An Information Theory Based Method for Quantifying the Predictability of Human Mobility

ZHIWEN YU*, Northwestern Polytechnical University, China
MINLING DANG, Northwestern Polytechnical University, China
QILONG WU, Northwestern Polytechnical University, China
LIMING CHEN, Ulster University, UK
YUJIN XIE, Northwestern Polytechnical University, China
YU WANG, Northwestern Polytechnical University, China
BIN GUO, Northwestern Polytechnical University, China

Research on human mobility drives the development of economy and society. How to predict when and where one will go accurately is one of the core research questions. Existing work is mainly concerned with performance of mobility prediction models. Since accuracy of predict models doesn't indicate whether or not one's mobility is inherently easy to predict, there has not been a definite conclusion about that to what extent can our predictions of human mobility be accurate. To help solve this problem, we describe the formalized definition of predictability of human mobility, propose a model based on additive Markov chain to measure the probability of exploration, and further develop an information theory based method for quantifying the predictability considering exploration of human mobility. Then we extend our method by using mutual information in order to measure the predictability considering external influencing factors, which has not been studied before. Experiments on simulation data and three real-world datasets show that our method yields a tighter upper bound on predictability of human mobility than previous work, and that predictability increased slightly when considering external factors such as weather and temperature.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing theory, concepts and paradigms**;

Additional Key Words and Phrases: Human mobility, Predictability, Information entropy, Human behavior prediction

1 INTRODUCTION

Move behavior, as the most basic human activity, appears in every aspect of social life. Understanding the internal principles of human mobility will profoundly promote the development of economy and society [43]. For instance, temporal and spatial patterns on human mobility are of great theoretical significance in disease transmission, traffic flow control, abnormal behavior monitoring [40], personalized recommendation system [42] and so on. The rapid development of mobile location services, wireless communication and mobile Internet technologies

*Corresponding author.

Authors' addresses: ZHIWEN YU, Northwestern Polytechnical University, Xi'an, China, zhiwenyu@nwpu.edu.cn; MINLING DANG, Northwestern Polytechnical University, Xi'an, China, yixi@mail.nwpu.edu.cn; QILONG WU, Northwestern Polytechnical University, Xi'an, China, qilongwu@mail.nwpu.edu.cn; LIMING CHEN, Ulster University, Belfast, UK, chen@ulster.ac.uk; YUJIN XIE, Northwestern Polytechnical University, Xi'an, China, xieyujin@mail.nwpu.edu.cn; YU WANG, Northwestern Polytechnical University, Xi'an, China, wyu@mail.nwpu.edu.cn; BIN GUO, Northwestern Polytechnical University, Xi'an, China, guob@nwpu.edu.cn.

1556-4681/2023/5-ART
https://doi.org/10.1145/3597500

has made it easier to obtain large-scale, long-term data about individual mobility at much fine-grained temporal and spatial levels, which sparked a wave of research into human mobility.

One main area that existing research on human mobility has focused on is about prediction algorithm, which includes building specific prediction models, adjusting optimization strategies and other methods to improve performance of models. However, human behavior has its inherent uncertainty, which makes it impossible for an algorithm to predict human behavior completely accurately, no matter how optimized it is. Therefore accuracy of the predictions that have been made so far does not tell us how accurately human mobility can be predicted. Predictability, as a metric of how predictable a time series of mobility is, indicates the upper bound of accuracy of predictions on it. Under this definition, predictability should be an internal feature of data and independent of any specific predictive model. For example, no matter how good a predictive algorithm is, the accuracy will be theoretically no more than 85% while the predictability is 85%.

Previous research have reported two other kinds of the definition of predictability. One can be seen as the magnitude of error propagation, which is related to the performance of the specific algorithm used. The growth rate of the error is mainly represented using Lyapunov exponents [24], perturbations are added to the model, and its error growth rate is calculated if the initial conditions are different from a small random perturbation. And it is not easy to split the predictability problem into components related to the initial error and components related to the model error. The other is based on complexity measurement. Their value has no exact physical meaning and can only help us understand the difficulty of prediction from a qualitative point of view. Permutation entropy is used to quantify a special human mobility behavior - the outbreak of infectious diseases[29]. The permutation entropy is regarded as the probability of the unpredictable part of the sequence. However, because the calculation of permutation entropy needs to set parameters: dimensions, and time delay, meanwhile, the results deviate greatly from the theoretical values and tend to zero as the sequence length increases. Therefore, this method is not suitable for quantifying the predictability of human mobility. Predictability that meets the quantitative definition we mentioned in the previous paragraph was first measured by Song [32]. But their method is a general method that can be used for any time series. They got gross to overestimate without considering any characteristics of human mobility, like exploration. What's more, a large number of factors affecting human mobility (like weather) have not been taken into consideration in existing methods for predictability quantification.

It is still a challenge to quantify the predictability of human mobility. To provide better insights, we summarize the difficulties as follows:

(1) Measures of predictability in previous research were either not quantitative or not accurate. No one has yet developed a specific method to more accurately measure the predictability combined with the characteristics of human mobility.
(2) External factors (e.g. temperature, weather, etc.) have a significant impact on human behavior but have been mostly ignored in previous research. It is an open issue to investigate how to quantify the influence of external factors on the predictability of human behavior.

