Serendipitous Recommendation based on Big Context

Andrew Koster, Fernando Koch and Yeun Bae Kim

SAMSUNG Research Institute
{andrew.k, fernando.koch, kimybae}@samsung.com

Abstract. Context-awareness is an essential requirement in crafting recommender systems that provide serendipity, *i.e.* "pleasant surprises", independently of human command. These solutions must be able to infer interactions based on data from sensors and recognised activities in order to infer what is useful information and when to deliver it. For that, we are devising advanced models of context inference based on the analysis of users' signals during everyday activities. In this paper, we present a proof-of-concept platform that allows for the application of techniques of deep learning and context analytics to derive patterns in spatio-temporal context signals. We call this composition *Big Context*. We argue that by understanding how people and things are connected, one can devise novel forms of interactions that provide a more pleasant user experience. In this work, we introduce our method and platform, and illustrate some of the possible techniques using a prototype application that provides serendipitous recommendations.

1 Introduction

Serendipitous interfaces are an emerging paradigm in user experience. The idea is to create a self-governing recommender system that provides relevant information [18] as a "pleasant surprise" in the absence of intentional commands. Examples of such interfaces are: SAMSUNG S-Health[1], when it counts how many steps a user walked during the day and offers a congratulatory notification once a (supposedly healthy) threshold has been surpassed, and; Google Now[2], when it keeps track of your location and daily activities to provide contextualised advise. These sort of systems encompass: (i) sensors that collect multi-dimensional data; (ii) mechanisms of context inference in mobile computing; and (iii) a deliberation process to infer what is *useful information* and when to deliver it.

We hypothesise that it is possible to create inference models that classify and understand user behaviour based on analysis of events emitted by the handling of everyday things, such as smart phones, Smart TVs, fridges, and air conditioning. We investigate how to apply deep learning and context analytics

[1] http://www.samsung.com/global/microsite/galaxys4/lifecare.html#page=shealth
[2] http://www.google.com/landing/now

to derive patterns and correlations in spatiotemporal context signals — we call this configuration *Big Context*. The leading questions in our investigation are:

- How to infer meaningful context events out of multi-dimension signals collected from a variety of sensors?
- How to reason explanation about facts or incidents with multi-aspectual proofs based on pre-classified events?
- How to design deliberation mechanisms aiming at context support and proactive interactions?

A brief illustration scenario is as follows. Let us assume that smart phones are instrumented to emit signals s_i containing information such as time, location, type of object, type of interaction and parameters of the interaction, which are captured and stored in repositories $\mathbf{S}_u$ for each user u. Then, sequences of signals $S \subseteq \mathbf{S}_u$ can be classified as context events e_i by applying techniques of sequence labelling algorithms, such as a Hidden Markov Model [4] and Conditional Random Fields (CRF) [12]. For instance, to classify sequences $s_1, \ldots, s_n$ as car parking signature, stored in a spatiotemporal repository $\mathbf{E}$. Moreover, techniques for learning probabilistic models for collections of discrete data (*e.g.* Latent Dirichlet Allocation [1]) are applied to identify patterns from the excess of spatiotemporal events that relate to salient contextual situations.

We expect to understand the context and deliberate to provide a "pleasant surprise". In this case, by informing that there is a high possibility that a parking space is to be freed up soon (around a area). The system infers this information by classifying the social behaviour upon historic context events from $\mathbf{E}$. Different scenarios may be drawn up, considering signals incoming from alternative objects and application domains, such as the provision of contextualised education material, introduced in [10].

In what follows, we introduce *Big Context* and related concepts, present the architecture of the "Sensible Lives Platform", and propose a proof-of-concept demonstration.

2 On Big Context and Related Concepts

Dey (2001) [3] characterises context in the following way: "Context is any information that can be used to characterise the situation of an entity. An entity is a person, place or object that is considered relevant to the interaction between a user and an application, including the user and the application themselves." Current approaches to context acquisition tend to focus on a bridge between the high-level context descriptions in applications and the low-level data that is collected from sensors [6, 11, 20]. We claim that current methods are insufficient to handle highly diversified views that can be derived from the multitude of sensors present on everyday devices. The ACE system [15] is nearest to our work in this sense, because it discovers the relations between contexts. We will discuss this further in Section 3.3, where we discuss our approach to the semantic modelling of contexts.

