

Predicting Calls – New Service for an Intelligent Phone

Santi Phithakkitnukoon and Ram Dantu

Network Security Laboratory, Department of Computer Science and Engineering,
University of North Texas, Denton, TX 76203, USA
{santi, rdantu}@unt.edu

Abstract. Predicting future calls can be the next advanced feature of the intelligent phone as the phone service providers are looking to offer new services to their customers. Call prediction can be useful to many applications such as planning daily schedule and attending unwanted communications (e.g. voice spam). Predicting calls is a very challenging task. We believe that this is a new area of research. In this paper, we propose a Call Predictor (CP) that computes the probability of receiving calls and makes call prediction based on caller's behavior and reciprocity. The proposed call predictor is tested with the actual call logs. The experimental results show that the call predictor performs reasonably well with false positive rate of 2.4416%, false negative rate of 2.9191%, and error rate of 5.3606%.

Keywords: Caller, Callee, Communications, Incoming calls, Outgoing calls, Arrival time, Inter-arrival time, Inter-arrival/departure time, Reciprocity, Behavior, Kernel density estimation, Probability density function (pdf), Call matrix, Receiving call probability.

1 Introduction

Prediction plays an important role in various applications. The prediction is widely applied in the areas such as weather, environmental, economic, stock, disaster (earthquake, flooding), network traffic, and call center forecasting [1, 2, 3, 4]. Companies use predictions of demands for making investment and efficient resource allocation. The call centers predict workload so that they can get the right number of staff in place to handle it. Network traffic prediction is used to assess future network capacity requirement and to plan network development so as to better use of network resources and to provide better quality of services. Prediction is also applied in the human behavior study by combining the computer technology and social networks [5, 6, 7, 8]. There is also some work reported on telephone telepathy based on psychology [9].

Predicting the expected calls for a busy business executive can be very useful for scheduling a day. Match making services can use calling patterns for the compatibility studies [10]. Moreover, the prediction of incoming calls can be used to avoid unwanted calls and schedule a time for wanted calls. For example, the problem of spam in VoIP networks has to be solved in real time compared to email systems. Compare receiving an email spam at 2:00 AM that sits in the inbox until you open it the next morning to receiving a junk phone call that must be answered immediately.

Over the past few years, there has been a rapid development and deployment of new advanced phone features, including internet access, e-mail access, scheduling software, built-in camera, contact management, accelerometers, and navigation software as well as the ability to read documents in variety of format such as PDF and Microsoft Office. In 2005, Google filed a patent including detail about the Google Phone (GPhone) that could predict what a user is searching for or the words they are typing in a text messages by taking into account the user's location, previous searching/messaging history, and time of the day. However, none of these features offers ability to predict future calls.

Let us consider a simple caller-callee scenario shown in Fig. 1. In order to have an efficient scheduling of transactions, the caller wants to know the willingness of taking calls of the callee, which could be determined by the callee's presence. At the same time, the callee wants to know (predict) the incoming calling pattern (calling schedule) of the caller. *This raises two interesting problems; (1) predicting the incoming calling pattern of the caller, and (2) determining the presence of the callee. In this paper, we attempt to solve the first problem. The second problem and its solution will be addressed in our future work.*



Fig. 1. A simple caller-callee scenario

To the best of our knowledge, no scientific research has been reported in predicting the incoming calls for phone services. Predicting of incoming calls using just the call history is a challenging task. We believe that this is a new area of research. One way of predicting incoming calls from specified callers is to compute the probability of receiving calls associated with them. In this paper, we present a model for predicting the next-day calls based on caller and user's past history. Section 2 presents the architecture of a Call Predictor (CP) that computes the probability of receiving calls from a specified caller and makes the next-day call prediction. Section 3 presents the CP framework which describes our real-life data sets and carries out the receiving call probability computation. The performance of the proposed CP is then measured and discussed in Section 4.

2 Call Predictor

The Call Predictor (CP) for computing the probability of receiving calls from a specified caller and making next-day call prediction can be deployed either in conjunction with perimeter controllers such as voice spam filters or firewalls, or in end systems such as multimedia phones. The basic architecture of the CP is shown Fig. 2.

