Analysis of the Contextual Behaviour of Mobile Subscribers

Hannu Verkasalo and Borja Jimenez Salmeron

Helsinki University of Technology
hannu.verkasalo@tkk.fi

Abstract. In this paper, contextual behavior of mobile subscribers is studied with data collected straight from smartphones. The paper develops an approach to study how people use mobile devices in different contexts, by proposing an algorithm that works with device-based sensor data. This approach consists of context detection and data analysis. The context detection algorithm analyses cellular network radio logs in modeling the location of people. This paper then analyses usage patterns over different contexts. Demonstration of the contextual modeling with a sample of Finnish smartphone users proves that the applications of the approach are numerous.

Keywords: context modeling; context detection; mobile phone usage.

1 Motivation

The programmability of smartphones has facilitated new research topics, such as the modeling of contextual behavior of people [1]. The issue how people use their mobile phones in different situations is a new topic in market research.

Earlier mobile end-user research (see [14] and [15]) has focused on the analysis of application usage patterns or data service adoption, for example. However, the utilization of handset-based research data facilitates additional dimensions of end-user research. More specifically, analysis of service usage can be conducted over contexts, and locations. This essentially means that in addition to time stamps, also location data is used in the analysis. This gears end-user research towards more context-specific directions. This is a natural evolution path of mobile end-user research, as mobile services themselves are supposed to deliver context-specific value to end-users [16].

The scope of this paper is the analysis of context-specific data collected from mobile handsets, and analysis of that data from perspective of end-user research. The research problem of the paper is: "*What kinds of end-user research approaches are being facilitated by the automatic collection and processing of contextual data from mobile phones?*"

2 Background

Several studies related to context and context-aware systems conclude that a definition of context varies depending on the particular situation and the aim of the study.

R. Mehmood et al. (Eds.): EuropeComm 2009, LNICST 16, pp. 252–266, 2009.

Context-aware systems are used for different purposes such as sensors-based context-awareness for PDA interfaces, adaptive mobile phone application or mobile context-aware tour applications (see [2], [3], [4]). On the other hand, smartphones facilitate contextual computing easier than earlier, with programmable software platforms. Although attempts to create a standardized definition of use-context have been made, context is an abstract concept generally adapted to every specific application that models it [5]. Based on earlier research, context can be understood as something related to all situational factors around the user in a particular use case, a group of variables, parameters and characteristics that specifies the situation of an entity in its current environment where time and location are key aspects for its classification (see e.g. [6], [19], [20]).

Rule induction and machine learning are both valuable data mining techniques transforming data on cell-id transitions to information of location and context. The application of this discipline in the use of algorithms that automatically process huge amounts of cell-id location data has been proven as an accurate method to model end-user context. Once the detection of different contexts, automation of data processing or the visualization of results are solved research questions, the interest focuses on the analysis of these data and the extraction of results. [6]

Context modeling can be applied to social sciences as well in matters such as group behaviour (families, friends, etc) through application usage (study of the likings and the connections among people by listing their music preferences) [7]. The analysis of contextual behaviour of mobile subscribers also covers relevant aspects related to marketing like user segmentation. This paper focuses on how to utilize the data outputs of a context detection algorithm in order to deliver results and interpret them (intensity analysis based on propensity to use is presented as a method to extract reliable conclusions).

Most of the earlier research tends to use location and context as synonyms, understanding the process of context detection as geographical positioning. Vast majority of the applications and tools developed are like sensors that provide additional support rather than noticing the current location (e.g. context phone applications informing about friends in close locations), although some attempts to identify present location and predict future movements have been carried out in the last years (e.g. [8], [9]). But context is much more than mere location and the translation of contexts into geographical locations becomes a secondary objective from the academic point of view.

Previous research on context-aware applications and services pinpoints the difficulty of modeling and identifying context. However, the user's necessity to know contextual information of the person they want to contact to has been presented in several papers (e.g. [10], [11], [12]). Privacy matters regarding to the personal data used for the new tools, methods and experiments in social and context-aware systems must be taken into account as well [13].

The present paper introduces an algorithm to detect end-user context through transitional cell-id logs. With the results obtained of applying this algorithm to real data from the Finnish market, the paper analyses the contextual behaviour of mobile subscribers and explores possible applications in different studies where the contextual perspective can provide with additional and valuable information (e.g. propensity to use mobile services, visualization of movements and service usage distribution). Conclusions are presented last, with special focus on the future research.