We are researching alternative techniques for multi-dimensional context analysis. We aim at new models to recognise patterns in data and then match these to contextual situations without previous understanding about these patterns or the configuration of analysis rules. We call this configuration *Big Context*.

This development seeks a new kind of context support mechanism that provides highly adaptive context inference, even to unanticipated situations where emerging patterns can be discovered. It will support real-time context-aware inference towards proactive deliberation and serendipity. Insofar as we know, there is only one similar platform, CQue [16], which aims at integrating information from various different classifiers in order to improve accuracy, reduce battery usage and provide better privacy. The main distinction is that CQue is intended to run on the mobile phone, and thus aims mainly at the detection of contexts that can be learned through analysis of a single user's information. By running the Big Context platform as a cloud-based service, we are able to leverage multiple users' contextual information to infer social context, as well as combine data from many users to better train classifiers, allowing us to recognise the occurence of rare events.

Figure 1 depicts the architecture of the Big Context platform. We integrate multiple different learning techniques, which we will describe in the next section, in an encompassing framework for collecting and reasoning upon multi-dimensional sensors. The composition includes interface support for deliberation mechanisms aiming at context support and pro-active interactions. It operates as follows:

- *Input Interface* receives signals captured by sensors in everyday life's devices.
- *Noise Filtering and Fusion* pre-processes this data eliminating entries that are either repetitive or not relevant; these signals are stored in a short-term repository for future reference.
- *Probabilistic Activity Recognition* analyses a sequence of signals, identifying the impact of each signal upon the sequence given the observations; recognised activity patterns are stored in a long-term repository for historic context analysis.
- *Probabilistic Situation Modelling* implements techniques of context modelling based on generative probabilistic models for collections of discrete data; recognised situation patterns are stored in a long-term repository for historic context analysis.
- *Semantic Context Classification* provides support to reason about facts or incidents with multi-aspectual proofs, using configurable *Context Analysis Rules* and *Semantic Rules*. The generated entries are stored in a long-term repository for historic context analysis.
- *Context Query Language* provides a query language to access the information stored in the four repositories being populated by the methods aforementioned.

The service can be executed in a combination between local- and server-based processing. For instance, in the prototype being presented in the next section,