Motivated by the above observations, we propose a quantification method for the predictability of human mobility based on information theory and additive Markov chain and show how external factors affect the predictability. The main contributions of this work include:

- We describe the formalized definition of the predictability of human mobility, propose a model based on an additive Markov chain to measure the probability of exploration, and further develop an information theory-based method for quantifying the predictability considering the exploration of human mobility. We provide the complete process of mathematical derivation as a theoretical foundation for the proposed method.

- We analyze the influence of external factors on predictability and extend our method by using mutual information in order to measure the predictability considering external influencing factors, which has not been studied before.
- We evaluate our method on simulation data and three real-world datasets. The results demonstrate that our method yields a tighter upper bound on the predictability of human mobility than previous work, and that predictability increased slightly when considering external factors such as weather and temperature.

The rest of this paper is organized as follows. Section 2 introduces related work. Section 3 describes the concept of predictability and related problems. Section 4 shows the method we proposed and its derivation. We evaluate the method and analyze experimental results in Section 5. Finally, we conclude our paper in Section 6.

2 RELATED WORK

2.1 Human Behavior Prediction

Research on the predictability of time series is inseparable from forecasting algorithms. The performance of forecasting algorithms can be used to evaluate the quality of the proposed predictability quantification method to some extent. Modeling time-series data for the purposes of prediction dates back to Yule's 1927 invention of autoregression [45]. Since then, a lot of strategies have been developed for a wide variety of prediction tasks. Human behavior prediction can be seen as a time series forecasting problem. And there are many existing time series models, including univariate Autoregressive (AR), univariate Moving Average (MA), Autoregressive Integrated Moving Average (ARIMA) model [20] [30] and Vector Autoregressive model [49]. These classic models can be used to capture the dependency of time series. In addition to these models, Markov Chains is a popular model, which assumed that the probability of the next behavior only depends on the current behavior. Lu [21] applied the Markov Chain model to a mobile dataset from Côte d'Ivoire, and their prediction goal was to estimate the final location in a day. Yan [38]proposes a weighted Markov prediction model based on mobile user classification, the trajectory information of a user is extracted by analyzing real mobile communication data, and all users are classified with machine learning algorithms, the step threshold and the weighting coefficients of the weighted Markov prediction model are optimized, the result improves the performance of the weighted Markov prediction model. Another popular prediction framework is based on the Naive Bayes model, in which the probability of next position is decomposed into independent probabilities of multiple context variables. Gao [12] applied this method with time and location features to the Nokia Data Challenge dataset [18] and obtained about 50% accuracy. A number of more complex methods have also been explored in the literature, including non-linear time series [21], Principal Component Analysis [27], Gaussian Mixtures [4] and Dynamic Bayesian Networks[27].

Moreover, the study regularity of human mobility further provides a theoretical thesis for the prediction algorithm. Human mobility spans the entire space and has different regularity at different spatio-temporal scales, Austin [1] collected a data set of almost 15 million observations from 19 adults spanning up to 5 years of unobtrusive longitudinal home activity monitoring, finding that in-home mobility is not well represented by a universal scaling law, but that significant structure (predictability and regularity) is uncovered when explicitly accounting for contextual data in a model of in-home mobility, and mobility patterns of older individuals in their home also show a high degree of predictability and regularity in human home-space mobility when context is taken into account. Human movement outside the home space will contact different individuals, Wang [35] proposed a hybrid predictive model integrating both the regularity and conformity of human mobility as well as their mutual reinforcement. In addition, it learned location profiles from heterogeneous mobility datasets based on a gravity model, capturing users' regular movement patterns and their occasional visits influenced by others. Within this regular movement pattern there is a special phenomenon, Leng [19] collected a data set of almost 15 million observations from 19 adults spanning up to 5 years of unobtrusive longitudinal home activity monitoring, finding that in-home mobility is not well represented by a universal scaling law, but that significant structure (predictability and regularity) is uncovered when explicitly accounting for contextual data in a model of in-home mobility, and mobility patterns of older individuals in their home also show a high degree of predictability and regularity in human home-space mobility when the context is taken into account. As human mobility expands to the urban scale, Oliveira [23] presented a system model, which unifies different datasets into a common representation of urban scenarios and analysed the visiting patterns. It has results about human mobility: people have a tendency to use the shortest path when moving around, and their mobility is confined, in addition, regular patterns found in human mobility are not restricted by the scale of the dataset.

