CONTEXT DIMENSIONALITY REDUCTION FOR MOBILE PERSONAL INFORMATION ACCESS

Andreas Komninos[1], Athanasios Plessas[2,3], Vassilios Stefanis[2,3] and John Garofalakis[2,3]
[1] *Mobile and Ubiquitous Computing Group, Glasgow Caledonian University, Cowcaddens Road, Glasgow, U.K.*
[2] *Department of Computer Engineering and Informatics, University of Patras, Patras, Greece*
[3] *Computer Technology Institute "Diophantus", Rion, Patras, Greece*
andreas.komninos@gcu.ac.uk, {plessas, stefanis, garofala}@ceid.upatras.gr

Keywords: Context awareness, Dimensionality Reduction, Mobile Personal Information Access.

Abstract: We propose an application of the Fastmap algorithm that could provide a breakthrough in the efforts to present mobile personal information to the user in context, and describe our vision for context-driven interfaces generated by this method that will support the richness of data stored in personal devices.

1 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 in daily situations the user may require access to multiple types of personal information, 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. For example consider work on visualizing and retrieval from large mobile photo galleries (Hsu et al., 2009) or ever expanding music collections (Tolos, Tato and Kemp, 2005) or contact lists. What can a user do with all this data? Surely users want the data, but is it really useful in its raw form as a singular item and can it be used to help them achieve their goals?

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, that perhaps 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. Thus, mobile devices should support not just the retrieval of contacts, but also of information that is somehow relevant to these contacts. 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 (and consequently UI design for mobile information access). We choose to start with the concept of contacts as a problem domain, largely due to its importance in everyday interaction with mobile devices.

2 BACKGROUND

Our motivation stems from the observation of user behaviour in accessing non-contextualised information from single repositories, namely the mobile contact list. In recent work (Bergman et. al., 2011), we demonstrated some interesting user behaviours, which we believe, highlight the issue of non-contextualisation. Firstly, we observed that our 18 young users (<25 yrs. old) from varied backgrounds tend to keep fairly large repositories of contacts (m=92.47 sd=56.93), which, as literature suggests (Boardman and Sasse, 2004), are bound to get bigger as they grow older (just one of our users reported actively deleting unused contacts). In these collections, a large percentage of contacts (av=39%, sd=17%) has either not been used in over 6 months or not at all. Viewing the collections as a whole (n=754 contacts), over 47% of all contacts had not been used in over 6 months or at any other point, while just 15% were identified as frequently used contacts, the remaining 38% having been used between 1 and 6 months previous to our study period. We also found a highly significant correlation between the size of the contact list and the number of unused contacts (r=0.81). In a carefully designed experiment, we presented our users with a contextualised UI which placed "important" contacts in an alphabetic list, followed by an alphabetic list of all other contacts and asked them to perform several retrieval tasks on this augmented UI and on a classic (Nokia) UI. Importance in this experiment was set as a single-value vector of frequency of use, with 6 months or more being a threshold for considering a contact as not important. We found that users provided very positive subjective feedback to their experience using this approach (86% found it easier to use than the traditional UI and 64% would like to have this feature on their next device). This was backed up by significant performance metrics. On average users required less button presses to successfully "find" a contact (mean reduction=1.96, sd=3.56) and significantly less time with the augmented UI (m=4,424ms, sd=1,872ms) than the traditional UI (m=5,204ms, sd=2,829ms).

While these findings assumed a naïve classification of importance (just frequency of use), they demonstrate quite effectively how contextualisation on just a single dimension can have a very positive effect on the retrievability of personal mobile information. In the past, other researchers have demonstrated either issues with the usability of contact and call lists (Böcker and Suwita, 1999) (Klockar et al., 2003) or improvement in usability through the introduction of user activity, proximity & state context (Oulasvirta et al., 2005), social context (Gaur, 2008) (Rhee et al., 2006) and temporal context (Jung et al., 2008). Further work by Ankolekar et al. (Ankolekar et al., 2009) discusses how a combination of contextual cues might offer usability advantages but leaves the categorisation of contacts to users and does not present any tangible research into one of the most oft-used applications of mobile devices. The volume of research in contact list use remains low and the technology behind contact list UIs in today's devices is mainly based on alphabetic lists. Currently, only the Android OS provides a feature of presenting a contact list by use frequency, though again this is a simplistic view of importance, considering just one type of context (fig. 1).

