Using communication frequency and recency context to facilitate mobile contact list retrieval

Athanasios Plessas[1,2]**, Vassilios Stefanis***[1,2]**, Andreas Komninos**[3]**, John Garofalakis**[1,2]
1. *Department of Computer Engineering & Informatics, University of Patras, Greece*
2. Computer Technology Institute and Press "Diophantus", Greece
3. Glasgow Caledonian University, United Kingdom

ABSTRACT
As mobile contact lists get bigger and bigger the cognitive load on the user increases while trying to retrieve the next contact to start a communication session. In this paper we focus on the task of retrieving a contact when the purpose is to start a phone call, examining mobile users' call logs and showing that it is possible to accurately predict the next contact to be called using relatively simple heuristics and algorithms that describe usage context. We present and discuss the results of the proposed method applied on a dataset collected from an experiment we organised involving 25 mobile users.

Keywords*:* Mobile context, mobile contact list, call prediction, call log analysis, mobile contact retrieval.

INTRODUCTION
Technological advances of the last decade have turned mobile phones to small multi-purpose personal computers being equipped with camera, GPS receiver, accelerometer, Bluetooth and other sensors. These devices are now used, among others, to access the World Wide Web, to transfer files, produce multimedia content, as email clients and digital calendars. However, mobile phones remain primarily communication devices (LaRue et al., 2010), supporting the communication needs of their users within their social networks with several tools. As such, it is reasonable that a common task for their owners include searching for a contact in a phonebook or selecting one from a recent-call list (Lee, Seo & Lee, 2010) in order to start a new communication session using one of the provided methods. As contact lists get increasingly bigger and since a significant percentage of contacts are never used (Bergman et al., 2012), the cognitive load on the user increases while trying to retrieve a contact from this repository. This effort is also obstructed by the limitation of the relatively small screen that mobile phones are equipped with. Furthermore, since call logs impart information about use and not *lack of use*, mobile devices have become good at supporting communication but provide little support for the task of managing social relationships (i.e. deciding who to contact and how frequently), leaving decisions entirely to the users.

At the same time, mobile devices collect a significant amount of data and information about the user's context, including location, the current date and time, the orientation of the device, whether the user of the device is on move and his speed, the user's current task (e.g. on the phone, messaging), whether the vibration or the silent mode are enabled etc. (Komninos et al., 2011b). The user considers her mobile device a "trusted device", usually having it close to her, sometimes operating 24 hours per day. Devices also contain a lot of personal information related to the user's social environment (Toninelli et al., 2008). These are either generated automatically by the device (e.g. a 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) or consist of user-generated content (e.g. SMS/MMS and multimedia files, browser's history, calendar events etc.). Therefore, a mobile device could also be 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 (Gilbert, Cuervo & Cox, 2009).

Our work is based on the hypothesis that context mined from personal interactions with a mobile device can be used to aid personal mobile information retrieval tasks. In this paper we attempt to address the problem of contact retrieval when performing an outgoing call, by predicting which contact is the most probable to be called at any time. Though communication often takes place on mobile phones through not just phonecalls and sms, but other networks such as Skype, Facebook, Twitter etc., it is not yet possible to collect all such communication data from the various apps due to varying data access permissions. Such permissions however are available for the most basic communication modes, i.e. phonecalls and short text messages.

While contacts are just one aspect of mobile personal information, considering modern devices' capacity to store information (several gigabytes are available on most devices) and the fact that additional storage capacity is afforded by cloud services, personal information management is likely to pose significant challenges to users in the future. Hence, methods for managing context with a view to inform retrieval tasks can be applied to multiple retrieval situations. We believe that a solution to such problems can be informed by mobile social context as mobile users seem to adopt different behaviour patterns under different contexts. As an example consider the following scenarios derived from our experiment participants that involve two different context dimensions, frequency and recency:

Scenario 1: George is from Greece but works and lives in the UK. George's contact list is in both English and Greek, and the constant switching between languages makes searching for contacts quite bothersome. Thus, he relies on his call log for retrieving contacts but, given the large number of calls he receives each day he frequently has to scroll up and down a lot before he can find a contact to call. Though not optimal, he prefers this style of interaction as he perceives it to be less annoying than switching languages and searching.

Scenario 2: Maria is a PhD student at the University. She rarely calls her supervisor on his mobile phone, however, today he called her to arrange a meeting. After the meeting, Maria tried hard to remember the name of a paper her supervisor recommended but she couldn't. She had to phone him again, as he was out of the office for the whole day.

The context dimensions that could be considered include location, frequency, recency, time of day, day of week and personal preference (user indicated favorite contacts). As a first step in our research, we consider only two context dimensions, frequency and recency of use of each contact for reasons that we explain in the methodology section. The paper focuses on examining the effect of each dimension on the success of predicting the likelihood of a contact being called. We show that although a simple idea, the combination of these context dimensions provides better prediction results than traditional access modes available in mobile phones (list of recent calls, list of more frequently used contacts), something that has implications in the design of better interfaces for communication support.

RELATED WORK

To the best of our knowledge, although the idea of taking advantage of context to provide adaptive services to mobile users is not new, little research has been conducted on predicting the next call a mobile user is going to make and providing a rearranged contact sub-list to replace traditional methods of contact repository access.

In (Lee, Seo & Lee, 2010) an algorithm that builds an adaptive speed-call list based on call logs is presented. Based on the observation that outgoing communication follows a periodical pattern, 5 dimensions (day of week, weekend/weekday spans, time of day, day-parts of a day, 1-hour slots of a day) are proposed as recommendation conditions. Whenever a user presses the call button, the algorithm computes the Bernoulli probabilities of each dimension for each contact, sorts all contacts according to the respective conditions maximum probabilities and creates the speed-call list. However, the probabilities of the proposed dimensions are considered separately and are not combined as in our approach.

In (Barzaiq & Loke, 2011) a similar approach to predict outgoing calls analyzing mobile phone historical call log data is described. Three dimensions are proposed to capture frequency and regularity of communication behaviour. An important finding from this research is that combining factors rather than exploiting each one independently leads to better results. The proposed algorithm analyzes historical data from a period of two years, a decision that adds computational load to the device and seems to be unnecessary since only a recent portion of communication history is needed to predict future behaviour (Phithakkitnukoon & Dantu, 2008). Moreover, the weights that are assigned to each dimension seem to be arbitrarily decided and the success rate is quite low (below 40%) for the period of 5 weeks that the experiment was running, following however an upward trend.

Another attempt to predict outgoing calls of mobile users is described in (Phithakkitnukoon et al., 2011). The researchers have implemented a call predictor for both incoming and outgoing calls. The outgoing call predictor constructs a probabilistic model capturing the user’s behaviour based on call departure and inter-departure times. The prediction algorithm provides good results (for example a success rate around 70% for a prediction list with 5 entries). Although the researchers prove that only recent history is needed to predict future communication behaviour, it is not clear whether the predictor takes into account all historical call data or only a recent portion of the call log. The presented probabilistic approach is promising; nevertheless it seems difficult to incorporate other mobile and social contextual dimensions, such as location or personal preference.

In (Subramanya, 2012) a framework for the analysis of mobile phone call data to determine the characteristics, the personality and lifestyle of mobile user and enhance its experience is presented. An elementary call log data analysis that gathers simple statistics is described along with a preliminary analysis of the call log data of a typical homemaker that reveals that the frequency of communication and the indication of a contact as ‘favorite’ could play an important role in providing a personalized communication experience. Moreover, the researcher presents his findings regarding how the calls are distributed within the week and according to geographic location but he does not suggest any further use of this information. The researcher denotes in the future work section the need for an analysis of data from multiple users.

