Investigation of Context Prediction Accuracy for Different Context Abstraction Levels

Stephan Sigg, *Member*, *IEEE*, Dawud Gordon, *Student Member*, *IEEE*, Georg von Zengen, Michael Beigl, *Member*, *IEEE*, Sandra Haseloff, and Klaus David, *Member*, *IEEE*

Abstract—Context prediction is the task of inferring information about the progression of an observed context time series based on its previous behaviour. Prediction methods can be applied at several abstraction levels in the context processing chain. In a theoretical analysis as well as by means of experiments we show that the nature of the input data, the quality of the output, and finally the flow of processing operations used to make a prediction, are correlated. A comprehensive discussion of basic concepts in context prediction domains and a study on the effects of the context abstraction level on the context prediction accuracy in context prediction scenarios is provided. We develop a set of formulae that link scenario-dependent parameters to a probability for the context prediction accuracy. It is demonstrated that the results achieved in our theoretical analysis can also be confirmed in simulations as well as in experimental studies.

Index Terms—Pervasive computing, stochastic processes, location-dependent and sensitive, performance evaluation of algorithms and systems, time series analysis.

1 Introduction

With distinct frameworks and architectures for context computing, different representations, processing orders, and hierarchies of context abstraction are proposed. In 1994, for instance, Schilit designed an architecture that communicates context changes to applications [1] and presented a distributed structure for context-aware systems. In this architecture, agents provide context information that is aggregated from multiple context sources. An aggregation and hierarchy is mentioned but a detailed description is not provided.

These thoughts are further developed in the context toolkit that was introduced in 2000 [2]. It constitutes a conceptual framework to support the development of context-aware applications and distinguishes between context sensing and context computing. Context sensing describes the process of acquiring information about contexts using sensors while context computing refers to the interpretation of acquired contexts.

Later on, Schmidt presented a "working model for context-aware mobile computing" as an extensible tree structure [3]. The proposed hierarchy of features starts with distinguishing human factors and the physical environment and expands from there.

In 2004, the distributed middleware framework Solar was presented by Chen [4]. It provides means to derive higher level contexts from lower level sensor and aggregated data from a multilayered directed acyclic information fusion graph of event processing operators that represents the underlying context structure [5], [6].

The abstraction levels of context in distinct stages of context processing architectures are frequently referred to by the notions high-level, low-level and raw data. A rough distinction between low-level and higher level contexts is made by Dey [2], Schilit and Theimer [7]. Following this discussion, low-level context is used synonymously for data directly obtained from sensors, while high-level context is context information that is further processed. This processing can, for example, be semantic reasoning, an interpretation, data calibration or noise removal.

Mäntyjärvi distinguishes between context information that describes an action or a condition [8], where following his notion, the lowest abstraction level, raw data, would be $24°C$ or 70 percent humidity, for example. Following his notion, the lowest abstraction level, raw data, can be, for example, $24°C$ or 70 percent humidity. For low-level contexts, these values are further processed to conditions like "warm" or "high humidity." Finally, a high-level context is an activity such as, for instance, "having lunch."

For all these distinctions, higher level contexts are derived by further processing lower level context data. We propose an alternative distinction on context abstraction that is based on the amount of processing applied to contexts in Section 2. Following this model, the context abstraction rises with the amount of processing applied. In particular, we do not restrict the number of distinct context

- *S. Sigg and G. von Zengen are with TU Braunschweig, Muehlenpfordtstrasse 23, Braunschweig 38106, Germany. E-mail: sigg@ibr.cs.tu-bs.de, g.vonzengen@tu-bs.de.*
- *D. Gordon is with the Telecooperation Office (TecO), Vincenz-Prießnitz-str. 3, Karlsruhe 76131, Germany. E-mail: gordon@teco.edu.*
- *M. Beigl is with Pervasive Computing Systems, Karlsruhe Institute of Technology.*
- *S. Haseloff is with Alexander von Humboldt-Stiftung, Nationale Kontaktstelle Mobilitaet, Jean-Paul-Str. 12, Bonn 53173, Germany. E-mail: sandra.haseloff@avh.de.*
- *K. David is with the University of Kassel, Wilhelmshöher Allee 73, Kassel 34121, Germany. E-mail: david@uni-kassel.de.*

Manuscript received 5 Mar. 2008; revised 5 Aug. 2010; accepted 2 Dec. 2010; published online 8 Aug. 2011.
For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2008-03-0073.
Digital Object Identifier no. 10.1109/TMC.2011.170.

abstractions to any finite set such as, for instance, "raw data," "low level," or "high level," but expect a fine-grained transition among context abstractions. This computation-centric view allows a comparison of contexts that is not based on subjective classifications but rather on the amount of processing that is actually applied.

We claim that, depending on the probability of error for context processing operations, the order in which these operations are applied might impact the overall probability of error for context processing. Based on our definition of context abstraction, we develop a probabilistic model to estimate the accuracy of context values computed along different context processing chains and conditioned on various input parameters. In particular, and in contrast to many studies on the provision of a high context accuracy in the literature, we do not improve the accuracy of a specific context processing operation but instead consider the impact of environmental factors as well as the order in which context processing operations are applied on the context processing accuracy. The results derived in our study enable new design alternatives in the development of context aware applications. Consider, for instance, the integration of a low-power sensor for context processing in a resource-restricted device. While the resource restrictions of the device might not allow the selection of highly accurate but computationally complex processing operations, the probabilistic model derived can provide the optimum processing order of feasible processing operations that maximizes the expected context processing accuracy.

For the example operations "Context acquisition," "Context interpretation," and "Context prediction," we show that the order in which processing operations are applied, the probabilities of error for these operations, the prediction horizon, the context history size and finally the input dimension all impact the overall context accuracy.

Similar to the common understanding [9], [10], [11], [12], we model context prediction as a single context processing operation. Typically, it is deployed as one of the last context processing operations. We take a more general approach and assume that context prediction can be applied at an arbitrary context abstraction level.

Section 3 details existing approaches to context prediction in the literature. Section 4 discusses impacts of context processing order on the context prediction accuracy and Section 5 presents results from experimental studies and simulations that support our analytical findings. For these studies, various algorithms introduced in Section 3 are applied in different environmental settings and at different positions in the context processing chain. Section 6 summarizes our results.

2 An Operational Context Hierarchy

We classify the level of context abstraction by the amount of processing applied to the data. With an increasing number of processing operations applied to context data the context abstraction level rises. We denote various levels of context abstraction as $Cal_i; i \in \mathbb{N}$ and require $Cal_i > Cal_j \Leftrightarrow i > j$. When we are able to quantify the amount of context abstraction induced by an individual context processing operation, this concept becomes operational.

In Section 4, for instance, we associate processing operations with context abstractions proportional to the error probability of these operations. Each processing operation applied to context data also contains the probability of an error and possibly also the probability to correct prior errors.

For the remainder of our work, we consider the three context processing operations "acquisition," "interpretation," and "prediction." In particular, we study the impact of applying context prediction at various stages in the context processing chain. The following section details prominent approaches to context prediction.

3 Algorithms for Context Prediction

The task of context prediction is defined as follows [13]:

Definition 3.1 (Context prediction). *Let $k, n, i \in \mathbb{N}$ and t_i describe any interval in time. Furthermore, let T be a context time series. Given a probabilistic process $\pi(t_i) \rightarrow T$, context prediction is the task of learning and applying a prediction function $f_{t_i} : T_{t_{i-k+1},t_i} \rightarrow T_{t_{i+1},t_{i+n}}$ that approximates $\pi(t_i)$.*

For context prediction, we therefore assume that the observed context time series follows a probabilistic process. Through the approximation of this process, an estimation of the continuation of this time series is possible.

In the literature, context prediction is usually applied at the end of the context processing chain (see for instance [9], [10], [14]). Observe that this decision also impacts the type of input data expected. Typically, contexts of low abstraction levels tend to be numerical while with higher context abstraction context might become symbolical. Consequently, not all prediction approaches are applicable at arbitrary context abstraction levels.

Several authors have studied aspects of future context with the aim of enabling proactive behavior in applications. In the MavHome project [15], movement of inhabitants of a smart home are predicted by a pattern matching approach [16]. The algorithm identifies frequent sequences of length 3 or greater in the recent history of symbolically represented inhabitant contexts and provides the most frequent continuation of these sequences as predictions. Gradual changes in inhabitant behavior are addressed by weighting observed patterns. Related approaches that also utilize exact matching of observed sequences are the ONISI system [17], the IPAM algorithm to predict UNIX command line instructions [12], [18] as well as the IPHYS method [19].

In many context-aware applications, however, exact pattern matching may lead to inferior prediction accuracies, as typical patterns can incorporate measurement errors and slightly changed context durations or sequence orders. To cope with these issues, an approximate pattern matching method is proposed in [20]. This alignment prediction approach is especially well suited to finding typical context patterns in a time series of contexts. This time series can be constituted of numerical and nonnumerical context data alike. When k is the maximum length of any context pattern, the overall running time of this prediction approach is $\mathcal{O}((k^2)|S'|) = \mathcal{O}(k^3)$ [21].

Other approaches for context prediction are the stochastic ARMA and Kalman filter-based methods. The author of [9] derived in his studies that ARMA processes are able to achieve excellent results in context prediction. The method is applicable to one dimensional as well as multidimensional data sets and has a computational complexity of $\mathcal{O}(k \log(k))$ [22]. It is, however, only applicable to contexts of numeric context data types.