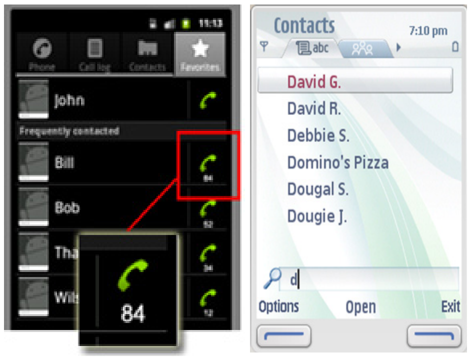

Figure 1: The Android contact list UI, augmented with a single-dimension context cue (frequency of use) [left] and a standard Symbian contact list UI, restricted to alphabetically ordered lists [right].

3 THE NEED FOR A K-DIMENSIONAL APPROACH TO CONTEXT MAPPING

Mobile devices collect a significant amount of data and information about the user's context. Such information includes location (absolute or relative), 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 (e.g. on the phone, messaging), whether the vibration or the silent mode are enabled (Beach et al., 2010) 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 (Toninelli et al., 2008). 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 could also be aware of the social environment of the user (social context). In addition, mobile applications can take advantage of social data from online social networks in order to enrich a contact with more information (Bentley et al., 2010). The combination of social and mobile context results in a dynamically defined social context, termed the mobile social context (Gilbert et al., 2009). Therefore, a truly context aware mobile information access application has to consider social context as part of its context representation.

It is therefore quite apparent that true contextualisation is much more complex than in our previous experiment's assumption, and that it requires k-dimensional space in order to be defined. As shown in equations (1), (2) and (3) we use a vector model to represent the context of an information item i in a k-dimensional space as a k-tuple, where d_n is the context atom value for dimension n. Its importance can be considered as the sum of the weighted distances between the vector atoms describing current context properties and an item's context atoms ($\Delta\vec{C}(i)$).

$$\vec{C}(i) = (d_1, d_2, ..., d_n) \quad (1)$$

$$I_i = \sum_{i=1}^{n} \left(w_{d_n} \times \left(1 - \Delta\vec{C}(i)\right)\right) \quad (2)$$

$$\Delta\vec{C}(i) = \left|\vec{C}(n) - \vec{C}(i)\right| \quad (3)$$

Researchers in the past have often attempted to combine the user's input with the user's context in order to provide a richer user experience (context-aware applications), e.g. (Yoon et al., 2008). Most approaches tend to focus on narrow objectives, as the capture of context for general use is still considered very difficult. In literature, context has been represented in the form of vectors, e.g. (Du and Wang, 2008) while other researchers have adopted an ontological approach, e.g. (Korpipaa et al., 2004), often using naïve Bayes classifiers to solving the issue of defining context space. 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. On the other hand, ontologies tend to be inflexible and too strict to be applicable to wider ranges of problems, requiring careful, manual approaches to their construction.

4 MAPPING CONTEXT IN K-DIMENSIONAL SPACE

In the previous section we outlined the disadvantages of the vector and ontological approaches for determining context. In our view, the vector approach can offer a more flexible solution to context acquisition and representation, hence our work relates to overcoming its computational complexity and vector weighting issues. In (Komninos and Liarokapis, 2009) 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 (equations (1), (2) and (3)). 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 (see scenario in figure 2). The example scenario introduces the problem of context-derived specific importance (vs. general importance) and shows that a decent approximation to the calculation of this term is very difficult, due to the infinite variability of context itself and the complexity of its dimensions. A possible solution however could come from the field of Databases and IR, where dimensionality reduction is a technique often used to automatically extract important features from complex data and reduce the complexity of searches in multidimensional spaces.

Take the example of "John", a contact that the user calls every day and "Jane" another contact that is only called once a year. "John" can be considered generally important. However if the user is running late for a meeting with Jane Doe that will take place in 10 minutes, then Jane becomes undeniably more important than anyone else, and her importance should rise and decay naturally as time flows around the scheduled event.

Figure 2: The Importance Scenario.

5 DIMENSIONALITY REDUCTION (DR) IN M-PIM ACCESS

DR, as the name suggests, is an algorithmic technique for reducing the dimensionality of data, applied in several computer science fields such as databases, information retrieval, data mining, recommendation systems, signal processing etc. Real-world data usually has a high dimensionality, a fact that affects data processing performance (the so called "curse of dimensionality" originally appearing in (Bellman, 1957) 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 $x=(x_1,\ldots,x_p)^T$, find a lower dimensional representation of it, $s=(s_1,\ldots,s_k)^T$ with $k\leq p$, that captures the content in the original data, according to some criterion (Fodor, 1992). There are two categories of methods in order to solve the problem: a) feature selection, where an optimal subset of features (dimensions) is chosen and b) feature extraction, where existing features are combined and transformed to new ones.