With the success of neural networks and machine learning in computer vision and natural language understanding, people are gradually applying them to time series forecasting problems [39]. Recurrent Neural Network(RNN) has achieved great success in sequence learning, which can be used to mine long short-term correlations of sequences. It can also be combined with **C**onvolutional **N**eural **N**etworks, the influence of external factors on time sequence is further considered to improve the accuracy of sequence prediction. Dang [6] introduce the graph convolution and dual-attentive mechanism to handle the sparsity and inaccuracy of the trajectory data and the high-order sequential nature in the problem of human mobility prediction, GCDAN achieves significant performance gain compared with state-of-the-art baselines. There are also many studies on human behavior based on the above models, including the mining of human behavior patterns [15] [44] [46]. And there are some studies[2] [25] that predict the movement trajectory of users based on the historical location of human beings, as well as predict traffic flow[22] [11] and traffic congestion[10]. Kong[16] adopt a multi-pattern approach to predict the bus passenger flow by taking advantage of graph learning, it proposes a multi-pattern passenger flow prediction framework, MPGCN, based on Graph Convolutional Network (GCN), to learn human mobility knowledge from fixed travel behaviors, this work is based on frequent and consistent travel, but it may not be suitable for the infrequently used transport mode. The basic idea of these studies is to construct feature models by machine learning algorithms to predict human behavior. The main drawback is that these studies cannot give a reasonable upper limit of human behavior prediction, and therefore cannot prove whether the algorithm or model proposed is really good.

2.2 Time Series Predictability

Quantifying predictability has been studied before in many fields, such as climate [17] and stock returns [28]. Most solutions can be divided into two categories: model-based error analysis method [14] and model-free information theory-based method [7].

The methods in the first category mainly quantify predictability by analyzing the error of a specific predict model [8]. They cannot draw a conclusion about the degree to which a time series can be predictable, which can be used to evaluate other forecast methods. The method proposed in this paper falls into the second category in which methods are to mine the inherent predictability of a time series by analyzing the characteristics of the sequence itself [13]. Existing model-free information theory-based methods can be further divided into two categories — sequence complexity-based and information entropy-based [32].

2.2.1 Predictability Based on Sequence Complexity. Intuitively, we generally believe that for a time series, the more complex the less predictable, which means high complexity corresponds to low predictability, and low complexity corresponds to high predictability conversely. Inspired by this, a measure of predictability based on permutation entropy [29], [26] has been studied, where permutation entropy is considered as the complexity of the sequence. This process involves several steps — calculate the permutation entropy of a sequence, convert the permutation entropy to a 0-1 interval (normalization), and then the predictability of the corresponding sequence is obtained by one subtract complexity. However, its result is only from a qualitative perspective and not a direct measure of predictability.

2.2.2 Predictability Based on Information Entropy. In this kind of method, the intuitive understanding of predictability remains the same, that is, the more complex a time series is, the lower the predictability is. The difference is that predictability is linked to a specific metric — the upper limit of accuracy that predictions based on a time series can achieve. The method proposed in [32] establishes for the first time a rigorous quantitative relationship between this specific metric and the sequence uncertainty (entropy). This method modeled an individual's movement pattern as a stochastic process defined a new inherent uncertainty of a time series called real entropy, and proposed a formula for calculating the upper bound of predictability from real entropy. This work was widely followed up in analyzing the predictability of various time series. Some problems remained open, for example, articles [36], [31] pointed out that the upper bound of predictability in [32] is an over-estimate, and also the effects of external factors on human behavior predictability has never been considered.

The work presented here differs substantially from previous research in that it integrated preferential return and random exploration [5] existed in human mobility into the model of the upper bound of predictability in [32]. Our refinement results in a tighter and more realistic upper bound. In addition, we also consider the influence of external factors (weather and temperature, etc.) on predictability as an extension of this method.

3 PROBLEM DESCRIPTION

Here we present relevant definitions and problem descriptions.

Definition 1 (Predictability). *Given a time series $X = \{X_1, X_2, ..., X_i, ..., X_n\}$ where X_i denotes the status at time i and a prediction algorithm set $F = \{f(\theta)\}$, if $\exists\ P_A \in [0, 1]$, such that $p_f \leq P_A$ for $\forall\ f \in F$, where p_f is the overall predictive accuracy of the entire sequence by f, and meanwhile $\exists\ f \in F$ such that $P_A - p_f < \varepsilon(\forall \varepsilon > 0)$, then P_A is defined as predictability for the time series X.*

The above is the formalized definition of predictability for time series $X = \{X_1, X_2, ..., X_i, ..., X_n\}$. What we want to emphasize is that predictability is an inherent feature of a time series that characterizes the ability to be predicted, in other words, the upper bound on the extent to which X can be accurately predicted. Predictability does not depend on a particular sequence prediction algorithm. A sequence is completely unpredictable while its predictability is 0 and is completely deterministic while its predictability is 1.

Problem 1 (Human Mobility Predictability). *Let $\{1, 2, ..., t, ..., T\}$ be time points in a time range, given a user's location series $X = \{P_1, P_2, ..., P_t, ..., P_T\}$ during the time period, how to quantify the overall predictability of this location series?*

Many existing human mobility prediction algorithms consider not only regularity or correlation within sequences but also external influence factors. Therefore, simply quantifying the predictability of a mobility sequence itself cannot evaluate these algorithms, which leads to the following problem.

Problem 2 (Human Mobility Predictability Considering External Factors). *Given location series $X = \{P_1, P_2, ..., P_t, ..., P_T\}$ and external factor variation series $Y = \{Y_1, Y_2, ..., Y_t, ..., Y_T\}$ during the time period $\{1, 2, ..., t, ..., T\}$, how to quantify the overall predictability of this location series?*

Note that unless otherwise specified, symbols used in this article are derived from Table I.