The Kalman filter is a stochastic method designed for forecasting numerical time series. Examples for applications of the Kalman filter technique to context-aware scenarios are [23], [24], [25].

The Kalman filter computes a prediction based on an arbitrary long history of observed contexts. The computational load of the method is $\mathcal{O}(k^4)$ [26]. It is not applicable to non-numeric contexts.

In [27], a high prediction accuracy of a principle component analysis (PCA) [28] based prediction approach is reported on a data set with three context classifications (home, work, elsewhere). The PCA is a statistical technique to identify patterns in high-dimensional data. Basically, the eigenvectors and eigenvalues of the covariance matrix of input data are computed. Eigenvalues indicate the significance of the corresponding eigenvector in describing the data. It is then transformed to a new basis spanned by the most relevant eigenvectors—the principal components. For context prediction, the PCA is applied to binary indicator feature vectors of the input data. The runtime of the method is dependent on the number of distinct contexts $|\mathcal{C}|$ in a scenario, as the length of the binary feature vector increases with this value. When M patterns are utilized, the runtime of the method is $\mathcal{O}(M \cdot (k \cdot |\mathcal{C}|)^2)$ for nonnumeric context patterns and $\mathcal{O}(M \cdot k^2)$ in scenarios with only numeric input patterns (no transformation to binary indicator feature vectors required) [29]. Especially in scenarios with nonnumeric input patterns, the method is well suited when the number of distinct contexts $|\mathcal{C}|$ is reasonable.

For the PCA-prediction approach, a priori knowledge of the length and occurrence time of common behavior patterns is required. When typical patterns do not reappear at similar times, the prediction accuracy is reduced. While many patterns in ubiquitous settings are, in fact, very static in nature (e.g., people sleep at night, have breakfast, work, lunch, work, and finally come back home), other patterns might not follow such a strong scheme, as, for example, having phone calls or meetings.

A popular prediction approach is the prediction by Markov models. It can be applied to numerical and nonnumerical data alike. However, a prediction that reaches farther into the future implicitly utilises already predicted data which might decrease the prediction accuracy. The computational complexity is $\mathcal{O}(k \cdot |C|^2)$.

In [30], [31], Libo Song et al. study the accuracy of Markov and compression-based prediction [32], prediction by partial matching [33] and sampled pattern matching [34] approaches of mobility patterns using a huge data set sampled at Dartmouth campus. The Markov approach achieved the best prediction accuracy for the next context (in this case WLAN access points).

Despite numerous studies on context prediction, a concise investigation of the various parameters that impact the prediction accuracy aside from the algorithm applied was not conducted. In this study, we provide a comprehensive consideration of various aspects that impact the context prediction accuracy; in particular, we consider the order in which processing operations are applied to the context data. We show that the context prediction accuracy differs for a given context prediction algorithm depending on when it is applied in the context processing chain. The extent of this impact is, among other aspects, further dependent on design parameters as the length of the context history, the prediction horizon or the number of context sources utilized.

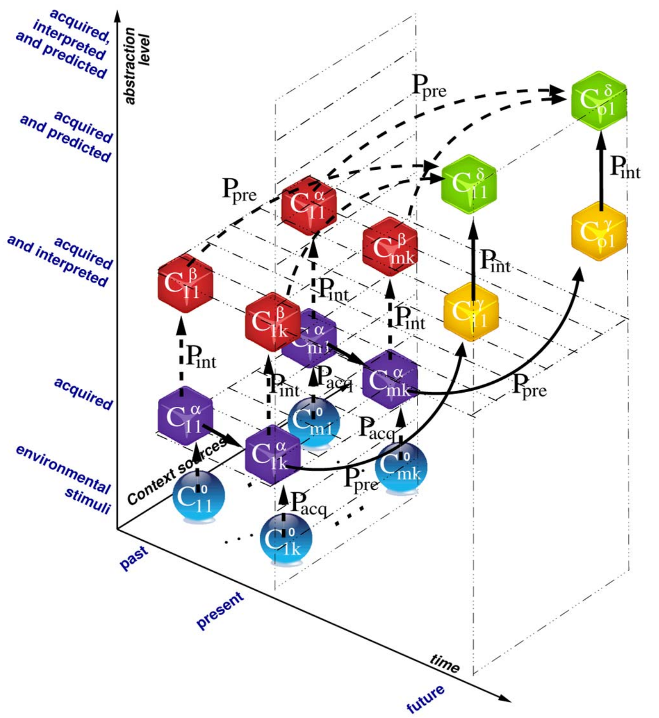


Fig. 1. Context prediction applied at various stages in the context processing chain.

4 Context Abstraction and Accuracy

This section presents our context abstraction model and discusses the effects of prediction accuracy on various orders of context processing operations (see Fig. 1).

The figure distinguishes between two context prediction schemes in which the prediction is applied on a higher context abstraction level (hl) and on a lower abstraction level (ll) The ll-prediction scheme applies the prediction operation prior to the context interpretation while the hl-context prediction scheme applies these modules conversely. In both cases, predicted and interpreted contexts of the same abstraction level are achieved from the data acquired from various context sources.

We make several assumptions on context data and processing which we assume to be reasonable for many application scenarios. In particular, we develop our model for the three context processing operations acquisition, interpretation, and prediction. As the number or type of processing operations increases, the derived formulas must to be adapted analogous to the analysis detailed.

Context acquisition preserves dimensionality: We assume that context acquisition is an m to m operation. For every single value obtained from a context source, a separate context acquisition step is applied that computes exactly one ll-context.

Context interpretation alters dimensionality:Context interpretation is not applied overlapping or combining time intervals. However, it might alter the time series dimension m at a lower context abstraction level to an o-dimensional context time series at a higher context abstraction. In a scenario in which the dimension is not altered, the two variables o and m collapse to one single variable.

Context prediction preserves dimensionality: We model a q-dimensional time series prediction by a q-fold one-dimensional prediction. If required, this condition can be relaxed by introducing an additional variable to describe the number of predictions that are applied.

Error probability known and constant: For acquisition, interpretation and prediction, we assume that the error probability is known and constant for each application of an operation. Probabilities for distinct types of operations are independent of each other.

Processing operations are identical: We assume that the processing operations utilized impose an identical error probability on the input values regardless of the context abstraction level on which they are applied.

Number of context values is constant: The number of possible context values is constant among context types of one abstraction level. In a scenario in which the number of distinct values differs for different context types, this can be modeled by individual variables that describe the number of possible context values for each context type.

Uniform probability distribution: Errors that occur in the interpretation or prediction steps are independently and identically distributed.

No mixed abstraction level processing: Processing operations utilize contexts of exactly one context abstraction level at one time.

Assume $i, k, m, o, v_l, v_h \in \mathbb{N}^{\setminus 0}$. For our discussion, k represents the length of the context history while m and o describe the dimensionality of the context time series before and after the context interpretation, respectively. A context may take one of v_l values in advance and one of v_h values after context interpretation is applied. The number of different configurations for a time series element of the context history at one point in time is therefore v_l^m before and v_h^o after context interpretation is applied. During context processing, sources of error are the context acquisition, the context interpretation, and the context prediction. Corresponding error probabilities are

- P_{acq}. The probability that no error occurs in the context acquisition step.
- P_{int}. The probability that no error occurs in the context interpretation step.
- P_{pre}. The probability that no error occurs in the context prediction step. $P_{pre}(i)$ expresses the probability that no error occurs in the prediction of the ith context.

We derive the probability that an arbitrary predicted time interval is without error for context prediction applied before (ll-context prediction) and after (hl-context prediction) the context interpretation in the following sections. For ease of presentation, we denote contexts prior to the interpretation as ll-contexts and otherwise as hl-contexts.

4.1 Prediction after the Context Interpretation

The context acquisition is the first processing operation applied to the sampled context information. For all k time series elements in the context history, every one of the m values is processed in the context acquisition (cf. Fig. 1). Since P_{acq} describes the probability that error does not occur in one of these operations, the probability that error does not occur in any of the $k \cdot m$ context acquisition steps is P_{acq}^{km}.

In the context interpretation, the m ll-contexts of every one of the k context time series elements in the ll-context history are interpreted to o hl-contexts that constitute a time series element of the hl-context time series. Altogether, $k \cdot o$ context interpretation steps are applied. Since P_{int} describes the probability that error does not occur in one of these steps, the probability that error does not occur in the whole context interpretation process is consequently P_{int}^{ko}. Finally, $P_{pre}(i)$ describes the probability that the prediction of the ith context is without error. Since the ith time series element consists of o context elements, $P_{pre}^o(i)$ is the probability that error does not occur in the context prediction. The approximated probability P_{hl}^{approx} that no errors occur in the hl-context prediction process of one specific hl-time series is then given as

$$P_{hl}^{approx} = P_{acq}^{km} P_{int}^{ko} P_{pre}^{o}(i). \tag{1}$$

In this approximation, we do not take into account that errors occurring in one processing step might be corrected by succeeding operations. The probability P_{cor}^{int} that an error which occurs in a context acquisition step is corrected by an error that occurs in the context interpretation step is

$$P_{cor}^{int} = \left(1 - P_{acq}^{m}\right)\left(1 - P_{int}^{o}\right) \frac{1}{v_h^o - 1}. \tag{2}$$

In this formula, $1 - P_{acq}^m$ is the probability that an error occurs in one of the m context acquisition steps that are related to one context time series element and $1 - P_{int}^o$ describes the probability that an error occurs in one of the o context interpretation steps. With probability $\frac{1}{v_h^o - 1}$, the specific error required for a correction is observed from all $v_h^o - 1$ equally probable interpretation errors. Since v_h values are possible for every one of the o hl-contexts in one time series element, the number of possible hl-time series elements is v_h^o. Consequently, the number of possible errors is $v_h^o - 1$ since one element represents the correct interpretation that is without error.