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. Feature selection for context DR is not practical, as it is neither possible to know the best-describing features of the context vector nor their weights in advance. Since, as already presented, 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.

FastMap algorithm:

1. Find two objects that are far away.
2. Project all points on the line the two objects define, to get the first coordinate.
3. Project all objects on a hyperplane perpendicular to the line the two objects define.
4. Repeat k-1 times

Figure 3: The FastMap Algorithm.

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 (Faloutsos and Lin, 1995). 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 (Faloutsos and Lin, 1995) show that the algorithm performs well for visualization and clustering. The algorithm functions as shown in figure 3 and its complexity is $O(Nk^2)$, where k are the dimensions of the target space. A further advantage of this approach is that context vector "atoms" need to be defined by application developers just once – the recursive and atom-agnostic nature of the algorithm allow it to work for any context description vector.

6 CONCLUSIONS

In the previous sections, we underlined the significant role that mobile social context can play in retrieving and presenting information to the user from repositories within the mobile device that may contain large volumes of data. We have proposed the use of a vector model for the representation of context and since finding weights (that may change dynamically) is a very complex task, we introduced the idea to apply the technique of dimensionality reduction. This technique is characterised by its ability to automatically produce meaningful clusters of related information and thus can make the contextualised visualization of personal information items in 2 or 3 dimensions feasible.

We believe that our technique as described above can be very useful in order to build rich singular, two or three-dimensional information retrieval interfaces that will support data from multiple information repositories and present them in context, as demonstrated in the mock-ups presented in this paper (figures 4 & 5).

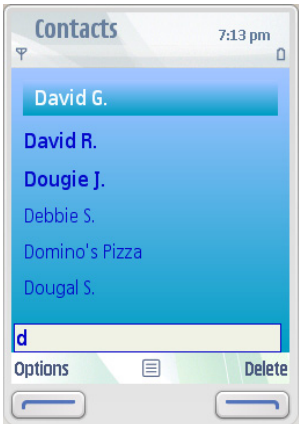

Figure 4: Hybrid one-dimensional mapping of importance (important items are highlighted in bold but ordered alphabetically so as to maintain user familiarity with existing UIs).

In figure 4 a one-dimensional hybrid interface is presented, where contacts are sorted alphabetically, but for each letter the important contacts are on top of the list and the remaining follow in alphabetical order. In figure 5 we show some concept renderings of 2-d and 3-d retrieval interfaces. In the case of 2 dimensions, the dimension of time can be preserved and as a result the user is able to retrieve the most important contacts during each time period. Finally, in the case of 3 dimensions an example of how this technique extends beyond the domain of contact lists is presented. The respective figure (5b) shows how several personal information items (contacts, e-mails, SMSs etc.) could be projected in a 3-dimensional space (with possible dimensions presented on the axes of time, importance and distance from current location).

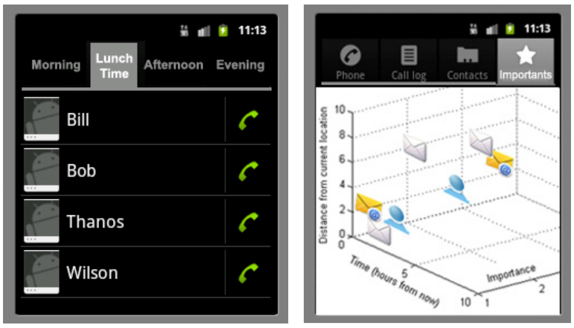

Figure 5a (left) and 5b (right): Retrieval UIs using 2D (left) and 3D (right) projections of item (contacts, e-mails, photos etc.) importance combined with retained, unprojected dimensions (time, distance from current location).

At this point in time our work focuses on a context-enabled contact list and following trials will extend to support a richer information space that will include all types of media and information pertinent to those contacts, enabling a new mode of context-based search and retrieval for mobile devices.

ACKNOWLEDGEMENTS

This work has in part been supported by the National Strategic Reference Framework (NSRF) (Regional Operational Programme – Western Greece) under the title “Advanced Systems and Services over Wireless and Mobile Networks” (number 312179).

