005.

[20] D.B. Leake, *Case-Based Reasoning*, 2003.

[21] S. Lee, J. Chang and S.G. Lee, Survey and trend analysis of context-aware systems, *Information – An International Interdisciplinary Journal* **14**(2) (2011), 527–548.

[22] J.A. Lino, B. Salem and M. Rauterberg, Responsive environments: User experiences for ambient intelligence, *Journal of Ambient Intelligence and Smart Environments* **2**(4) (2010), 347–367.

[23] L. Liu, F. Lecue, N. Mehandjiev and L. Xu, Using context similarity for service recommendation, in: *Semantic Computing (ICSC), IEEE Fourth International Conference*, 2010, pp. 277–284.

[24] M. Miraoui and C. Tadj, A service-oriented definition of context for pervasive computing, in: *Proceedings of the 16th International Conference on Computing*, IEEE Computer Society Press, Mexico City, 2007.

[25] M. Miraoui, C. Tadj and H. Belgacem, Dynamic context-aware adaptation of mobile phone incoming call indication using context similarity, in: *World Congress Computer and Information Technology (WCCIT)*, IEEE, 2013, pp. 1–5.

[26] K. Ning and D. O'Sullivan, Context modeling and measuring for context aware knowledge management, *International Journal of Machine Learning and Computing* **2**(3) (2012), 287–291. doi:10.7763/IJMLC.2012.V2.132.

[27] M. Perttunen, J. Riekki and O. Lassila, Context representation and reasoning in pervasive computing: A review, *International Journal of Multimedia and Ubiquitous Engineering* **4**(4) (2009).

[28] M. Petit, L'informatique contextuelle, Technical report, South Britany University, Vannes, 2005.

[29] G. Sancho, Adaptation d'architectures logicielles collaboratives dans les environnements ubiquitaires, Contribution à l'interopérabilité par la sémantique, Ph.D. Dissertation, EDSYS, France, 2010.

[30] K. Saruladha, G. Aghila and S. Raj, A survey of semantic similarity methods for ontology based information retrieval,

in: *Machine Learning and Computing (ICMLC), 2010 Second International Conference on*, IEEE, February 2010, pp. 297–301.

[31] B. Schilit, N. Adams and R. Want, Context-aware computing applications, in: *IEEE Workshop on Mobile Computing Systems and Applications*, Santa Cruz, CA, 1994.

[32] A. Schmidt, M. Beigl and H.-W. Gellersen, There is more to context than location, *Computers & Graphics* **23**(6) (1999), 893–901. doi:10.1016/S0097-8493(99)00120-X.

[33] J. Simonin and N. Carbonell, Interfaces adaptatives, Adaptation dynamique à l'utilisateur courant, arXiv preprint arXiv:0708.3742, 2007.

[34] T. Slimani, B. Ben Yaghlane and K. Mellouli, Une extension de mesure de similarité entre les concepts d'une ontologie, in: *4th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications*, Tunisia, 2007, pp. 1–10.

[35] E.S. Smirnov, On exact methods in systematics, *Systematic Zoology* **17**(1) (1968), 1–13. doi:10.2307/2412390.

[36] C. Stanfill and D. Waltz, Toward memory-based reasoning, *Commun. ACM* **29**(12) (1986), 1213–1228. doi:10.1145/7902.7906.

[37] F. Topcu, Context modeling and reasoning techniques, in: *SNET Seminar in the ST*, 2011.

[38] Z. Wu and M. Palmer, Verb semantics and lexical selection, in: *32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, New Mexico, 1994, pp. 133–138. doi:10.3115/981732.981751.

[39] B. Yuan and J. Herbert, Context-aware hybrid reasoning framework for pervasive healthcare, *Personal and Ubiquitous Computing* **18**(4) (2014), 865–881. doi:10.1007/s00779-013-0696-5.

[40] D. Zhang, H. Huang, C.F. Lai, X. Liang, Q. Zou and M. Guo, Survey on context-awareness in ubiquitous media, *Multimedia Tools and Applications* **67**(1) (2013), 179–211. doi:10.1007/s11042-011-0940-9.