3 Context Detection Algorithm

Figure 1 illustrates the context detection process of the algorithm used for modeling contexts [6]. The goal of this algorithm is to extract contextual information of mobile subscribers from data logs of their cell-id transitions and classify every cell visited for every user into abroad, home, office or on the move context. The logic of the algorithm can be divided in three main steps: pre-processing of the input data, cluster extraction and context detection. Finally, the algorithm generates several output files about the context of the user at every moment during the panel. The unique part of the algorithm is that it is developed specifically for smartphone-based sensor data, and it can be used easily to study other phenomena like how people use devices, in a post-analysis mode.

The algorithm has been developed as a main supportive aspect in the modeling of mobile end-user context and it has been applied in several studies regarding the "Modeling of Mobile Internet Usage and Business" project (e.g. [6], [17], [18]).

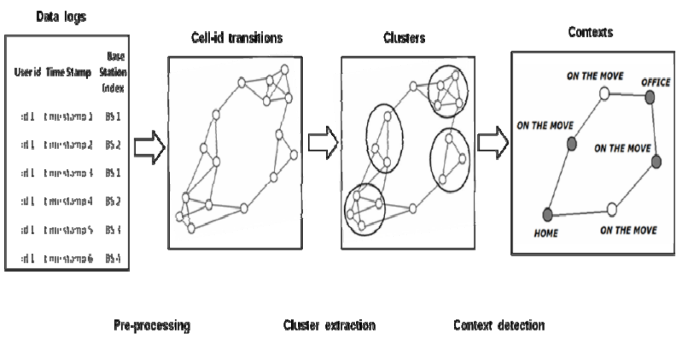

Fig. 1. Context modeling process

3.1 Pre-processing

The preparation of the data logs used as input for the algorithm and time-based calculus are the main parts of this stage. The handset-based end-user research method applied not only measures the usage frequencies, durations and volumes of all terminal features and applications, but also stores transitional information regarding base station connections. The algorithm cleans these logs from missing variables, wrong data or duplicated cases and prepares a file containing the map of every user's movements. This log file is a sequence of timestamps and cell-ids. Every time a user changes the position, a new connection to the base station covering the present cell takes place and this transition is stored in the data logs mentioned. At the end of the day, all the movements are mapped in these logs.

Once the algorithm has pre-processed the data, new variables reflecting the time spent at every cell are calculated. These variables store not only total time spent for a user on every cell id, but also time spent during weekends, at nights or during working hours (night and working hours are defined in the context detection section). The time-based calculus is the key for context detection and it is used in the process of cluster extraction as well.

3.2 Extraction of Clusters

The objective of the cluster extraction process is to group all the cells that are physically close to each other into clusters, utilizing specialized clustering heuristics developed for this purpose and appropriate for the nature of cell-id logs. The coverage of base stations generally overlaps in their limits because the zones without network coverage have to be minimized in order to provide service at every place. Power looses, interferences or movements inside households placed in these overlapping zones that force the handsets to connect to a different base station for a short period of time produce long sequences of transitions among close cells.

The data log used contains many sequences of changes between particular cells, changes that are very short in time mostly. Thus, it can be observed a significant number of cases where a handset connected to a base station changes to a new base station in the following case and presents another transition to the first one afterwards (it can be seen as a "*sandwich*"). These transitions between base stations usually appear many times in the file, what means they are not just a bug because of a dead battery, interference or a power loose of one of the base stations involved. A repeated sequence of cells like that is relevant in the way that it permits the extraction of user behaviour and contextual information.

The algorithm searches for this kind of situations in the file, detecting the exact number of repetitions of these transitions between every pair of cells. If the number of transitions is higher than a pre-determined threshold, both cells are grouped into one renaming their cell-ids into the same. At the end, all the cells are grouped into clusters. These clusters will consist of a non determinate number of cells or, in some cases, the clusters will be formed by a single cell (generally for most of cells detected as on the move context since users do not stay on them for long periods of time what minimizes transitions among them).

3.3 Classification of Cell-Id Clusters

The cluster extraction is the first step in the context classification since the context detection algorithm works with groups of cells. In this step, every cluster is classified into abroad, home, office or on the move context. The algorithm follows a set of rules in order to identify to which group of contexts every cluster belongs to as described in the decision tree on Figure 2.

