ABSTRACT
The mobile phone offers a unique opportunity to predict a person’s behaviour automatically for advanced ubiquitous services. In this note, we analyse cellular data collected as part of the Reality Mining project and use information-theoretic concepts to answer three questions (i) What time points in the day help predict a mobile phone user’s activity at another time point? (ii) What time points in history are most useful to predict his future activities? and (iii) How difficult is it to predict his activity at a given time from another user’s activity at another time?

Author Keywords
Machine learning, human behaviour, mobile phone.

ACM Classification Keywords
H.1.1 Systems and Information Theory: Information theory.

General Terms
Experimentation, Human Factors, Measurement.

INTRODUCTION AND RELATED WORK
The idea that past information can help predict human behaviour is not new. For example, [6], [5] and [7] have noticed that prediction can be improved by looking at certain periodic behaviours. Using mobile phones, [2], [3] and [9] have also observed periodicity in human behaviour and [8] has investigated the effect of history length on prediction accuracy. In this note, we take an information-theoretic approach to quantify how difficult it is to predict a mobile phone user’s activity when his or another user’s activities at selected time points are known. By analysing cellular data from the publicly available Reality Mining dataset, we do not just show periodicity in human behaviour but in information-theoretic quantities that measure uncertainty about a user’s activity when his activities at selected time points are known. The distinction between the two notions is important since our results do not hold on strictly periodic signals. In addition, human behaviour does not have the same cycles as the quantities that we analyse. We also investigate what exact time points help predict a user’s activity throughout the day, explain why these time points help and examine the impact of activity granularity on prediction. Lastly, we look at how predictions can be made from another user’s activities.

The closest work to ours are [11] and [1]. In [11], information theory is used to investigate whether knowing the cells visited by a user in the past can help determine the current location of that user. However, their approach only considers cells where phone calls were made. In contrast, we follow the user continuously throughout the day. In addition, our analysis takes into account the meaning of locations, not just cells. This allows us to find periodic phenomena in information-theoretic quantities. Different information-theoretic concepts were also used in [1], but applied to one mobile phone user only. The analytical techniques introduced in this note are new and validated on data from thirty users. Predictability is analysed at a temporal resolution twice as fine as previously and up to three weeks in the future.

MOBILE PHONE DATASET
In order to ensure repeatability, we carry out our experiments on the publicly available Reality Mining dataset [2]. Although focused on academic mobile phone users, this remarkable resource has served as a basis for a large body of work and is one of the most studied mobile phone datasets. It was recorded over the course of nine months by 94 mobile phone users from the MIT Media Lab and Sloan Business School (students and staff members). Every time a user changed cell tower, the identifier of his new serving cell tower was recorded. Using the participants’ feedback during the Reality Mining experiment, previous work has clustered Cell IDs into a small number of categories corresponding to different user activities. We consider in the following the same four categories as in [4], namely, the user being at home, at work, at some other location or receiving no signal. In [4], a subset of 30 participants and 121 consecutive days is identified as being the subset of the dataset containing the most cell tower information. We consider the same users and time period (from 26th August to 24th December 2004) and look up, for each half-hour, the cluster in which each user spent the longest.
nant clusters over half-hour intervals eliminates noise in data and provides a relatively fine temporal resolution. For each user, we obtain a sequence of 5808 consecutive data points that are grouped by days on Figure 1. The distribution of all users’ activities is provided on Figure 2 for information.