REFERENCES

Ankolekar, A., Szabo, G., Luon, Y., Huberman, B. A., Wilkinson, D., Wu, F., 2009. Friendlee: a mobile application for your social life. In *Proceedings of MobileHCI '09*, Bonn, Germany.

Beach, A., Gartrell, M., Xing, X., Han, R., Lv, Q., Mishra, S., Seada, K., 2010. Fusing Mobile, Sensor, and Social Data To Fully Enable Context-Aware Computing. *In Proceedings of HotMobile 2010*, Annapolis, Maryland, USA.

Bellman, R., 1957. *Dynamic Programming*, Princeton University Press. Princeton, NJ, USA.

Bentley, F., Kames, J., Ahmed, R., Zivin, R., Schwendimann, L., 2010. Contacts 3.0: bringing together research and design teams to reinvent the phonebook. *In Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, Atlanta, USA

Bergman O., Komninos A., Liarokapis D., Clarke J., 2011. "You never call: Demoting unused contacts on mobile phones using DMTR", *Personal & Ubiquitous Computing*, Online First, DOI 10.1007/s00779-011-0411-3.

Boardman, R., Sasse, M. A., 2004. "Stuff goes into the computer and doesn't come out": a cross-tool study of personal information management. *In Proceedings of CHI'04*, Vienna, Austria.

Böcker, M., Suwita, A., 1999. Evaluating the Siemens C10 mobile phone – Beyond “Quick and Dirty” Usability Testing. In *Proceedings of HFT’99*, Copenhagen, Denmark.

Du, W., Wang, L., 2008. Context-aware application programming for mobile devices. In *Proceedings of C3S2E Conference,* Montreal, Canada.

Faloutsos, C., Lin, K. I., 1995. Fastmap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. In Proceedings of ACM SIGMOD '95, ACM, New York, USA.

Fodor, I. K., 1992. A survey of dimension reduction techniques. *LLNL Technical Report*, UCRL-ID-148494.

Hsu, S.-H., Cubauld, P., Jumpertz, S. 2009. Phorigami: A Photo Browser Based on Meta-categorization and Origami Visualization. *Human-Computer Interaction: Novel Interaction Methods and Techniques Lecture Notes in Computer Science*, Volume 5611/2009, 801-810, Springer.

Gaur, S., 2008. Mobile Phone Contact. *In: Ylä-Jääski A, Takkinen L (eds) Technical Reports in Computer Science and Engineering*.

Gilbert, P., Cuervo, E., Cox, L., 2009. Experimenting in Mobile Social Contexts Using JellyNets. In *Proceedings of HotMobile 2009*, Santa Cruz, USA.

Jung, Y., Anttila, A., Blom, J., 2008. Designing for the evolution of mobile contacts application. In *Proceedings of MobileHCI'08*, Amsterdam, Netherlands.

Klockar, T., Carr, D., Hedman, A., Johansson, T., Bengtsson, F., 2003 Usability of Mobile Phones. In the *19th International Symposium on Human Factors in Telecommunications*, Berlin, Germany.

Komninos, A., Liarokapis, D., 2009 The Use of Mobile Contact List applications and a Context-Oriented Framework to Support their Design. In *Proceedings of MobileHCI'09*, Bonn, Germany.

Korpipaa, P., Hakkila, J., Kela, J., Ronkainen, S., Kansala, I., 2004 Utilising context ontology in mobile device application personalisation. In *Proceedings of MUM'04*, ACM, New York, USA.

Oulasvirta, A., Raento, M., Tiitta, S., 2005. ContextContacts: re-designing SmartPhone's contact book to support mobile awareness and collaboration. In *Proceedings of MobileHCI'05*, Salzburg, Austria.

Rhee, Y., Kim, J., Chung, A., 2006. Your phone automatically caches your life. In *ACM Interactions 13, 4*.

Tolos, M., Tato, R., Kemp, T., 2005. Mood-based navigation through large collections of musical data. *Consumer Communications and Networking Conference, 2005*.

Toninelli, A., Khushraj, D., Lassila, O., Montanari, R., 2008. Towards Socially Aware Mobile Phones, In *Proceedings of SDoW 2008,* Karlsruhe, Germany.

Yoon, Y., Ahn, Y., Lee, G., Hong, S., Kim, M., 2008. Context-Aware Photo Selection for Promoting Photo Consumption on a Mobile Phone. In *Proceedings of MobileHCI’08*, Amsterdam, Netherlands.