Several other works exist, focusing on call records and other mobile data mining and analysis out of the scope of predicting the next call. In (Jeon et al., 2008) log data from mobile devices are used to classify their users according to their device usage pattern, a study that could lead to improvements in mobile user interfaces. Other researchers (Salovaara et al., 2011) have studied mobile usage logs in order to investigate occasional unavailability in a mobile communication context. In (Phithakkitnukoon & Dantu, 2010) a study of the impact of the mobile social closeness to the similarity in calling patterns and reciprocity is presented. Finally, mobile call data records are analyzed in (Calabrese et al., 2011) to investigate the relationship between people's calls and their physical location and in (Candia et al., 2011) to investigate various aspects of human dynamics and social interactions.

METHODOLOGY

This chapter first shortly presents the mobile handset-based data collection method and introduces the data collected, as well as some preliminary observations. Second, the context-based prediction procedure is presented.

Mobile Dataset and Users

In order to extract real communication data from mobile phones, we developed an Android application that extracts the contact list (and “starred” status), the call log, the SMS log and the existing contact groups from the mobile device in a text file. The application was delivered to 42 subjects with Android smartphones, however only 25 datasets were considered as valid, since some were incomplete (e.g. extremely small number of records in call log, coverage period of log being too short, too few contacts in their contact list). More specifically, the exact criteria used to include datasets from our analysis were the number of contacts (more than 20 were required), the number of calls in the call log (more that 150 were required) and the period that the log covered (more than 15 days were required). Concerning the 25 subjects that we take into account in our analysis, 22 of them were male and 3 female, while their age ranges were from 19 to 39 years old and they were from varied backgrounds, though most were Computer Science students as they were recruited after a public announcement circulated through the email list and the forum of the university community. In total, the participants’ contact lists contained 4185 entries. We found that on average, each contact list contained 167.4 entries (mean=167.4, stdev=87.60, min=33, max=344). An interesting point is that 13 of the participants had used a feature of the Android OS that allows users to indicate their personal preferences (marking a contact as starred) and promotes the preferred contacts at the top of the contact list’s favorite tab. The extracted logs covered a different time period in days for each mobile phone (mean=52.80, stdev=35.23, min=18, max=170). On average each user made 449.88 calls (stdev=98.12, min=182, max=500). We should stress here that the Android platform limits the call log history to 500 calls. The characteristics of the users are shown in Table 1.

Table 1. Users, contact lists and call logs

USER ID	CONTACTS	STARRED	CALLS	DAYS
2	286	8	499	45
4	293	10	500	20
5	70	8	366	52
6	247		500	51
9	203	5	500	47
13	59		500	74
14	239		500	37
17	130		500	75
18	72		383	170
19	69		500	58
22	149		500	45
24	291		499	64
26	202	2	490	103

33	183	9	500	24
34	125		500	39
35	65	8	498	18
36	128	11	182	21
37	132		203	30
38	96		270	37
39	234	2	499	35
40	88	3	500	77
41	206		499	115
42	344	7	500	41
43	231	9	359	19
44	43	5	500	23
Mean	167.4		449.88	52.8
SD	87.60		98.12	35.23

Lee et al. (2010) find that users in their study fell within two groups based on their perceived “socialness”. For the rest of the paper we use the term “socialness” not literally but to express the pattern of incoming and outgoing communication from the user’s mobile device. Having selected a suitable sample of users, we then wanted to see if we could organize these users into clusters, based on their communication behaviour and perceived “socialness”. The need for this categorisation is made greater because of the nature of the call logs, which have varying lengths and densities. Since it is not desirable to dilute the call logs by normalising or otherwise massaging the raw data, arranging the users into categories is a step towards ensuring the integrity of our conclusions. The obvious first step would be to look at the sparsity of communication, i.e. the number of calls made per day, however, this would not show emergent behaviour in terms of the “socialness” of a user. A user could make lots of calls per day, however these could be to a very few distinct numbers. We thus analyzed each user’s call log to examine the percentage of calls made to each of their contacts as follows: For each user, we took a list of all contacts to whom calls had been placed for their entire call log duration. For each of these contacts, we calculated the percentage of calls made over the total calls made by the user. Having sorted these contacts and their percentages in descending order, we calculated for each contact the difference from the previous contact (e.g. if there were three contacts with a percentage of outgoing calls of 80%, 15% and 5%, the differences would be 80%, 65% and 10%). We then calculated the means of all differences and used that as a metric (Socialness-Metric or S-Metric) to determine “socialness”. The lower the metric, the more “social” a user is, since this means the user calls more people and with less frequency. We finally ran a k-means clustering algorithm on these users, testing the results for k=1,2,3 and 4 clusters (since the empirical rule of thumbs indicated that k should be between 3 and 4) and an attempt to run the algorithm for three groups showed much better results by optimally reducing the within clusters sum of squares. The final groupings are depicted in Table 2. Group 1 is the “least social” users, i.e. people who tend to make most of their calls to just a handful of numbers and we hypothesized that they are therefore most likely to exhibit regular predictable behaviour. Group 2 are the “averagely social” group, while Group 3 are the “most social” users, in constant communication with a variety of contacts and thus likely to be most difficult to predict. As it can be seen, Group 1 has only three users, thus data for this group is reported henceforth with some reservation on its significance.

Table 2. User Groups

Group 1		Group 2		Group 3	
UID	**S-Metric**	**UID**	**S-Metric**	**UID**	**S-Metric**
13	0.072	17	0.027	2	0.011
35	0.078	18	0.042	4	0.007
37	0.150	22	0.029	5	0.009
		26	0.036	6	0.012
		33	0.033	9	0.021
		34	0.033	14	0.007
		39	0.037	19	0.014
		42	0.024	24	0.011
		44	0.024	36	0.018
				38	0.021
				40	0.009
				41	0.007
				43	0.016

Prediction Procedure

Our first step was to determine which context dimensions we would use for our analysis. Such dimensions could include location, frequency, recency, time of day, day of week and personal preference (user indicated favorite contacts). Obviously, location was not available in our dataset as the android OS does not record user location for phone call events. In previous work (Stefanis et. al, 2012), we found that the role of personal preference (“starred” contacts in android) is ambiguous since there is no correlation between the “starred” status of a contact and the probability of a call being made to her, hence we discarded this dimension. Temporal context (time of day and day of week) can be a function of other dimensions such as social relationship between user and contact, cultural norms, user and contact activity (Palen 2001), an indicator of user or contact location etc.. Furthermore, Peters and Allouch (2005) found that the use of mobiles blurs the boundaries between work and social life, which has been traditionally temporally separated. Thus, it cannot safely be examined on its own without knowledge of these other types of context, which of course are not available in our dataset. A frequently used model on communication logs for business analytics is the Recency, Frequency and Monetary model (e.g. Cheng & Sun 2012). However, in our case the cost of communication cannot be part of the analysis since this type of context data has to be manually provided by the user for each contact (e.g. intra-operator network call costs, national landline/mobile costs for other operator networks, international, sms and mms charges etc.). Providing this information would entail a huge interaction cost for each user.

Therefore, the safest dimensions that could be used as a first step to establish a baseline performance for a predictive system are frequency and recency of communication.

In order to evaluate the role of the frequency and recency metrics in predicting the likelihood of placing a call to a contact, we used the extracted datasets from our users to perform a series of predictions, using the concept of a sliding *training window*. This training window is defined as a subset within the call log datasets and is used to make predictions regarding the next call, since

Phithakkitnukoon & Dantu, (2008) showed that only an amount of recent data from the historical dataset is adequate to predict future behaviour. As such, we define the training window *t* to be of a fixed temporal size, measured in *t-days*, i.e. temporal periods of 864000 seconds (10 days). All calls made within this time are used as training data, upon which we attempt to predict the person to be called next. Within this training window *t* we also define a fixed recency window *r* measured in hours, which contains all the calls made in a fixed time period from the start of the training window. All calls within this recency window *r* are used to give temporal significance to contacts, in contrast with training window *t* which is used to capture the overall (historical) significance of a contact.