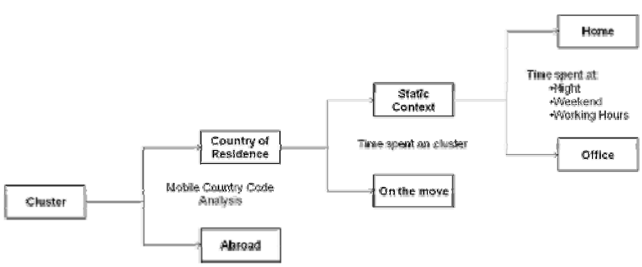

Fig. 2. Decision tree for the context detection process

Abroad context detection is carried out by examination of the mobile country code present in every report. After the identification of the country of residence for every user (the most frequent country code visited), the algorithm searches and marks every visit abroad.

For the rest of the contexts, time-based heuristics are required. First, the algorithm classifies between static and on the move contexts by comparing the amount of time spent at every cluster to a pre-established threshold. If the context is a static one, the next step sets the cluster under analysis as home or office checking and comparing the amount of time spent at nights, weekends and working hours to every corresponding threshold.

The decision of the thresholds as well as the definition of working and night hour is not free of subjectivity. In spite of this, a rigorous sensitivity analysis has been carried out to determine the thresholds that optimize the results [6].

4 Analysis

The context detection algorithm described above generates several output files storing contextual information of mobile subscribers. The analysis section focuses on how to interpret these outputs in order to provide valuable results that can be applied to several fields such as marketing (user segmentation), social sciences (behavioral analysis) or adoption of new technologies and services (technology and service usage analysis) among others [21].

4.1 Dataset

The data used in this paper comes from the context detection algorithm's output files commented in section 3. This algorithm transforms cell-id transitions contained in handset-based data logs of mobile subscribers into contextual information as already described [6, 14, 22].

All the data utilized comes from a Finnish panel provided by Nokia. It can be considered a representative sample of the Finnish market composed of 576 users. From those, 211 users with home and office context properly detected were finally indentified (for this purpose, questionnaires regarding context/location and demographic information were used) [6]. Figures illustrating following sections have been carried out using information from these 211 users.

4.2 Modeling of Movements

The context detection algorithm presented in the previous section is able not only to identify end-user contexts but also to generate a result file storing the physical movements of every user analyzed. These movements can be easily translated into a network plot. Despite the conversion of base stations into geographical locations is not possible because of privacy issues, these visualizations can help in the context detection by giving a new type of information. The advantage of a graph is that it offers the possibility to understand a big mass of data at first sight using nodes, arrows, sizes and colours. The nodes are clusters of network cells, the arrows represent the transitions between these clusters (movements of the user), the size of the nodes symbolize the importance of the clusters in terms of time (total amount of time spent for the user in a cluster) and the code of colours for the nodes informs about the most frequent hours of the day in which the user spends his time in the cluster represented by that node.

Although the result is not a geographical map, it provides information about movements (these visualization show all the "paths" that a user follows over the time of the study), relevant context environments (groups of connected clusters) and the most visited locations (represented with the biggest nodes). Besides, these figures can be used to test the process of context detection (e.g. the environment of a cell classified as office should be, generally, green and orange because these are the colours assigned to the block of the day when people usually work at office; office context should not appear in the weekends visualizations; the node representing home context should be the biggest, etc).

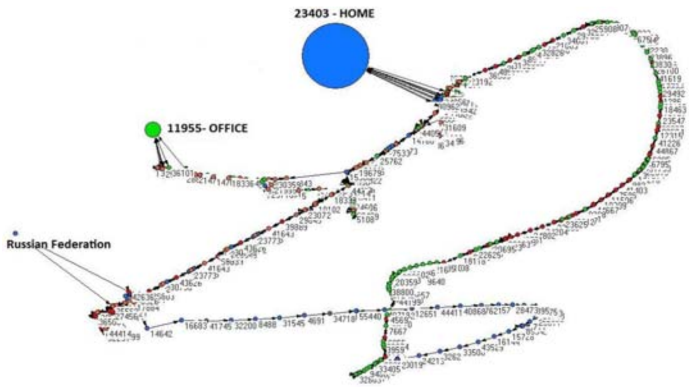

Fig. 3. Movement's network for a single user

The analysis of the average number of different cells visited per hour shows how during weekdays there are two peaks in the number of different cells visited at 7 a.m. and 4 p.m. corresponding to the hours where users typically go to work. There is a noticeable fall between these two peaks corresponding to the standard working hours. Weekend's analysis presents, however, a different curve closer to a Gaussian with its peak of movement at 3 p.m. and a bigger number of movements per hour in average.