We additionally consider the correcting influence of the context prediction error. The probability $P_{hl}(i)$ that a time series element i is accurately predicted if the prediction is based on the hl-context time series is then

$$\begin{aligned} P_{hl}(i) = {} & \left(P_{acq}^{m} P_{int}^{o} + P_{cor}^{int}\right)^k P_{pre}^{o}(i) \\ & + \left(1 - \left(P_{acq}^{m} P_{int}^{o} + P_{cor}^{int}\right)^k\right) \frac{1 - P_{pre}^{o}(i)}{v_h^o - 1}. \end{aligned} \tag{3}$$

4.2 Prediction Prior to the Context Interpretation

For ll-context prediction, context prediction is applied in advance of context interpretation. The probability that the ith time series element is correctly predicted is described by

TABLE 1
Classification Accuracies of the C4.5 Decision Tree on Output Data from a Microvibration Sensor [35]

	Bus	Bike	Walk	Jog	Lift	Type	Stair
accuracy	0.27	0.49	0.58	0.79	0.26	0.91	0.47
frequency	0.05	0.05	0.1	0.02	0.01	0.74	0.02

$$P_{ll}^{approx} = P_{acq}^{km} P_{pre}^{m}(i) P_{int}^{o}. \quad (4)$$

In analogy to the discussion above, we obtain the probability $P_{ll}(i)$ that time series element i is correctly predicted as

$$\begin{aligned} P_{ll}(i) = &\left(P_{acq}^{k} P_{pre}(i) + P_{cor}^{pre}\right)^{m} P_{int}^{o} \\ &+ \left(1 - \left(P_{acq}^{k} P_{pre}(i) + P_{cor}^{pre}\right)^{m}\right) \frac{1 - P_{int}^{o}}{v_h^{o} - 1}. \end{aligned} \quad (5)$$

4.3 Application Scenario

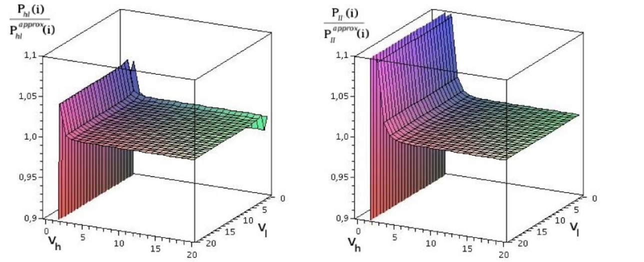
The following example shall demonstrate the application of these formulas in a practical setting. Assume the development of continuous limited prediction capability as an enhancement for a wearable device as, for instance, a wrist watch. The watch shall be equipped with a low-energy microvibration sensor (MVS) we studied in [35]. For the prediction, the alignment algorithm we presented in [21] is utilized. Seven situations ($v_h = 7$) shall be recognized by the C4.5 decision tree. Table 1 details the accuracy for the classification of these situations as observed in [35] together with expected occurrence frequencies.

For simplicity, we utilize the mean normalized classification accuracy of $P_{int} = 0.8012$ as obtained from the values detailed in the table. Since only one microvibration sensor is utilized, we have $m = o = 1$. Similar to [35], we cumulate the binary ticks of the sensor and cut the resulting integer time series in the range $[0, 99]$ ($v_l = 100$) into distinct samples during the context acquisition. Assume $P_{acq} = 0.99$ due to processing noise. Also, due to resource restrictions, the context history is limited to $k = 5$ values. Assume that for the alignment prediction an evaluation of training data has provided $P_{pre} = 0.83$.

By substituting these values in (3) and (5), we obtain $P_{ll}(i) \approx 0.64$ and $P_{hl}(i) \approx 0.28$. Consequently, we expect a higher prediction accuracy when context interpretation is applied after context prediction in this setting.

4.4 Discussion

Having derived the context prediction accuracies for ll- and hl-context prediction schemes, we now discuss the possible impact of the context abstraction level on the context prediction accuracy. We explore this impact by a comparison of $P_{ll}(i)$ and $P_{hl}(i)$. These formulas are hard to grasp due to the multitude of parameters involved. However, for $v_l \to \infty$ and $v_h \to \infty$, the hl-and ll-prediction accuracies can be approximated by $P_{ll}^{approx}(i)$ and $P_{hl}^{approx}(i)$. Fig. 2 shows a comparison between the approximated and the exact probabilities. From the figure, we observe that for sufficiently large values of v_l and v_h, observations made for $P_{ll}^{approx}(i)$ and $P_{hl}^{approx}(i)$ are also valid for $P_{ll}(i)$ and $P_{hl}(i)$. We therefore initially discuss $P_{ll}^{approx}(i)$ and $P_{hl}^{approx}(i)$ before considering the more exact formulas $P_{ll}(i)$ and $P_{hl}(i)$. First of all, we observe that the influence of acquisition errors is equal for hl-and ll-context prediction schemes, since the factor P_{acq}^{km} appears in both formulas.



(a) $\frac{P_{hl}(i)}{P_{hl}^{approx}(i)}$ (b) $\frac{P_{ll}(i)}{P_{ll}^{approx}(i)}$

Fig. 2. Comparison of the approximated and exact probability of prediction errors for $k = m = o = v_l = v_h = 5$ and $P_{acq} = 0.99, P_{int} = P_{pre} = 0.9$.

The fraction of these probabilities yields

$$\frac{P_{hl}^{approx}(i)}{P_{ll}^{approx}(i)} = P_{int}^{k} P_{pre}^{o-m}(i). \quad (6)$$

Clearly, this term is smaller than 1 for all possible configurations other than $P_{int} = P_{pre}(i) = 1$. Consequently, for sufficiently large values of v_l and v_h, context prediction based on ll-context elements is superior to context prediction based on hl-context elements.

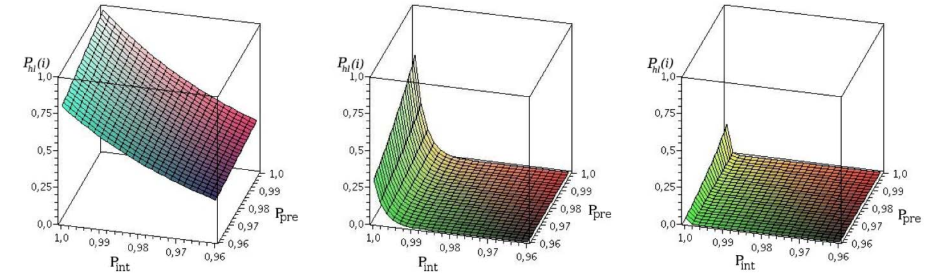
However, this observation is only true for high values of v_l and v_h. We therefore also study $P_{hl}(i)$ and $P_{ll}(i)$.

In Fig. 3, the probabilities that a prediction based on hl- and ll-context elements has no erroneously predicted context elements are depicted for several values of $P_{pre}(i)$ and P_{int}. We observe that the probability for a correct prediction decreases with increasing v_h, k, v_l, m, and o as expected.

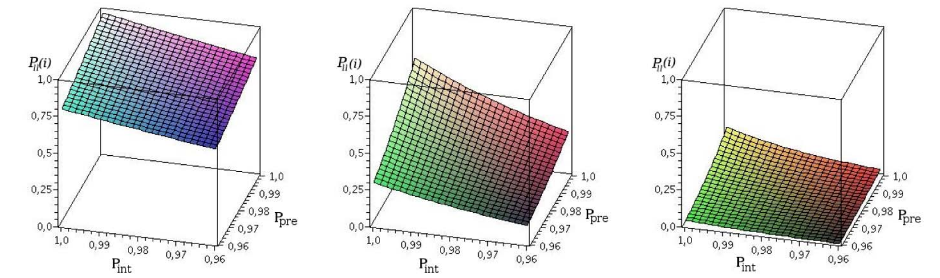
For ll-context prediction, the degradation is less harsh as for hl-context prediction. We therefore conclude that the ll-prediction scheme is better capable of dealing with this configuration of the input parameters v_h, k, v_l, m, and o.

Fig. 4 illustrates the predominance of the ll-context prediction scheme above the hl-prediction scheme. In these figures, only the points below 1.0, at which the hl-context prediction scheme is superior, are depicted. The ll-context prediction has a smaller error probability for all but low values of $P_{pre}(i)$. The number of points where the ll-context prediction is superior increases for higher values of v_h, k, v_l, m, and o.

Impact of the acquisition accuracy: When P_{acq} is decreased, the overall probability for a correct prediction decreases as expected for both prediction schemes (see Fig. 5). Note that for Figs. 5b and 5d the scaling of the Z-axis is different since otherwise details are hardly visible. The impact is serious for both prediction schemes. The error probability of the acquisition process is therefore a highly critical input to the overall prediction process regardless of the prediction scheme utilized. However, the slope of the probability plane and the probability of error is higher for the hl-context prediction scheme.