The prediction procedure works as follows: Suppose we wanted to predict which contact was called at position x in the call log. This interaction must represent an *outgoing* call, as we intend to support the contact retrieval task and as a result we do not make predictions for incoming calls (Figure 1). We then pick up all the calls (incoming and outgoing) in the call log that have taken place within the specified number of t-days from the timestamp of call x. We use both incoming and outgoing calls to capture the effect of reciprocity (Phithakkitnukoon & Dantu, 2010). This training window *t* can have thus a varied number of calls, which are used as training data on which the prediction is made. Once a prediction has been made, we record the outcome and move to the next call. This way, we work through the user's call log, call by call, and try to predict each one (obviously the earliest calls are only used as training data and not for predictions).

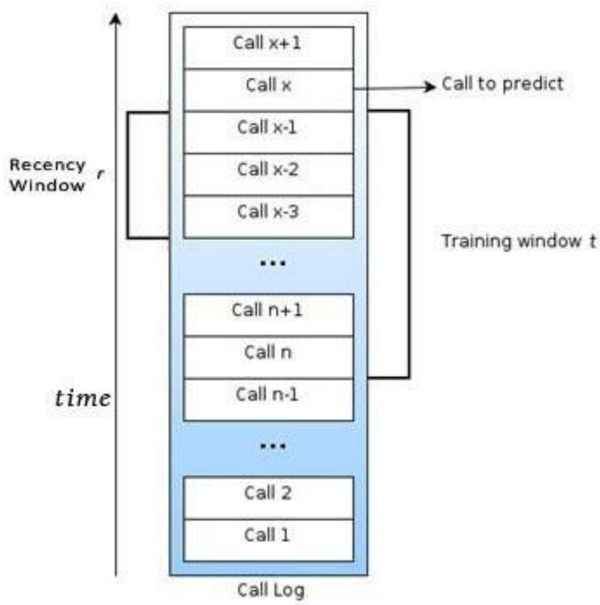

Figure 1. Operation of the sliding training window

The Prediction algorithm

In our approach, personal information items such as contacts are represented as context augmented vectors $(x_1,x_2,...,x_n)$ where x_i is the value of a context dimension i that characterizes the item. Our technique is based upon the context dimensions of contact use frequency and recency. Other dimensions of contact context can, of course, be incorporated in a predictive algorithm, but the purpose of this paper is to investigate the role of these two dimensions of context, thus we focus solely on these. For each contact in the user's contact list, we assign a score, comprised of the sum of a weighted score F(c) that reflects the frequency with which the contact has been used in the given training window, and a second weighted score R(c) that reflects the temporal distance of the latest use of that contact within the training window. The equation used to assign a score to each contact is:

$$\Pi(c) = w_f \times F(c) + w_r \times R(c)$$

where Π(c) is the score assigned to the contact c, w_f and w_r are the weights for the frequency score F(c) and recency score R(c) respectively. F(c) is calculated as the percentage of the communications (incoming and outgoing) within the training window between the user and the contact. R(c) is calculated as the percentage of the time interval between the start of a defined recency timeframe until the most recent communication between the user and the contact over the entire duration of the recency timeframe. In the case that there is no contact between them within the recency timeframe, R(c) is zero. For each call thus, we can pick the top n contacts based on this score and offer these as likely candidates for our prediction.

EXPERIMENTAL RESULTS

Experimental Considerations

Prior to proceeding with our experiments, we needed to determine an appropriate length for the training window that would be used, as well as to find a suitable temporal threshold for the recency score. The latter threshold is desirable as we have empirically found that recency of communication is influential, with a decayed effect, for a period of 6 hours. Experimenting more with our dataset, we observed slightly better results for a timeframe for 12 hours, so the temporal threshold was set to 12 hours. Phithakkitnukoon et al. (2011) demonstrated that the accuracy of his call predictor did not improve in line with the size of the training data in his prediction work. This is a reasonable outcome, since as people change behavioural patterns and perhaps interact more closely with different “social” groups during the course of time, older interaction data becomes not only redundant, but can be detrimental to the success of predictions. In our case, because the length of our call logs was 52 days on average, we could not experiment with too large a training window. The training window should be long enough to provide adequate data, however, a training window of more than 2 weeks would likely fail to capture dynamic changes in a user’s calling behaviour (e.g. taking a week off to go on holidays). We thus examined the performance of our technique, using equal weights for the frequency and the recency components and compared the success of the technique with all possible combinations of a training window of 10 and 15 days and a recency threshold of 6 and 12 hours. By examining the success means for all users, we found that our technique gave highest scores with a 10 day training window and 12 hours recency threshold, though the performance was not much better than in other combinations.

Finally, we needed to consider suitable suggestion list lengths for our experiments. Phithakkitnukoon et al. (2011) consider several sizes of prediction suggestions (up to 20), though we felt that a mobile interface that would offer quick access to a likely desirable contact should not display more than 8 suggestions (Stefanis et al., 2012), as this would force the user to further interact with the interface by scrolling, thus detracting from the usability of such a system. We decided to perform experiments for 1 (straight hit/miss), 3 (reasonable number for a mobile screen widget), 5 (to directly compare with (Phithakkitnukoon et al., 2011), (Lee, Seo & Lee, 2010) and (Barzaiq & Loke, 2011) and 8 (maximum that would fit on a mobile screen) suggestions.

Baseline Experiments

Apart from performing a search on the alphabetically sorted phonebook, two other methods are usually available on a typical smartphone when its user wants to retrieve a contact in order to start a phone call: using the list with the most frequently called contacts or the list with the most recent calls. These two methods correspond to executions of the prediction algorithm with pairs of weights (w_f=1,w_r=0) and (w_f=0,w_r=1) respectively. We consider these executions as a baseline experiment, the results of which, compared with the results of the executions where both dimensions are combined, reveal the improvement of our approach over the traditional contact retrieval methods. In Figure 2 and Figure 3 the average performance of the two methods for each user group for different suggestion list lengths is shown. If we equate these performances to the use of the call log screen and the frequently used contacts screens on an real device, we would thus expect that most “social” users (Groups 2 & 3) could only expect to find the person they really want to call at the top of the call log around 40% of the time and at the top of their most used contacts around 20-35% of the time. Additionally, for these groups, performance peaks at around 60% as they look further into the call log history without much improvement after looking at the first 3 entries, while there seems to be an almost linear increase in the likelihood to find the contact that they want to call, as the size of the most frequently called list grows. Group 1 (the least “social” users) exhibits a similar behaviour, but with much better performance, as expected.