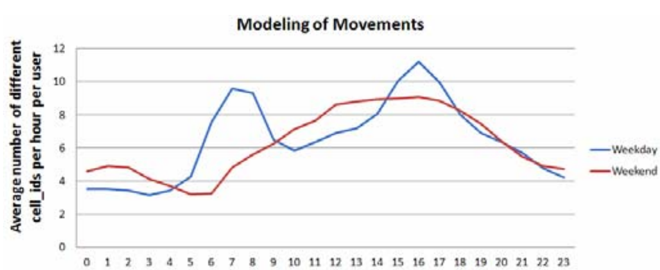

Fig. 4. Average number of different cells visited per hour

4.3 Time Distribution

Additional conclusions can be inferred through the visualization of the proportion of time spent on context. First, there is a logical increase at office context presence during standard working hours (i.e. from 7 a.m. to 4 p.m.) corresponding to a consequent fall at home. Next, the presence at office context during night time does not reach absolute 0% because of the big number of users analyzed (not every one of these users e.g. has to have a day shift) and the algorithm's restrictions (the conditions used in the context detection process tend to be flexible rather than strongly restrictive in order to identify all the possible office contexts) [6]. Last, on the move context presence reaches a peak from 4 p.m. to 7 p.m. corresponding to the time when most of the working users leave office. Curiously, there is no similar peak in the morning when users are supposed to go to work. It can be explained by the randomly chosen plans and paths coming up at the end of a working day.

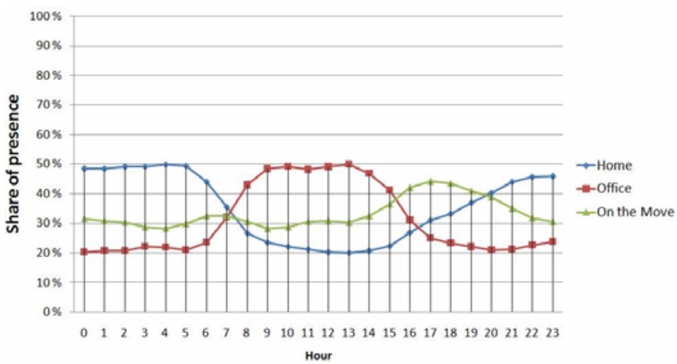

Fig. 5. Proportion of time spent on context

4.4 Distribution of Application Usage

The time distribution of application usage (in launches per hour) gives another perspective of the mobile subscriber behavior. Wide use of voice calls and messaging along the day (excluding night time) with a peak of usage at 3 p.m., the increase in the

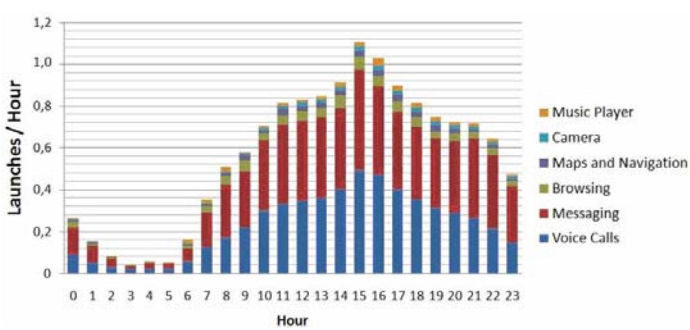

Fig. 6. Distribution of launches per hour between applications and hours of weekdays

usage of maps and navigation applications at hours when users are supposed to go to office or come back home from there, and the similar usage at any time of the rest of the applications (excluding again night time from 0 p.m. to 7 a.m.) are some initial conclusions extracted from figure 6. The comparison of weekday's and weekend's usage (Figure 12 in Appendix A) shows that from 10 a.m. to 11 p.m. there is less variation at weekends although general usage decreases when comparing it to weekdays.

4.5 Propensity to Use Mobile Services

Basic service usage analysis under contextual perspective have shown how deceptive the results can be when considering absolute usage times or relative rates normalized using the total amount of time spent for the user under analysis [6].

The classification of the context into home, office and on the move suggests the use of different variables to analyze results since the time spent at home is, for obvious reasons, bigger than the time spent at the other two contexts although this fact does not imply that the service usage is more intense at home. On the contrary, graphical results have shown that the intensity of usage is bigger on the move and at office contexts rather than at home in all of the services and applications analyzed.