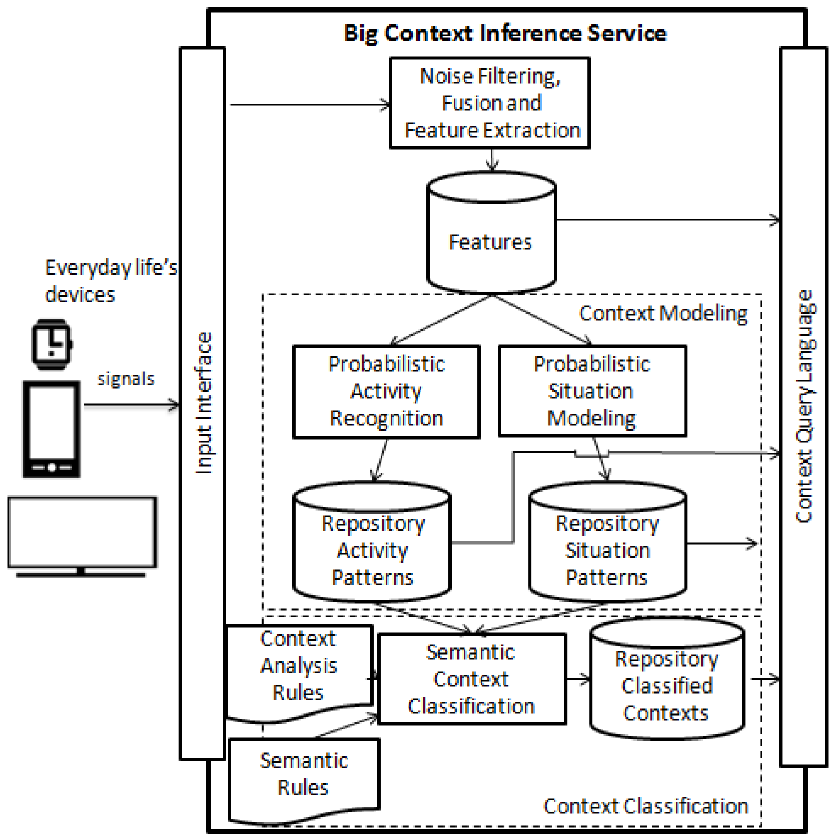

Fig. 1. Architecture of Big Context

we execute the methods for noise filtering and probabilistic activity recognition on the mobile device, whilst executing the information sharing services on the server.

This composition supports external applications to query for shared context information band intertwine local processing and global context. We claim that this setup greatly facilitates the implementation of serendipity, providing recognition of surrounding to support automatic processing.

Next we describe the main elements that compose this mechanism. Throughout this paper, we will demonstrate the concepts with an application that recognises parking spaces. We describe the algorithms we use for this specific situation, and briefly survey some of the other methods that can be used for the task at hand. The mobile app and its functioning is described in Section 4.

3 Context Recognition Methods

3.1 Activity recognition

Activity features are not obvious when collecting data from multi-dimensional sensors. In order to categorise contextual patterns we need to identify powerful distinctive features out of data that is not easily explainable. Activity recognition is a temporal classification problem: the application must analyse a sequence of signals, identifying the impact of each signal upon the sequence given the observations [19].

For instance, let us consider the method for detecting car parking signatures from observed movements of user's device captured through sensors like accelerometer and gyroscope. The application collects and indexes timestamped signals $s_0, \ldots, s_t$ over a period of time t, and classifies discrete features to facilitate the machine learning. That is, features are considered together within a timeframe forming linear sequence of events $e_0, \ldots, e_t$, which represent discrete words w_i in a domain; for instance, distinct movements during car parking movement such as "reverse acceleration" (w_1), "slight right turn" (w_2), "slight left turn" (w_3), and so on.

Out of the various context modelling tasks, activity recognition is probably the one that has received the most attention so far. There are many different works focusing on recognising many different activities, using various different sensors. It is beyond the scope of this work to perform a sufficient survey, for which we refer to Mannini and Sabatini [14] regarding activity recognition using inertial sensors, and Poppe [17] for vision-based activity recognition.

A majority of the methods for recognising activity based on IMU sensor data, like the data we collect from drivers, use a method based on graphical models. We follow suit, and employ Conditional Random Fields (CRF) [12], one of the methods that places the least assumptions on the conditional independence relationship between variables. Nevertheless, an assumption underlying all graphical models is that the labels have the Markov property $Pr\{e_{t+1} =$

$e', s_{t+1} = s|e_t, s_t\}$, that is the conditional probability distribution of future labels depends only upon the present labels, not on the sequence of events e_i and signals s_j that preceded it.

The aim of using CRF is to model the joint probability distribution for all the labels, given the observed features. It models this conditional probability distribution of the labels given the features as follows:

$$Pr(\mathbf{E}|\mathbf{S}) = \frac{1}{Z}\prod_{i=0}^{t} exp\big(\sum_{k=1}^{K}\theta_k f_k(e_{i-1}, e_i, s_i)\big),$$

where Z is a normalisation factor, $\theta_k \in \mathbb{R}$ are parameters of the model and f_k is a real-valued *feature function* for each of the K features we are interested in, such as w_1, w_2 and w_3 of the above example.

This technique yields the most likely explanation for the sequence of events. That is, it computes the joint probability of the entire sequence of hidden states that generated a particular sequence of observations within a distinct feature, such as "parking in signature", "parking out signature", or "non parking signature". Then, we can form a repository of historical activity patterns $\mathbf{E}$ where each record contains timestamp, location, user identification and distinct feature[3] by capturing the information from multiple users over time.