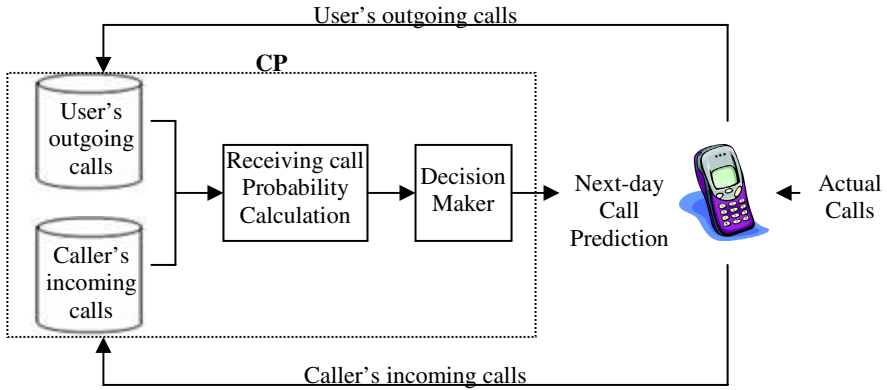


Fig. 2. Architecture of Call Predictor (CP). The CP calculates the probability of receiving next-day calls from specified callers based on the past call history (incoming and outgoing calls) and makes next-day call prediction. The call database is updated with the actual call activities.

For any time that phone user requests for a call prediction of a particular caller, the CP computes the probability of receiving calls of the next 24 hours based on the caller's past history (*Caller's incoming calls*) and the previous outgoing calls from the phone user to the caller (*User's outgoing calls*). Both of these histories are maintained by the CP by logging the call specific information for every call received and made by the user. The computed receiving call probability is checked with a preconfigured threshold value to make a decision as to predict "call" or "no call" for each of the next 24 hours.

3 Call Prediction Framework

To predict the future incoming calls, the behavior learning models must be used. These models should incorporate mechanisms for capturing the caller's behavior (based on call arrival time and inter-arrival time), the user's behavior (based on call departure time), reciprocity (based on call inter-arrival/departure time), the probability model of receiving calls from caller, and finally, the next-day call prediction.

3.1 Real-Life Data Sets

Every day calls on the cellular network include calls from different sections of our social life. We receive calls from family members, friends, supervisors, neighbors, and strangers. Every person exhibits a unique calling pattern. These calling patterns can be analyzed for predicting the future calls to the callee.

To study calling pattern, we collected the actual call logs of 20 individuals at our university. These 20 individuals are faculties, staffs, and students. We are in process of collecting many more call logs. The details of the data collecting process are given

in [11]. We found it difficult to collect the data set because many people are unwilling to give their call logs due to privacy issues. Nevertheless, the collected datasets include people with different types of calling patterns and call distributions.

As part of the data collecting process, each individual downloaded three months of detail telephone call records from his/her online accounts on the cellular service provider's website. Each call record in the dataset had the 5-tuple information as shown below:

Call record: (date, start time, type, caller id, talk time) where

date - date of communication

start time - the start time of the communication

type - type of call i.e., "Incoming" or "Outgoing"

caller id - the caller identifier

talk time - amount of time spend by caller and the individual during the call

We then used the collected data for deriving the traffic profiles for each caller who called the individuals. To derive the profile, we inferred the arrival time (time of receiving a call), inter-arrival time (elapsed time between adjacent incoming calls), and inter-arrival/departure time (elapsed time between adjacent incoming and going calls).