Analysis of the propensity to use of mobile services and applications considers time spent on every context individually what makes it a better measurement of user activity giving a more realistic perspective and a more accurate result. The following descriptive figures regarding propensity to use certain applications consider the time spent on every application at every specific context and hour of the day.

Individual studies of applications over contexts at weekdays show that voice is the most used application at home and on the move, closely followed by SMS. However, SMS is preferred to voice at office context where substitutes such as land lines or meetings probably explain the fall in voice call usage. The figures describing propensity to use camera, music player or browsing show that although usage is much lower compared to voice or SMS services, the propensity to use these applications at on the move and office is considerably higher than at home (see Figures 7, 8 and Appendix B). Results are logical considering that camera and music player are basically on the move applications and that alternative ways to access the Internet at home and office makes browsing another clear on the move application.

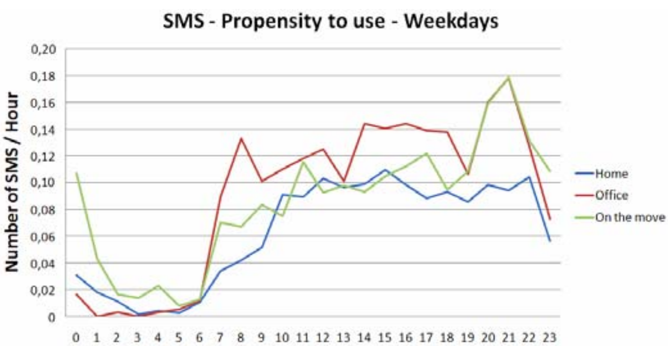

Fig. 7. Propensity to use "SMS" applications during weekdays (number of SMS sent/Hour)

Propensity to use maps and navigation tools at weekdays illustrates that this is again a clear on the move application reaching two peaks at 11 a.m. and 5 p.m. (see Figure 8). First one can be explained by movements related to job during working hours and the second illustrates the time when most users go back home.

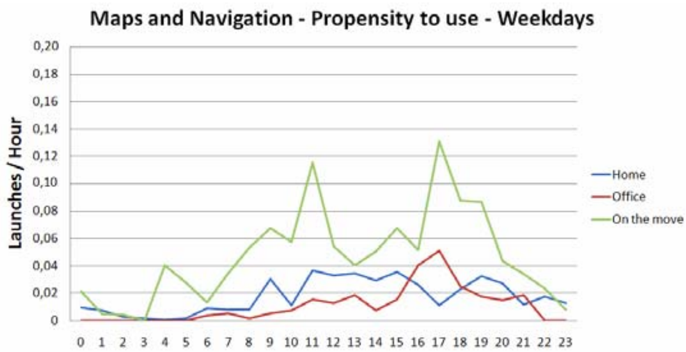

Fig. 8. Propensity to use "Maps and Navigations" applications during weekdays

Further analysis over maps and navigation, voice and music player applications has been carried out. Figures 17, 18 and 19 in Appendix C compare weekdays and weekends usage at home and on the move contexts. For all applications, usage at on the move is higher than at home what proves that this context is the most active considering the time spent on it (here lies the relevance of the propensity to use parameter). Focusing at on the move context, usage at weekdays tends to be higher than at weekends in all applications but in maps and navigation. As it is obvious, needs regarding this application during weekends becomes bigger as Figure 4 indicated in section 4.1 (the number of different cells visited per hour at weekends is higher than at weekdays what means that the users move more frequently at these moments and, consequently, the use of maps and navigation applications increase).

