Dynamic Cost-Sensitive Feature Acquisition Framework for Mobile Context-Aware Applications

Eran Fainman, Prof. Bracha Shapira, Prof. Lior Rokach

Mobile context-aware applications have become popular in the last years due to the high adoption of smart mobile devices among world's population [1]. Most of today's mobile devices have sensors that can be used to extract many features to describe user's contexts [2]. One of the major inherited challenges in context-aware applications for mobile devices is to determine a sampling policy. A sampling policy describes which of the sensors are sampled and what the sampling triggers are. The sampling policy has major effects on the application's resource consumption (in terms of memory and CPU), response time, context inference accuracy, communication transportations volumes, required storage (either in the client and the server sides) and data manipulations capabilities [3]. In this paper we suggest a dynamic sampling approach to minimize the sampling's costs with minimal accuracy loss. We assume a given set of features, each of them can be independently extracted from a sensor's stream, with a given fixed acquisition cost. A sampling process is performed at each time a sampling trigger is met. The sampling process is performed sequentially, i.e., features are acquired one by one. This paper proposes a feature acquisition framework that enables to involve machine learning techniques in the sampling process. The suggested framework acquires features in two steps: *(1) initial step* – in which all of the features are acquired when a sampling process is performed, *(2) dynamic step* – in which dynamic feature acquisition is performed in each sampling process. This paper explores the suggested framework with four configurations: *(1) Pure random* sampling - in which the initial dataset is initialized with an empty set and in the dynamic step, all of the features have the same fixed probability to be sampled in each sampling process. *(2) Cost-sensitive stochastic sampling* – in which the initial dataset is initialized with an empty set and in the dynamic step, each feature is acquired with probability that is calculated with a normalized version of its inverted acquisition cost (therefore, features with low cost are likely to be acquired more). *(3) Variance-sensitive stochastic sampling* – in which different sizes of initial dataset (*initial step*) are set. Variance calculation for each feature is performed based on the initial dataset. In the dynamic step, each feature is acquired with probability that is calculated with a normalized version of its variance (therefore, features with high variance are likely to be acquired more). A similar version of this type of configuration is described in [5]. *(4) Dynamic cost-sensitive tree based configuration* – in which different sizes of initial dataset (*initial step*) are set. In the dynamic step, a cost-sensitive version of C4.5 tree [4] is built in the beginning of each sampling process (tree is built based on the already seen observations). Then, the sampling process is performed by acquiring the features that are necessary to classify the new observation. The remained features are sampled with same fixed probability. Thus, the tree represents hypothesizes that previous observations have shown and new observations support or weaken them. The evaluation of the framework and its configurations is performed on the following datasets: *(1) Conjure* – a dataset that describes the problem of identifying real and fake SMSs based on mobile sensors (10,000 instances, 153 features, 2 labels). *(2) HPAT* – a UCI's dataset of Human Activity Recognition built from the recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors (10,029 instances, 561 features, 12 labels). The evaluation process was performed as follow: datasets were split to training set (70%) and test set (30%). The framework was assigned with the training set and emitted a sparser dataset. A cost-sensitive version of C4.5 tree [4] was built based on the emitted dataset and a weighted precision measure for the test set was calculated. The experimental results (as described in Appendix 1 - Figure 1 and Figure 2) show that the framework with dynamic cost-sensitive tree-based configuration leads to approximately 80% sampling cost saving with less than 20% weighted precision loss in the Conjure dataset and approximately 90% sampling cost saving with less than 12% weighted precision loss in the HPAT dataset.

The contribution of this research is two-fold:

- We propose a novel framework for feature acquisition in mobile context-aware applications. The proposed framework enables to involve machine learning algorithms in order to decrease the amount of collected data with minimal accuracy loss. Furthermore, it implies dynamic learning that enables to improve the sampling process for new observations.
- We present a sensitivity analysis for different configurations of this framework including the resultants features acquisition's cost savings and weighted precision loss measures for two datasets.

This research lays foundations for future works in which diverse machine learning algorithms are involved in the dynamic step of the framework and more datasets are tested to evaluate the framework.

References:

[1] Baldauf M, Dustdar S, Rosenberg F. A survey on context-aware systems. *International Journal of Ad Hoc and Ubiquitous Computing*. 2007;2(4):263-277.

[2] Hoseini-Tabatabaei SA, Gluhak A, Tafazolli R. A survey on smartphone-based systems for opportunistic user context recognition. *ACM Computing Surveys (CSUR)*. 2013;45(3):27.

[3] Lane ND, Miluzzo E, Lu H, Peebles D, Choudhury T, Campbell AT. A survey of mobile phone sensing. *Communications Magazine, IEEE*. 2010;48(9):140-150.

[4] Lauwereins S, Meert W, Gemmeke J, Verhelst M. Ultra-low-power voice-activity-detector through context-and resource-cost-aware feature selection in decision trees. . 2014:1-6.

[5] Wood AL, Merrett GV, Gunn SR, Al-Hashimi BM, Shadbolt NR, Hall W. Adaptive sampling in context-aware systems: A machine learning approach. . 2012.