3.2 Probability Computation

In our daily life, when we receive a phone call, at the moment of the first phone ring before we look at the caller ID, we often guess who the caller might be. We base this estimation on:

- *Caller's behavior*: Each caller tends to have a unique calling pattern. These patterns can be observed through history of calling time (we normally expect a call from a caller who has history of making several calls at some particular time, for example, your spouse likes to call you while you drive to work in the morning and after work in the evening therefore when your phone rings while you are on the way to work or back home, you likely to guess that it is a phone call from your spouse), periodicity of call history (we can expect that a caller who calls periodically will repeat the same pattern, for example, your friend calls you at about 2:00 PM every Tuesday therefore you expect a call from him/her at about 2:00 PM for next Tuesday).
- *Reciprocity*: The communication activity patterns between the caller and the user in the past. These patterns can be observed in terms of number of user's outgoing calls per caller's incoming call and call inter-arrival/departure time.

Therefore, we believe that receiving a call is influenced by caller's past incoming calls and call interaction history between caller and phone user. The patterns of caller's incoming calls can be observed from call arrival time and inter-arrival time. The patterns of call interaction between caller and phone user can be observed from the number of outgoing calls per incoming call and the inter-arrival/departure time.

The calling pattern based on caller’s call arrival time can be captured by using nonparametric density estimation. The most popular method for density estimation is the kernel density estimation (also known as the Parzen window estimator [12]) which is given by Eq. (1).

$$a(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right). \tag{1}$$

$K(u)$ is kernel function and h is the bandwidth or smoothing parameter. The most widely used kernel is the Gaussian of zero mean and unit variance which is defined by Eq. (2).

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}. \tag{2}$$

The choice of the function and h is crucial. Several optimal bandwidth selection techniques have been proposed [13, 14]. In this paper, we use AMISE optimal bandwidth selection using the Sheather Jones Solve-the-equation plug-in method [15]. Fig. 3(a) shows an example histogram of call arrival time. It should be noted that the widow of observation is shifted to start at 5:00 AM and end at 4:59 AM in order to capture the entire calling pattern in the middle. The corresponding estimated probability density function (pdf) using kernel density estimation is shown in Fig. 3(b).

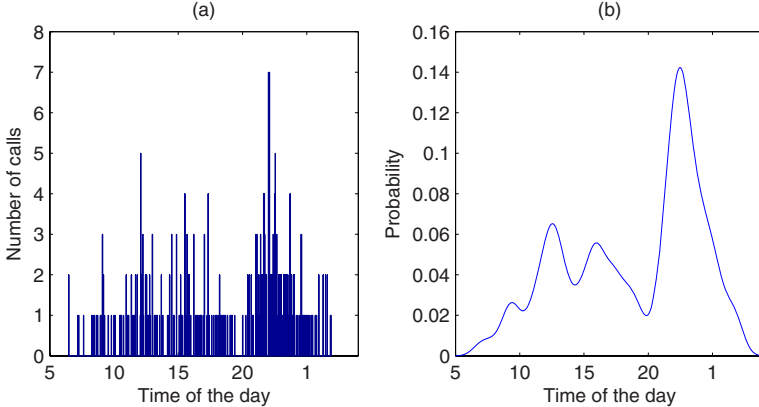


Fig. 3. (a) An example histogram of call arrival time. (b) The estimated probability density function using kernel density estimation of the example histogram of call arrival time shown in Fig. 3(a). Note that observation window is 5:00 AM to 4:59 AM.

We define a *Call Matrix* as a matrix whose entries are call indicators where rows are hours of the day and columns are days of observation. The call indicator (CI) indicates if there is at least one incoming call or going call or both incoming and outgoing call or no call. CI 's values and its indications are given in Eq. (3) and an example Call Matrix of 15 days of observation is shown in Fig. 4.

$$CI = \begin{cases} 0, \text{ no call} \\ 1, \text{ at least one incoming call} \\ 2, \text{ at least one outgoing call} \\ 3, \text{ at least one incoming call and one outgoing call.} \end{cases} \quad (3)$$