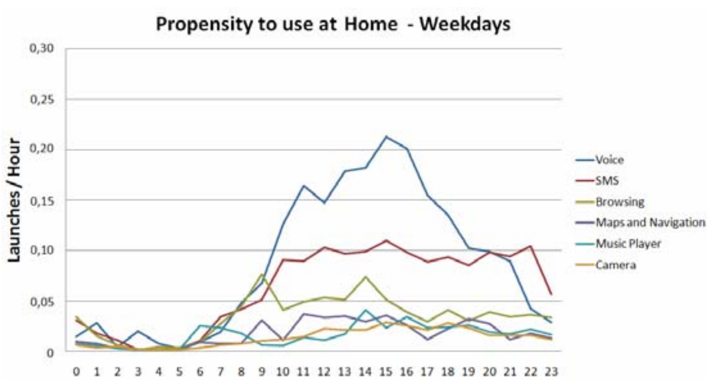

Fig. 9. Propensity to use applications at home context during weekdays

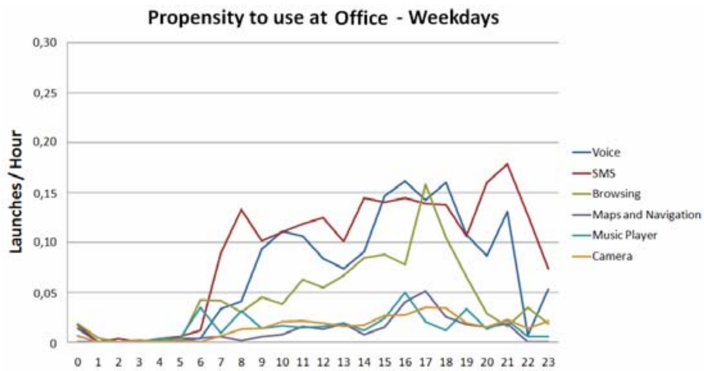

Fig. 10. Propensity to use applications at office context during weekdays

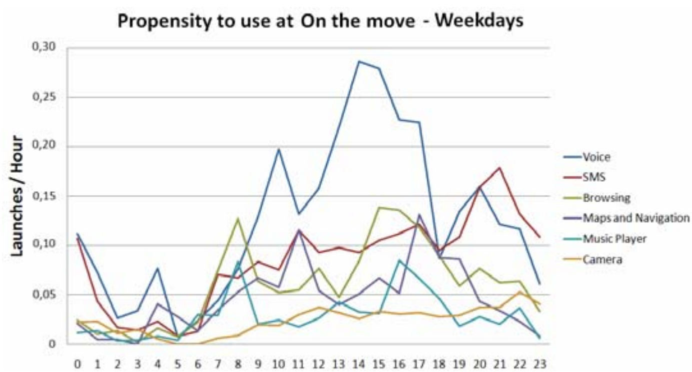

Fig. 11. Propensity to use applications at on the move context during weekdays

Figures 9, 10 and 11 compare the propensity to use all the applications analyzed at every specific context during weekdays. Figure 9 illustrates how the usage of voice and SMS imposes on the rest at home context. As commented before, alternative access to the Internet, substitute leisure devices and the lack of movement explain the fall at home of browsing, music player or camera and maps and navigation applications, respectively.

Figure 10 shows how usage of voice and SMS at office context decreases probably motivated for land lines and e-mail as replacement of SMS for communicating. The rest of applications present low rates of usage similar as the ones at home context.

In Figure 11, propensity to use applications on the move is illustrated. General usage of all the studied applications increases at every hour while on the move. Interesting conclusions can be drawn from the peaks reached at specific hours for certain applications. For example, maps and navigation presents an important peak at 5 p.m. when most users go back home. On the other hand, music player reaches two peaks at 8 a.m. and at 4 p.m., times when most of users are on the way to their offices.

5 Conclusion

This conclusion summarizes the main results, and proposes an agenda for further work in this area.

5.1 Results

This paper is focused on the application of the results of the context detection algorithm presented in section 3. The algorithm provides with a new dimension: context (e.g. [18], {21]). How to use the contextual dimension in standard service usage analysis among others is the matter of study.

The combination of the modeling of mobile subscriber movements, time distribution over contexts and time distribution of application usage lead to the use of a new variable for the analysis of the context-specific behavior of mobile subscribers: the propensity to use. This parameter has been proven as a suitable method to interpret results where context has to be considered. Propensity refers to the intensity to use mobile applications at specific contexts. Graphical plots have demonstrated how on the move context is the most active or, in other words, how the propensity to use every one of the applications analyzed is higher at on the move. The figures show also that propensity to use at weekdays is higher than at weekends in all applications but maps and navigations (the needs of location information increase during weekends when users move more frequently). Visualizations comparing propensity to use all services at different contexts illustrates the dominance of voice and messaging applications over the rest at any hour of the day in every context analyzed. On the other hand, general usage of all applications decreases at office context, particularly voice and SMS.

Although results prove that mobile subscribers use applications more actively on the move, home can be considered as the most important context considering total time of use. Despite this, the variable propensity to use applications has a big value in service usage analysis for academia and industry (e.g. adoption of new services or matters related to segmentation).

5.2 Future Research

The addition of contextual information improves classic service usage analysis helping in the understanding of mobile-subscribers behavior. Thus, a more detailed context analysis (e.g. regarding the number of sub-contexts detected) is an interesting line of research as far as wireless network traffic amounts continue increasing and intelligence of wirelessly connected devices evolves. If the importance of indoor areas increases, a deeper classification of contexts becomes relevant.