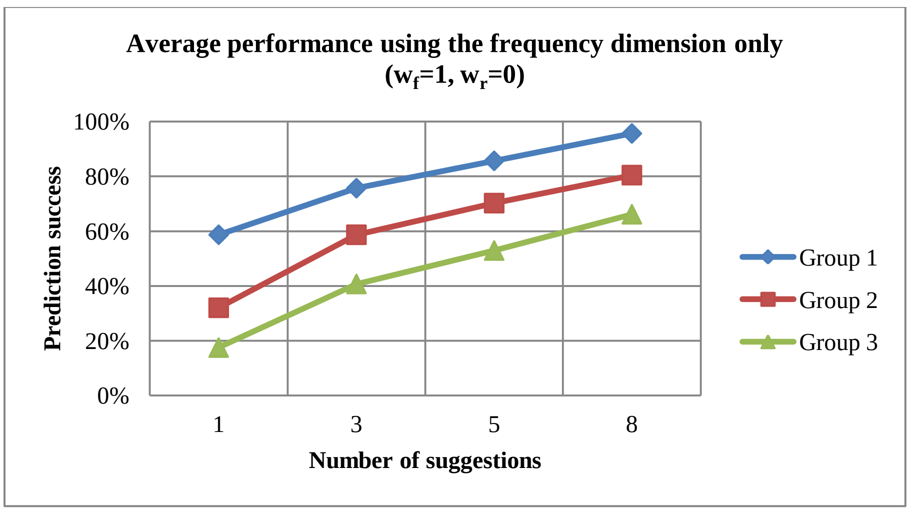

Figure 2. Performance using only the frequency dimension (equivalent to a list of most frequently used contacts)

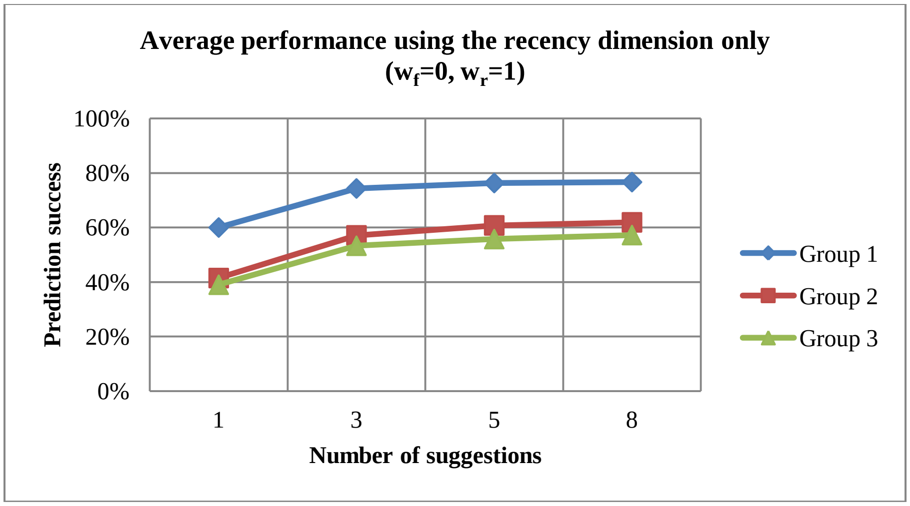

Figure 3. Performance using only the recency dimension (equivalent to using the standard call log)

Actual Experiments

Our experiments are divided into two distinct sets that explore the relationship between the importance of the Frequency and the Recency criteria, as discussed in the previous sections. The methodology of the experiments remains precisely the same except that in the first run (Set A) we are only interested in knowing whether the actual called contact is in the list of suggestions (a "hit"), while in second run (Set B), we keep a track of the position in the suggestion list that the contact is found, in the case of a hit. In this case, a suitable score to rate the quality of the prediction is given, which ranges in increments of one unit between [1 .. n] that reflect the number of positions available within the suggestion list (higher is better).

Experiment set A – hit or miss

The following graphs show the success rates for all users. As an example, we show the complete graph for the success performance of one suggestion (Figure 4) but to conserve space, we show the average success for each suggestion list size for all users in Figure 5. Finally, Figure 6 Figure 6 shows the performance for all suggestion list sizes per group, while Figure 7 shows the precise breakdown for each list size and for each group. The first conclusion that is immediately obvious is that using the Frequency or the Recency dimensions alone offers worse performance than any combination of weight. This indicates that the standard mobile device screens that provide a call log and a most frequently used contacts view, are less than optimal and that an interface that would provide call suggestions based on both metrics, is much more effective. It is clear also from these results that Group 1 has consistently the best performance, while Groups 2 and 3 follow. This confirms our hypothesis, as Group 1 exhibits the most predictable behaviours (frequent calls to a very small number of users). Additionally, we observe that as the size of the suggestions list grows, the role of the weights becomes less important. For a small suggestion list (1-3 suggestions), the weight of the Recency dimension seems to play a more important role in obtaining a "hit", which is a clear indicator that call recency is more important that call frequency for determining the importance of a contact.

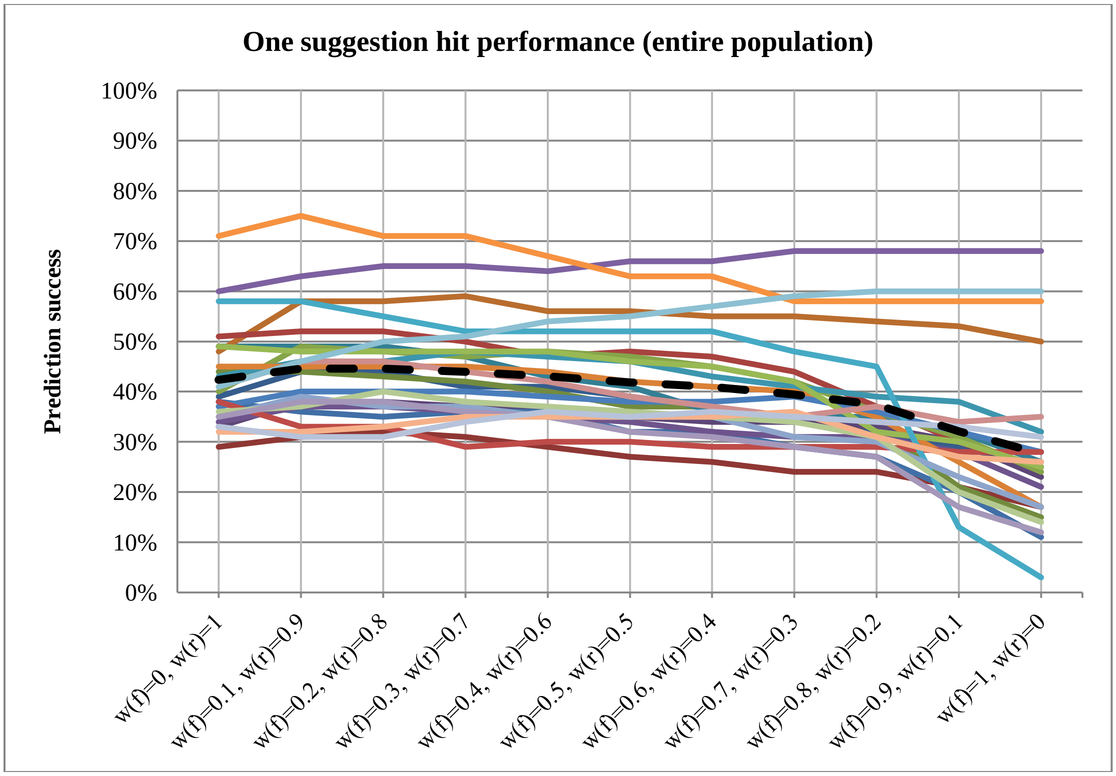

Figure 4. Overview of the one suggestion hit performance for all users. The population average is denoted with the dashed black line.