(a) $P_{hl}(i)$ for $k = m = o = v_l = v_h = 5$ and $P_{acq} = 0.99$ (b) $P_{hl}(i)$ for $k = m = o = v_l = v_h = 20$ and $P_{acq} = 0.99$ (c) $P_{hl}(i)$ for $k = m = o = v_l = v_h = 40$ and $P_{acq} = 0.99$



(d) $P_{ll}(i)$ for $k = m = o = v_l = v_h = 5$ and $P_{acq} = 0.99$ (e) $P_{ll}(i)$ for $k = m = o = v_l = v_h = 20$ and $P_{acq} = 0.99$ (f) $P_{ll}(i)$ for $k = m = o = v_l = v_h = 40$ and $P_{acq} = 0.99$

Fig. 3. Hl- and ll-context prediction accuracies.

Impact of the number of context values: Next, we consider the impact of the v_l possible ll-context values. Fig. 6 shows the probability that error does not occur in the hl- and ll-prediction schemes. We observe that the effect is minor. This property is similar for the v_h of different hl-context time series.

Impact of the ll-time series dimension: For the m context sources utilized, as well as the dimension of the ll-context time series, the context prediction accuracy decreases with an increasing number of context sources for both prediction schemes (see Fig. 7). The ll-context prediction performs better for configurations with higher values of $P_{pre}(i)$, whereas for hl-context prediction the accuracy is better for higher values of P_{int}.

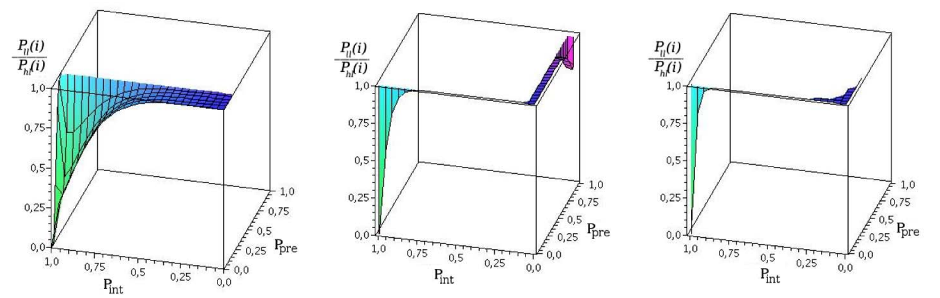
From Figs. 7e and 7f, we observe that the ll-context prediction scheme is advantageous for roughly $P_{pre}(i) > P_{int}$. Therefore, for high values of m the ratio of P_{int} to $P_{pre}(i)$ determines which context prediction scheme is beneficial.

Impact of the hl-time series dimension: The number of parallel hl-context time series o has significant impact on the context prediction accuracy. Fig. 8 shows that the impact is more significant for hl-context prediction.

Impact of the context prediction horizon: For the value k that describes the context prediction horizon, we again observe that the hl-context prediction scheme has a greater probability of error (see Fig. 9). This property intensifies as the size of the context history increases.

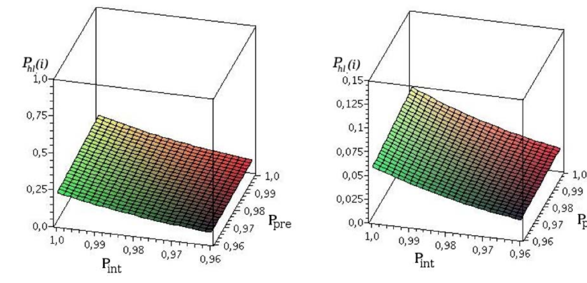
5 Experimental and Simulation Studies

In the following sections, we present results of experimental studies and simulations that confirm our findings on the impact of the order of context processing operations. For all

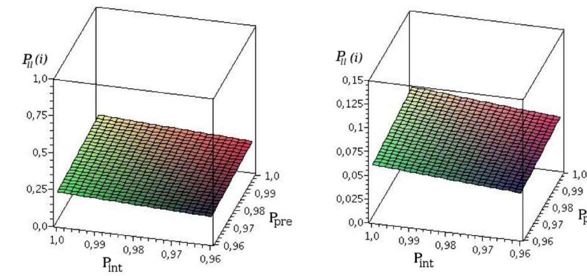


(a) $\frac{P_{ll}(i)}{P_{hl}(i)}$ for $k = m = o = v_l = v_h = 5$ and $P_{acq} = 0.99$ (b) $\frac{P_{ll}(i)}{P_{hl}(i)}$ for $k = m = o = v_l = v_h = 20$ and $P_{acq} = 0.99$ (c) $\frac{P_{ll}(i)}{P_{hl}(i)}$ for $k = m = o = v_l = v_h = 40$ and $P_{acq} = 0.99$

Fig. 4. Regions in the probability space where the hl-context prediction scheme outperforms the ll-context prediction scheme.

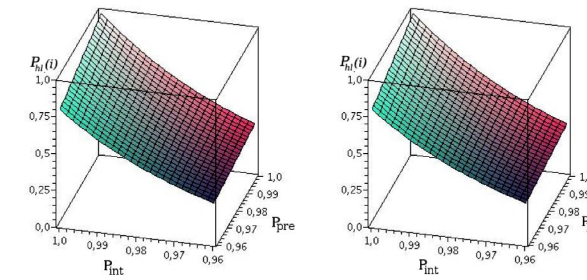


(a) $P_{hl}(i)$ for $P_{acq} = 0.95$ (b) $P_{hl}(i)$ for $P_{acq} = 0.9$

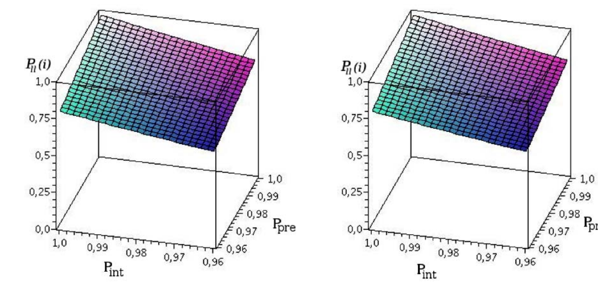


(c) $P_{ll}(i)$ for $P_{acq} = 0.95$ (d) $P_{hl}(i)$ for $P_{acq} = 0.9$

Fig. 5. Probability planes for hl- and ll-context prediction ($k = m = o = v_l, v_h = 5$).



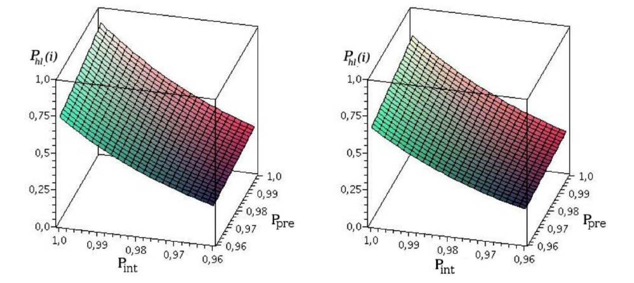
(a) $P_{hl}(i)$ for $P_{acq} = 0.99$, $v_l = 20$ and $m = o = k = v_h = 5$ (b) $P_{hl}(i)$ for $P_{acq} = 0.99$, $v_l = 40$ and $m = o = k = v_h = 5$



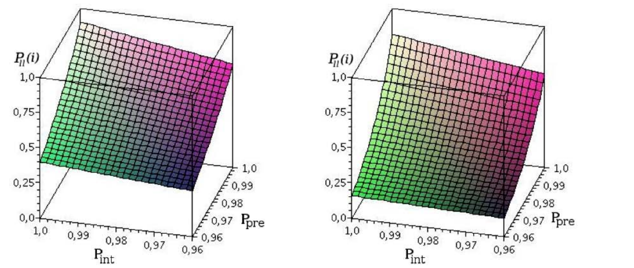
(c) $P_{ll}(i)$ for $P_{acq} = 0.99$, $v_l = 20$ and $m = o = k = v_h = 5$ (d) $P_{ll}(i)$ for $P_{acq} = 0.99$, $v_l = 40$ and $m = o = k = v_h = 5$

Fig. 6. Probability planes for hl-and ll-context prediction when the number of ll-context types is varied.

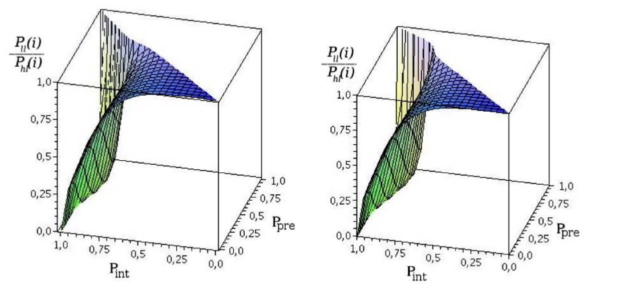
studies, we utilize the approximate pattern matching approach described in [21], [20] for its simplicity and because it is applicable to hl- and ll-context data alike. This method finds typical context patterns in observed sequences by approximate pattern matching. Suitable algorithms for this task are detailed in [36]. We apply the approach first detailed in [37].



