Chapter 20
Smart Phone: Predicting the Next Call

Huiqi Zhang and Ram Dantu

Abstract Prediction of incoming calls can be useful in many applications such as social networks, (personal, business) calendar and avoiding voice spam. Predicting incoming calls using just the context is a challenging task. We believe that this is a new area of research in context-aware ambient intelligence. In this paper, we propose a call prediction scheme and investigate prediction based on callers' behavior and history. We present Holt-Winters method to predict calls from frequent and periodic callers. The Holt-Winters method shows high accuracy. Prediction and efficient scheduling of calls can improve the security, productivity and ultimately the quality of life.

20.1 Introduction

Prediction plays an important role in various applications. Several schemes have been widely deployed for predicting weather, environment, economics, stock, market, earthquakes, flooding, network traffic and call center traffic [1–5, 7–14]. Companies use predictions of demands for making investments 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 access future network capacity requirements and to plan network development for optimum use network resources and improve quality of services. Prediction is also applied in the human behavior study [6] by combining the computer technology and social networks [10, 15, 16].

Over the past few years, there has been a rapid development and deployment of new strategic services based on the IP protocol, including Voice over IP (VoIP) and IP-based media distribution (IPTV). These services operate on private and public IP networks, and share their network with other types of traffic such as web traffic. Not

H. Zhang (✉) · R. Dantu
Computer Science and Engineering, University of North Texas, Denton, TX 76201, USA
e-mail: hz0019@unt.edu

R. Dantu
e-mail: rdantu@unt.edu

L. Cao, P.S. Yu (eds.), *Behavior Computing*,
DOI 10.1007/978-1-4471-2969-1_20,

only will VoIP reduce communication costs and provide enhanced and more flexible communication experiences to people, but it will also pave the road for innovative, valued added, highly-personalized services. We can expect, for example, that interactive multimedia and broadcast video services will be reusing the infrastructure that is being deployed for VoIP. Such trends will result in what we can call IP-based multimedia communications infrastructure, encompassing both the equivalence of conventional phone conversations and advanced communication and content distribution services. For example, we expect VoWiFi (voice over WiFi) and mobile-TV services will be widely used in the next five years. Hence, it is very conceivable that the people will be using or wearing several communication devices simultaneously. Unwanted interruptions due to these devices can waste people's time and can be of serious impact to the productivity. On the other hand, efficient scheduling of transactions can improve the security, productivity, and ultimately quality of life.

Therefore, we need context-based services for accepting or rejecting the incoming transactions. For example, predicting the wanted and unwanted calls from customers can improve the revenues for sales people. Predicting the call volume in E911 agencies can improve the efficient deployment of the resources. Predicting the expected calls for a busy business executive (personal or business) can be very useful for scheduling a day. Match making services can use calling patterns and calling behavior for the compatibility studies [17]. 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 e-mail systems. Compare receiving an e-mail spam at 2:00 a.m. that sit in the inbox until you open it the next morning to receiving a junk phone call that must be answered immediately. There was some work reported on telephone telepathy based on psychology. So far no scientific research has been reported in predicting the incoming calls for context-based services. Predicting of incoming calls using just the context is a challenging task. We believe that this is a new area of research. In this paper we predict calls from our social network.

Real-Life Data Sets Every day calls on the cellular network include calls from different sections of our social life. We believe calls from family members, friends, supervisors, neighbors, and strangers. Every person exhibits a unique traffic pattern. Calls to our home from neighbors and business associates may not be as frequent as those from family members and friends. Similarly, we talk for longer periods to family members and friends compared to neighbors and distance relatives. These traffic patterns can be analyzed for inferring the closeness to the callee. This closeness represents the social closeness of the callee with the caller on the cellular communication network.

To study closeness the people have with their callers, we collected the calling patterns of 20 individuals at our university. We are in process of collecting calling patterns from 20 more people. The details of the survey are given in Dantu et al. [18]. We found that it is difficult to collect the data set because many people are unwilling to given their calling patterns due to privacy issues. Nevertheless, the collected datasets include people with different type of calling patterns and call distributions. As part of the survey, each individual downloaded two months of detailed cell phone records from his online accounts on the cellular service provider's website. Each call record in the dataset had the 5-tuple information *Call record: (date, start time, type, caller id, talk-time)* where: date is the date of communication; start time is the start time of the communication; type is the type of call, i.e., "Incoming" or "Outgoing"; caller id is the caller identifier; and talk-time is the amount of time spend by caller and the individual during the call.