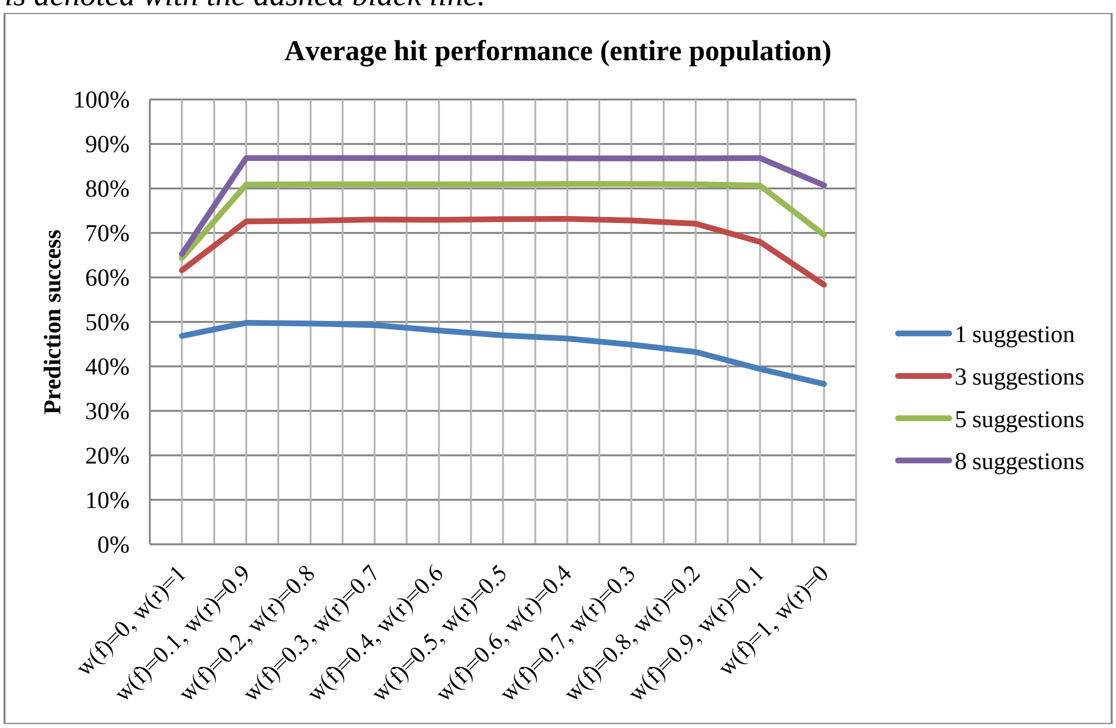

Figure 5. Average hit performance of the entire population, broken down by suggestion list size.

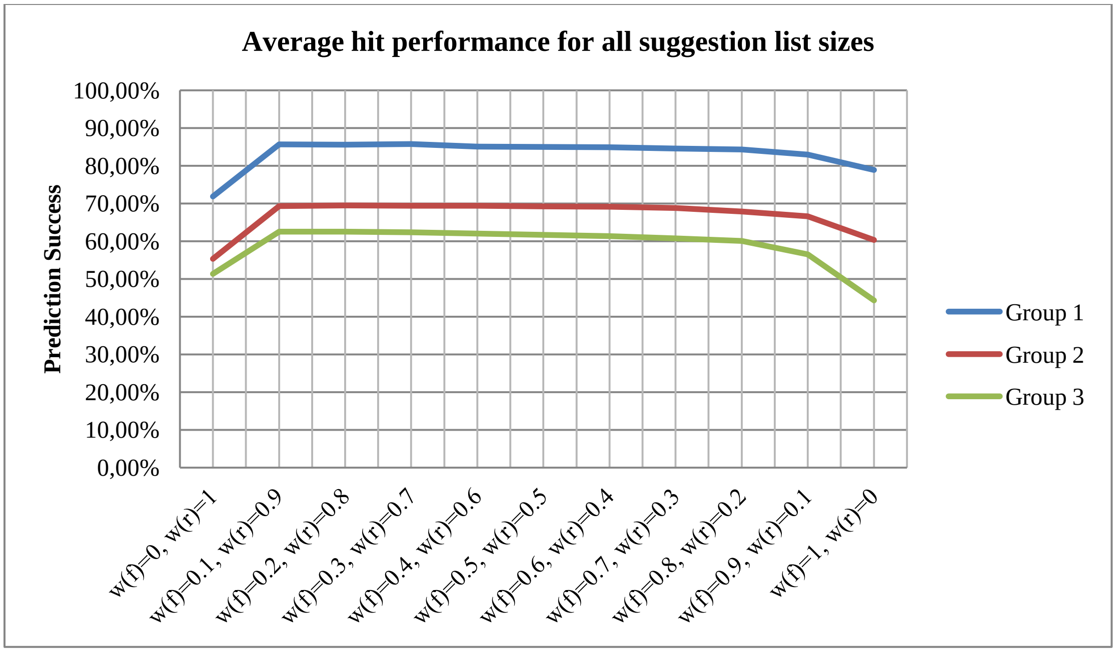

Figure 6. Average hit performance for all list sizes, broken down by group.

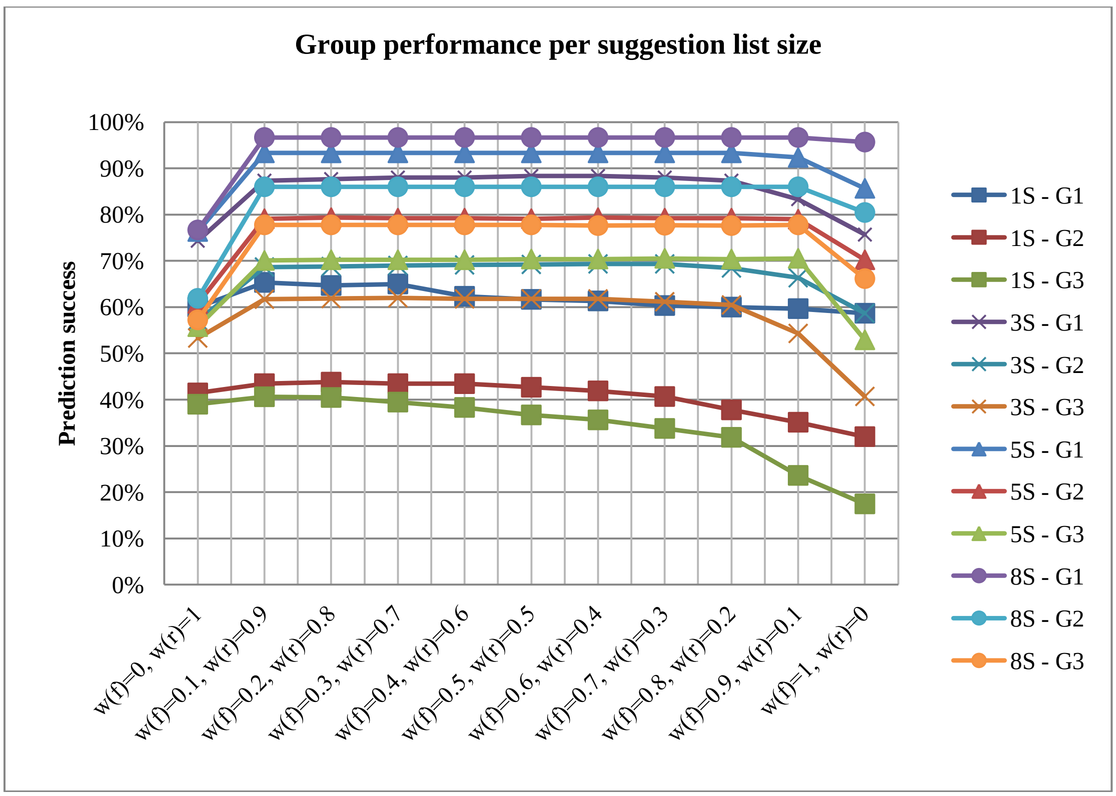

Figure 7. Breakdown of each group's performance for all suggestion list sizes.

Experiment set B – Scored performances

Our first experiment set showed that for list sizes greater than 3, the weighting balance of the frequency and recency dimensions is practically immaterial as the performance remains more or less constant. To investigate the quality of the suggestions (i.e. how close was the contact that was actually called to the top of the suggestion list, thus likely to be seen sooner by the user), we performed the second set of experiments as previously described, for all suggestion list sizes apart from the one (as this is equivalent to the one suggestion hit-or-miss experiment reported earlier). In this case, we notice that generally, the algorithm offers good placement of the actual correct predictions within the suggestion list, which is on average, quite close to the top in each case (Figures 8, 9 and 10). Again, we note that the recency dimension seems to offer better performance when weighed favourably over the frequency dimension. We note also that users of Group 1 enjoy the best performance, which is followed by the performance experienced by Group 2 and Group 3, again confirming our earlier hypothesis.