(a) $P_{hl}(i)$ for $P_{acq} = 0.99$, $m = 20$ and $o = k = v_l = v_h = 5$ (b) $P_{hl}(i)$ for $P_{acq} = 0.99$, $m = 40$ and $o = k = v_l = v_h = 5$



(c) $P_{ll}(i)$ for $P_{acq} = 0.99$, $m = 20$ and $o = k = v_l = v_h = 5$ (d) $P_{ll}(i)$ for $P_{acq} = 0.99$, $m = 40$ and $o = k = v_l = v_h = 5$



(e) $\frac{P_{ll}(i)}{P_{hl}(i)}$ for $P_{acq} = 0.99$, $m = 20$ and $o = k = v_l = v_h = 5$ (f) $\frac{P_{ll}(i)}{P_{hl}(i)}$ for $P_{acq} = 0.99$, $m = 20$ and $o = k = v_l = v_h = 5$

Fig. 7. Probability planes for hl- and ll-context prediction when the dimensions of the ll-context time series is varied.

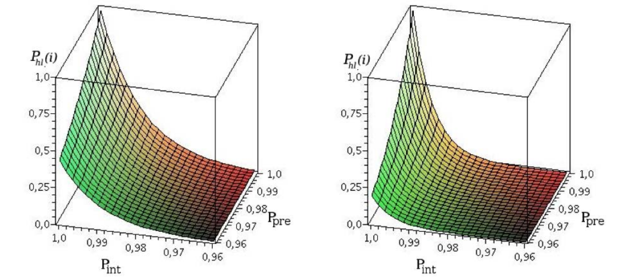
Two metrics commonly utilized to measure the accuracy of predictions are the root of the mean square error (RMSE) and the mean absolute error (BIAS). For a predicted time series of length n, these metrics are defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(p_i - d_i)^2}{n}}, \tag{7}$$

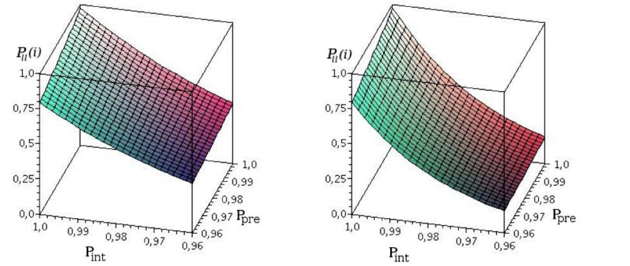
$$BIAS = \frac{\sum_{i=1}^{n}|p_i - d_i|}{n}. \tag{8}$$

In these formulas, p_i denotes the predicted value at time i while d_i is the value that actually occurs at time i.

Sections 5.1 and 5.2 detail experimental studies in which we equipped test subjects with measurement hardware. In Section 5.1, the context sequence of a group of users is predicted based on input from temperature, light, and vibration sensors. The results show that the prediction accuracy of hl-prediction approaches is more sensitive to changes in the prediction horizon than the ll-prediction scheme.

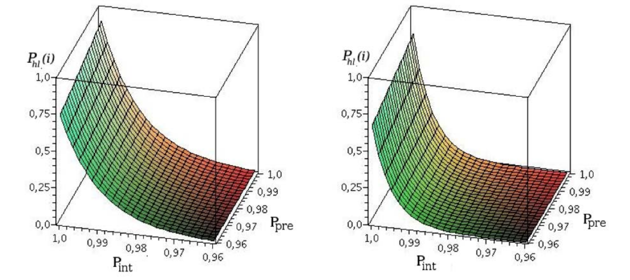


(a) $P_{hl}(i)$ for $P_{acq} = 0.99$, $o = 20$ and $v_l = v_h = m = k = 5$ (b) $P_{hl}(i)$ for $P_{acq} = 0.99$, $o = 40$ and $v_l = v_h = m = k = 5$

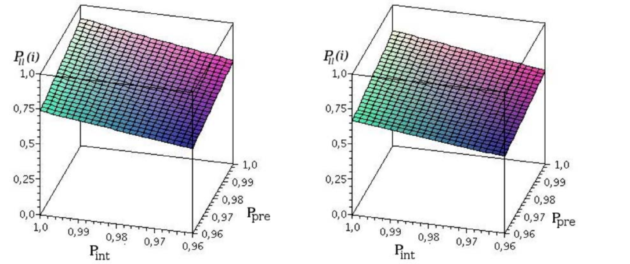


(c) $P_{ll}(i)$ for $P_{acq} = 0.99$, $o = 20$ and $v_l = v_h = m = k = 5$ (d) $P_{ll}(i)$ for $P_{acq} = 0.99$, $o = 40$ and $v_l = v_h = m = k = 5$

Fig. 8. Probability planes for hl-context prediction when the dimensions of the hl-context time series is varied.



(a) $P_{hl}(i)$ for $P_{acq} = 0.99$, $k = 20$ and $v_l = v_h = k = m = 5$ (b) $P_{hl}(i)$ for $P_{acq} = 0.99$, $k = 40$ and $v_l = v_h = k = m = 5$



(c) $P_{ll}(i)$ for $P_{acq} = 0.99$, $k = 20$ and $v_l = v_h = k = m = 5$ (d) $P_{ll}(i)$ for $P_{acq} = 0.99$, $k = 40$ and $v_l = v_h = k = m = 5$

Fig. 9. Probability planes for hl- and ll-context prediction when the context history size is varied.

In Section 5.2, the GPS-trajectory of a mobile user is predicted over an experiment duration of 21 days. In addition to the impact of the context horizon, we can observe how the length of the context history also impacts the prediction accuracy.

Further effects predicted by our analysis result from the accuracy of the context interpretation method and the number of context sources utilized. Since the interpretation error might also depend on the complexity of an observed context and the consideration of new context sources inherently affects the context interpretation error, we were not able to consider these aspects separately in experimental studies. Therefore, in Sections 5.2 and 5.4, we present simulations on synthetic data sets in which we could isolate these aspects.

5.1 Impact of the Prediction Horizon

In an experimental study with five test subjects, we consider the impact of the prediction horizon on the context prediction accuracy. As detailed in Section 4.4, we expect the hl-context prediction approach to be more seriously impacted than the ll-context prediction method (cf. Fig. 9). We prepared five subjects with our Akiba measurement nodes that are equipped with a microvibration sensor (we utilized the MVS0608.02 from Sensolute (http://www.sensolute.de)), an external ADXL335 3D accelerometer from Analog Devices Inc (http://www.analog.com), a temperature sensor (TC1047 from Microchip Technology, Inc., http://www.microchip.com) and the APDS-9003 Light photo sensor from Avago Technologies (http://www.avagotech.com). The Akiba node was provided with access to a microSD card to store the sensed information. Fig. 10 details the experimental setting and a schematic of the Microvibration Sensor.

Unlike the signal produced by an analog acceleration sensor, the output of the MVS is a digital binary vector. The interesting information from these signals are the unary transitions between the two states of the signal, as opposed to the state of the signal itself at any given time. This information is accumulated over a sample window to generate a time series for further processing that generally represents the frequency of state transitions of the MVS. Fig. 11 details this procedure.

The five subjects had to repeatedly complete predefined sequences of actions. These were recorded by the Akiba measurement nodes and utilized for context prediction. Data samples have been recorded every 0.01 seconds from all sources.

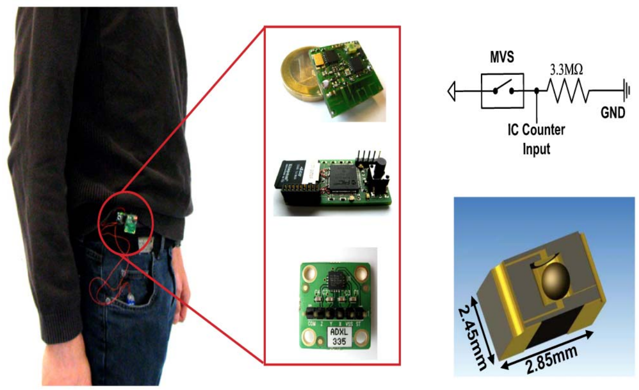


Fig. 10. A subject equipped with our measurement device for the experiment. The MVS and the accelerometer are attached at similar places on either side of the Akiba node so that measurement data are related.

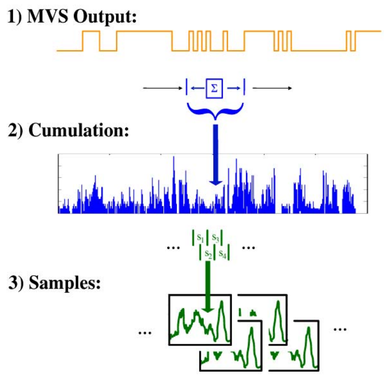


Fig. 11. The processing sequence for the MVS sensor output.

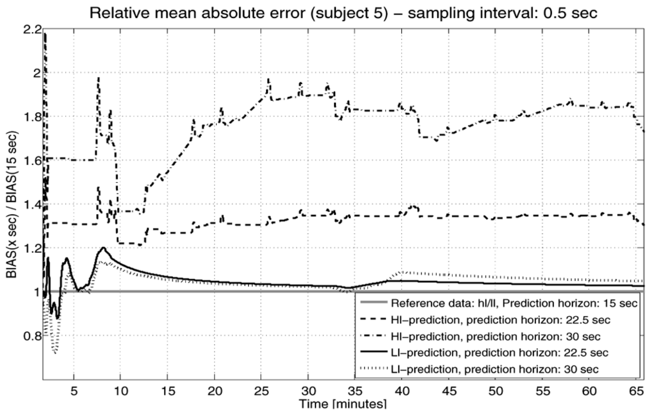


Fig. 13. Relative mean absolute error for one of the subjects over the course of an experiment.