Hour of the day	24	0	0	1	0	2	1	0	3	0	0	1	1	0	3	0	<input type="checkbox"/>
	23	1	2	2	0	0	2	2	0	1	3	0	0	2	1	2	<input type="checkbox"/>
	22	2	0	1	2	0	0	1	2	0	0	1	2	0	0	1	<input type="checkbox"/>
	21	0	0	1	0	0	0	2	1	0	0	0	0	0	1	0	<input type="checkbox"/>
	20	0	0	0	2	0	0	2	1	0	0	2	1	0	0	2	<input type="checkbox"/>
	19	1	0	3	2	1	0	3	1	2	0	2	1	3	2	1	<input type="checkbox"/>
	18	0	2	0	2	1	3	2	0	1	1	1	0	0	2	2	<input type="checkbox"/>
	17	2	3	0	0	0	2	3	1	0	0	2	1	2	0	0	<input type="checkbox"/>
	16	0	1	1	2	0	0	0	3	2	2	0	0	1	1	2	<input type="checkbox"/>
	15	2	3	2	0	1	0	0	0	0	2	2	3	1	1	1	<input type="checkbox"/>
	14	0	0	0	2	0	0	0	2	0	0	2	2	0	1	1	<input type="checkbox"/>
	13	0	0	2	3	0	0	1	1	0	0	1	1	0	0	2	<input type="checkbox"/>
	12	1	1	0	1	1	0	3	3	0	0	2	1	1	1	1	<input type="checkbox"/>
	11	2	3	0	0	0	2	2	0	1	0	0	1	2	0	1	<input type="checkbox"/>
	10	0	1	0	2	0	0	0	0	2	0	0	0	2	0	0	<input type="checkbox"/>
9	0	1	2	2	1	0	0	2	3	0	0	2	0	1	1	<input type="checkbox"/>	
8	0	0	0	1	2	0	2	0	0	0	1	1	0	0	2	<input type="checkbox"/>	
7	0	0	2	0	0	2	0	0	0	0	0	2	1	0	0	<input type="checkbox"/>	
6	0	0	0	0	0	0	0	2	0	0	0	0	2	0	0	<input type="checkbox"/>	
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	<input type="checkbox"/>	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	<input type="checkbox"/>	
3	0	0	2	0	0	0	0	0	3	0	0	1	0	0	2	<input type="checkbox"/>	
2	1	0	0	1	0	0	2	0	2	1	0	0	0	2	1	<input type="checkbox"/>	
1	0	1	0	0	1	2	1	0	0	1	2	2	1	1	1	<input type="checkbox"/>	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	↑
		Day of observation														Predicting	

Fig. 4. An example Call Matrix of 15 days of observation

The caller’s behavior can also be observed through the call inter-arrival time. However, the inter-arrival time in our normal sense is the elapsed time between temporally adjacent calls, which we believe that it does not accurately represent the caller’s behavior based on inter-arrival time. Due to the human nature that requires state of natural rest, sleeping time causes the inaccuracy in the average inter-arrival time. In fact, it increases the average inter-arrival time from the true value. Therefore,

we believe that the more accurate angle to observe calling pattern based on inter-arrival time is to scan over each hour of the day through days of observation, i.e. capturing inter-arrival time patterns by observing each row of the Call Matrix.

Let a random variable X_k be inter-arrival time of k^{th} hour where $k = 1, 2, 3, \dots, 24$. A Normal distribution $N(\mu_k, \sigma_k^2)$ is assumed for the call inter-arrival time since no information is available that $Pr(X_k \leq \mu_k - c) < Pr(X_k \leq \mu_k + c)$ or vice versa therefore it can be safely assumed that $Pr(X_k \leq \mu_k - c) = Pr(X_k \leq \mu_k + c)$ where μ_k is the mean and σ_k^2 is the variance of inter-arrival time of k^{th} hour which can be calculated by Eq. (4) and Eq. (5).

$$\mu_k = \frac{1}{N-1} \sum_{n=1}^{N-1} x_k(n). \quad (4)$$

$$\sigma_k^2 = \frac{1}{N-1} \sum_{n=1}^{N-1} (x_k(n) - \mu_k)^2. \quad (5)$$

N is the total number of calls and $x_k(n)$ is the n^{th} inter-arrival time. The inter-arrival time is now treated as a random variable X_k that consists of number of small random variables $\{x_k(1), x_k(2), x_k(3), \dots, x_k(N-1)\}$, is normal random variable which has probability density function (pdf) given by Eq. (6).

$$i_k(x_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-(x_k - \mu_k)^2 / 2\sigma_k^2}. \quad (6)$$

For example, if a caller calls on average every 3 days, the chances of receiving a call one day earlier (day 2) or day one later (day 4) are the same.