Figure 1. User 39's activities, grouped by days

Figure 2. Distribution of the thirty users’ activities throughout the day

BACKGROUND

Recall that we are concerned with quantifying relationships between mobile phone users’ activities which can be modelled as discrete random variables. An intuitive way to measure the dependency between two discrete random variables $X$ and $Y$ with respective supports $\mathcal{X}$ and $\mathcal{Y}$ would be to represent their outcomes by scalars and compute their correlation coefficient. However, this approach has two major drawbacks (i) the value of the correlation coefficient depends on the scalars chosen and (ii) correlation only measures linear dependency. An alternative approach is to consider the unweighted $L_1$ metric, defined as

$$H(Y|X) = -\sum_{(x,y)\in \mathcal{X}\times \mathcal{Y}} P_{XY}(x,y) \log_2 P_Y(y|x).$$

This quantity, expressed in bits, is the difference between the joint entropy $H(X,Y)$ and the entropy $H(X)$. It equals zero if the value of $Y$ is completely determined by the value of $X$ and equals $H(Y)$ if and only if $X$ and $Y$ are independent. Intuitively, $H(Y|X)$ measures uncertainty expressed as the average number of bits of information that are needed to determine the value of $Y$ when the value of $X$ is known. It therefore tells us how difficult it is to predict the value of $Y$ given the value of $X$. In the rest of this note, all discrete random variables represent the activities of mobile phone users at different time points. Unless stated otherwise, their support is $S = \{\text{no signal, other, home, work}\}$ where symbols represent the activities mentioned in the previous section.

PREDICTING ACTIVITIES THROUGHOUT THE DAY

Our first objective is to determine what time points in the day help predict a user’s activity at another time point. In order to analyse relationships between activities from one day to the next, we consider a time window spanning two consecutive days. Let $X_t$ represent a user’s activity during a fixed half-hour $t$ in this window. $t$ varies from 1 (from 00:00 to 00:30 on the first day) to $n = 96$ (from 23:30 to 00:00, two days later). We define the routine matrix of a user as the $n \times n$ matrix whose $(i,j)$th term equals $H(X_i|X_j)$. In other words, the element at position $(i,j)$ of the routine matrix represents the uncertainty that remains on the user’s activity at half-hour $i$ when his activity at half-hour $j$ is known. Figure 3 shows the average routine matrix $R$ of all thirty users. The two dark horizontal stripes have an average entropy lower than 1 bit, which confirms the intuition that the users’ activities are easier to predict at night than during the day (as an indication, drawing an activity at random among four from a uniform distribution leads to an entropy of 2 bits). For a fixed predicted half-hour, one can compare how the knowledge of different half-hours reduces uncertainty by following the horizontal line that goes through it. Because the entropy of a random variable conditional on itself is zero, the diagonal elements of the matrix are black. The diagonal is surrounded by a very dark area because users usually spend several consecutive half-hours on the same activity. The width of this area varies greatly throughout the day. It is largest during sleeping (n1|n1 and n2|n2) and working hours (d1|d1 and d2|d2), when the users do the same thing for long periods of time. Dark areas off the diagonal correspond to long-range dependencies in data. The darkness of areas n2|n1 and n1|n2 show that uncertainty on users’ activities at night is greatly reduced if their activities the night before or after, respectively, are known. This happens because users typically spend several consecutive nights at locations home or other. The darkness of areas n1|d1, n1|d2, n2|d1 and n2|d2 is a consequence of working routines. Knowing that the users were at work during the day is a strong indication that they were or will be home the previous or following night, respectively.