We used the call records for predicting the next caller. In Sect. 20.2 we describe a prediction technique based on a statistical model. In Sect. 20.3 we proposed a prediction scheme. In Sect. 20.4 the experiment and validation results are presented. Section 20.5 is the conclusion.

20.2 A Call Prediction Model for Frequent and Periodic Callers Using Holt-Winters Method

We used both Holt-Winters and ARIMA (Autoregressive Integrated Moving Average) [19] models which are the most general class of models for predicting a time series to predict the incoming calls and compared the accuracy of the results for these two methods. We found that there was no obvious difference between them for the accuracy since Holt-Winters method is the subset of ARIMA method. Therefore we choose the Holt-Winters method to predict incoming calls since the ARIMA method become the Holt-Winters method by adjusting the parameters of it.

Holt-Winters method [19] uses mathematical recursive functions to predict the future quantitative behavior. It uses a time series model to make predictions and assumes that the future behavior will follow the same pattern as the past. In our case we have patterns corresponding to some periodicity. This means that we use a factor in the equations that uses information from past days to make a prediction of what will happen in the future. The Holt-Winters method equations are given by:

$$Y_t = \alpha \frac{x_t}{I_{t-l}} + (1-\alpha)(Y_{t-1} - b_{t-1}) \tag{20.1}$$

$$b_t = \gamma(Y_t - Y_{t-1}) + (1-\gamma)b_{t-1} \tag{20.2}$$

$$I_t = \beta \frac{x_t}{Y_t} + (1-\beta)I_{t-l} \tag{20.3}$$

$$P_{t+m} = (Y_t - mb_t)I_{t-l+m} \tag{20.4}$$

where x_t is the observed value at time t; Y_t is the smoothed observation at time t; b_t is the trend smoothing at time t; I_t is the seasonal smoothing at time t. ($t-l$ is used in (20.1) to (20.4) to present the use of information from previous periods); l is the number of periods that complete 1 season; P_{t+m} is the prediction at m periods ahead; m is the number of periods ahead we want to predict; α is the overall smoothing parameter; β is the seasonal smoothing parameter; γ is the trend smoothing parameter.

α is the short term parameter. A large value of α will give a large weight to measurements very near in the past, while a small value of alpha will give more weight to measurements further in the past. γ is the trend parameter. A large value of γ will give more weight to the difference of the last smoothed observations; while a small value of γ will use information further in the past. β is the seasonal parameter. A large value of β will give more weight to the present relation between the observation and the smoothed observation, and small values of β will give more weight to past days relation between the observation and the smoothed observation.

Some values in (20.1) to (20.4) correspond to nonexistent periods of time in the case of Y_{t-1} or b_{t-1} When $t = 0$, we don't have any values for Y or b. So we need initial values. The same happen for the seasonal index I when $t < l$. The initial values of Y for the observation smoothing is assumed to be $S_0 = 0$. The initial values for the trend smoothing b is given in (20.5). This value is mainly an average of the differences in the observations of the first two periods divided by the length of a period.

$$b_0 = \frac{(x_{l+1} + x_{l+2} + \cdots + x_{l+l} - x_1 - x_2 - \cdots - x_l)}{l^2} \tag{20.5}$$

The initial values for the seasonal smoothing are given by (20.6). This value of the seasonal index is basically an average of the observed values of every period and an average of the day n of every period divided by the average of the observed values for its period.

$$I_{k-l} = \sum_{i=0}^{n} \frac{x_{k+il}}{\sum_{j=k+il}^{(i+1)l} x_j} \quad \text{for } k = 1, 2, \ldots, l-1, l \tag{20.6}$$

To find the parameters α, β and γ we use the minimum mean square error, where the error is the difference between the prediction and the observed values at time t. The minimization equation is given by (20.7) using $0 = \alpha, \beta, \gamma = 1$.

$$\text{Min}\left(\frac{1}{N}\sum (P_t - x_t)^2\right) \tag{20.7}$$

20.3 Prediction Scheme

Holt-Winters method can be used to predict one-step-ahead value of time series, which can be extended to k-step-ahead. We can use the mathematical expression to explain what is one-step-ahead and k-step ahead prediction. Let Yt to be the time series that we want to predict its performance. The k-step ahead prediction can be defined with $Yt + k$. This means $Yt + k$ denotes the k-step prediction made at origin time t. When $k = 1$, it is one-step-ahead prediction.