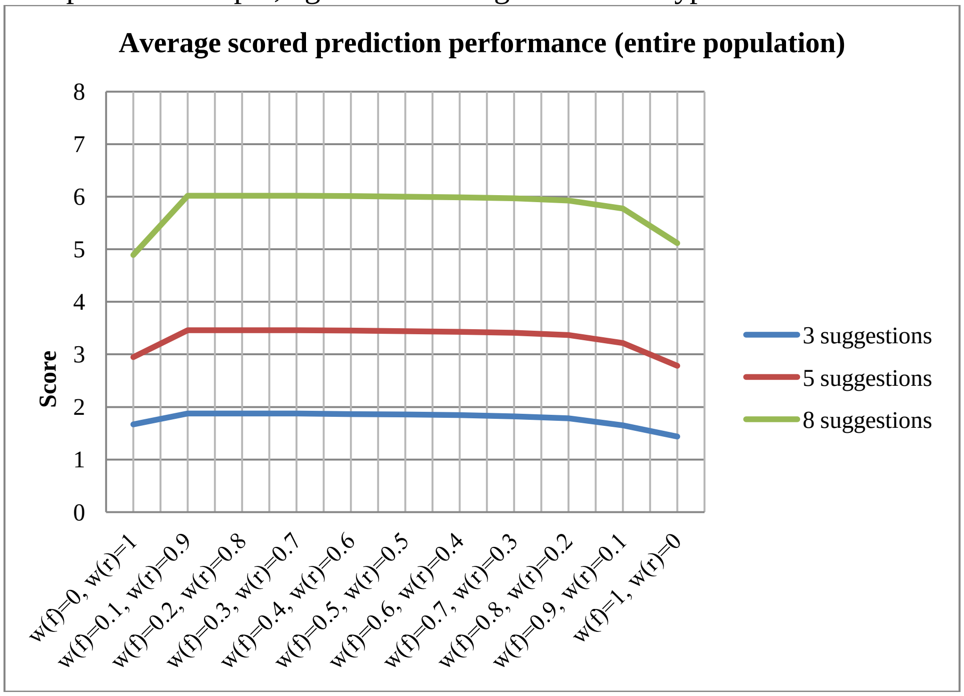

Figure 8. Average scored prediction performance for the entire population.

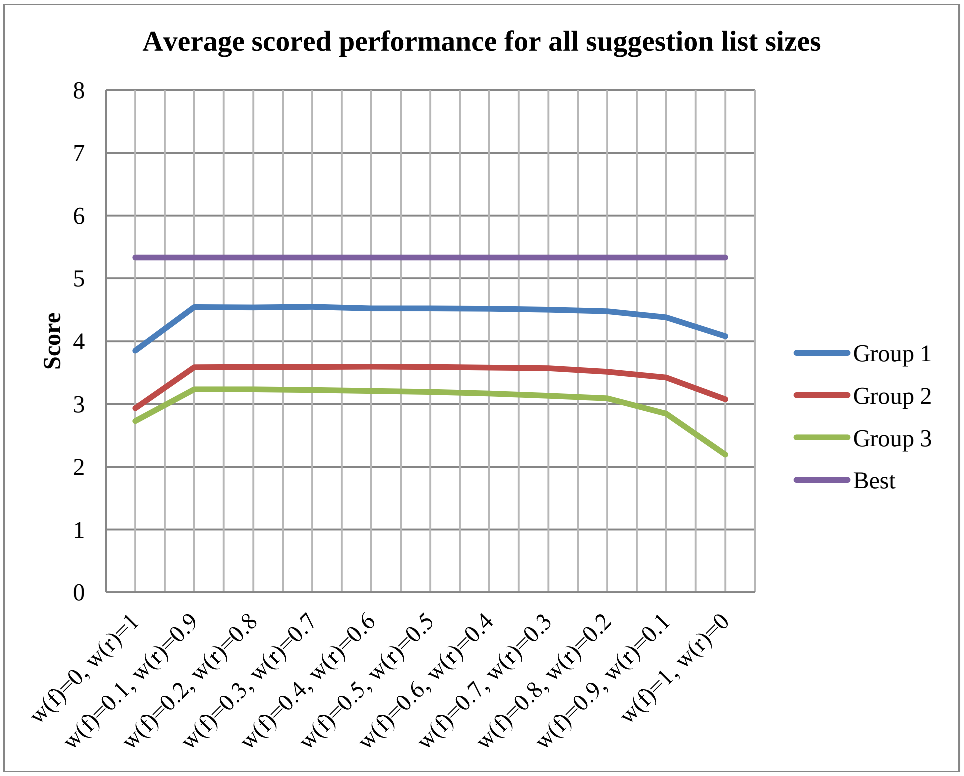

Figure 9. Average scored performance for all suggestion list sizes (the purple line shows the theoretic optimal average score of 5.33).

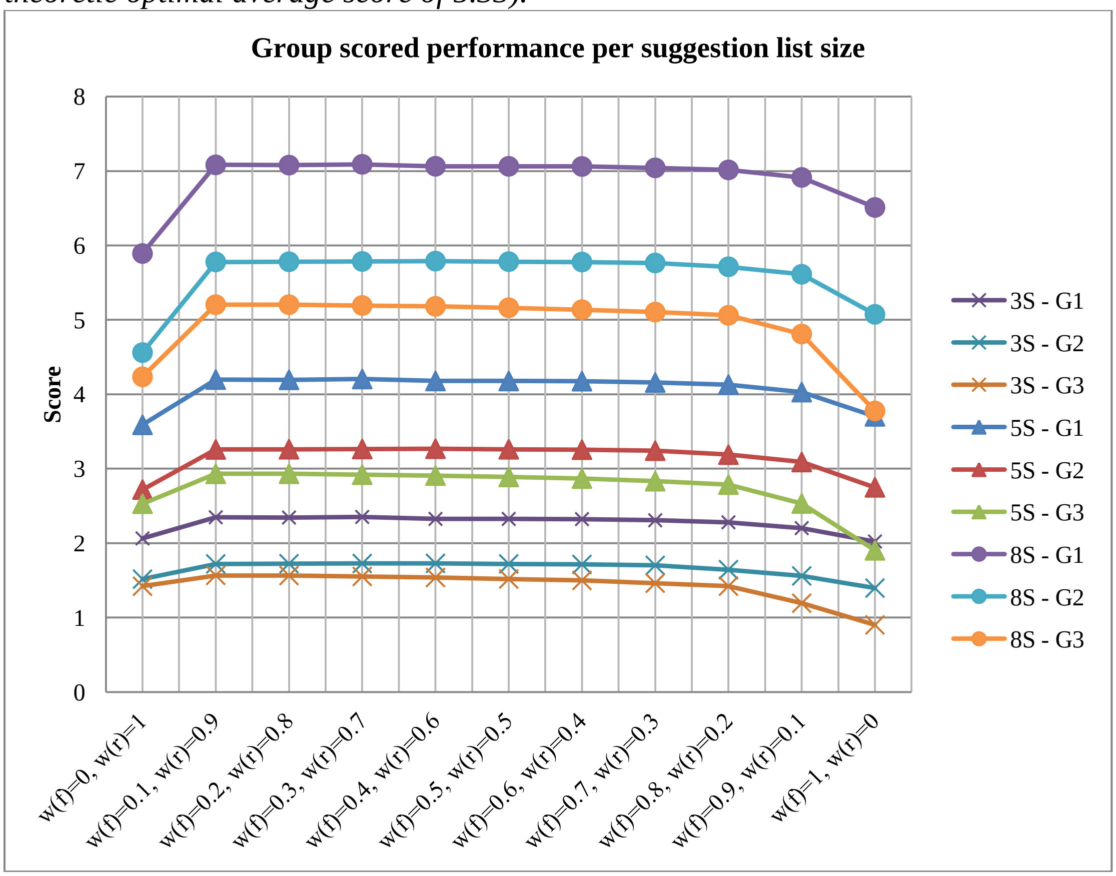

Figure 10. Breakdown of each group's scored performance for all suggestion list sizes.

