Context mining and integration into Predictive Web Analytics

Julia Kiseleva
Eindhoven University of Technology
P.O. Box 513, NL-5600 MB, Eindhoven
the Netherlands
j.kiseleva@tue.nl

Supervised by: Mykola Pechenizkiy
Eindhoven University of Technology
P.O. Box 513, NL-5600 MB, Eindhoven
the Netherlands
m.pechenizkiy@tue.nl

ABSTRACT

Predictive Web Analytics is aimed at understanding behavioural patterns of users of various web-based applications: e-commerce, ubiquitous and mobile computing, and computational advertising. Within these applications business decisions often rely on two types of predictions: an overall or particular user segment demand predictions and individualised recommendations for visitors. Visitor behaviour is inherently sensitive to the context, which can be defined as a collection of external factors. Context-awareness allows integrating external explanatory information into the learning process and adapting user behaviour accordingly. The importance of context-awareness has been recognised by researchers and practitioners in many disciplines, including recommendation systems, information retrieval, personalisation, data mining, and marketing. We focus on studying ways of context discovery and its integration into predictive analytics.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval - Information Filtering

General Terms

Algorithms, Performance, Experimentation

Keywords

User modeling, advertising, context

1. INTRODUCTION

Motivation. Context often has a significant impact on the way humans (or machines) act and on how they interpret things; furthermore, a change in context causes a transformation in the experience that is going to be lived. A goal of predictive web analytics is to foretell users' interests based on the discovered behaviour patterns and it can be considered as a particular case of a context-aware application. Predictive Web Analytics is a rather distributed research field. Increasing volumes of data and web related commercial activities provide opportunities to model and predict user needs more precisely. An importance of 'contextual' information has been recognised by researchers and practitioners in many disciplines, including e-commerce personalisation, recommendation systems, information retrieval, ubiquitous and mobile computing, and marketing. 'Context-aware' systems adapt to users' operations and thus aim at improving the usability and effectiveness by taking context into account. Although there is a large amount of the papers, a holistic framework to build 'context-aware' application has not been presented yet.

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

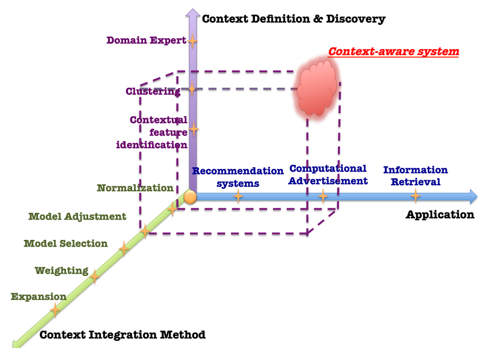

Figure 1: Categorization of the context-aware systems.

Goal. Our research aims to develop a generic framework and corresponding techniques for introducing the context-awareness in Predictive Web Analytics and accounting for the practical needs within the considered application areas.

In the following, we describe the research background (Section 2), research goals (Section 3), research methodology (Section 4), and current results (Section 5).

2. RESEARCH BACKGROUND

To summarise the categorisation of the context-aware systems we proposed 3-dimensional space with the following axes: a way to *define and discover context*, a *context integration method*, an *application* (Figure 1). The context-aware system in Figure 1 is presented as 'cloud' to emphasize that several ways to define and discover context can be used, several integration methods can be implemented, and several applications can utilized in one context-aware system.

Context definition & discovery. Many interpretations of the notion of context have emerged in various fields of research like psychology, philosophy, and computer science [4]. In literature, a context was presented as additional (situational) information: a user's location [1], helping to identify

Context	Year
Location	1992
Taxonomy of explicit context	1999
Predictive features vs. contextual	2002
Hidden context: (clustering, mixture models)	2004
Contextual bandits	2007
History of previous interaction	2008
Independence of predicted class	2011
Two level prediction model	2012
Focus on Context Discovery	2012 -

Timeline

Figure 2: The evolution of context definition

people near the user and objects around [9], current date, season, and weather [5]. Lately the user's emotional status was added to the context-aware application in [8] and the definition was broaden to *'any information that can characterise and is relevant to the interaction between a user and an application'*.