As previously mentioned that receiving a call is influenced by not just caller's behavior but also reciprocity, one way to observe the calling patterns based on reciprocity is to monitor the number of outgoing calls per incoming call. This can give us a good approximation of when the next incoming call can be expected. A normal distribution $N(\mu_k, \sigma_k^2)$ is also assumed for the same reason as in the inter-arrival time case, where Y_k is a random variable representing the number of outgoing calls per incoming call of the k^{th} hour where μ_k is the mean and σ_k^2 is the variance of which can be calculated by Eq. (7) and Eq. (8).

$$\mu_k = \frac{1}{M} \sum_{n=1}^{M-1} y_k(n). \quad (7)$$

$$\sigma_k^2 = \frac{1}{M} \sum_{n=1}^{M-1} (y_k(n) - \mu_k)^2. \quad (8)$$

M is the total number of incoming calls of k^{th} hour and $y_k(n)$ is the number of outgoing calls between the n^{th} and $(n+1)^{\text{th}}$ incoming call. Therefore, the pdf is given by Eq. (9).

$$n_k(y_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-(y_k - \mu_k)^2 / 2\sigma_k^2}. \quad (9)$$

An example of calculating $n_k(y_k)$ is shown in Fig. 5.

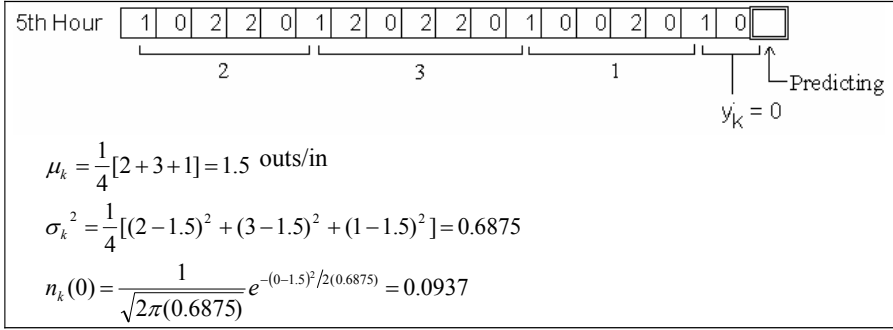


Fig. 5. An example of calculating $n_k(y_k)$ for one hour slot (5th hour) of 18 days of observation

Another angle to observe the calling patterns based on reciprocity is to monitor the inter-arrival/departure time. This gives us the chance (probability) of receiving a call from the caller given the time of the last outgoing call to the caller.

Let Z_k be a random variable mapping to the inter-arrival/departure time of the k^{th} hour. A normal distribution $N(\mu_k, \sigma_k^2)$ is also assumed for the the same reason previously mentioned. The mean (μ_k) and variance (σ_k^2) are given by Eq. (10) and Eq. (11).

$$\mu_k = \frac{1}{L-1} \sum_{n=1}^{L-1} z_k(n). \quad (10)$$

$$\sigma_k^2 = \frac{1}{L-1} \sum_{n=1}^{L-1} (z_k(n) - \mu_k)^2. \quad (11)$$

L is total number of incoming calls of k^{th} hour and $z_k(n)$ is the average inter-arrival/departure time of the n^{th} incoming call to all right-hand-side outgoing calls (in the Call Matrix's row) before reaching the $(n+1)^{th}$ incoming call (an example is illustrated in Fig. 6). The pdf of inter-arrival/departure time is given in Eq. (12).

$$t_k(z_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-(z_k - \mu_k)^2 / 2\sigma_k^2}. \quad (12)$$

An example of calculating $t_k(z_k)$ is shown in Fig. 6.