VALIDATION OF RESULTS

It is apparent from figures 2-10 that in most cases, there are measurable performance improvements through the combination of recency and frequency in our approach. We do not observe a measurable difference for these combinations, apart from the case of Group 1 and Group 3 where performance seems to peak when weighting is biased in favour of the recency metric. We compared the performance of the system when considering just frequency (w_f=1) or just recency (w_r=1) against the agnostic situation where we assume no definite knowledge of the user's "socialness" and thus set the criteria weighting without bias (w_f=w_r=0.5). An analysis of variance (ANOVA) shows that for most cases under the three conditions (w_r=1, w_r=0.5 and w_r=0), the null hypothesis (*H0*: $\mu_{wr=0}=\mu_{wr=0.5}=\mu_{wr=1}$) is rejected. The only situations where the null hypothesis is not clearly rejected is for Group 1, however this is a group with just 3 users and thus the data reported for it cannot be statistically significant.

Table 3 below shows the results of the ANOVA.

Table 3. ANOVA results (for list size =1, scored performance is the same as hit performance)

		Hit Performance		Scored Performance	
Group	**List size**	**F**	***p***	**F**	***p***
1	1	0.08892	>0.05	-	-
1	3	3.46667	>0.05	1.78063	>0.05
1	5	3.90619	>0.05	2.55229	>0.05
1	8	7.14375	<0.05	3.61975	>0.05
2	1	4.27903	<0.05	-	-
2	3	9.33996	<0.01	7.13402	<0.01
2	5	18.71707	<0.01	8.04515	<0.01
2	8	34.19153	<0.01	16.73811	<0.01
3	1	37.18531	<0.01	-	-
3	3	23.82456	<0.01	32.42174	<0.01
3	5	18.28601	<0.01	25.97022	<0.01
3	8	28.22977	<0.01	18.40532	<0.01

DISCUSSION OF RESULTS

In the previous section we presented in detail the results of our experiments. At first, a significant finding is that by combining the dimensions of frequency and recency we achieve better prediction results than by considering each dimension separately, which is in line with the findings of Barzaiq and Loke (2011). This is particularly promising since it supports our belief that adding more dimensions to the context vector could possibly provide an even better prediction success rate, allowing thus for new contact retrieval methods and interfaces, more effective than the two list types that smartphones currently support (recent calls, more frequently called contacts). Furthermore, as it was expected and in line with Phithakkitnukoon et al. (2011), the larger the suggestion list, the higher the prediction success rate is. However, the intensity of this positive effect decreases as the suggestion list grows larger and having in mind that small screens of mobile devices usually fit 8 lines of information, it seems that there is no point in providing more suggestions. Another interesting observation is that as the size of the suggestion list increases, the role of the weights becomes less important.

Moreover, the observation of Lee et al. (2010), that the existence of groups of users with different "social" communication behaviour influences the prediction performance, was also confirmed from our experiments (though we find three distinct groups instead of the two mentioned in that study). The variance of the results due to the different communication pattern of each group is a concrete indication that weights shouldn't be static, but dynamic for each user. In addition to this, we believe that the weights should be dynamic even for the same user under different contexts. For example location could play a more important role than time of day in determining the next contact likely to be called when the user is travelling to another city.

Our results in predicting the next contact to be called can be contrasted against the findings of Lee et al (2010), Barzaiq & Loke (2011) and Phithakkitnukook et al. (2011), particularly for predicting using 5 suggestions. Lee et al. (2010) achieve performances greater than 75% for just one type of user (easily predictable ones) while the other 2 groups that emerge in their study do not exceed 50% and 30% average success respectively. Barzaiq & Loke (2010) achieve a 40% success average for 5 suggestions after 5 weeks worth of training and adapting their system. Finally, Phithakkitnukook et al. (2011) achieve a 70% average for their 5 suggestions performance. The algorithms used in all cases are much more complex in nature than our own technique, which achieves an average success for the entire population of approximately 80% (Figure 5), while the performance even for Group 3 who are the most "social" and thus unpredictable users, hovers around 70% (Figure 11). Finally, the experiment set B shows that the algorithm offers good ranking for the predictions in within each suggestion list.

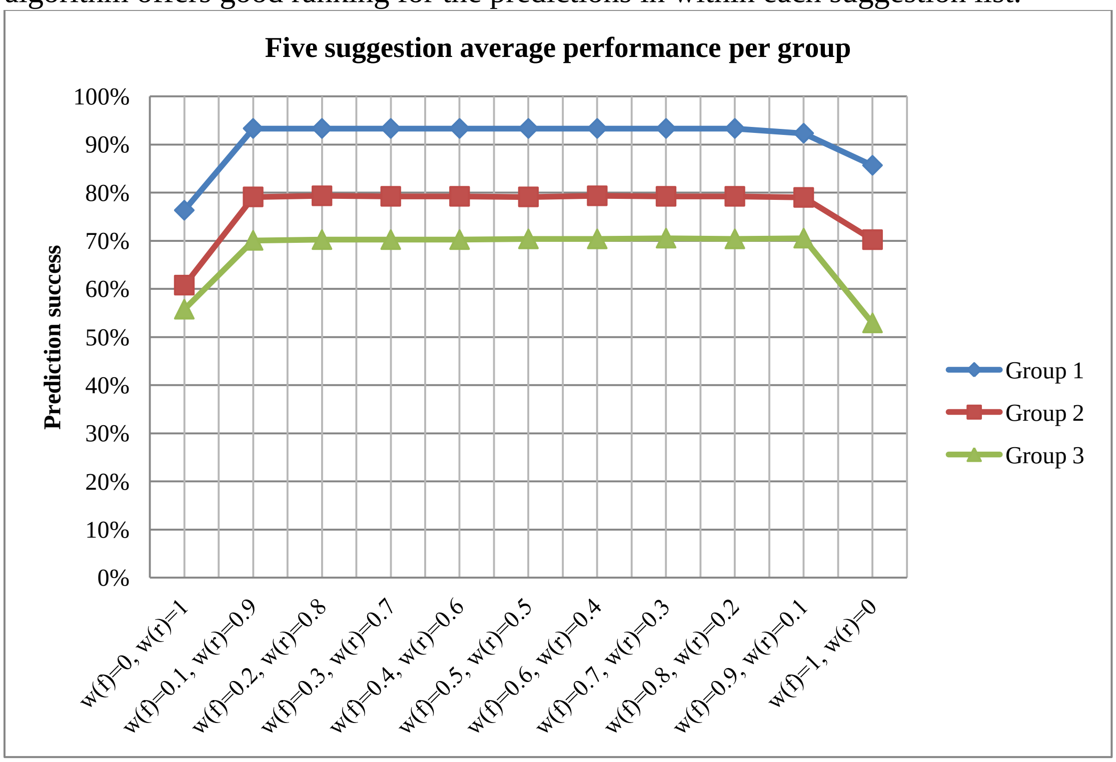

Figure 11. Breakdown of each group's performance for a suggestion list size of five.

