

Application of Dimensionality Reduction Techniques for Mobile Social Context

Andreas Komninos
Glasgow Caledonian
University
ako2@gcu.ac.uk

Athanasios Plessas
University of Patras &
RACTI
plessas@ceid.upatras.gr

Vassilios Stefanis
University of Patras &
RACTI
stefanis@ceid.upatras.gr

John Garofalakis
University of Patras &
RACTI
garofala@ceid.upatras.gr

ABSTRACT

We propose the application of a dimensionality reduction algorithm that could provide a breakthrough in the efforts to retrieve and present mobile personal information to the user in context.

Author Keywords Context awareness, Dimensionality Reduction, Mobile Personal Information Access

ACM Classification Keywords H.3.3 Information Search and Retrieval: Retrieval Models

General Terms Algorithms, Human Factors

INTRODUCTION

Advances in mobile hardware technology in terms of storage space and data generation modalities (efficiency-enhancing input UIs, sensor-generated data and meta-data) allow today's mobile device user to quickly generate and store large volumes of personal data. Much of this data is organized in structured repositories (e.g. contact list, photo gallery, message inbox), affording the user a means of learnable, procedurised retrieval. Even though multiple types of personal information are meaningful and helpful to the user in daily situations, in most devices the structured repositories remain mostly "walled gardens", requiring the user to sequentially visit each of them in order to assemble the information pieces she needs in one coherent mentally held collection, relevant to their current activity. Naturally, the cognitive load on the user increases with the number of information items that have to be retrieved from the repositories. But still, even considering each "walled garden" individually, it quickly becomes apparent that current retrieval methods are largely inefficient, in view of the ever-expanding storage space and richness of personal information stored in mobile devices.

Mobile devices such as smartphones, are still primarily information access devices and communication devices. Much of the activity on mobile devices, especially mobile phones, is in support of Human-to-Human interaction. Consequently, many of the actions that someone might perform on their mobile device involve the act of looking up a contact, which carries at that time some importance to the user, in order to initiate some form of communication (call, sms, email, comment on facebook etc.). Not

forgetting that communication is by definition the exchange of information, information items are an important part of it. It is critical here to underline that our work extends beyond the notion of contact importance – in fact, we believe that importance as a multidimensional vector can provide a complex query "key" on the large, rich dataset in a user's mobile device for any type of personal information item (e.g. a photograph, or a set of SMS messages or stored Word documents) so as to answer the question of who to communicate with and what to communicate. As such our aim is to construct an extensible, flexible approach to the definition of importance in a k-dimensional context space, which can be applied to any mobile information retrieval problem. We choose to start with the concept of contacts as a problem domain, largely due to its importance in everyday interaction with mobile devices.

MOBILE SOCIAL CONTEXT

Mobile devices collect a significant amount of data and information about the user's context (location, the current time, whether the user of the device is on move and their speed, the orientation of the device, the user's current task etc.). The user considers her mobile device a "trusted device". She usually has the device close to her, sometimes operating 24 hours per day. Devices contain a lot of personal information related to the user's social environment [1]. These are often generated automatically by the device (e.g. a smartphone's phone list saves the calls that have been made, the time of the day for each call and the duration of each call for the past few days or even weeks). Moreover, mobile devices store user-generated content (e.g. SMS/MMS and audio files, browser's history, calendar with user's events etc). Therefore, a mobile device is also aware of the social environment of the user (social context). The combination of social and mobile context results in a dynamically defined social context, termed the *mobile social context* [2]. Therefore, a truly context aware application has to consider social context as part of its context representation.

In literature, context has often been represented in the form of vectors [3]. Vector based approaches carry the disadvantage of computational complexity in similarity searches (each dimension needs to be compared separately) and that weighed vectors require an empirical (hence error-prone or narrowly applicable) estimation of the weights.