Most of the existing works assumed that context is explicit and given by a domain expert. In machine learning, context was considered as *contextual features* in supervised concept learning [18].

The contextual features are useful for classification only when they are considered in combination with other features. For example, in medical diagnosis problems, the patient's gender, age, and weight are often available. These features are contextual, since they (typically) do not influence the diagnosis when they are considered in isolation.

Later it was discovered that a context may not necessarily be present in form of a single variable in the feature space. It can be hidden in the data. Turney in [17] formulated the problem of recovering implicit context information and proposed two techniques: input data clustering and time sequence. According to [14], a context has *temporal* (when to deliver), *spatial* (where), and *technological* (how) dimensions. In terms of interactive system, [13] has shown that it was useful to consider the history of user interaction (changes in these entities). In the recent paper [21] a context was defined as an artefact in the data that does not directly predict the class label, e.g. accent in speech recognition. In [20] context-aware systems were proposed like two level prediction model for food sales. The timeline of the main milestones related to the research of context in predictive modeling is presented in Figure 2.

We are particularly interested in e-commerce applications so we narrow the overview to the following:

Application 1: Recommendation System. It was shown that the *situation* in which a choice is made is an important information for recommender systems [2]. E.g., using a temporal context in a travel recommender system would provide a vacation recommendation in the winter that can be very different from the one in the summer. Similarly, in the case of personalised content delivery on a Web site, it is important to determine what content needs to be recommended to a customer. The *purchase intent* of a customer is considered as contextual information in an e-commerce application because different purchasing intents may lead to different types of behaviour [2]. The *purchase intent* usually is considered as a **hidden context** which has to be derived. Then it can be used to **select 'right' model**. The context-aware recommenders utilize the information about a situation to make predictions. The authors in [13] defined **a hierarchy of a context** in the recommendation system they used the obtained contextual features **to expand feature space.** The other effective method for a context-aware rating prediction is Multiverse Recommendation based on the Tucker tensor factorization model [16]. The authors presented probabilistic model for generating personalised recommendations of items to users of a web service. The developed system was called Matchbox. The system makes use of **explicit context** information in the form of a user (e.g. age and gender) and meta data of an item (e.g. author and manufacturer) in combination with collaborative filtering information from previous user behaviour in order to predict the value of an item for a user. The contextual information is integrated into the prediction process using **a feature set expansion** manner to produce the better recommendations. [15] proposed a novel approach applying Factorization Machines to model contextual information and to provide context-aware rating predictions using **context explicitly specified** by a user to **the set of predictive features**.

Application 2: Computational Advertising. Revenue from advertising depends on the relevance of the displayed ads to user behaviour. Proper understanding of user's interests and delights is critical to effective *behavioural targeting* [3]. You could show an advertisement next to a sports page only for people who have recently visited Yahoo Autos because that shows that the user is also interested in this topic (**the history of an user interactions is used as context**). In general, targeting methods match the users within a given context to an appropriate ad. The context of the user usually consists the following **explicit factors: the page the user is currently visiting, the time, and the user's historical online behaviour**. Three types of targeting methods are popular in the advertising industry: **property**, **user segment**, and **behavioural targeting**. Property targeting refers to placing ads on specific web pages where interested users will appear, such as showing online brokerage ads on financial related pages. Although this reaches users who visit these finance pages, it may miss users who use some other web sites for their financial information. User segment targeting focuses on the gender and age of a user, and is only capable of targeting broad groups. Behavioural targeting involves using historical online information about the user to aid the publisher in showing them relevant ads wherever they appear. Whereas property targeting targets pages, and user segment targeting targets generic groups, behavioural targeting targets individuals. Online advertising usually uses context for **model selection**. As with other document retrieval systems, relevance is provided by scoring the match between individual ads (documents) and the content of the page where the ads are shown (query). In [7] authors show how this match can be improved by augmenting the ad-page scoring function with extra parameters from a logistic regression model on the words in the pages and ads.