We use the one-step-ahead prediction to explain the prediction scheme, and then we may extend it to be the k-step ahead prediction.

In one-step-ahead prediction scenario, we first set the prediction step to be one, which means each time we only predict value at one time unit. According to the Holt-Winters method, we have known all the parameters from the historical actual time series, and we also know the last one of the historical time series data. From the basic idea of the prediction, we use the minimum mean square error (MMSE) prediction method to predict its performance at one time unit.

To extend the horizon of time series, the k-step-ahead prediction value can be computed recursively. For example, to obtain the two-step-ahead prediction value, the one-step-ahead predicted value is computed first. Then it is used with the other lagged values to compute the two-step-ahead predicted value. This procedure is repeated to generate subsequent k predicted values. This is similar as the one-step-ahead prediction. The call prediction algorithm is presented as follows.

Step 1: Predict the number of incoming calls on the next day using the all recorded calls of a caller and find what day is that day.

Step 2: If the prediction result is not 0, compute the correspondent frequencies of the incoming calls in the 3 time intervals, that is morning (8–12 O'clock), afternoon (13–17 O'clock) and evening (18–24 O'clock) for all those weekdays/weekend using the all recorded calls of those days (say, calls of all Saturdays in a certain period of time).

Step 3: Choose the time interval with the highest, second high and third high call frequency and using all the incoming call times in that time interval to predict the time of incoming calls respectively.

20.4 Experimental Results and Validation

The following information is available for call prediction of callers.

1. The number of incoming calls per day during a certain period of time.
2. The number of incoming calls from Monday thru Sunday respectively (7 data sets).
3. Divide the total incoming call time from Monday thru Sunday (7 data sets) into 3 time intervals, that is morning (8–12 O'clock). Afternoon (13–17 O'clock) and evening (18–24 O'clock) respectively.

Because of the limited space we only choose the data set on all previous Fridays for the receiver 1 as an example to show how to predict the incoming calls on the next day which is Friday.

Figures 20.1, 20.2, 20.3 and Table 20.1 show the prediction results of the calls using our call prediction algorithm and Holt-Winters procedure, where the red dots denote the predicted values and green dots denote the observation values on all Fridays for the receiver 1.

Step 1: Predict the number of incoming calls for next day and this day is Friday. Figure 20.1 shows that there will be 1 incoming call (predicted value is 1.2) and the observation is 2 calls. In Fig. 20.1 the x-axis is the days and y-axis is the number of calls.

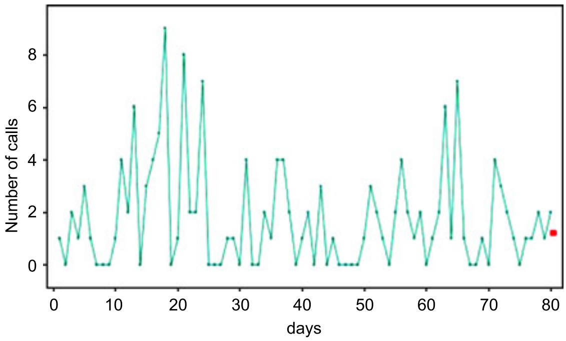

Fig. 20.1 The predicted and actual number of calls by caller 1

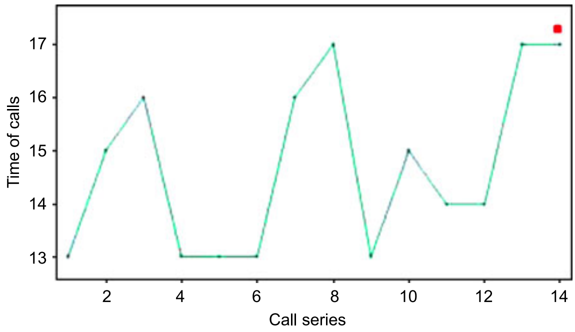

Fig. 20.2 The 1st predicted call time and actual call time of user 1 on Friday. The *red dot* indicates the predicted value and *green dots* indicate the actual values (Color figure online)