4 MEASURE PREDICTABILITY OF HUMAN MOBILITY

Here we present the framework for analyzing the predictability of human mobility.

4.1 Predictability without External Factors

The predictability P_A calculated by our method consists of two parts — P_S, the predictability of time series without considering exploration, which is quantified by the method proposed in [32], and P_E, the reduction in predictability due to exploration in human mobility. What is called exploration is a visit to a location that has

Table 1. Description of Notation

Symbol	Description
X	Location series reflecting human mobility
H_T	Historical sequence in time of $1 - T$
N	Size of action space
P_E	Probability of exploration
P_S	The predictability without considering exploration
P_A	The predictability considering exploration
$E(X)$	Entropy of time series X
Y	External factors
F	Set of prediction algorithms

never been visited before and therefore doesn't appear in the historical series, which represents the completely unpredictable part of predicting human mobility.

For the prediction of movement at time $t+1$, given the past history $H_t = \{X_1, X_2, ..., X_t\}$ where X_i is the user's behavior at time i, the probability distribution at time $t+1$ based on history is $P(X_{t+1} = x|H_t)$, which is the probability that the next location is x. Therefore, probability of correctly predicting is $P(X_{t+1} = \hat{X}_{t+1}|H_t)$ while the real location at time $t+1$ is $\hat{X}_{t+1}$.

Let x_{ML} be the most likely location the person is going to in the prediction at time $t+1$ given history H_t and p_{ML} be the corresponding probability. We have

$$p_{ML} = \max_x \{P(X_{t+1} = x|H_t)\} \tag{1}$$

For any prediction algorithm, $f \in F$ let $P_f(X_{t+1} = \hat{X}_{t+1}|H_t)$ be the probability distribution of locations at the next moment predicted by the algorithm f. Let $P_r(X_{t+1}|H_t)$ be the real probability distribution. Then the probability of correctly predicting is given as

$$P(X_{t+1} = \hat{X}_{t+1}|H_t) = \sum_x P_r(x|H_t) \times P_f(x|H_t) \tag{2}$$

Since $p_{ML} \geq P_t(x|H_t)$ and $\sum_x p_f(x|H_t) = 1$, we have

$$\begin{aligned} P(X_{t+1} = \hat{X}_{t+1}|H_t) &\leq \sum_x p_{ML} \times p_f(x|H_t) \\ &= p_{ML} \end{aligned} \tag{3}$$

Equation (3) shows that the accuracy of any prediction algorithm based on historical series is lower than p_{ML}, which means p_{ML} is the maximal accuracy.

Theorem 1. *Given a location series, there exist an algorithm f for that $p_{ML} - p_f < \varepsilon(\varepsilon > 0)$.*

Proof. Let us consider a predictor $\hat{f}$ — at any time step $t+1$ according to the given historical sequence H_t its prediction output is x_{ML}, which means $P_{\hat{f}}(x_{ML}|H_t) = 1$ and $P_{\hat{f}}(x|H_t) = 0 (x \neq x_{ML})$. Then the accuracy of $\hat{f}$ is given as

$$
\begin{aligned}
P_{\hat{f}}(X_{t+1} = \hat{X}_{t+1}|H_t) &= \sum_x P_r(x|H_t) \times P_{\hat{f}}(x|H_t) \\
&= p_{ML}
\end{aligned} \tag{4}
$$

This proves that at time $t+1$ given a particular history H_t, the maximal accuracy of prediction p_{ML} is theoretical achievable. Then the predictability at time $t+1$ for any history is given as

$$
P_S(t) = \sum_{H_T} p(H_t) \times p_{ML}(H_t) \tag{5}
$$

where $p(H_t)$ is the probability of a particular history H_t. Then the overall predictability P_S for any time step is calculated by

$$
P_S = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} P_S(t) \tag{6}
$$

As a direct measure of the amount of information, entropy can measure the complexity of time series from a quantitative perspective. With the chain rule, the entropy of human mobility is defined as

$$
\begin{aligned}
E(X) &= -\sum_{x\in X} p(x)\log_2 p(x) \\
&= \lim_{T\to\infty} \frac{1}{n} E(X_1, X_2, ..., X_T) \\
&= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E(X_{t+1}|H_t)
\end{aligned} \tag{7}
$$

where $p(x) = P_r\{X_t = x\}$ is the probability that $X_t = x$ given H_t.

Based on the history behavior H_t, $P_E(X_{t+1}|H_t)$ which is the probability of exploration at the time $t+1$ indicates the probability of visiting a new location that never visited before, that is, $X_{t+1} \notin H_t$. So $P_{NE}(X_{t+1}|H_t) = 1 - P_E(X_{t+1}|H_t)$ is the probability of visiting an old location, which can be calculated based on additive Markov chain [34].

$$
P_{NE}(X_{t+1}|H_t) = \sum_{i=1}^{t} AC(X_{t+1}, X_i, t+1-i) \tag{8}
$$

where $AC(X_{t+1}, X_i, t+1-i)$ is the **A**dditive **C**ontribution of X_i to $P_{NE}(X_{t+1}|H_t)$. In the First-Order Markov model, we regard the additive contribution of X_t as the transition probability of X_t to X_{t+1}. Compared with X_1, X_t has a greater impact on the movement at time $t+1$, so we give a weight coefficient $W(X_i)$ to every X_i in H_t, which represents the series delay weight with the delay rate $g \geq 0$.

$$
W(X_i) = 2^{-g\cdot(t-i)} \tag{9}
$$

In experiments, we give an appropriate value $g = 2$, which makes the results obtained by our method are closer to the theoretical predictability on simulation data.