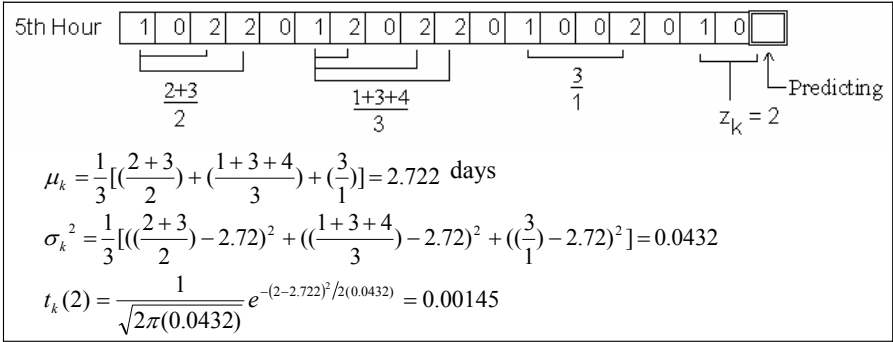


Fig. 6. An example of calculating $t_k(z_k)$ for one hour slot (5th hour) of 18 days of observation

From Eq.(1), (6), (9), and (12), we can infer the probability of receiving a call from “*Caller A*” of k^{th} hour ($P_A(k)$) as the average of the probability of receiving a call based on the caller’s behavior (arrival time and inter-arrival time) and the reciprocity (number of outgoing calls per incoming call and inter-arrival/departure time), which is given by Eq. (13) where $k = 1, 2, 3, \dots, 24$.

$$P_A(k) = \frac{1}{4} \left[\sum_{j=60(k-1)+1}^{60k} a_k(j) + i_k(x_k) + n_k(y_k) + t_k(z_k) \right]. \tag{13}$$

There is another group of callers who never receive any calls back from the user, i.e. no reciprocity. More likely these callers are telemarketers or voice spammers. Since there is no history of call interaction between the callers and the user, the Eq. (13) reduces to the averaging over the probability based on only the caller’s behavior, which is given by Eq. (14). Likewise, for the regular callers where some hour slots (rows of Call Matrix) have no reciprocity, Eq. (13) also reduces to Eq. (14).

$$P_A(k) = \frac{1}{2} \left[\sum_{j=60(k-1)+1}^{60k} a_k(j) + i_k(x_k) \right]. \tag{14}$$

To present the accuracy of the receiving call probability model, a phone user is randomly selected to represent all the individuals. Fig. 7 shows 30 consecutive days of receiving call probability calculation for an arbitrary caller where the receiving call probability is represented with a green surface and the actual calls during these 30 days of observation are represented with vertical black pulses.

It can be observed from Fig. 7 that most of the calls are received when the computed receiving-call probability is high. At the same time, no call is received during 0:00 AM to 9:00 AM period where the probability of receiving call is low.