EMERGENT ISSUES IN SUPPORTING COMMUNICATIONS THROUGH CONTEXT-AWARE UIS

Our results are a first step towards understanding user mobile communication behaviour and the retrieval tasks associated with this activity. This understanding is essential in order to design adaptive user interfaces that can have a predictable and desirable effect on observed behaviour. While aware of the convenience and utility of an intelligent retrieval aid such as an adaptive contact list, we must however consider the possibility that limiting users' options might lead to a channeling of behaviour (e.g. users gradually reducing communication with contacts other than those suggested) due to altered perceptions of cost/benefit values assigned to the retrieval task. To our knowledge there is no study about the effect of the availability of calling lists or frequently used number lists to the actual calling behaviour of users. Does this information affect the behaviour of users, i.e. does it exert influence on the maintenance of strong social links with other users by making access to calling them easier? Would a system that always predicts the right person to call next, prevent users from making contact with other, less important users?

Peters and Allouch (2005) found that initial gratifications of mobile use such as social interaction appear to decrease over time, giving way to more task-oriented use. Perhaps the introduction of "false positives", or reminders for contacts that used to be important but have not been contacted for some time, could encourage users who are not particularly "social", to communicate more often with a wider variety of contacts and maintain a level of social connectedness. And then, is the concept of "calling" the optimal means of contacting someone? Would an interface that suggested not only the person but also the mode of contact to something "more appropriate" than just text or talk (e.g. facebook message) be desirable, or help increase participation in social networks rather than one-to-one communication? In this sense, the discovery of distinct groups of users in terms of their communication behaviour is fortunate, as, for example, the most "social" group's behaviour could be used as a baseline, and further research could be undertaken on how close a system can bring to this behaviour users from other groups. We thus see our work of understanding and predicting communication patterns as an essential first step into designing persuasive user interfaces.

CONCLUSIONS AND FURTHER WORK

In this paper a context based approach for predicting the most probable contact to be called is presented. Although we start with the contact list and the task of facilitating contacts retrieval from it as a problem domain, we believe that this approach could extend and apply to other personal mobile information retrieval problems that involve context as well. Our main contributions are two-fold. First, we demonstrate that context-aware personal information management applications can perform satisfactorily without a need for significant complexity in their context discovery algorithms. This performance however cannot be achieved through universal and static weighting of context vector elements, but context cue importance must be weighted in light of the individual user's profile and also according to other available context cues. Our second contribution is the guidelines for user interface design that derive from our investigations. We show that it is possible to achieve very high performance by presenting retrieval options that fits the confines of a single mobile screen real estate. In the domain of personal information retrieval, up to 8 options for the user are enough to achieve close to optimal performance.

Our work is of course by no means complete. As a first step we intend to introduce more contextual dimensions to our algorithm, since in this work we focus on behavioural factors (such as frequency and recency) of communication. Such dimensions could include personal preference (e.g. the feature of denoting a contact as "starred" - favorite in Android phones), location, time of day, weekdays/weekends or day of week, user activity (e.g. tasks stored in a calendar application) etc. In this case we should also solve the uneven dimensionality problem of context vectors, in case the value of a context dimension is not known under a context situation.

Moreover, it is in our intentions to create an application for mobile platforms in order to test the algorithm and different mobile user interfaces as a replacement to traditional contact list access methods under real-life conditions. To test the efficiency of our approach we plan to conduct a longitudinal field trial once a prototype app is complete.

To conclude, the previously discussed observation about the dynamic nature of dimension weights for different users and different contexts is an indication that a more generic approach that would not involve manual adjustment of weights is needed. In a previous work (Komninos et al., 2011a) we proposed the application of a dimensionality reduction technique to context augmented personal information items, such as entries in a contact list, in order to extract a small number of features that could accurately represent the original items and their relationships. Our future experiments include the application of this technique to the available datasets for the problem of predicting the next contact to be called. We hope that this work could provide us with valuable insight and understanding of mobile users' behaviour, allowing us next to proceed with the design and experimentation of novel persuasive mobile user interfaces.

REFERENCES

Barzaiq, O. O., & Loke, S. W. (2011). Adapting the mobile phone for task efficiency: the case of predicting outgoing calls using frequency and regularity of historical calls. *Personal and Ubiquitous Computing*, *15*(8), 857-870.

Bergman, O., Komninos, A., Liarokapis, D., & Clarke, J. (2012). You never call: Demoting unused contacts on mobile phones using DMTR. *Personal and Ubiquitous Computing*, *16*(6), 757-766.

Calabrese, F., Smoreda, Z., Blondel, V. D., & Ratti, C. (2011). Interplay between telecommunications and face-to-face interactions: A study using mobile phone data. *PloS one*, *6*(7), e20814.

Candia, J., González, M. C., Wang, P., Schoenharl, T., Madey, G., & Barabási, A. L. (2008). Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, *41*(22), 224015.

Cheng, L. C., & Sun, L. M. (2012). Exploring consumer adoption of new services by analyzing the behavior of 3G subscribers: An empirical case study. *Electronic Commerce Research and Applications*, *11*(2), 89-100.

Gilbert, P., Cuervo, E., & Cox, L. P. (2009, February). Experimenting in mobile social contexts using jellynets. In *Proceedings of the 10th workshop on Mobile Computing Systems and Applications* (p. 16). ACM.

Jeon, M. H., Na, D. Y., Ahn, J. H., & Hong, J. Y. (2008, September). User segmentation & UI optimization through mobile phone log analysis. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services* (pp. 495-496). ACM.

Komninos, A., Plessas, A., Stefanis, V., & Garofalakis, J. (2011, September). Application of dimensionality reduction techniques for mobile social context. In *Proceedings of the 13th international conference on Ubiquitous computing* (pp. 583-584). ACM.

Komninos, A., Plessas, A., Stefanis, V. & Garofalakis, J. (2011). Context Dimensionality Reduction for Mobile Personal Information Access. In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (KDIR 2011)* (pp. 493-498). SciTePress.

LaRue, E. M., Mitchell, A. M., Terhorst, L., & Karimi, H. A. (2010). Assessing mobile phone communication utility preferences in a social support network. *Telematics and Informatics*, *27*(4), 363-369.

Lee, S., Seo, J., & Lee, G. (2010, April). An adaptive speed-call list algorithm and its evaluation with ESM. In *Proceedings of the 28th international conference on Human factors in computing systems* (pp. 2019-2022). ACM.

Palen, L., Salzman, M., & Youngs, E. (2001). Discovery and integration of mobile communications in everyday life. *Personal and ubiquitous computing*, *5*(2), 109-122.

Peters, O., & Allouch, S. B. (2005). Always connected: a longitudinal field study of mobile communication. *Telematics and Informatics*, *22*(3), 239-256.

Phithakkitnukoon, S., & Dantu, R. (2008). Adequacy of data for characterizing caller behavior. In *Proceedings of KDD Inter. Workshop on Social Network Mining and Analysis (SNAKDD 2008)*. ACM.

Phithakkitnukoon, S. & Dantu, R. (2010). Mobile social closeness and similarity in calling patterns. In *Proceedings of the 7th IEEE Conference on Consumer Communications & Networking Conference (CCNC 2010)* (pp. 1-5).

Phithakkitnukoon, S., Dantu, R., Claxton, R., & Eagle, N. (2011). Behavior-based adaptive call predictor. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, *6*(3), 21.

Salovaara, A., Lindqvist, A., Hasu, T., & Häkkilä, J. (2011, August). The phone rings but the user doesn't answer: unavailability in mobile communication. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 503-512). ACM.

Stefanis, V., Plessas, A., Komninos, A., & Garofalakis, J. (2012, September). Patterns of usage and context in interaction with communication support applications in mobile devices. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services* (pp. 25-34). ACM.

Subramanya, S.R. (2012). Analysis of Mobile Phone Call Data to Determine User Characteristics and to Enhance User Experience. In *International Journal of Computer Applications (IJCA)*, n. 9, 9-13.

Toninelli, A., Khushraj, D., Lassila, O., & Montanari, R. (2008, October). Towards socially aware mobile phones. In *First Workshop on Social Data on the Web (SDoW). Citeseer*.