Application 3: Information Retrieval. Context of a search query often provides a search engine with meaningful hints for answering the current query better and can be utilised for ranking. Given a query, a search engine returns the matched documents in a ranked list to meet the user's information need. Understanding **users' search in-**

tent expressed through their search queries is crucial to Web search. A web query classification has been widely studied for this purpose. [6] incorporates context information into the problem of query classification by using conditional random fields models (context is used **to expand a feature space**). This approach uses *neighbouring queries and their corresponding clicked Web pages in search sessions* as a context. Context-aware search adapts search results to individual search needs using contexts. While personalised search considers **individual users long and/or short histories**, context-aware search focuses on short histories of all users. The method from paper [19] adopts a learning-to-rank approach and integrates the ranking principles into a state-of-the-art ranking model by **encoding the context as a feature of the model**. The experimental results clearly show that this context-aware ranking approach improves the ranking of a commercial search engine.

3. RESEARCH GOALS

Our research aims to develop a generic framework and corresponding techniques for integrating context awareness into predictive analytics.

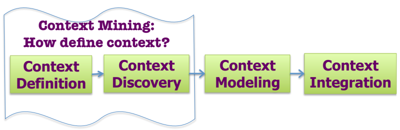

Figure 3: The design of context-awareness

Taking a broad approach in terms of relevant research content we aim for a complete solution that will allow the deployment in web analytics. Our main question is *"How can we effectively integrate context awareness into predictive web analytics in order to achieve better users' behaviour prediction accuracy?"*. We have divided it into the following sub-questions:

Question 1: How to define the context in predictive web analytics?

Question 2: How to connect context with the prediction process in predictive web analytics?

We consider four main steps to design a context-aware system as presented in Figure 3.

3.1 Context Definition & Discovery

Let $X \in \Re^p$ be an object of interest with a label $y \in \mathcal{Z}$. The ultimate task is to learn a label prediction function $y = \mathcal{L}(X)$. We want to incorporate a contextual information into the prediction process to improve the quality of the anticipations.

Let $C = \{C_1, C_2, \ldots, C_k\}$ be a set of *contextual categories* (Figure 4) and $L = \{L_1, L_2, \ldots, L_m\}$ be a set of individual learning procedures (defining e.g. the selection of training instances, input feature space, classification technique and its parametrization) or already learnt models. We define context awareness to the design of the prediction system to restrict the space of search of $\mathcal{L}$ as shown in Figure 4. Our goal is to find the function to map from contextual features to contextual categories. Let $\mathcal{G} : F_s \to C_i$ be a mapping from contextual features to contextual categories. The two key ingredients of a context-aware learning design are: defining

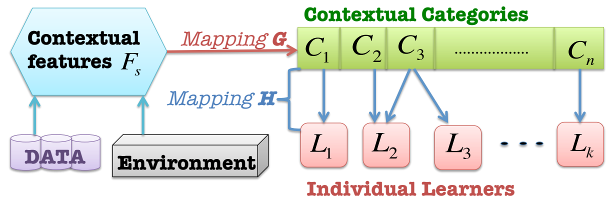

Figure 4: An example of context-aware system design

the contextual features F_s with a mapping $\mathcal{G}$, and fixing the mapping $\mathcal{H} : C \to L$.

There are tree general strategies to discover contextual categories:

Discovery from existing feature set: If context is *predefined* i.e. context categories are known from domain expert we only need to learn a mapping from contextual features to contextual categories and a mapping from contextual categories to learning models. E.g. the system designer knows based on his analytical experience that a user **location** is a contextual feature for our domain. So as a contextual feature F_s we can use the IP address of the user. We have an list of contextual categories based on the geographical regions: $\{C^{geo}\}_{i=1}^{n}$. The mapping function $\mathcal{G}$ is known $\mathcal{G} : IP \to C_i^{geo}$.

To identify context we can apply the definition of *contextual* and *context-sensitive* features from [18] to the existing source of features. Based on discovered set of features we identify the contextual categories using the similar strategy as before.