3.2 Situation modelling

We expect that patterns also emerge at a global context level considering the relation between events across multiple users. These patterns constitute the situation a user is in, and situation modelling is generally approached in two ways: using a probabilistic method, or a rule-based semantic method. The latter is problematic, as it requires a prior description of the possible situations that we wish to model, whereas the former can be *generative* and discover novel situations. While the application of such methods to situation modelling is relatively new [5, 13], similar methods have a rich history in understanding text. Examples of this type of method are Latent Dirichlet Allocation (LDA) [1, 9], Topical N-gram Models [21] and Beta Process Hidden Markov Models [7, 8].

These techniques allows us to classify and predict situations that involve subsets of context events and address questions like: (i) what is the probability that a frequency of context events over a period of time and region represents a situation x, such as a user looking for a parking space (x_1) or a user walking towards his car (x_2); or (ii) what is the probability of situation x being caused by a (subset of) context events E?

Currently, we do not have enough data to perform this kind of modelling in the parking place scenario, but when we gather more data we propose to use LDA, one of the most common generative probabilistic methods in topic modelling, to discern the various situations we are interested in. The LDA method

[3] Note: for privacy concerns, one can prevent to store user identification on $\mathbf{E}$; however, this approach limits the possibility of individual situation analysis at global context level.

operates as follows. The set $E_{R,T} \subseteq \mathbf{E}$ represents the spatiotemporal area of events to be considered, over a period of time T and within a spatial region R. We define the following notation:

- K denotes the number of situational descriptors;
- M specifies the number of areas being classified;
- N is the number of events per area;
- α is the parameter of the Dirichlet prior on the per-area distribution of situations;
- β is the parameter of the Dirichlet prior on the per-situation distribution of events;
- θ_m is the situation distribution for each area E_{R_m,T_m};
- ϕ_x is the distribution of context events for situation x;
- e_{mn} is the n^{th} context event in the m^{th} area, and;
- x_{mn} is the situation that e_{mn} belongs to.

Then the joint probability distribution we are interested in is (full explanation at [1]):

$$Pr(\mathbf{E}, \mathbf{X}, \theta | \alpha, \beta) = \prod_{k=1}^{K} Pr(\phi_k | \beta) \times \prod_{m=1}^{M} Pr(\theta_m | \alpha) \prod_{n=1}^{N} Pr(x_{mn} | \theta_m) Pr(e_{mn} | \phi_{x_{mn}})$$

The method allows for the inference of various different probabilities, allowing us to answer questions like the ones above. For instance, an event might be the user driving slowly. In isolation, this activity could mean that he is lost, stuck in traffic, or searching for a parking space. However, in conjunction with other events the situation becomes clear. For instance, if many other users are driving slowly in the same area, the situation is most likely a traffic jam, while if this is a recurrent pattern for the user, in isolation, over various days, he is most likely to be searching for a parking space.

3.3 Semantic context classification

Patterns also emerge at a social level where people use their devices differently depending on their social environment. For that, rule-based classification methods provide a powerful tool to reason explanation about facts or incidents with multi-aspectual proofs based on pre-classified events. Moreover, these methods allow for the composition of *context intelligence* through the discovery of new facts and rules/patterns from networked context facts by exploring causality residing inside the relation between events, as presented in [2].

Consider the question: What is the possibility that a parking space is to be freed up soon around this area?. The reasoning implies the analysis of historical behaviour and other parameters of the social settings. For instance, let us assume the semantics rules in the knowledge space that represent the number of parking spaces available, such as: "parking at position", "leaving and entering parking", "startend of office hours", "parking spaces become available or unavailable", and so on. Then, we can compose logic formulae between the elements as:

$who\_am\_I(?U)$
$local\_query(current\_location(?L))$
$server\_query(parking\_spaces(?Y, L))$
$possible\_office\_arrival\_time(U, ?TO, ?\Theta_x)$
$my\_office\_location(U, ?LO)$
$possible\_parking\_availability(LO, TO, ?Z, ?\Theta_z)$

Thus, the reasoning can infer the possibility of finding parking places at my office's location around the time I am about to arrive at my office. In this reasoning, it estimates the possible office arrival time TO with a certainty Θ_x e.g. based on historical information and/or driving information. it also infers the office location LO and thus is able to estimate the possible parking availability Z near that area with a certainty factor Θ_z. Intuitively, the more networked elements are being considered for the assertion, the more likely the prediction will align with observable features.