Figure 3. Average routine matrix of the thirty users

The high-entropy areas on Figure 3 can be explained through decomposition. Figure 4 is a four by four grid of matrices. Matrix $R_{x_2|x_1}$ at position $(x_1, x_2)$ in this grid represents the contribution of the pair of activities $(x_1, x_2)$ to $R$. It is itself the $n \times n$ matrix whose $(i,j)$th element equals $-P_{X_i,X_j}(x_i|x_j)\log_2 P_{X_i,X_j}(x_i|x_j)$. Hence, we have...
\[ R = \sum_{(x,y) \in S^2} R_{xy}. \] The bright areas \( d1|n1, d1|n2, d2|n1 \) and \( d2|n2 \) are mostly explained by \( R_{\text{other}|\text{home}}, R_{\text{home}|\text{home}} \) and \( R_{\text{work}|\text{work}} \) respectively. In other words, knowing that users were home at night does not help determine their activities during the previous nor the following time of day. Lastly, the brightness of \( d1|d2 \) and \( d2|d1 \) is mostly due to \( R_{\text{home}|\text{work}} \) and \( R_{\text{work}|\text{work}} \). Even when it is known that the users have been working during a certain day, a lot of uncertainty remains on whether they have done or will do so on the previous or following day, respectively. Over the time period that we considered, the average routine matrix does not vary much from one month to the next. However it does vary when only certain days of the week are considered. For example, the average routine matrix computed over Saturdays and their following Sundays only is less entropic at \( d1|n1, d1|n2, d1|d2, d2|n1, d2|d1 \) and \( d2|n2 \) because users rarely go to work during the weekend. Variations in routine matrices across users also exist but are not explained by their affiliations (Media Lab/Sloan students/staff members).

**PREDICTING FROM EARLIER ACTIVITIES**

What time points in history are most useful to predict a user’s future activities? In order to answer this question, we compute the entropy \( H(X_{t_1}|X_{t_1-\tau}) \) of a user’s activity at half-hour \( t \) conditional on his activity at half-hour \( t - \tau \), for \( \tau \) (the lag) varying between \( 1 \) and \( 3 \times 7 \times 48 = 1008 \) (3 weeks earlier). We call routine profile the graph of \( H(X_{t_1}|X_{t_1-\tau}) \) against \( \tau \), and represent on Figure 5 its average for the thirty users (top-most curve). The initial growth of entropy shows how much harder it is to make predictions as the range of predictions increases from 1 to 12 hours. If a user’s recent activities are unknown, one should then seek local minima of the routine profile. Local minima can be found at lags that are multiples of one week because of weekly periodicity in human behaviour. The graph also exhibits twelve-hour-periodic oscillations that result from the superimposition of several periodic contributions to \( H(X_{t_1}|X_{t_1-\tau}) \). Figure 6 shows the contributions \( -P_{X_{t_1}X_{t_1-\tau}}(x,y) \log_2 P_{X_{t_1}X_{t_1-\tau}}(x|y) \) against \( \tau \) for all activity pairs \( x|y \). While most contributions just grow very rapidly, some of them are daily periodic, as shown in the frequency domain by Figure 7. The main periodic contributions are other\text{home}, work\text{home} and work\text{work}. The latter has a phase difference of twelve hours with the former two, which creates the oscillations found on the routine profile. How do oscillations vary with the selected division of activities? Figure 5 also shows how the routine profile changes when two of the initial activities \( x \) and \( y \) are merged into one \( x + y \). The twelve-hour-periodic oscillations and the weekly minima disappear when the distinction between home and work is lost (bottom-most curve). Therefore, the routine profiles of a few users in the dataset with inaccurate Cell ID labels do not have these features either.

**PREDICTING FROM ANOTHER USER’S ACTIVITIES**

We have so far focused on predicting activities from just one earlier time point. What two time points provide the most information on a user’s future activity? We can generalise the concept of routine profile to more than one earlier activity by plotting the conditional entropy \( H(X_{t_1}|X_{t_1-\tau_1}, X_{t_1-\tau_2}) \) against \( \tau_1 \) (the first lag) and \( \tau_2 \) (the second lag) (Figure 8). This symmetric plot features a darker grid with twelve-hour steps along both directions. Diagonal points are much brighter than non-diagonal ones. This shows the advantage of considering two time points that are far away from each other to make predictions. By carefully selecting time points on the darker grid, one can reduce uncertainty on users’ activities even without any recent knowledge. For example, we circled on Figure 8 a local minimum at exactly \( \tau_1 = 4 \times 48 \) and \( \tau_2 = 7 \times 48 \) half-hours.