MAPPING CONTEXT IN K-DIMENSIONAL SPACE AND DIMENSIONALITY REDUCTION

In our view, despite its drawbacks, the vector approach can offer a flexible solution to context acquisition and representation, hence our work relates to overcoming its computational complexity and vector weighting issues. In [4] we proposed four criteria for the determination of a m-PIM (mobile personal information management) item importance (namely contacts). From these criteria, we can derive the following context dimensions for the contact list problem domain, on which a contact can be mapped:

- Frequency (e.g. Frequency of use in last n months)
- Recency (e.g. days since last use)
- Location (e.g. Geographic areas from which at least n% of uses are made)
- Time of Day (e.g. Time segment of day in which at least n% of uses are made)
- Task (e.g. Boolean measure of existence of a scheduled task involving a contact within a certain window of time [for example today] or temporal distance between now and such scheduled task?)
- Personal preference (e.g. scale of 1-5 of explicit user rating of importance for a given contact)

In order to estimate a “match” signifying importance between these types of contextual information and the user’s current context, a typical approach would be to measure the distance between current context and contextual data (e.g. time of day now vs. usual time of day of contact use) and combine this with static context (e.g. explicit importance rating). The derived metrics would need to be weighted and the sum of these weighted metrics could then be used to infer “importance” for a single contact under any context. This approach though is not without challenges: Firstly, one must determine appropriate weights for each context type. Subsequently, it is easy to realize that this would be a futile attempt, as the weights of each context type are naturally dynamic and can vary under different use contexts. A possible solution however could be the application of dimensionality reduction, an algorithmic technique for reducing the dimensionality of data, applied in several computer science fields. Real-world data usually has a high dimensionality, a fact that affects data processing performance (the so called “curse of dimensionality” that suggests exponential dependence of an algorithm on the dimension of the input). The idea is to transform data from a high-dimensional space to a low-dimensional space, preserving some critical relationships among elements of the data set. In mathematical terms, given a p-dimensional object $x=(x_1,\dots,x_p)^T$, find a lower dimensional representation of it, $s=(s_1,\dots,s_k)^T$ with $k\leq p$, that

captures the content in the original data, according to some criterion.

Our research idea is to perform DR to context augmented personal information items, such as entries in a contact list, an idea that has not yet been proposed and applied in scientific literature as far as we know. Since, as already discussed, context augmented personal information items can be represented as multidimensional vectors, we find it highly appealing to try to extract a small number of features that could accurately represent the original items and their relationships, so as to enable quick and accurate similarity searches for related personal information items. Furthermore, after reducing the dimensions of the items, it might be desirable to map them to a 1-d, 2-d or 3-d space, as often done in high-dimensional data projections, since visualization tends to reveal existing groups of objects.

There is a wide range of algorithms with diverse characteristics that achieve dimensionality reduction, following different approaches. An interesting method that could be appropriate for the case of mobile phones due to its simplicity and computational efficiency is the FastMap algorithm [5]. FastMap is a fast algorithm that maps high-dimensional objects into lower-dimensional spaces, while preserving well distances between objects and the structure of the data set, as a result preserving also dis-similarities between objects. Experiments presented in [5] show that the algorithm performs also well for visualization and clustering.

REFERENCES

1. Toninelli, A., Khushraj, D., Lassila, O., Montanari, R. Towards Socially Aware Mobile Phones, *SDoW 2008*, (2008).
2. Gilbert, P., Cuervo, E., Cox, L. Experimenting in Mobile Social Contexts Using JellyNets. *HotMobile 2009*, USA (2009).
3. Du, W., Wang, L. Context-aware application programming for mobile devices. *In Proc. C3S2E Conference*, (2008), 215-227.
4. Komninos, A., Liarokapis, D. The Use of Mobile Contact List applications and a Context-Oriented Framework to Support their Design. *In Proc. MobileHCI09*, Germany, (2009).
5. Faloutsos, C., Lin, K. I. Fastmap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. *In Proc. ACM SIGMOD '95*, USA, (1995), 163-174.