Hidden context: A hidden context discovery relied on an automatic pattern understanding methods e.g. *clustering*, *subgroup discovery* [12], *mixture models*, or more sophisticated techniques [21]. Usually the context identification techniques require two mechanisms: (1) how to group the training data X into k context and (2) how to assign an unseen instance to one of the contexts. So we will consider a context of users' requests during a session $\{s_i\}$ as F_s.

External sources: The external factors like 'weather' can be considered as context.

3.2 Context Modeling

We consider context as a secondary label/classification, describing an object of interests. To improve prediction we need to know this secondary labels. Particularly we can use a concept of *context spaces* presented in [10].

3.3 Context Integration

Types of an context-awareness integration into prediction is shown in Figure 5. Turney [17] lists five strategies for using contextual information:

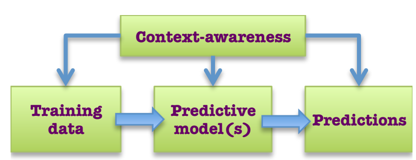

Figure 5: Types of an context-awareness integration.

Contextual normalization: The context (contextual features) can be used to normalize the primary context-sensitive features, prior to using the prediction model. The purpose is to process context-sensitive features in a way that reduces their sensitivity to the context.

Contextual expansion: A feature space composed of primary features F_p can be expanded with contextual features F_s. The contextual features can be treated by a learning process in the same manner as the primary features.

Contextual prediction model selection: The prediction can proceed in two steps: first select a specialized predictive model, based on the context information. Then apply this model to the primary features. For our educational prorate example we will use this strategy: first the system identifies the region of a user and then applies specialized predictive model for this geographical position.

Contextual prediction adjustment: First predict, using only the primary features. Then make an adjustment to the prediction, based on the context.

Contextual weighting: The context can be used to weight the primary features, prior to prediction. The goal of weighting is to assign more importance to features that, in a given context, are more useful for prediction.

4. RESEARCH METHODOLOGY

The research of Context Aware Predictive Analytics involves basic and applied research aimed to benefit from each other. The multi-methodological approach (conceptual theoretical, constructive and experimental) will be adopted.

Development. In the **conceptual-theoretical** approach, conceptual basics and formalisms of the generic framework will be developed. First, a taxonomy of context-aware approaches will be built. Then the applicability and limitations of the existing techniques w.r.t. the properties of the real application problems will be identified.

Implementation. In the **constructive** approach the developed techniques will be embedded into the prototype to test it through the experimentation approach and to facilitate the subsequent refinement of the theory and the prototype in an iterative manner.

Evaluation. The traditional experimental data mining research paradigm will be used for the **internal evaluation** of the developed framework and corresponding techniques on a set of reference and online real-world datasets. Progressive evaluation (time-wise) and cross validation (object-wise) procedures will be employed.

External validation of our work will be performed through the integration of the developed techniques into web analytics systems. We will employ traditional A/B and multivariate testing procedures providing reliable estimates of the performance of the alternative approaches. We have the commitment from the participating companies to conduct such procedures. The thorough continuous evaluation process consisting of internal and external validation procedures is an essential part of both theory building and testing.

5. CURRENT RESULTS

In this section we briefly summarize one of recently conducted studies about temporal context discovery (TCD). For the detailed discussion please refer to [11]. Users usually perform actions in a sequential manner on a website and the set of all actions are given. The goal is to predict the next action each user will perform given historical data about users' activities. We assume that under one context users perform a specific set of actions and when the context is switched to another one, another set of actions are performed. That means a context is defined as an external factor which is associated with a specific set of user actions. For instance, in a web session, there are sets of actions associated with the context like "search" while there are other sets of actions associated with the context like "buying". Our ultimate goal is to improve prediction of next user's activity on the website.

The general representation of users' historical behavior is given as a log with web sessions: $\mathfrak{D} = \{s_1, s_2, \ldots, s_n\}$. Given a set $\mathfrak{A} = \{a_1, \ldots, a_m\}$ of *event types*, an *event* is a pair (a, t), where t is an occurrence time of event. The web session of the user n is an ordered sequence of events:

$$s = \langle (a_1, t_1), (a_2, t_2), \ldots, (a_n, t_n) \rangle \quad (1)$$

such that all $i \in [1, n]$ and $t_i < t_{i+1}$ for all $i \in [1, n-1]$. Note that we do not have $t_i = t_{i+1}$, i.e. several events cannot occur at the same time. The session can be divided into segments which are related to users' intents. The collection of all sessions can be represented as a user navigation graph.