The integration of the context identification service in handsets that leads to the adaptation of devices to the current context is an important topic for future research. But before this, improvement on the accuracy of the context detection algorithm and the measurement client used to collect data can be done.

As it has been demonstrated along this paper, contextual information can be used in studies regarding adoption of new technologies. Propensity to use applications provides valuable support to network operators informing about where, when and how much their mobile subscribers use the applications and services installed on their devices. But the context detection process does not only cover service usage analysis neither is a supportive tool just for academia. The interest for industry is increasing for obvious reasons. Context detection can be applied to different fields such -as targeted marketing or user segmentation. The usage of contextual information to support person-situation segmentation and the search of killer applications under a contextual perspective are some examples of possible applications [18].

References

[1] Verkasalo, H., Hämmäinen, H.: Handset-Based Monitoring of Mobile Subscribers. Department of Electrical and Telecommunications Engineering. Helsinki Mobility Roundtable. Helsinki School of Economics (2006)

[2] Schmidt, A., Beigl, M., Gellersen, H.-W.: There is more to Context than Location. In: Proceedings of Workshop on Interactive Applications of Mobile Computing, Rostock, Germany (November 1998)

[3] Esbjörnsson, M., Weilenmann, A.: Mobile Phone Talk in Context. In: Dey, A.K., Kokinov, B., Leake, D.B., Turner, R. (eds.) CONTEXT 2005. LNCS (LNAI), vol. 3554, pp. 140–154. Springer, Heidelberg (2005)

[4] Long, S., Aust, D., Abowd, G.D., Atkeson, C.G.: Rapid prototyping of mobile context-aware applications: The cyberguide case study. In: Proceedings of the 2nd annual international conference on Mobile computing and networking, New York, United States, pp. 97–107 (1996)

[5] ISO 13407, Human-centred design processes for interactive systems. International Standard, the International Organization for Standarization (1999)

[6] Jimenez, B.: Modeling of Mobile End-User Context. M.Sc. Thesis, Helsinki University of Technology (May 2008)

[7] Eagle, N.: Machine Perception and Learning of Complex Social Systems. Doctoral dissertation, Massachusetts Institute of Technology (2005)

[8] Raento, M., Oulasvirta, A., Petit, R., Toivonen, H.: ContextPhone - A prototyping platform for context-aware mobile applications. IEEE Pervasive Computing 4(2), 51–59 (2005)

[9] Laasonen, K., Raento, M., Toivoinen, H.: Adaptative On-Device Location Recognition. In: Second International Conference on Pervasive Computing, Vienna, April 23 (2004)

[10] Kankainen, A., Tiitta, S.: Exploring everyday needs of teenagers related to context-aware mobile services. In: Proceedings of HFT 2003, Berlin, Germany, pp. 19–26 (2003)

[11] Tamminen, S., Oulasvirta, A., Toiskallio, K., Kankainen, A.: Understanding mobile contexts. In: Proceedings of Mobile HCI 2003, Udine, Italy, pp. 18–35 (2003); A revised version submitted to Personal and Ubiquitous Computing

[12] Oulasvirta, A., Kurvinen, E., Kankainen, T.: Understanding contexts by being there: case studies in bodystorming. Personal and Ubiquitous Computing 7, 125–134 (2003)

[13] Raento, M.: Exploring Privacy for Ubiquitous Computing: Tools, Methods and Experiments. Department of Computer Science, Series of publications A, Report A-2007-2 (2007)