Let $tp(X_i, X_j)$ be the transition probability from location X_i to X_j, and we get $tp(X_i, X_j)$ by the frequency appeared in history sequence.

$$
tp(X_i, X_j) = \frac{fre(X_i \to X_j)}{fre(X_i)} \tag{10}
$$

$$
AC(X_{t+1}, X_i, t+1-i) = \frac{W(X_i) \cdot tp(X_i, X_{t+1})}{\sum_{i=1}^{t} W(X_i)} \tag{11}
$$

By uniting equations (8) and (11), we get

$$\begin{aligned} P_{NE}(X_{t+1}|H_t) &= \sum_{i=1}^{t} AC(X_{t+1}, X_i, t+1-i) \\ &= \sum_{i=1}^{t} \frac{W(X_i) \cdot tp(X_i, X_{t+1})}{\sum_{i=1}^{t} W(X_i)} \end{aligned} \tag{12}$$

To get P_E on the entire time series, we take the average of probabilities of exploration of every step.

$$\begin{aligned} P_E &= \frac{1}{T-1} \sum_{t=1}^{T-1} P_E(X_{t+1}|H_t) \\ &= \frac{1}{T-1} \sum_{t=1}^{T-1} (1 - P_{NE}(X_{t+1}|H_t)) \end{aligned} \tag{13}$$

Algorithm 1 Exploration Probability Computation

Require: $H_T = \{X_1, X_2, ..., X_T\}, g$
Ensure: $P_E(X_{t+1}|H_t)$
set $M = 0$, $D = 0$, $P_E = 0$
for $t = 1, ..., T-1$ **do**
  for each $X_i \in H_t$ **do**
    $tp(X_i, X_{t+1}) = fre(X_i \rightarrow X_{t+1}) / fre(X_i)$
    $w = 2^{-g \cdot (t-i)}$
    $M += w \cdot tp(X_i, X_{t+1})$
    $D += w$
  end for
  $P_{NE} = M/D$
  $P_E += (1 - P_{NE})$
end for
$P_E = P_E/(T-1)$
return P_E

Next, we calculate the probability P_S, which represents the upper bound of successfully predicting human movement for any predictor without considering exploration.

Given a history H_t, the real probability distribution of the next location $P_r(X_{t+1}|H_t)$ can be expressed as

$$P_r(X_{t+1}|H_t) = (p(X_1), p(X_2), ..., p_{ML}, ..., p(X_N)) \tag{14}$$

where N is the size of the action space. Then, we create a new distribution $P_t(X_{t+1}|H_t)$ where all locations are uniformly distributed except p_{ML}.

$$P_t(X_{t+1}|H_t) = (\frac{1-p_{ML}}{N}, \frac{1-p_{ML}}{N}, ..., p_{ML}, ..., \frac{1-p_{ML}}{N}) \tag{15}$$

According to the property of information entropy, the more uniform the distribution is (the more scattered the values are), the higher the entropy is. Therefore we have

$$E_r(X_{t+1}|H_t) \leq E_t(X_{t+1}|H_t) \tag{16}$$

To simplify the formula, let $p = p_{ML}(H_T)$. The entropy of the new distribution we create at time $t + 1$ given H_t can be calculated as

$$\begin{aligned} E_t(X_{t+1}|H_t) &= -\sum_{x \in X} p_x \times \log_2(p_x) \\ &= -p\log_2 p - \sum \frac{1-p}{N-1}\log_2(\frac{1-p}{N-1}) \\ &= -p\log_2 p - (1-p)\log_2(\frac{1-p}{N-1}) \\ &= -[p\log_2 p + (1-p)\log_2(1-p)] \\ &+ (1-p)\log_2(N-1) \end{aligned} \tag{17}$$

Now we define $f(p)$ as

$$f(p) = -[p\log_2 p + (1-p)\log_2(1-p)] + (1-p)\log_2(N-1) \tag{18}$$

Then we have

$$E_r(X_{t+1}|H_t) \leq E_t(X_{t+1}|H_t) = f(p) \tag{19}$$

which is satisfied for the particular history H_t. So the entropy at time $t + 1$ for any history is given as

$$\begin{aligned} E_r(t) &= \sum_{H_t} p(H_t) \times E_r(X_{t+1}|H_t) \\ &\leq \sum_{H_t} p(H_t) \times f(p) \\ &\leq f(\sum_{H_t} p(H_t) \times p) \\ &= f(\sum_{H_t} p(H_t) \times p_{ML}(H_t)) \\ &= f(P_S(t)) \end{aligned} \tag{20}$$

where we're using Jensen's inequality and the concave property of $f(p)$. Similarly, we continue to get the entropy of the entire location series $E(X)$.