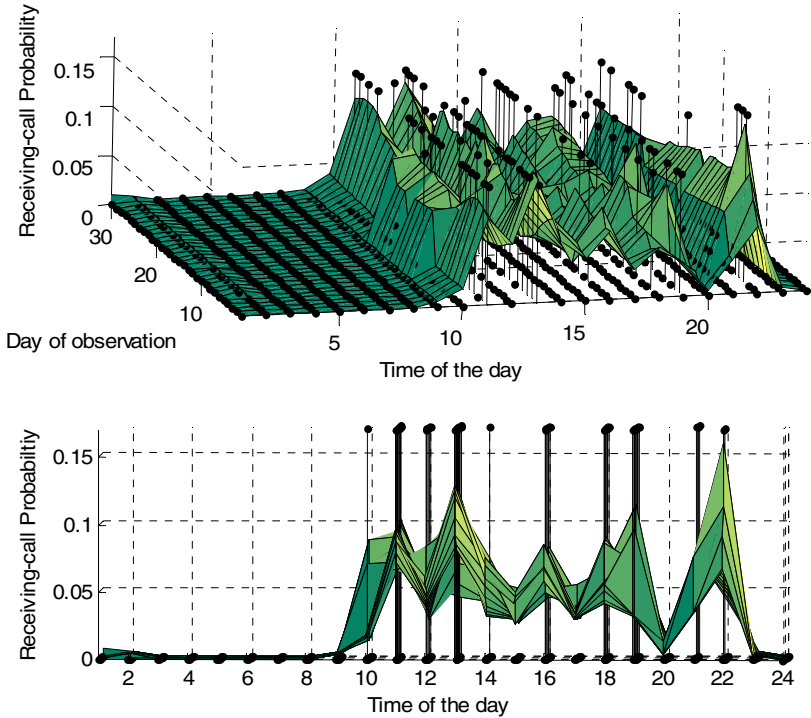


Fig. 7. A randomly selected phone user with 30 consecutive days of computed receiving-call probability of an arbitrary caller plotted with the actual received calls represented with vertical pulses. Top figure is the 3-dimensional view. Bottom figure is the front view (looking from the first day of observation).

4 Performance Analysis

The CP is tested with the actual call logs. Its performance is then measured by false positives, false negative, and error rate. A false positive is considered when a call is predicted but no call is received during that hour. A false negative is considered when no call is predicted but a call is received during that hour. Error rate is defined as a ratio of the number of fault predictions to the total number of predictions.

An experiment is conducted with 20 phone users (as mentioned in Section 3.1). The call logs of the first 2 months are used to train the CP, which is then tested with the call logs of the following month (next 30 days). Each of the 30 days of testing, the new prediction is consequently made by the CP at midnight (0 AM) with all available call history (up to that day) taken into account. The computed receiving call probability is checked with a threshold value to make a decision as to predict “call” or “no call” for each of the next 24 hours. The average number of calls per day is computed and rounded to the next largest integer M . The threshold is dynamically set as M hour slots are selected to make “Call” prediction and the rest of the $(24-M)$ time slots are predicted “No Call.” The experimental results are shown in Table 1.

Table 1. The experimental results of 20 phone users

Phone user	Number of predictions	Number of fault predictions	False positive (%)	False negative (%)	Error rate (%)	Number of fault predictions per day	Average tolerance (hours)
1	6,432	332	2.5683	2.9214	5.4896	1.3175	1.9070
2	14,472	503	1.4486	2.2042	3.6528	0.8767	1.4618
3	1,968	133	4.0278	2.7183	6.746	1.6190	2.5676
4	13,512	609	1.9651	2.5916	4.5568	1.0936	1.8395
5	8,136	547	3.9371	4.2595	8.1966	1.9672	2.8694
6	5,616	579	6.4966	5.6342	12.1308	2.9114	2.7092
7	6,000	211	2.4096	1.6697	4.0793	0.9790	1.9995
8	10,178	178	1.1033	0.8860	1.9893	0.4774	1.3011
9	1,776	224	6.4342	8.1774	14.6117	3.5068	2.1220
10	8,352	659	3.4785	4.7221	8.2005	1.9681	2.8823
11	17,400	870	2.6798	2.5612	5.2409	1.2578	1.8337
12	2,088	67	2.6235	0.6944	3.3179	0.7963	1.5622
13	7,416	374	2.6365	2.9602	5.5968	1.3432	2.0133
14	3,720	167	2.6730	3.1831	5.8561	1.4054	2.3980
15	7,632	357	2.2900	3.0674	5.3574	1.2857	2.4452
16	19,416	1,090	2.7982	3.3569	6.1551	1.4772	2.7210
17	6,840	652	3.8854	5.6129	9.4984	2.2796	1.6221
18	2,808	216	3.5417	4.2014	7.7431	1.8583	1.8703
19	2,208	181	3.5779	4.6196	8.1975	1.9674	2.7001
20	5,040	146	1.0417	1.8750	2.9167	0.7000	1.6675

There are total of 151,008 predictions made with 8,095 total fault predictions. The average false positive is 2.4416%, the average false negative is 2.9191%, and the average error rate is 5.3606%. Therefore the overall average number of fault predictions per day (24 predictions) is 1.2866 and the average tolerance is 2.1246 hours. The average tolerance is a measure of how far off (in hours) the predicted call from the actual call when fault prediction occurs.