[14] Verkasalo, H.: Handset-Based Monitoring of Mobile Customer Behaviour. Master's Thesis Series. Networking Laboratory. Department of Electrical and Telecommunications Engineering. Helsinki University of Technology, Espoo, Finland (September 2005)
[15] Verkasalo, H.: A Cross-Country Comparison of Mobile Service and Handset Usage. Licentiate's thesis, Helsinki University of Technology, Networking Laboratory, Finland (2007)
[16] Heinonen, K., Pura, M.: Classifying Mobile Services. Presented at Helsinki Mobility Roundtable, Helsinki, June 1-2 (2006)
[17] Modeling of Mobile Internet Usage and Business Project (2008), `http://www.netlab.tkk.fi/tutkimus/momi/`
[18] Uronen, M.: Market Segmentation Approaches in the Mobile Service Market. Master's Thesis, Networking Laboratory, Helsinki University of Technology (2008)
[19] Schmidt, A., Beigl, M., Gellersen, H.W.: There is more to Context than Location. In: Proceedings of Workshop on Interactive Applications of Mobile Computing, Rostock, Germany (November 1998)
[20] Lee, I., Kim, J., Kim, J.: Use Contexts for the Mobile Internet: A longitudinal Study Monitoring Actual Use of Mobile Internet Services. International Journal of Human-Computer interaction 18(3), 269–292 (2005)
[21] Smura, T.: Access alternatives to mobile services and content: analysis of handset-based smartphone usage data. In: ITS 17th Biennial Conference, Montreal, Canada, June 24-27 (2008)
[22] Verkasalo, H., Hämäinen, H.: Handset-Based Monitoring of Mobile Subscribers. Department of Electrical and Telecommunications Engineering. Helsinki Mobility Roundtable. Helsinki School of Economics (2006)
[23] Kivi, A.: Measuring Mobile User Behavior and Service Usage: Methods, Measurements Points, and Future Outlook. In: Proceedings of the 6th Global Mobility Roundtable, Los Angeles, California, U.S, June 1-2 (2007)

Appendix A – Distribution of Application Usage

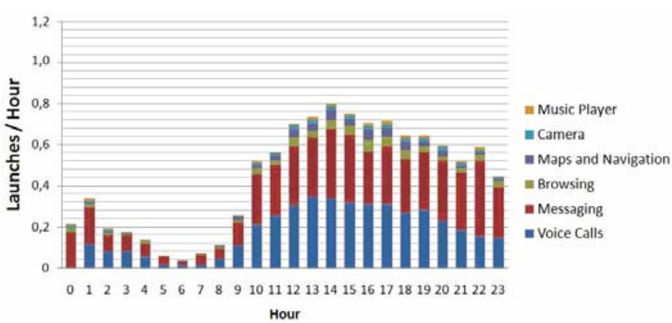

Fig. 12. Distribution of launches per hour between applications and hours of weekends

Appendix B – Propensity to Use Services

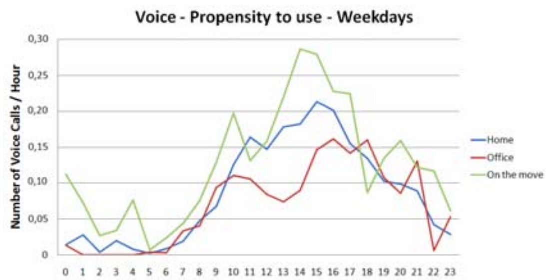

Fig. 13. Propensity to use "Voice" applications during weekdays (number of voice calls emitted/Hour)

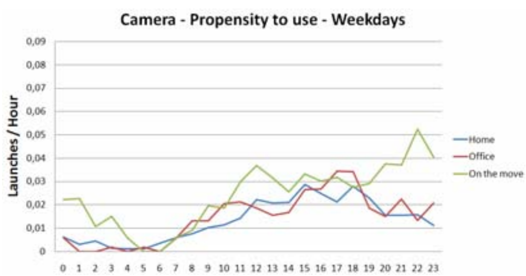

Fig. 14. Propensity to use "Camera" applications during weekdays

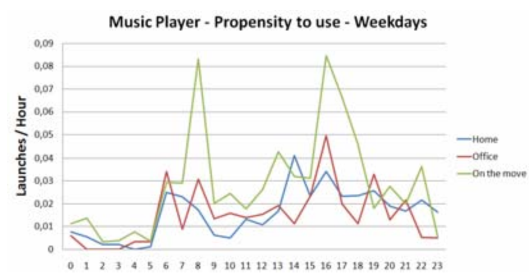

Fig. 15. Propensity to use "Music Player" applications during weekdays

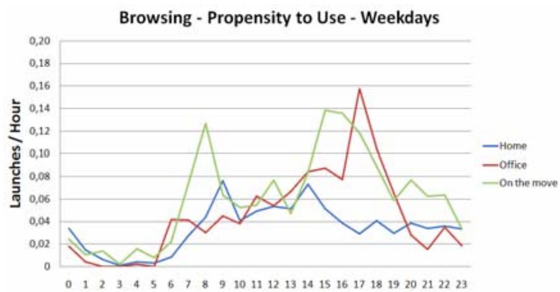

Fig. 16. Propensity to use "Browsing" applications during weekdays