$$\begin{aligned} E(X) &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E_r(t) \\ &\leq \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} f(P_S(t)) \\ &\leq f(\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} P_S(t)) \\ &= f(P_S) \end{aligned} \tag{21}$$

There exists ${P_S}^{up}$ for which the equation below holds:

$$E(X) = f({P_S}^{up}) \leq f(P_S) \tag{22}$$

As the function f monotonically decreases with P_S, we have

$${P_S}^{up} \geq P_S \tag{23}$$

So the solution ${P_S}^{up}$ of (24) is the upper bound of predictability P_S.

$$\begin{aligned} E(X) &= f({P_S}^{up}) \\ &= -{P_S}^{up}\log_2({P_S}^{up}) - (1-{P_S}^{up})\log_2(1-{P_S}^{up}) \\ &+ (1-{P_S}^{up})\log_2(N-1) \end{aligned} \tag{24}$$

As mentioned before, P_E represents the completely unpredictable part of predicting human mobility, which is the reduction in predictability when considering exploration in human mobility. Then we can get a tighter upper bound of predictability ${P_A}^{up}$ by

$${P_A}^{up} = {P_S}^{up} - P_E \tag{25}$$

Based on the formulas above, given a historical series of human mobility, we can calculate the upper bound of its predictability.

4.2 Predictability Considering External Factors

The predictability without considering external influencing factors can only evaluate those prediction algorithms utilizing regularity or correlation within sequences, such as Markov Chain and autoregressive model. Actually, many prediction algorithms proposed, such as Neural Network [39], utilize the information of external factors (e.g. weather, holiday) to improve the accuracy of sequence prediction further. Therefore, a noticeable problem is how to quantify the predictability considering external factors.

The definition of predictability considering external factors is the upper bound of the accurate rate that a prediction algorithm utilizing external factor variation data can achieve theoretically.

We have explained that the predictability of time series is directly related to information entropy as it can capture the uncertainty of time series. The purpose of taking external factors into consideration is to complement original information. For example, a fisherman goes to the sea on sunny days and rests on rainy days. Considering external factors would reduce uncertainty, and therefore increase predictability. Existing work has also analyzed the predictability of adding external factors, which was as contextual information to reduce the uncertainty of the time series. Actually, researchers have studied this issue from different aspects. For example, Context-Transition Entropy is used to calculate predictability by considering external factors[47], which increases the certainty of predictability of time series, it is larger than the upper bound of predictability proposed by song et al[32]. In this paper, the exploration behavior of human movement is considered, and the exploration rate of humans is added to the calculation of predictability, without changing the method of calculating the entropy of predictability. Therefore, our method makes the predictability upper bound tighter. When considering the predictability of external factors, partly researchers focus primarily on the routine component with the goal of showing that there are patterns in one's hard-to-predict [9]. In order to quantify predictability, this paper creates a baseline sequence, by comparing it with the person's actual mobility trace to obtain the deviation of predictability between the original sequence and the baseline. The deviation refers to the impact of the exploration part of the human routine on predictability. However, in real life, human mobility is not entirely affected by exploration which is the probability of exploration at the time indicates the probability of visiting a new location that never visited before, but may also be due to other factors, such as weather, holidays, and so on. We incorporate its probability into the calculated predictability of the next position, using a higher-order Markov chain to obtain this value. There are also different ways to deal with the time series which adding external factors, it is not incorporate external factors directly into the formula when calculating predictability [33], Instead, using sequence-splitting and sequence-merging methods to process sequences, but we directly add Mutual Information to calculate the influence of external factors on predictability. Because adding external factors according to experience will increase predictability, the original method of predictability has expanded the upper bound of prediction, so the increase of exploration makes it closer to real human movement.

To measure the extent of uncertainty reduced by external factors, we are going to introduce Mutual Information (MI). In probability theory and information theory, MI quantifies the "amount of information" obtained about a random variable through observing another random variable. In other word, MI is the decreasing quantity of entropy of a random variable knowing another one.

Based on the chain rule of information entropy, the mutual information of random variables$X$ and Y is

$$\begin{aligned} I(X;Y) &= E(X) - E(X|Y) \\ &= E(X) + E(Y) - E(X,Y) \end{aligned} \tag{26}$$

where $E(X,Y)$ is joint entropy. Given a time series of external factor Y, the conditional entropy of location sequence X is

$$\begin{aligned} E(X|Y) &= E(X) - I(X;Y) \\ &= E(X,Y) - E(Y) \end{aligned} \tag{27}$$

By replacing $E(X)$ in (24) with $E(X|Y)$, we get the quantitative formula for predictability with external factors:

$$\begin{aligned} E(X|Y) &= f({P_S}^{up}) \\ &= -{P_S}^{up}\log_2({P_S}^{up}) - (1-{P_S}^{up})\log_2(1-{P_S}^{up}) \\ &+ (1-{P_S}^{up})\log_2(N-1) \end{aligned} \tag{28}$$

Note that exploration should also be considered. So the upper bound of predictability with external factors is calculated by uniting (25) and (28).

5 EXPERIMENTS

In this section, we will introduce the datasets, settings and results of our experiments. In addition, we discuss the evaluation results from some typical predictors.