Each triple in this sequence details the measurement from the light sensor, the temperature sensor, and the MVS.

For activity recognition, the WEKA data mining toolkit [38] was used to train a C4.5 decision tree [39] which we selected for its prevalence in the activity recognition literature using acceleration data [40], [41], [42].

The context sequences followed by the subjects were (in this order)

- Sitting at a desk typing into a computer—Descending stairs from the office—taking the elevator to ground level—walking to the tram station—standing still, waiting for a tram—riding the tram home.
- Standing still, waiting for a tram—riding the tram to work—walking to the Institute—taking the elevator to the Institute floor—ascending stairs to the office.

Fig. 12 shows the prediction accuracy for 3 of the subjects by what amount a predicted hl-context varies from the actual values measured.

Although the prediction accuracy deviates slightly among distinct subjects, prediction accuracy achieved by various prediction horizons is reasonably accurate and deviates with increasing prediction horizon, as expected.

Furthermore, we observed that the hl-prediction approach was more seriously impacted by the accuracy loss due to an increased prediction horizon. Fig. 13 demonstrates this using the results from one of the test subjects.

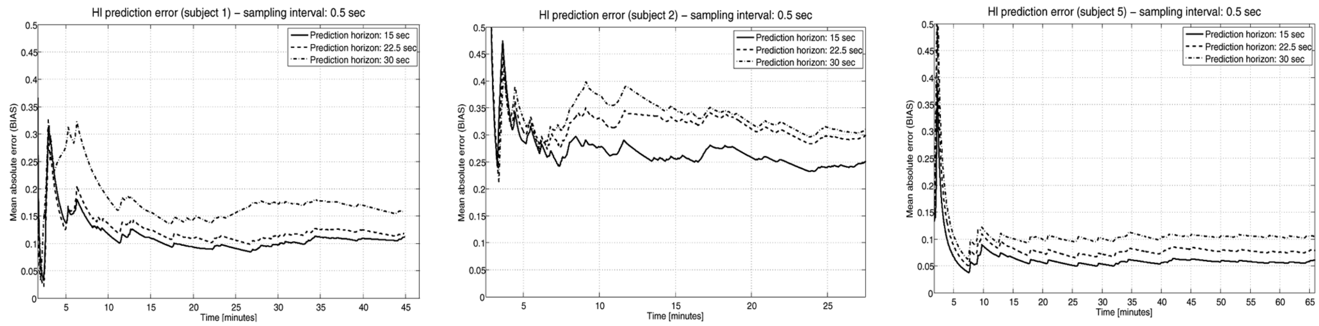
In the figure, the relative decrease in the accuracy is detailed compared to the accuracy at the smallest prediction horizon. Observe that the decrease in accuracy is several orders higher for the hl-prediction approach as predicted by our analytic consideration in Section 4.4 (cf. Fig. 3).

5.2 Impact of the Context History Size

We study influences of varying levels of context abstraction on the accuracy of hl- and ll-context prediction schemes on a sampled GPS trajectory of a mobile user. The sampling hardware consists of a mobile phone and a GPS-receiver. A python script running day and night on the phone was used to obtain the GPS-information from the GPS-receiver. The simulation data consist of three consecutive weeks of GPS samples. Every 2 minutes a GPS sample is taken. When no GPS is available (e.g., indoors), we utilize the last available sample to approximate the current position. For the simulation, we utilize the samples on an 8-minutes and 12-minute scale, respectively, to reduce sequences of idle periods where no significant movement is observed.

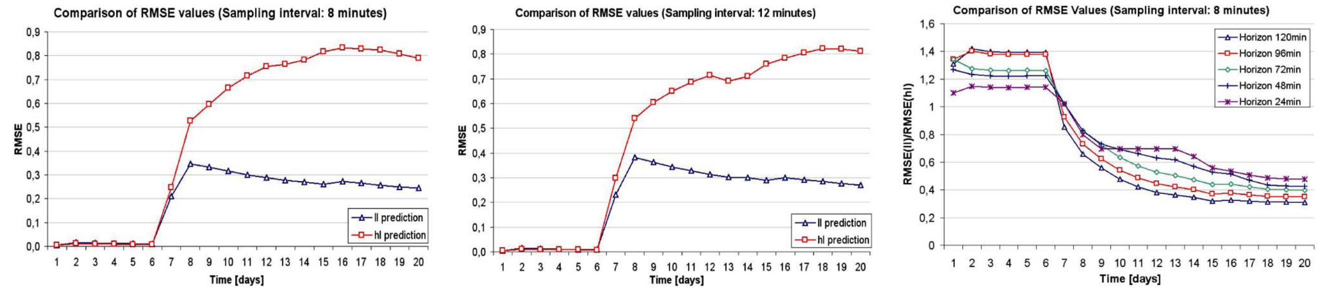
For ll-context prediction, we use the three-dimensional GPS-samples as input data. For hl-context prediction, we define a total of 36 hl-locations as, for instance, "Home," "Bakery," "University," or "Market." The hl-locations are specified by a GPS-center-location and a radius. A default hl-location named "Outdoors" is applied when no other location matches.

The context history covers a time interval of 40 minutes for the 8 minute sampling interval and 1 hour for the 12 minute sampling interval.



(a) Prediction accuracy achieved for subject 1 (b) Prediction accuracy achieved for subject 2 (c) Prediction accuracy achieved for subject 5

Fig. 12. Comparison of hl-context prediction accuracies for several subjects during the experiment.



(a) RMSE values at a sampling interval of 8 minutes (b) RMSE values at a sampling interval of 12 minutes (c) Various prediction horizons

Fig. 14. Context prediction accuracies for hl- and ll-context prediction schemes at various sampling intervals.

In Fig. 14, the results for the ll- and hl-context prediction algorithms are depicted. In the figures, the prediction horizon is set to 2 and 3 hours for both sampling intervals. The first six days of simulation show only minor prediction errors as this period corresponded with the work schedule of the individual who went to work the same way every day with little variance. When the first weekend started, the behavior changed and new time series occurred which had not been previously observed. Therefore, the RMSE values increase drastically. At the time of the second weekend, we again observe an increase in the RMSE values, although it is less harsh than the first one.

The ll-context prediction scheme performs better than the hl-context prediction scheme and outperforms the hl-context prediction approximately by factor 3 (see Figs. 14a and 14b) regardless of the sampling interval.

However, during the first week of simulation the hl-context prediction scheme performs better. Due to the higher context abstraction level of the hl-context history, the patterns observed in this period are more general and of a simpler structure. At times when only few, easily distinguishable patterns are present, the higher context abstraction level simplifies the distinction of time series. However, with the introduction of further context time series that are harder to distinguish, the higher context abstraction level becomes a hindrance.

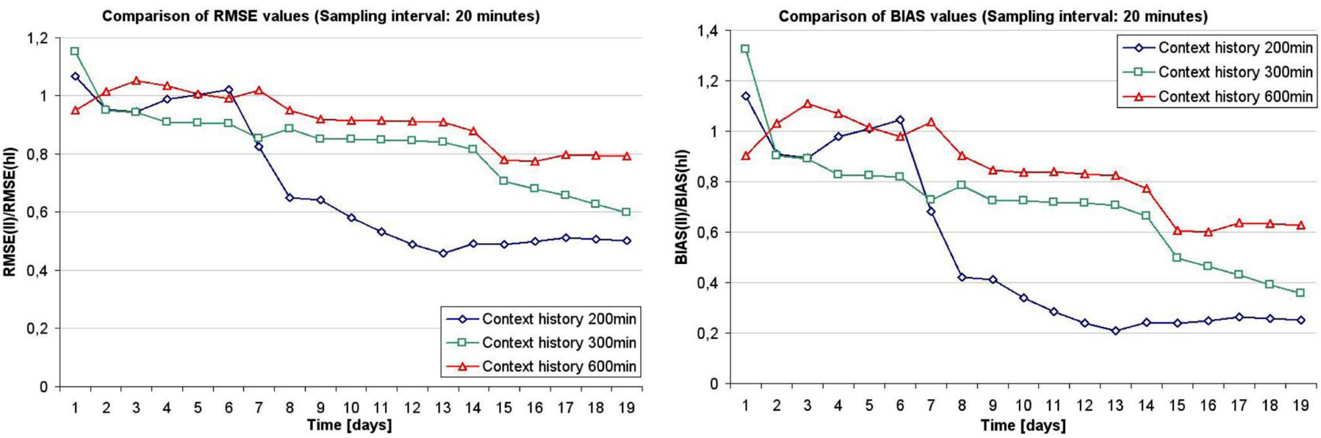
When the context prediction horizon is modified, these general trends stay evident (cf. Fig. 14c). Additionally, with an increasing context prediction horizon, the advantage of the ll-prediction algorithm over the hl-context prediction algorithm increases (compare also Fig. 3).

Finally, we modify the context history length. For these experiments, we chose a sampling interval of 20 minutes and a context history length of 200, 300, and 600 minutes, respectively.

From Fig. 15, we observe that with an increasing context history length, the performance gain of the ll-context prediction scheme over the hl-context prediction scheme decreases (compare also Fig. 9).

