Pervasive and Mobile Computing 16 (2015) 286–304

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

Continuous discovery of co-location contexts from Bluetooth data

Thuong Nguyen*, Sunil Gupta, Svetha Venkatesh, Dinh Phung

Centre for Pattern Recognition and Data Analytics, Deakin University, Australia

ARTICLE INFO

Article history:
Available online 24 December 2014

Keywords:
Nonparametric
Indian buffet process
Incremental
Particle filter
Co-location context

ABSTRACT

The discovery of contexts is important for context-aware applications in pervasive computing. This is a challenging problem because of the stream nature of data, the complexity and changing nature of contexts. We propose a Bayesian nonparametric model for the detection of co-location contexts from Bluetooth signals. By using an Indian buffet process as the prior distribution, the model can discover the number of contexts automatically. We introduce a novel fixed-lag particle filter that processes data incrementally. This sampling scheme is especially suitable for pervasive computing as the computational requirements remain constant in spite of growing data. We examine our model on a synthetic dataset and two real world datasets. To verify the discovered contexts, we compare them to the communities detected by the Louvain method, showing a strong correlation between the results of the two methods. Fixed-lag particle filter is compared with Gibbs sampling in terms of the normalized factorization error that shows a close performance between the two inference methods. As fixed-lag particle filter processes a small chunk of data when it comes and does not need to be restarted, its execution time is significantly shorter than that of Gibbs sampling.

1. Introduction

Context awareness is one of the most important features of a pervasive computing system [1]—the discovery of contexts allows the system to adapt to changes in user status, environment or social situation. Recent advances in wearable technology provide an opportunity to collect a wide range of signals for context discovery. Simple forms of context, such as user's location or acceleration, can be easily extracted from the sensors and used immediately. They are, however, not rich enough and need more interpretation [2] because the meaningful information is typically *embedded inside the data* and cannot be seen easily. The hidden patterns, e.g. daily routines, activity trajectory, social interaction, once extracted, can provide a rich form of information to build context-aware applications. These hidden patterns can be identified as context [3].

Context in pervasive computing can be classified into three categories: the status of user, the status of ambient environment, and the interaction with objects or people. In this paper, we focus on the third category. Specifically, we demonstrate our framework in detecting the co-location contexts in a community of people instrumented with Bluetooth devices.

Inferring these complex contexts, however, is a challenging task. From the pairwise proximity data collected from Bluetooth devices, co-location groups can be extracted using graph based approaches such as K-clique [4–6] or the Louvain method [7]. Input of these methods is a pairwise network that can be unweighted for K-clique and weighted for the Louvain

* Corresponding author.
E-mail addresses: ctng@deakin.edu.au (T. Nguyen), sunil.gupta@deakin.edu.au (S. Gupta), svetha.venkatesh@deakin.edu.au (S. Venkatesh), dinh.phung@deakin.edu.au (D. Phung).

http://dx.doi.org/10.1016/j.pmcj.2014.12.005
1574-1192/

method; the output is a set of fully connected sub-graphs. Each sub-graph is a group of users that are usually in proximity. These approaches, however, can extract only the co-location groups from the whole data and it is difficult to decide which group a user belongs to in a particular time interval. In other words, these approaches are designed for community detection and unsuitable for context-aware applications unless we further interpret the group that a user participates in during each time interval.

Existing approaches to extract contexts or hidden patterns from raw signals, typically borrow techniques from machine learning and data mining. For example, the principal component analysis (PCA) is used to extract the co-location groups and the group membership of user in [8]. Another example is a modified version of latent Dirichlet process (LDA) [9]. These techniques are, however, parametric models as they require the number of groups to be specified in advance. This number may dynamically change in stream data and thus needs to be estimated automatically. This is an extremely expensive process. More importantly, in pervasive computing, data stream in without clear demarcation of start and finishes [10]. Thus, there is a need of an incremental inference method to process data when they come in with efficient computation.

To address these problems, we propose a Bayesian nonparametric (BNP) model to extract co-location contexts from Bluetooth stream data of several users. We present the data as a matrix—time vs. user, where each entry is Bluetooth signals using count features. We cast the solution as a matrix factorization problem: the derived factors describe typical patterns in co-location contexts and the coefficient matrix indicates the presence or absence of the contexts during each time interval. We propose a Poisson–exponential model to perform the factorization, replacing the gamma distribution in [11] by an exponential distribution to enforce the sparsity of the factors. We employ the Bayesian nonparametric approach by specifying a prior distribution over the coefficient matrix using an Indian buffet process (IBP) that infers the number of factors automatically given the observed data. The data is modeled through a Poisson distribution parametrized by a product of factors and factor indicators. Typical inference method for this type of model is the Markov chain Monte Carlo method, e.g. Gibbs sampling [12,13]. This sampling scheme, however, requires all data to be presented and needs to be restarted when new data arrive. Such inference method is inefficient to use in pervasive computing applications. Instead, we propose a novel fixed-lag particle filter. This method presents the objective distribution as a set of samples. When new data arrive, the new samples are drawn from the old samples using the new data only. Thus the computational requirement for each update is fixed and does not grow over time when more data come.

We demonstrate the effectiveness of the model on three diverse datasets: a synthetic dataset, the Sociometric dataset [14], and the Reality Mining dataset [15]. We treat the Bluetooth time series across multiple users as stream data, as in online systems. The performance of fixed-lag particle filter is comparable to full-batch Gibbs sampling in terms of normalized error, while the execution time is more than 100 times shorter. The extracted co-location contexts are then compared to the co-location groups extracted by the Louvain method [7]. The visual and quantitative (using *purity* metric) comparison between the results of the two methods shows a strong correlation between them. Moreover, we show that unlike the Louvain method, our approach can track the evolution of user groups and membership.

Our main contributions are:

(i) A Bayesian nonparametric matrix factorization framework modeling data using Poisson distribution and the factors using exponential distribution—the number of factors is inferred automatically.
(ii) The development of a fixed-lag particle filter algorithm to permit incremental model updates—achieving comparable performance to Gibbs sampling while taking much shorter execution time.
(iii) A demonstration of the proposed framework to detect meaningful co-location contexts from Bluetooth data using two real-world datasets.

Although we demonstrate the discovery of co-location contexts from Bluetooth data, our model is generic to deal with any type of count data, and thus has a wide applicability.

The remaining part of this paper is structured as follows. Section 2 reviews the related work in high-level context discovery and a brief background on Bayesian nonparametric matrix factorization. Section 3 introduces our Poisson–exponential model. Section 4 explains two inference methods for the proposed model, Gibbs sampling and the proposed fixed-lag particle filter. Section 5 demonstrates the experimental results. In Section 6, we discuss the pros and cons of our model and some of its possible applications. Finally, we conclude our paper in Section 7.

2. Background

In this section, we provide a brief background of high level context discovery and Bayesian nonparametric matrix factorization using Indian buffet process.

2.1. High-level context discovery

The term *context-aware* was introduced by Schilit et al. [16,17]. Since then, there have been many attempts to define *context* but the most widely accepted one is propounded by Dey [3]. Context in this definition is "*any information that can be used to characterize the situation of an entity*". This definition is general enough to be applied in various scenarios of context-aware applications.

For a better view on context-aware applications, we classify contexts into *three* categories: the personal status of the user, the status of surrounding environment, and the interaction with dynamic objects.