5.1 Datasets

We do experiments on simulation data and three real-world datasets. Here we firstly introduce the generation method of simulation data, and then introduce the three real-world datasets.

The simulation data we generate is a Markov chain, where the agent chooses the location to visit at each time step with nearly the same probability distribution, which ensures that the theoretical predictability of the sequence generated is a definite value we preset. Our generation model has three inputs — the size of action space m (which is the number of locations available at each step), the length of the sequence generated n and the predictability preset p ($p \geq 1/m$). At each step, the agent visit x_{ML} with probability p and visit other $m-1$ locations with equal probability $(1-p)/(m-1)$. Note that x_{ML} at each step is determined by a deterministic rule — for example, let all the locations in action space be arranged in number order to be x_{ML} in turn. According to our definition of predictability, the theoratical predictability of the simulation sequence generated by this method is p. Compared with real-world data, simulation data has the advantage that theoretical predictability is known.

All three real-world datasets we adopted can reflect human mobility, two of which are based on Call Detail Records (CDRs) and the other one is based on GPS location information.

The first CDR dataset [3] is from Data for Development Challenge (D4D), collected in Côte d'Ivoire from 50,000 mobile phone users during 150 days. To protect user privacy, continuous records of the same user are limited to a maximum of two weeks. So we use a part of data from March 26,2012 to April 8,2012. Each record contains the user ID, timestamp and the ID of base station used for this communication. To improve data quality, we deleted data from users with too little action space ($N \leq 2$) and too little frequency of communication ($f < 0.5 calls/hour$). Then we rebuilt the dataset into hourly fixed-frequency continuous sampling data. If the sampling moment is not recorded, the position record of the previous moment is used, which means the user is set to be stationary by default.

The second CDR dataset [41] is mobile phone App network request data from Tsinghua University, which include anonymous cellular data captured from 1000 users by the Deep Packet Inspection (DPI) device during a week. Each record contains the user ID, timestamp, and the ID of the base station used for this network connection. Similarly, we resample the data with a frequency of 2 records/minute and assume the user still lacks a record.

Finally, we analyze the GPS-based dataset [48] collected in the Geolife project from April 2007 to August 2012. The dataset contains 182 individual trajectories and each trajectory contains the latitude, altitude, and timestamp of the collected data point. The majority of data in this dataset were collected in Beijing, and the remaining are from 36 cities in China as well as a few cities in the USA, South Korea, and Japan. Some individuals' trajectories are so short that we cannot get their statistical features, so we only consider users with enough track points ($n \geq 20$).

The trajectory information in the dataset is the latitude and longitude collected by GPS. In order to obtain a user's location sequence and to examine the user's mobility at the grid scale, we quantize the area into small discrete grid cells. We used a grid to rasterize the map of Beijing, and we tagged the label of each cell from 1 to 1200, as shown in Fig. 1.

What's more, we collected the holiday information of each timestamp and the weather conditions such as sunny and rainy of each cell in order to study the influence of external factors on predictability.

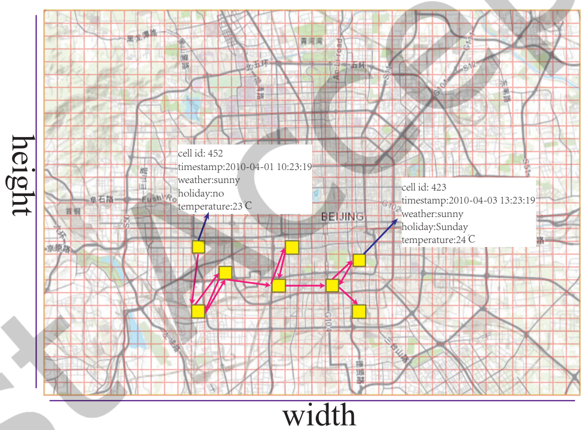

Fig. 1. Mobile behavior series: The map is partitioned into cells by a $width * height$ grid, and a user's longitude and latitude coordinates are mapped to a cell ID to get the user behavior sequence. At the same time, the corresponding weather information, holiday information, temperature information, and other external factors are obtained.

5.2 Experimental Settings

We conduct our experiments from the following aspects:

- Calculate the predictability of human mobility using our proposed method on simulation data and three real-world datasets, and compare the performance with baseline methods for quantifying predictability to explain that our method gives a tighter and more realistic upper bound of predictability.
- Calculate the predictability considering external factors (such as weather and holidays) to discuss the influence of external factors and to demonstrate the extensibility of our method on external factors.

- Compare our predictability with the accuracy of the state-of-the-art predictors on human mobility series to prove its basic reasonableness.

5.2.1 Baseline Methods for Comparison. Our method which quantifies the Predictability of human mobility Considering the Probability of Exploration (*PCPE*), is compared with four state-of-the-art methods for calculating predictability as follows:

- The method based on the Real Entropy and Fano's inequality in information theory (*RE-based*) proposed in [32], which is a classic method that firstly derives a rigorous quantitative relationship between predictability and entropy.
- The method based on the Real Entropy gives a refined upper bound on the predictability of human mobility (*RE-refined*) proposed in [31], which is a typical improvement on the above method. This method gives a more precise result to some extent by optimizing the measurement of action space.
- The method based on Multi-Scale Entropy (*MSE-based*) proposed in [37], which replaces the real entropy in the first baseline method with multi-scale entropy. This method has good performance in quantifying the predictability of travel time. The scale factor τ is set to 1, the embedding dimension m is set to 2 and the tolerance r is the standard deviation of sequence multiplied by 0.2 in our experiments.
- The method based on Permutation Entropy (*PE-based*) used in [29], [26], which captures the complexity between local sequences and the predictability of sequence is quantified by one subtract the complexity. The window size d is set to 3 and delay τ is set to 1 in our experiments.