In summary, we have observed that the ll-context prediction scheme is advantageous when compared to the hl-context prediction scheme on this location data set. Furthermore, we could observe that the impact of an increasing prediction horizon is more serious for hl- than for ll-context prediction, as it has been suggested by the analytical results in Section 4.

Finally, for an increasing context history size, the accuracy gap between the accuracies of the hl- and ll-context prediction schemes is narrowed. While ll-context prediction schemes better cope with short context histories, this advantage diminishes with an increasing context history size.



(a) Comparison of RMSE values (b) Comparison of BIAS values

Fig. 15. Comparison of ll- and hl-context prediction schemes.

TABLE 2
Context Prediction Accuracy

Int. error	accuracy (HL)	accuracy (LL)	Ratio
0.1	0.964	0.898	1.073
0.2	0.583	0.796	0.732
0.3	0.445	0.702	0.634
0.4	0.331	0.605	0.547
0.5	0.017	0.475	0.036

5.3 Impact of the Interpretation Error

In order to obtain a more complete understanding of the influence of the context interpretation error on prediction, we create synthetic context data with special properties. We also exclude the context acquisition step to focus on the impact of the context interpretation step only.

In the GPS simulation, the impact of the interpretation error is compounded with the impacts of the acquisition and prediction errors as well as with further side effects.

These effects are known as concept drift and can be summarized as hidden changes in contexts as described in [43]. In the GPS-example above, the change of habits as well as new colleagues or project partners might constitute a concept drift.

For interpretation, we provide a known mapping between the ll and hl-contexts. The error probability of this module is configurable.

The context interpretation error is varied in different simulation runs from 0.1 to 0.5. We decided for a uniform distribution of errors. In this simulation, we describe the accuracy by the fraction of accurately predicted contexts to the number of predicted contexts.

We utilize four distinct simple one-dimensional, real-valued, ll-context patterns with 41 elements each. Patterns 1 and 2 contain linearly increasing values from 0 to 20 and from 0 to 40, respectively, while for patterns 3 and 4 the values linearly decrease from 40 to 0 and from 20 to 0. In the course of the experiment, we repeatedly choose one of these patterns with a uniform distribution and feed it into the context prediction architecture.

The results of this simulation are illustrated in Table 2. With a context interpretation error of 0.2 or higher, the ll-prediction method achieves better accuracy. While it might be feasible for some applications to utilize the ll-context prediction scheme with low interpretation accuracies, the accuracy of the hl-context prediction scheme diminishes at such a fast pace that it becomes infeasible for arbitrary applications. The greater impact of the interpretation error on the prediction accuracy of the hl-prediction approach was also predicted by our analytic findings (cf. (6)).

5.4 Impact of the Input Dimension

In this section, we study the influence of a varying number of input data sources used and also vary the size of the context history. In this simulation, the same modules as in Section 5.3 are used. Furthermore, we increase the time series dimension, where a maximum of 10 dimensions are applied in the simulation. We use 12 different time series of data values for each dimension, resulting in 120 distinct time series overall. The acquisition and interpretation error probabilities are set to $P_{acq} = 0.98$ and $P_{int} = 0.94$, respectively.

TABLE 3
Error Probability Ratios (P_{ll}/P_{hl})

Context history size	5	10	15	20
Dimension 1	0.93	0.62	0.58	0.25
Dimension 5	0.96	0.90	0.69	0.64
Dimension 10	0.96	0.94	0.91	0.90

For the interpretation error, we assume a uniform distribution of possible errors, while we apply a Gaussian distribution for the acquisition error. The Gaussian distribution models the property that small errors are more reasonable than substantial errors in the acquisition module.

In each simulation run, we chose 12 context time series out of the pool of time series one after another following a uniformly random distribution and subsequently feed them into the architecture.

In Table 3, we depict the fraction of the results obtained by the context prediction based on ll-context elements to the results obtained by the hl-context prediction scheme.

With increasing time series dimension, the predominance of the ll-context prediction scheme above the hl-context prediction scheme diminishes while with increasing context history size the predominance of the ll-context prediction scheme above the hl-context prediction scheme increases (compare also Fig. 6).

The number of erroneous contexts in the input time series is higher for hl-context prediction schemes and increases with increasing context history length. With more errors in the input time series, the context prediction accuracy consequently decreases.

Another trend visible in Table 3 is that the dominance of the context prediction based on ll-context elements diminishes with increasing dimension of the context history.

6 Conclusion

We have studied the impact of the order of context processing operations on the accuracy of the processing result. We also considered the application of context prediction at various context abstraction levels with several examples.

The impact of distinct input parameters on the context prediction accuracy of hl- and ll-context prediction schemes was considered. These parameters are the length of the context history, the dimension of the observed context sequence, the dimension of the hl-context sequence as well as the number of distinct values for hl- and ll-contexts.

We could show that these parameters have a different impact on the prediction accuracy depending on the order in which the context processing operations acquisition, interpretation, and prediction are applied.

Also, the error probabilities for the context processing operations impact the prediction accuracy differently when the order of processing operations is altered. As a major contribution of our study, we derive probabilistic formulas that describe the overall error probability for a specific set and order of context processing operations.

Regarding the dimensions of the input time series, we observed that a higher prediction accuracy for an increased dimension of the input time series can be achieved when context prediction is applied after the context interpretation process.

A converse effect was observed for the length of the context history. With an increasing context history length, the prediction accuracy is higher when context prediction is applied prior to the context interpretation process.

Furthermore, the accuracy of the context interpretation has a significant impact on the context prediction accuracy. In particular, for increasing error probabilities of the context interpretation operation, we observed a tendency that the prediction accuracy is higher when prediction is applied prior to the context interpretation process.

In summary, in a scenario where the context interpretation operation can hardly cope with the noisy input data, it is more beneficial to apply context prediction in advance of the context interpretation process. When, however, context interpretation is highly accurate, the application of context prediction after the context interpretation might yield improved context prediction accuracy.

We also observed that the context prediction accuracy is tightly linked to the context acquisition accuracy. Consequently, the main focus of the application designer should be on the context acquisition procedure. Furthermore, designers of context prediction architectures have to consider the ratio of prediction to interpretation accuracy. The number of context types available, however, has only a minor influence on the context prediction accuracy.

For all these analytically derived results, we have conducted experimental and simulation studies to confirm the analytic findings. The experimental studies are situated in mobile Ubiquitous Computing settings. In a first study, a group of users completed predefined sequences of actions that have been sampled by temperature, light, and vibration sensors. In a second study, we monitored the GPS-trajectory of a mobile subject using latitude, longitude, and altitude as input data. The results from these experiments confirm the results from our theoretical analysis.

A major result we show in the theoretical analysis and by means of experiments is that the nature of the input data, the quality of the output and the construction of a flow of processing operations to achieve a prediction are correlated. In particular, we expect greater accuracy of context prediction when either the input data for context prediction that is pre-processed by other context processing operations has a high accuracy or when otherwise context prediction is applied in advance of further context processing operations.

Acknowledgments

The authors are grateful for the meticulous proofreading of Rostom Kilani who spotted errors in the captions of some figures in the process of revising the paper. They would like to acknowledge partial funding by the "Deutsche Forschungsgemeinschaft" (DFG) for the project "SenseCast." They would further like to acknowledge partial funding by the European Commission for the ICT project CHOSeN "Cooperative Hybrid Objects Sensor Networks" (Project number 224327, FP7-ICT-2007-2) within the 7th Framework Programme.