<u>Appendix 1 – Experimental results figures</u>:

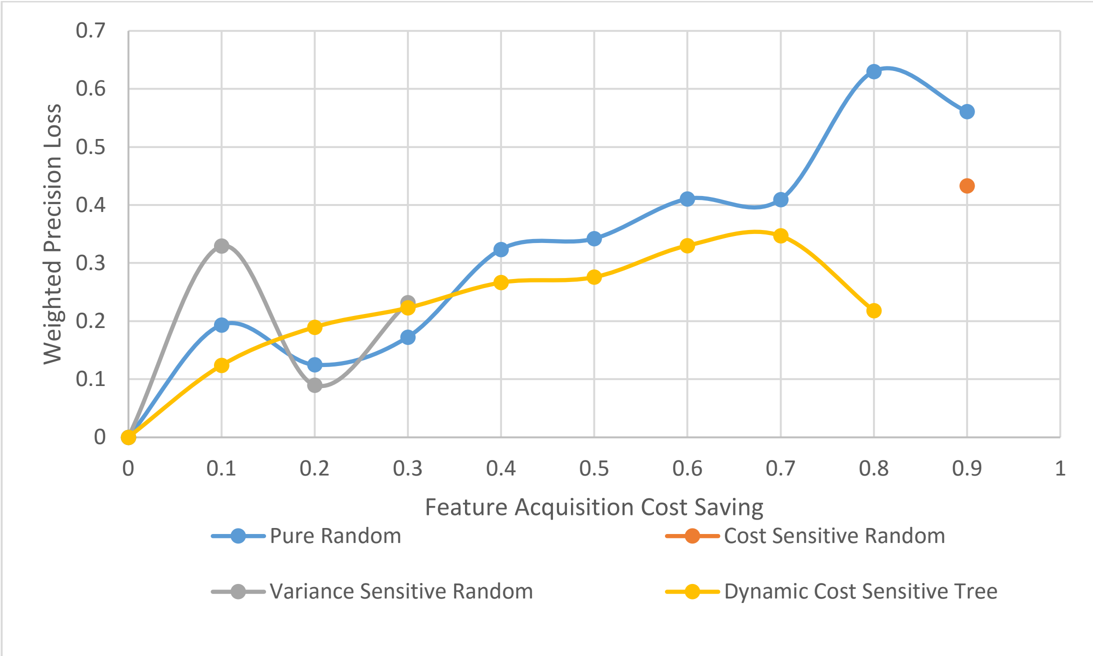

Figure 1 - Weighted precision loss over feature acquisition cost saving for Conjure dataset

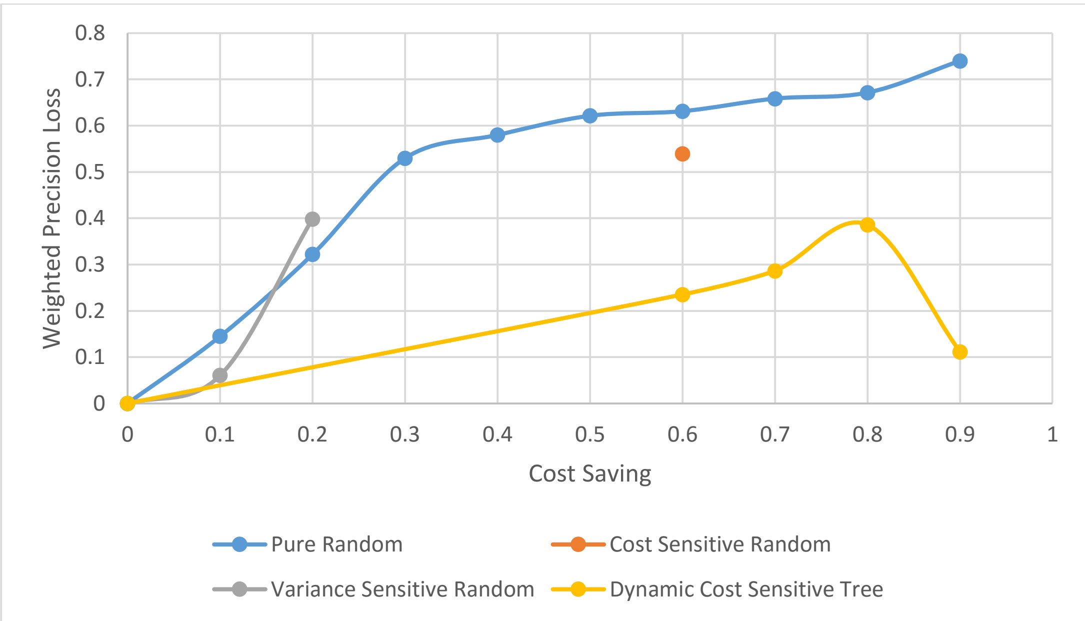

Figure 2 - Weighted precision loss over feature acquisition cost saving for HPAT dataset