5 Conclusion

In this paper, we propose a Call Predictor that computes receiving call probability and makes the next-24-hour call prediction. The receiving call probability is based the caller's behavior and reciprocity. The caller's behavior is measured by the caller's call arrival time and inter-arrival time. The reciprocity is measured by the number of outgoing calls per incoming call and the inter-arrival/departure time.

The kernel density estimation is used to estimate the probability model for the calling pattern based on caller's arrival time. The normal distributions are assumed for the inter-arrival time, number of outgoing calls per incoming call, and the inter-arrival/departure time. The final receiving call probability model is the average of the receiving call probabilities based on these four parameters.

To validate the model, the cell phone call records of real-life individuals at our university are used to test the call predictor. The results show that the call predictor exhibits a reasonably good performance with low false positives, false negatives, and error rate.

Clearly, there are still many parameters that need to be identified to capture the calling patterns. This work is intended to be the first piece of many more to come in this new area of predicting future calls which can be useful to many applications such as planning a daily schedule and preventing unwanted communications (e.g. voice spam). Also, the prediction technique proposed here is preliminary and other approaches need to be considered in order to minimize the number of false positives and negatives. We will continue to investigate other parameters to characterize the behaviors of the phone users and explore other prediction techniques to improve the performance of the call predictor as our future direction.

Acknowledgements. This work is supported by the National Science Foundation under grants CNS-0627754, CNS-0619871, and CNS-0551694.

References

1. Magalhães, M.H., Ballini, R., Molck, P., Gomide, F.: Combining forecasts for natural stream flow prediction. *IEEE annual meeting of the fuzzy information*, 390–394 (2004)
2. Guang, C., Jian, G., Wei, D.: Nonlinear-periodical network traffic behavioral forecast based on seasonal neural network model. *IEEE international conference on communications, circuits, and systems* 1(1), 683–687 (2004)
3. Tych, W., Pedregal, D.J., Young, P.C., Davies, J.: An unobserved component model for multi-rate forecasting of telephone call demand: the design of a forecasting support system. *International journal of forecasting* 18(4), 673–695 (2002)
4. Hansen, J.W., Nelson, R.D.: Neural networks and traditional time series methods: a synergistic combination in state economic forecasts. *IEEE Transaction on neural network* 8(4), 863–873 (1997)
5. Eagle, N., Pentland, A., Lazer, D.: Inferring social network structure using mobile phone data. In: *PNAS* (in submission, 2007)
6. Eagle, N., Pentland, A.: Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing* 10(4) (2006)
7. Eagle, N., Pentland, A.: Eigenbehaviors: Identifying structure in routine. In: *Proc. Roy. Soc. A*. (in submission, 2006)
8. Eagle, N., Pentland, A.: Social serendipity: Mobilizing social software. *IEEE Pervasive Computing* 4(2) (2005)
9. Shelldrake, R., Smart, P.: Testing for telepathy in connection with e-mails. *Perceptual and Motor Skills* 10, 771–786 (2005)
10. Eagle, N.: *Machine Perception and Learning of Complex Social Systems*, PhD Thesis, Massachusetts Institute of Technology (2005)
11. Dantu, R., Kolan, P.: Survey of Calling Patterns. University of North Texas Internal Survey (2006), <http://secnet.csci.unt.edu/nuisance/index.htm>
12. Parzen, E.: On estimation of a probability density function and mode. *Ann. Math. Statist.* 3(33), 1065–1076 (1962)
13. Jones, M., Marron, J.S., Sheather, S.J.: A brief survey of bandwidth selection for density estimation. *J. Amer. Stat. Assoc.* 433(91), 401–407 (1996)
14. Wand, M.P., Jones, M.C.: Multivariate plug-in bandwidth selection. *Computational Statistics* 9, 97–117 (1994)
15. Sheather, S.J., Jones, C.: A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B* 53, 683–690 (1991)