The use of a semantic layer in context-aware computing is not novel. As mentioned in the previous section, it is often used in situation modelling, but the use most similar to our proposal is found in ACE [15]. Their principal aim is to improve energy efficiency by recognising when sensors are needed, and when the context can be inferred through logical rules. For instance, if the user is driving (known because he is tethered to his hands-free set), then he is not at home, and thus there is no need to use the GPS to discover whether he is at home or not. These semantic relations are learned using a logical rule miner similar to the one we propose to use, however it learns Boolean rules. This is a problem, because everything else is based on probabilistic inference: they do not know the user is driving, they just infer it, possibly with a high probability. This is exacerbated, because the rules themselves are learned, there is a chance they are wrong, and thus come associated with some form of confidence value that the rule is correct. We thus learn probabilistic rules, which account for the inherent uncertainty in the domain.

4 Serendipitous Parking Recommendations

As a proof-of-concept implementation, we developed an application to find free parking spaces and also help the user locate his parked car. We intended to exploit concepts of non-intentional interactions, pro-active recommendations, and social connectedness. Figure 2 depicts the two screens in the wePark Application, and an overview of its classifier: (i) a widget that collects data (accelerometer, gyroscope, and GPS) on the background; (ii) the main screen that automatically presents free parking spaces around the user's location; and (iii) the classification process, explained below.

(a) Widget showing number of available spots.

(b) Map display of the available parking spots.

Fig. 2. Prototype wePark application

The application works as depicted in Figure 3 . The widget operates autonomously collecting (a) data samples and pre-process locally for noise filtering and probabilistic activity recognition. Features are being extracted based on acceleration data in the XY plane divided into nine regions, turns around the Z-axis, clockwise or counter clockwise, movement speed, and movement duration. The application extracts the features using information fusion techniques, followed by discretisation, yielding (b) a sequence of movement features. We then apply a Conditional Random Fields (CRF) model to classify the sequence of features into (c) probabilities of parking movement events.

Once the application collects a sufficient amount of data, the information is grouped and labelled to know which clusters are movements corresponding parking movements. These events are communicated to a server-based application that: stores these events in repository of activity patterns, so that the server knows where the parking spaces are being freed up/occupied; and executes the probabilistic situation modelling to create the global context considering the relation between events across multiple users.

When the user taps on the wePark widget, it opens the (ii) main screen (Figure 2 (ii)) that queries the server-based component about free parking spaces around the current location. Moreover, the widget also detects walking patterns and when the user is wandering around the area where his car is parked, the application notifies the user, providing direction information where his car is parked.