**Figure 5. Average routine profile of the thirty users**

We have so far focused on predicting activities from just one earlier time point. **What two time points provide the most information on a user’s future activity?** We can generalise the concept of routine profile to more than one earlier activity by plotting the conditional entropy \( H(X_{t_1}|X_{t_1-\tau_1}, X_{t_1-\tau_2}) \) against \( \tau_1 \) (the first lag) and \( \tau_2 \) (the second lag) (Figure 8). This symmetric plot features a darker grid with twelve-hour steps along both directions. Diagonal points are much brighter than non-diagonal ones. This shows the advantage of considering two time points that are far away from each other to make predictions. By carefully selecting time points on the darker grid, one can reduce uncertainty on users’ activities even without any recent knowledge. For example, we circled on Figure 8 a local minimum at exactly \( \tau_1 = 4 \times 48 \) and \( \tau_2 = 7 \times 48 \) half-hours.

**Figure 8. Average 2D routine profile of the thirty users**

We have so far focused on predicting activities from just one earlier time point. **What two time points provide the most information on a user’s future activity?** We can generalise the concept of routine profile to more than one earlier activity by plotting the conditional entropy \( H(X_{t_1}|X_{t_1-\tau_1}, X_{t_1-\tau_2}) \) against \( \tau_1 \) (the first lag) and \( \tau_2 \) (the second lag) (Figure 8). This symmetric plot features a darker grid with twelve-hour steps along both directions. Diagonal points are much brighter than non-diagonal ones. This shows the advantage of considering two time points that are far away from each other to make predictions. By carefully selecting time points on the darker grid, one can reduce uncertainty on users’ activities even without any recent knowledge. For example, we circled on Figure 8 a local minimum at exactly \( \tau_1 = 4 \times 48 \) and \( \tau_2 = 7 \times 48 \) half-hours.
matrix of a pair of users as the \( n \times n \) matrix whose \((i, j)\)th element equals the entropy \( H(X_i|Y_j) \) of the first user’s activity \( X_i \) at half-hour \( i \) conditional on the second user’s activity \( Y_j \) at half-hour \( j \). Six of the thirty users that we considered in our experiments were students from the MIT Sloan Business School. There are good reasons to believe that the behaviours of individuals belonging to the same group are related. For example, students attending the same classes have the same breaks. Figure 9 shows the average cross-routine matrix of any pair of different Sloan students. Its structure is very similar to the average routine matrix (Figure 3) except that diagonal elements are brighter during the day. A single user is subject to a certain inertia that makes him do the same thing for several consecutive half-hours even during the day. In contrast, nothing prevents one user to stop an activity shortly after another user started it. We also computed the average cross-routine matrix of pairs of users where the first user is a Sloan student and the second is not. The difference between this cross-routine matrix and the cross-routine matrix of Figure 9 is provided on Figure 10. It shows how much easier it is to predict a Sloan student’s activity from another Sloan student’s activity rather than from a user who does not belong to this group. The difference in entropy has a mean of 0.23 and a standard deviation of 0.06 bits. The greatest improvements are observed when predicting day activities from other day activities. Despite the high remaining entropy of day compared to night, knowing when a Sloan student attends classes does help determine whether another Sloan student does the same and can also be used to make predictions over the previous and the following days.

**CONCLUSION**

Predicting the behaviour of mobile phone users is an ambitious undertaking that will impact people’s lives in the coming years. In this note, we took an information-theoretic approach and quantified predictability in a principled way without referring to a specific predictor. We showed how to select time points to reduce uncertainty on a mobile phone user’s activity at a given time of the day and up to three weeks in the future. We also quantified how his activity at a certain time of the day can be determined from another user’s activity at another time. The results presented in this note could be extended by looking at a broader set of activities using other sensors and developing prediction systems.

**ACKNOWLEDGEMENTS**

We thank Nathan Eagle, Katayoun Farrahi and Daniel Gatica-Perez for their help with the Reality Mining dataset.

**REFERENCES**