References

[1] W.N. Schilit, "A System Architecture for Context-Aware Mobile Computing," PhD thesis, Columbia Univ., 1995.
[2] A.K. Dey, "Providing Architectural Support for Building Context-Aware Applications," PhD thesis, Georgia Inst. of Technology, 2000.
[3] A. Schmidt, "Ubiquitous Computing—Computing in Context," PhD thesis, Lancaster Univ., United Kingdom, 2002.
[4] G. Chen, "Solar: Building a Context Fusion Network for Pervasive Computing," PhD thesis, Hanover, New Hampshire, 2004.
[5] G. Chen, M. Li, and D. Kotz, "Design and Implementation of a Large-Scale Context Fusion Network," *Proc. First Ann. Int'l Conf. Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous '04)*, 2004.
[6] G. Chen and D. Kotz, "Context Aggregation and Dissemination in Ubiquitous Computing Systems," *Proc. Fourth IEEE Workshop Mobile Computing Systems and Applications (WMCSA '02)*, vol. 105, 2002.
[7] B.N. Schilit and M.M. Theimer, "Disseminating Active Map Information to Mobile Hosts," *IEEE Network*, vol. 5, no. 5, pp. 22-32, Sept./Oct. 1994.
[8] J. Mäntyjärvi, "Sensor-Based Context Recognition for Mobile Applications," PhD thesis, VTT Technical Research Centre of Finland, 2003.
[9] R.M. Mayrhofer, "An Architecture for Context Prediction," PhD thesis, Johannes Kepeler Univ. of Linz, Austria, 2004.
[10] P. Nurmi, M. Martin, and J.A. Flanagan, "Enabling Proactiveness through Context Prediction," *Proc. Workshop Context Awareness for Proactive Systems (CAPS '05)*, 2005.
[11] J. Petzold, "Zustandsprädiktoren zur Kontextvorhersate in Ubiquitären Systemen (in German)," PhD thesis, Univ. of Augsburg, 2005.
[12] B.D. Davison and H. Hirsh, "Predicting Sequences of User Actions," *Proc. Workshop Predicting the Future: AI Approaches to Time-Series Analysis (AAAI/ICML)*, 1998.
[13] S. Sigg, S.L. Lau, S. Haseloff, and K. David, "Approaching a Definition of Context Prediction," *Proc. Third Workshop Context Awareness for Proactive Systems (CAPS '07)*, 2007.
[14] R.M. Mayrhofer, H. Radi, and A. Ferscha, "Recognising and Predicting Context by Learning from User Behaviour," *Proc. Int'l Conf. Advances in Mobile Multimedia (MoMM '03)*, vol. 171, pp. 25-35, 2003.
[15] D.J. Cook, M. Youngblood, E. Heierman, K. Gopalratnam, S. Rao, A. Litvin, and F. Khawaja, "Mavhome: An Agent-Based Smart Home," *Proc. First IEEE Int'l Conf. Pervasive Computing and Comm. (PerCom '03)*, pp. 521-524, 2003.
[16] S.K. Das, D.J. Cook, A. Bhattacharya, E.O. Heierman, and T.Y. Lin, "The Role of Prediction Algorithms in the Mavhome Smart Home Architecture," *IEEE Wireless Comm.*, vol. 9, no. 6, pp. 77-84, Dec. 2002.
[17] P. Gorniak and D. Poole, "Predicting Future User Actions by Observing Unimodified Applications," *Proc. Conf. Am. Assoc. for Artificial Intelligence*, 2000.
[18] B. Korvemaker and R. Greiner, "Predicting Unix Command Lines: Adjusting to User Patterns," *Proc. 17th Nat'l Conf. Artificial Intelligence and 12th Conf. Innovative Applications of Artificial Intelligence*, pp. 230-235, 2000.
[19] A. Roy, S.K.D. Bhaumik, A. Bhattacharya, K. Basu, D.J. Cook, and S.K. Das, "Location Aware Resource Management in Smart Homes," *Proc. First IEEE Int'l Conf. Pervasive Computing and Comm. (PerCom '03)*, 2003.
[20] S. Sigg, S. Haseloff, and K. David, "A Novel Approach to Context Prediction in Ubicomp Environments," *Proc. 17th Ann. IEEE Int'l Symp. Personal, Indoor and Mobile Radio Comm. (PIMRC '06)*, 2006.
[21] S. Sigg, S. Haseloff, and K. David, "An Alignment Approach for Context Prediction Tasks in Ubicomp Environments," *IEEE Pervasive Computing*, vol. 9, no. 4, pp. 90-97, Oct.-Dec. 2010.
[22] J. Cadzow and K. Ogino, "Adaptive ARMA Spectral Estimation," *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing*, vol. 6, pp. 475-479, 1981.
[23] L. Capra and M. Musolesi, "Autonomic Trust Prediction for Pervasive Systems," *Proc. 20th Int'l Conf. Advanced Information Networking and Applications*, pp. 481-488, 2006.
[24] M. Musolesi and C. Mascolo, "Evaluating Context Information Predictability for Autonomic Communication," *Proc. Second IEEE Workshop Autonomic Comm. and Computing*, 2006.

[25] C. Chapman, M. Musolesi, W. Emmerich, and C. Mascolo, "Predictive Resource Scheduling in Computational Grids," *Proc. 21st Int'l Parallel and Distributed Processing Symp.*, 2007.
[26] M.J. Goris, D.A. Gray, and I.M. Mareels, "Reducing the Computational Load of a Kalman Filter," *IEE Electronics Letters*, vol. 33, no. 18, pp. 1539-1541, Aug. 1997.
[27] N. Eagle and A.S. Pentland, "Eigenbehaviors: Identifying Structure in Routine," *Behavioral Ecology and Sociobiology*, vol. 63, no. 7, pp. 1057-1066, 2009.
[28] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
[29] Q. Du and J.E. Fowler, "Low-Complexity Principal Component Analysis for Hyperspectral Image Compression," *J. High Performance Computing Applications*, vol. 22, pp. 438-448, 2008.
[30] L. Song, D. Kotz, R. Jain, and X. He, "Evaluating Next-Cell Predictors with Extensive Wi-Fi Mobility Data," *IEEE Trans. Mobile Computing*, vol. 5, no. 12, pp. 1633-1649, Dec. 2006.
[31] L. Song, U. Deshpanade, U.C. Kozat, D. Kotz, and R. Jain, "Predictability of Wlan Mobility and Its Effects on Bandwidth Provisioning," *Proc. 25th IEEE Int'l Conf. Computer Comm.*, 2006.
[32] K. Gopalratnam and D.J. Cook, "Active LeZi: An Incremental Parsing Algorithm for Sequential Prediction," *Int'l J. Artificial Intelligence Tools*, vol. 14, pp. 917-930, 2004.
[33] J. Cleary and I. Witten, "Data Compression Using Adaptive Coding and Partial String Matching," *IEEE Trans. Comm.*, vol. 32, no. 4, pp. 396-402, Apr. 1984.
[34] P. Jacquet, W. Szpankowski, and I. Apostol, "An Universal Predictor Based on Pattern Matching, Preliminary Results," *Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities*, pp. 75-85, Birkhauser, 2000.
[35] D. Gordon, H.R. Schmidtke, M. Beigl, and G. von Zengen, "A Novel Micro-Vibration Sensor for Activity Recognition: Potential and Limitations," *Proc. 14th Int'l Symp. Wearable Computers (ISWC '10)*, 2010.
[36] P.A. Pevzner, *Computational Molecular Biology—An Algorithmic Approach.* MIT, 2000.
[37] S.B. Needleman and C.D. Wunsch, "A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins," *J. Molecular Biology*, vol. 48, no. 3, pp. 443-453, 1970.
[38] I.H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, second ed. Morgan Kauman, 2005.
[39] J.R. Quinlan, *C4.5: Programs for Machine Learning*, first ed. Morgan Kaufmann, 1993.
[40] L. Bao and S.S. Intille, "Activity Recognition from User-Annotated Acceleration Data," *Proc. Second Int'l Conf. Pervasive Computing*, 2004.
[41] E. Miluzzo, N.D. Lane, K. Fodor, R. Peterson, S.B. Eisenman, M. Musolesi, X. Zheng, H. Lu, and A.T. Campbell, "Sensing Meets Mobile Social Networks: The Design, Implementation and Evaluation of the Cenceme Application," *Proc. Sixth ACM Conf. Embedded Network Sensor Systems*, pp. 337-350, 2008.
[42] E.M. Tapia, S.S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman, "Real-Time Recognition of Physical Activities and Their Intensities Using Wireless Accelerometers and a Heart Rate Monitor," *Proc. IEEE Int'l Symp. Wearable Computers*, 2007.
[43] M. Harries, K. Horn, and C. Sammut, "Learning in Time Ordered Domains with Hidden Changes in Context," *Proc. Workshop Predicting the Future: AI Approaches to Time-Series Problems*, pp. 29-33, 1998.

Stephan Sigg received a diploma in computer sciences from the University of Dortmund, Germany, in 2004 and the PhD degree in 2008 from the chair for communication technology at Kassel University, Germany. He is currently with the Pervasive Computing Systems group (TecO) at the Karlsruhe Institute of Technology (KIT), Germany. His research interests include the design, analysis, and optimization of algorithms for context aware and ubiquitous systems. He is a member of the IEEE.

Dawud Gordon received the BSc degree from the State University of New York at Albany in 2005 and the MSc degree from the Technische Universität Braunschweig. He is currently a research assistant and doctoral candidate at the Karlsruhe Institute of Technology (KIT), Germany. His research interests include context and activity recognition in embedded and distributed systems as well as machine learning and recognition algorithms. He is a student member of the IEEE.

Georg von Zengen is currently working toward the BSc degree in computer science from the Technische Universität Braunschweig, Germany, since 2008. Currently, his work in the Distributed and Ubiquitous Systems Group (DUS) at the Technische Universität Braunschweig focuses on activity recognition and the development of measurement hardware. He is also involved in the development of smart wireless sensor networks.

Michael Beigl received both the MSc and PhD (Dr.-Ing) degrees from the University of Karlsruhe in 1995 and 2000, respectively. He is a professor of Pervasive Computing Systems at the Karlsruhe Institute of Technology (KIT). Previously, he was a professor for Ubiquitous and Distributed Systems at the Technische Universität Braunschweig (2005-2009), research director of TecO, University of Karlsruhe (2000-2005), and visiting associate professor at Keio University (2005). His research interests include wireless sensor networks and systems, ubiquitous computing, and context awareness. He is a member of the IEEE.

Sandra Haseloff is a program director at the Alexander von Humboldt Foundation. Previously, she worked for the Fraunhofer Institute for Software and Systems Engineering (ISST) (1999-2005) and as a postdoctoral senior researcher at the Chair for Communication Technology at Kassel University, Germany. Her research interests include ubiquitous computing, mobile middleware, personalization, and context awareness.

Klaus David received a diploma and the PhD degree from the University of Siegen, Germany, in 1988 and 1992, respectively. He has 12 years of industrial experience with HP, Bell Northern Research, IMEC, T-Mobile (head of group), and IHP (head of department). He has been a professor since 1998 and has been the head of the Chair for Communication Technology (ComTec) at Kassel University, Germany, since 2000. His research interests include mobile applications and context awareness. He is a member of the IEEE.