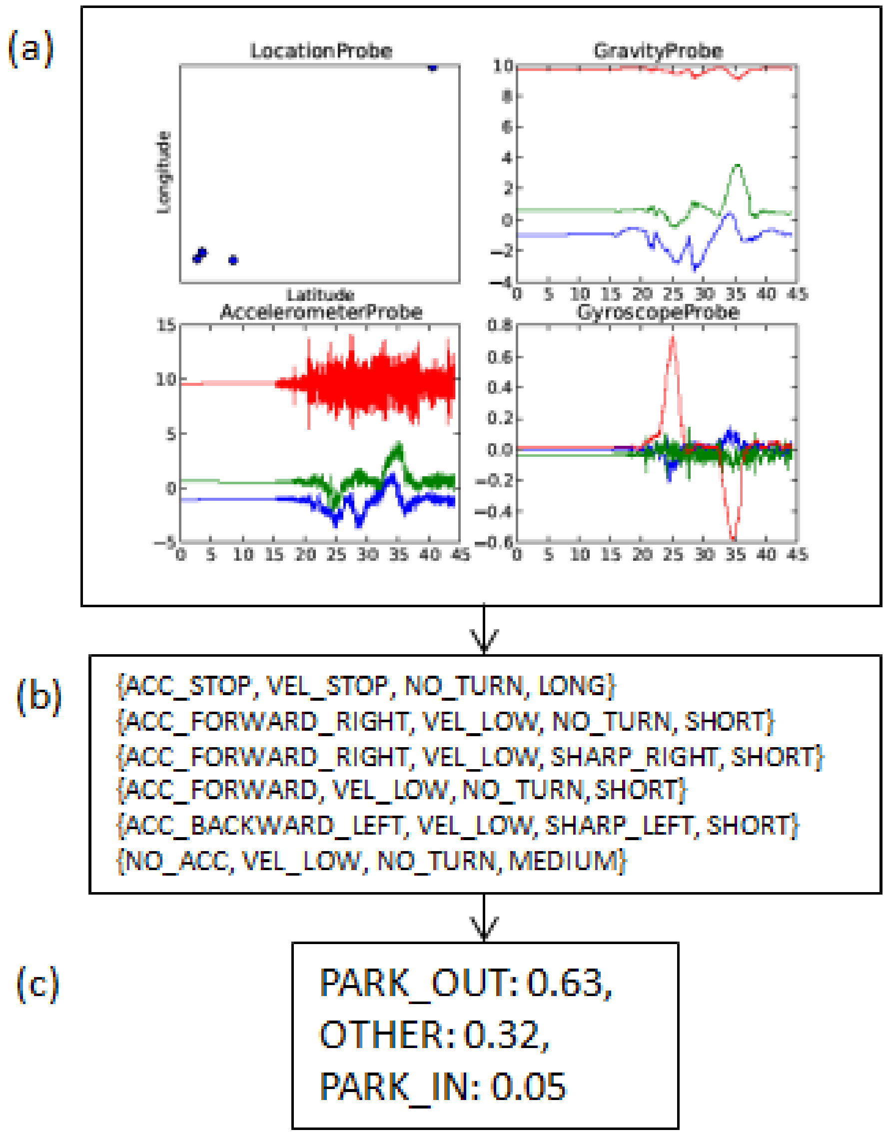

Fig. 3. Processing context signals for recognizing parking movements. (a) is a sample of sensor data, varying over time, collected from a 9DOF IMU, (b) shows the feature set that we extract from this data, and (c) the probability for each of the potential labels of this sample.

5 Discussion

We seek a new kind of user experience, understanding user activities and delivering relevant information pro-actively. We expect that these developments will improve the quality of mobile applications with a new paradigm of user experience. In this paper, we present our work towards *Big Context* and support to serendipitous recommendations using our *Platform for Sensible Lives.* Nevertheless, these methods come at a cost. Firstly, it is necessary to collect and maintain the database of raw contextual data required to discover contextual patterns. Secondly, it requires local processing and continuous access to sensors, which in the case of mobile devices implies in battery utilisation. There are technical solutions to mitigate this problem, such as SAMSUNG Software Development Kits (SDKs) for the latest device technologies (*viz.* Galaxy S5 and Samsung Gear 2), which includes the "Motion package" with features for continuous sensor monitoring with low power technology.

The development of Big Context and serendipitous interfaces is still experimental. We have ideas about initial applications, as described in this paper, and how to move forward with them. Nevertheless, the concept of Big Context is larger than the applications we can conceive of at this early stage. In future work we will refine the method of situation modelling, aiming to provide better support to group support in applications that relate context events across multiple users. Moreover, we will further develop the technology for semantic context classification, aiming at support to social connectedness by supporting applications that analyse patterns at a social level where people use their devices differently depending on their social environment.