DEFINITION 1 (USER NAVIGATION GRAPH). *A user navigation graph is a directed and weighted graph $G = (V, E)$, where V is a set of vertices corresponding to all possible user actions $\mathfrak{A}$ and E are the set of edges. Each edge e of G is associated a weight $w(e)$ indicating the transition probability between two incident vertices of the edges.*

Let $\Theta = C_1 \times C_2 \times \cdots \times C_N$ be the space of all possible contextual features associated with every data instance. Let $M : \Theta \times \mathfrak{D} \mapsto V$ be a predictive model that maps each test sequence $s \in \mathfrak{D}$ associated with the contextual information θ_s to the decision space V. Let $F(s, M(\theta_s, s)) : \mathfrak{D} \times V \mapsto \mathbb{R}$ be the function evaluates how good a model is.

Given a class of predictive model $\mathfrak{M}$, we are looking for h small models from $\mathfrak{M}$ such that a user's behaviour is explained better by single model compared to global model. Namely the set of individual models give better evaluation then global model. Let $M = \{M_1, M_2, \ldots, M_h\}$ be a set of individual learning procedures from class $\mathfrak{M}$.

We propose that the data are generated as follows: the events alphabet $\mathfrak{A}$ is produced by h temporal hidden contexts. Under that assumption, our goal is to decompose the web session, which is the sequence of users' actions on the site, into homogeneous pieces, such that the data in each segment can be described accurately by a temporal context C_i and a simple prediction model M_i. Formally, we have to decompose the event space into E clusters each corresponds to a hidden context. Given the original log of sequences $\mathfrak{D}$ that is randomly divided two disjoint $\mathfrak{D}_{train}$, $\mathfrak{D}_{validation}$ and $\mathfrak{D}_{test}$. A decomposition of the event space into h clusters uniquely splits each sequence in the data into segments. Let $T_{train} = \cup_{j=1}^{Freq(C_i)}(Seg_j \in \mathfrak{D}_{train})$ be the set of segments in the training set that corresponds the context C_i. We learn sets of predictive models $\{M_i\}_{i=1}^{h}$ based on h sets of sessions segments. We validate our set of models based on $T_{validation} = \cup_{j=1}^{Freq(C_i)}(Seg_j \in \mathfrak{D}_{validation})$. And we test resulted clusters based on $T_{test} = \cup_{j=1}^{Freq(C_i)}(Seg_j \in \mathfrak{D}_{test})$. The effectiveness of one learner can be defined as:

$$F_{c_i}(T_{validation_i}, M_i) = \sum_{a_j \in T_{validation_i}} F(a_j, M_i) \quad (2)$$

To find *best* decomposition E we use a hierarchical clustering and our objective function is:

$$EF^* = \arg\max_{E} \sum_{i=1}^{h} F_{c_i}(T_{validation_i}, M_i) - N_E, \qquad (3)$$

where N_E is number of transition points between one context to another context. At the transition point, we always make wrong prediction because a model is not aware about event from different clusters. The final accuracy was calculated using the T_{test}.

Table 1: The evaluation of the prediction accuracy for FOMM and TCD for discovered 7 clusters.

Model	FOMM (%)	TCD (7 clusters) (%)
Accuracy	40.6±0.25	50.2±3.24

We have run our experiments on a real dataset collected at *MastersPortal.eu* which is a web service that provides information about various study programmes in Europe. For our experiment we used data collected during May 2012. The results is presented in Table 1. We obtained the highest accuracy when we use 7 clusters. We used 1^{st}-order Markov model (FOMM) as a baseline.