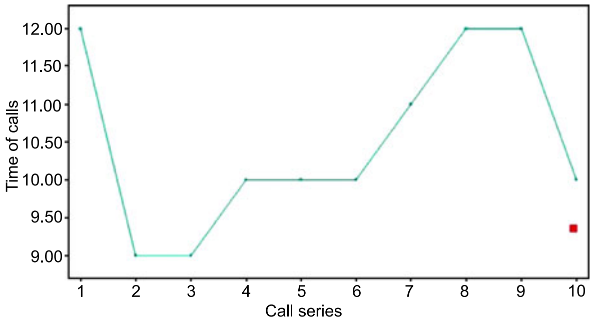

Fig. 20.3 The 2nd predicted call time and actual call time of user 1 on Friday

Step 2: Compute the correspondent frequencies of the incoming calls in the 3 time intervals on all previous Fridays. There are about 36% call time are in 8–12 O'clock interval, 50% call time are in 13–17 O'clock interval, and 14% call time are in 18–24 O'clock interval shown in Table 20.1 respectively.

The 13–17 O'clock time interval on all previous Fridays was selected to predict the incoming call time since there is the highest frequency in this time interval (50% of the all incoming calls in this time interval).

Next we predict the time of the incoming calls on the next day (Friday). Figure 20.2 shows that there will be an incoming call at about 17.3 O'clock which is

Table 20.1 Call frequency

Time interval	Frequency	Cumulative %
8–12	10	35.71%
13–17	14	85.71%
18–24	4	100.00%

Table 20.2 Prediction results

Receivers	Callers	Total # of days/calls	# of calls on weekdays	Prediction of # of calls	Prediction of time	Observation of # of calls	Observation of time
Receiver1	Caller1	80/148		1.2		2	
			14 (13–17 Fri.)		17.3		17
			10 (8–12 Fri.)		9.3		10
	Caller4	92/86		1.4		1	
			9 (18–24 Mon.)		23		22
	Caller2	87/181		−0.3		1	
			7 (8–12 Wed.)		11.0		11
Receiver2	Caller9	129/229		2.3		3	
			12 (18–24 Sat.)		21.0		22
	Caller10	116/218		1.5		2	
			15 (18–24 Fri.)		23.8		24
			17 (13–17 Fri.)		13.0		13
Receiver10	Caller1	84/234		2.5		4	
			10 (18–24 Sat.)		21.5		22
	Caller114	49/138		2.5		3	
			14 (18–24 Sun.)		22.3		22
	Caller3	75/109		0.9		1	
			7 (18–24 Fri.)		22.0		23
Receiver3	Caller10	116/282		2.4		2	
			29 (13–17 Fri.)		13.6		13
			11 (18–24 Fri.)		23.3		24

the predicted value (17:20 O'clock) and the observation is 17 O'clock. In Fig. 20.2 the x-axis is the call series on all Fridays (i.e. 1st call, 2nd call, ...) and y-axis is the incoming call time. Next the 8–12 O'clock time interval on Friday is computed to predict the second incoming call time. Figure 20.3 shows that the second predicted incoming call will be at about 9.3 O'clock (9:20 O'clock) and the observation is 10 O'clock.

The number of calls and corresponding call times for additional callers and receivers have also been predicted. The results are summarized as follows and shown in Table 20.2. Caller 4 for receiver 1: The number of call predicted is 1.4 and the observation is 1. The predicted call time is timeHour23Minute023 O'clock and the observation is timeHour22Minute022 O'clock. Caller 2 for receiver 1: The number of calls predicted is −0.3 and the observation is 1. This call time predicted is timeHour11Minute011 O'clock and the observation is timeHour11Minute011 O'clock. Caller 9 for receiver 2: The number of call predicted is 2.3 and the observation is 3. This call time predicted is 21 O'clock and the observation is 22 O'clock.