References

1. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of Machine Learning Research **3** (2003) 993–1022
2. Choi, K.S., Kim, Y.B.: Knowledge seeking activities for content intelligence. In: Digital Libraries: Technology and Management of Indigenous Knowledge for Global Access. Springer (2003) 415–426
3. Dey, A.: Understanding and Using Context. Personal and Ubiquitous Computing **5** (2001) 4–7
4. Elliott, R.J., Aggoun, L., Moore, J.B.: Hidden Markov Models. Springer (1995)
5. Fischer, Y., Beyerer, J.: Defining dynamic bayesian networks for probabilistic situation assessment. In: Proc. of International Conference on Information Fusion (FUSION), IEEE (2012) 888–895
6. Fonteles, A.M., Neto, B.J.A., Maia, M., Windson, V., Andrade, R.M.C.: An adaptive context acquisition framework to support mobile spatial and context-aware applications. In: Web and Wireless Geographical Information Systems. Volume 7820 of LNCS., Springer (2013) 100 – 116
7. Fox, E.B., Sudderth, E.B., Jordan, M.I., Willsky, A.S.: Sharing features among dynamical systems with beta processes. In: Conference on Neural Information Processing Systems (NIPS'09), Vancouver BC, Canada (2009) 549–557

8. Hughes, M., Suddereth, E.B.: Nonparametric discovery of activity patterns from video collections. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Providence RI, USA (2012) 25–32
9. Huỳnh, T., Fritz, M., Schiele, B.: Discovery of activity patterns using topic models. In: ACM International Joint Conference on Pervasive and Ubiquitous Computing (Ubicomp'08), Seoul, Korea, ACM (2008) 10–19
10. Koch, F., Rao, C.: Towards massively personal education through performance evaluation analytics. International Journal of Information and Education Technology **4**(4) (2014) 297–301
11. Kramer, D., Kocurova, A., Oussena, S., Clark, T., Komisarczuk, P.: An extensible, self contained, layered approach to context acquisition. In: Proceedings of the Third International Workshop on Middleware for Pervasive Mobile and Embedded Computing, ACM (2011)
12. Lafferty, J.D., McCallum, A., Pereira, F.C.N.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proceedings of the Eighteenth International Conference on Machine Learning (ICML), San Francisco, CA, USA, Morgan Kaufmann (2001) 282–289
13. Lopez, G., Brena, R.: Probabilistic situation modeling from ambient sensors in a health condition monitoring system. In: Proc. of UCAmI, Springer (2013) 175–182
14. Mannini, A., Sabatini, A.M.: Machine learning methods for classifying human physical activity from on-body accelerometers. Sensors **10**(2) (2010) 1154–1175
15. Nath, S.: ACE: Exploiting correlation for energy-efficient and continuous context sensing. In: Proc. of MobiSys'12, ACM (2012) 29–42
16. Parate, A., Chiu, M.C., Ganesan, D., Marlin, B.M.: Leveraging graphical models to improve accuracy and reduce privacy risks of mobile sensing. In: Proc. of MobiSys'13, ACM (2013) 83–96
17. Poppe, R.: A survey on vision-based human action recognition. Image and vision computing **28**(6) (2010) 976–990
18. Ricci, F., Rokach, L., Shapira, B.: Introduction to recommender systems handbook. In Ricci, F., Rokach, L., Shapira, B., Kantor, P.B., eds.: Recommender Systems Handbook. Springer (2011) 1–29
19. Vail, D.L., Veloso, M.M., Lafferty, J.D.: Conditional random fields for activity recognition. In: Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems, ACM (2007) 1331–1338
20. Wang, A.I., Ahmad, Q.K.: CAMF – context-aware machine learning framework for android. In: Proceedings of the International Conference on Software Engineering and Applications (SEA 2010), IASTED (2010) Article no. 725–003.
21. Wang, X., McCallum, A., Wei, X.: Topical n-grams: Phrase and topic discovery, with an application to information retrieval. In: IEEE International Conference on Data Mining (ICDM'07), Omaha NE, USA, IEEE (2007) 697–702