6. CONCLUSION AND FUTURE WORK

Starting from the prototype of context-aware system which is partially presented in Figure 4 we will study algorithmic aspects and analyze the performance of the two level decision making for alternative web analytics application scenarios. We will develop a generic framework including techniques for forming contextual categories and for linking them Context Awareness in Predictive Analytics with the predictors for integrating context awareness into predictive models. We will also produce a set of guidelines for using the proposed framework for designing new techniques. The techniques will be tested retrospectively on historical data as well as deployed and validated online in the field experiments. Taking a broad approach in terms of relevant research content we aim for a complete solution that would allow the deployment in web analytics.

Acknowledgements

This research has been partly supported by STW and it is a part of CAPA project. We would like to thank Thijs Putman from *MastersPortal.eu* for providing the anonymized dataset used in the experimental study and for allowing to make it publicly available.

7. REFERENCES

[1] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles. Towards a better understanding of context and context-awareness. 1999.

[2] G. Adomavicius and A. Tuzhilin. Context-aware recommender systems. *CARS*, 2010.

[3] M. Aly, A. Hatch, V. Josifovski, and V. K. Narayanan. Web-scale user modeling for targeting. In *WWW*, 2012.

[4] C. Bolchini, C. A. Curino, E. Quintarelli, F. A. Schreiber, and L. Tanca. A data-oriented survey of context models. In *SIGMOD*, 2007.

[5] P. Brown, J. Bovey, and X. Chen. Context-aware applications: From the laboratory to the marketplace. *IEEE Personal Comm*, 4:58–64, 1997.

[6] H. Cao, D. H. Hu, D. Shen, D. Jiang, J.-T. Sun, E. Chen, and Q. Yang. Context-aware query classification. In *SIGIR*, 2009.

[7] D. Chakrabarti, D. Agarwal, and V. Josifovski. Contextual advertising by combining relevance with click feedback. In *WWW*, 2008.

[8] A. Dey, G. Abowd, and D. Salber. A conceptual framework and a toolkit for supporting the rapid prototyping of contextaware applications. *Human Computer Interaction*, 2:97–166, 2001.

[9] R. Hull, P. Neaves, and J. Bedford-Roberts. Toward situated computing. pages 146–153, 1997.

[10] J. Indulska, R. Robinson, A. Rakotonirainy, K. Henricksen, M.-S. Chen, P. K. Chrysanthis, M. Sloman, and A. B. Zaslavsky. Experiences in using cc/pp in context-aware systems. *Mobile Data Management,Lecture Notes in Computer Science, Springer*, 2574, 2003.

[11] J. Kiseleva, H. T. Lam, M. Pechenizkiy, and T. Calders. Discovery temporal hidden contexts in web sessions for user trail prediction. In *TempWeb*, 2013.

[12] P. K. Novak, N. Lavrac, and G. I.Webb. Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. *Journal of Machine Learning Research (JMLR)*, 27:77–403, 2009.

[13] C. Palmisano, A. Tuzhilin, and M. Gorgoglione. Using context to improve predictive modeling of customers in personalization applications. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 20(11):1535–1549, November 2008.

[14] C. Prahalad. Predicts customer context is the next big thing. *AMA MwWorld*, 2004.

[15] S. Rendle, Z. Gantner, C. Freudenthaler, and L. Schmidt-Thieme. Fast context-aware recommendations with factorization machines. In *SIGIR*, volume 10, 2011.

[16] D. Stern, R. Herbrich, and T. Graepel. Matchbox: Large scale online bayesian recommendations. In *WWW*, pages 111–120, 2009.

[17] P. Turney. Exploiting context when learning to classify. 2002.

[18] P. Turney. The management of context-sensitive features: A review of strategies. In *CoRR*, 2002.

[19] B. Xiang, D. Jiang, J. Pei, X. Sun, E. Chen, and H. Li. Context-aware ranking in web search. In *SIGIR*, 2010.

[20] E. Zliobaite, J. Bakker, and M. Pechenizkiy. Beating the baseline prediction in food sales: How intelligent an intelligent predictor is? *Expert Syst. Appl. (ESWA)*, 39(1):806–815, 2012.

[21] I. Zliobaite. Identifying hidden contexts in classification. In *PAKDD (1)*, pages 277–288, 2011.