Predicted calls for other receivers and frequent callers on the weekdays/weekend are shown in Table 20.2. In Table 20.2 the "80/148" indicates that there are 148 incoming calls in 80 days in the "Total # of days/calls" column and the "14(13–17 Fri.)" indicates that there are 14 incoming calls in 13–17 O'clock time interval on all previous Fridays in the "# of calls on weekdays" column.

20.5 Conclusion

The results show that this approach can be used to predict the calls and are reasonable accurate. The prediction technique proposed are preliminary and other approaches need to be considered in order to improve accuracy. The sample size is only 20, but, represents typical profiles (socially active, socially inactive, socially moderate). We are also working on the call logs of 100 phone users from MIT Reality Mining Group over a period of 8 months (between October 2004 and May 2005) [17]. We are observing similar results issues are to be addressed by future work.

References

1. Gray, A.G., Thomson, P.J.: On a family of finite moving-average trend filters for the ends of series. J. Forecast. **21**(2), 125–149 (2002)
2. Taylor, G.W.: Smooth transition exponential smoothing. J. Forecast. **23**(6), 385–404 (2004)
3. Man, K.S.: Long memory time series and short term forecasts. Int. J. Forecast. **19**(3), 477–491 (2003)
4. Zou, H., Yang, Y.: Combining time series models for forecasting. Int. J. Forecast. **20**(1), 69–84 (2004)
5. Liu, H., Hall, S.G.: Creating high frequency national accounts with state space modeling A Monte Carlo experiment. J. Forecast. **20**(6), 441–449 (2001)

6. Cao, L.: In-depth behavior understanding and use: the behavior informatics approach. Inf. Sci. **180**, 3067–3085 (2010)
7. Janacek, G., Swift, L.: Time Series: Forecasting, Simulation, Applications. Ellis Horwood, Chichester (1993)
8. Magalhaes, M.H., Ballini, R., Molck, P., Gomide, F.: Combining forecasts for natural stream flow prediction. In: IEEE Annual Meeting of the Fuzzy Information, vol. 1, pp. 390–394 (2004)
9. Guang, C., Jian, G., Wei, D.: Nonlinear-periodical network traffic behavioral forecast based on seasonal neural network model. In: IEEE International Conference on Communications, Circuits and Systems, vol. 1, pp. 683–687 (2004)
10. Young, P.C.: Non-stationary time series analysis and forecasting. J. Prog. Environ. Sci. **1**, 3–48 (1999)
11. Young, P.C., Tych, W., Pedregal, D.J.: Stochastic unobserved component models for adaptive signal extraction and forecasting. In: IEEE Signal Processing Society Workshop, vol. 8, pp. 234–243 (1998)
12. Groschwitz, N.K., Polyzos, G.C.: A time series model of long term NSFNET backbone traffic. In: IEEE International Conference on Communications, vol. 3, pp. 1400–1404 (1994)
13. Hansen, J.W., Nelson, R.D.: Neural networks and traditional time series methods: a synergistic combination in state economic forecasts. IEEE Trans. Neural Netw. **8**(4), 863–873 (1997)
14. Tych, W., Pedregal, D.J., Young, P.C., Davies, J.: An unobserved component model for multirate forecasting of telephone call demand: the design of a forecasting support system. Int. J. Forecast. **18**(4), 673–695 (2002)
15. Harvey, A.C.: Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge (1989)
16. Nelson, M., Hill, T., Remus, B., O' Connor, M.: Can neural networks applied to time series forecasting learn seasonal patterns: an empirical investigation. In: IEEE International Conference on System Sciences, vol. 3, pp. 649–655 (1994)
17. Eagle, N., Pentland, A., Lazer, D.: Inferring social network structure using mobile phone data. PNAS (2006, in submission)
18. Kolan, P., Dantu, R.: Socio-Technical Defense Against Voice Spamming, ACM Trans. Auton. Adapt. Syst. December (2006, accepted)
19. Bowerman, B.L.: Forecasting, time series, and regression: an applied approach (2005)