5.2.2 Human Mobility Predictors. According to **Definition 1**, predictability is the upper limit of the probability that a sequence can be accurately predicted, and therefore it can be used to evaluate the performance of a prediction algorithm. Conversely, the accuracy of good prediction algorithms on this sequence can prove the basic reasonableness of predictability quantified to some extent. We use four different time-series prediction algorithms to evaluate our method. The details of these algorithms are as follows:

- Auto Regressive Integrated Moving Average model (ARIMA) can be used to forecast future values based on its own past values.
- A Markov Random Field (MRF) is an undirected probabilistic graphical model representing random variables and their conditional dependencies.
- The Recurrent Neural Network (RNN) is a class of neural networks that allows previous outputs to be used as inputs while having hidden states, and can be used to predict the next state based on a history sequence.
- Long-Short-Term-Memory network (LSTM) is a kind of time recurrent neural network, which is suitable for processing and predicting with very long intervals and delays in time series.

We have adjusted the parameters of these prediction algorithms based on the dataset used so that they can achieve the best prediction performance.

5.3 Results

5.3.1 Effectiveness of Predictability Measurement. Fig. 2 shows the result of the simulation experiment. We generate simulation data of eight different sequence lengths, which are 2^4 $2^1$1. As previously mentioned, the generated sequences are Markov chains whose theoretical predictability is a fixed value that can be preset. Without loss of generality, we set the theoretical predictability to 0.45, which is represented by the black straight line in the figure. The dots in different colors represent the results obtained by different methods as shown in the legend. Each value is the average of the results of 20 repetitions. Note that the predictability quantified by our method (green dots) is closest to the theoretical value, which indicates that our method is more accurate than others.

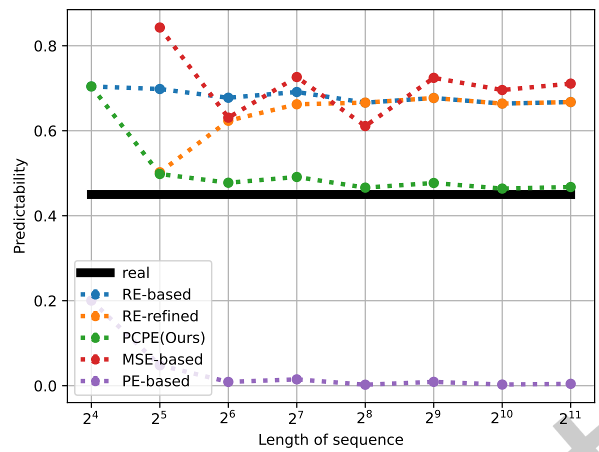

Fig. 2. The values of predictability calculated by different methods at different sequence lengths on simulation data, which consists of generated Markov chains whose theoretical predictability is 0.45. Each value is the average of the results of 20 repetitions.

In addition, the method based on Permutation Entropy (*PE-based*) gives results that deviate greatly from the theoretical values and tend to zero as the sequence length increases. In view of this, we believe that this method is not suitable for quantifying the predictability of human mobility and therefore only compare our method with the other three baseline methods on the real-world datasets.

The probability distributions of predictability on D4D dataset, Tsinghua dataset and Geolife dataset are shown in Fig. 3 respectively Fig. (a), Fig. (b), and Fig. (c). For any predictability value P, the higher the y value is in these figures, the more users there are whose predictability of mobility is P. In previous work, the predictability value with the highest probability was usually taken as the predictability of the whole dataset [32]. In this sense, our method gives the lowest result on every dataset. So we can conclude that our method quantifies a tighter upper bound of predictability of human mobility.

Note that our method is the combination of the *RE-based* method and our quantifying for the probability of exploration. Therefore, we can see intuitively how exploration reduces predictability by comparing the results obtained by the *RE-based* method and our method. As can be seen from the figures, when exploration probability is taken into account, predictability decreases by 5.1% on D4D dataset, 16% on Tsinghua dataset and 5.7% on Geolife dataset.

5.3.2 Influence of External Factors. With the development of deep learning, more and more prediction algorithms begin to consider the influence of external factors, such as weather, which will greatly affect people's travel. As a result, the application of simply measuring the predictability of the time series itself is limited. By considering the influence of external factors on the predictability of human mobility, it can be used to evaluate the quality of relevant prediction algorithms. After capturing weather, holiday, and temperature information, we can obtain predictability with external factors by (28).

We conduct the experiments using the *RE-based* method and our method *PCPE*. As shown in Fig. 4, the predictability calculated by our method increases 1%, 0.7%, and 0.5% respectively while considering weather,

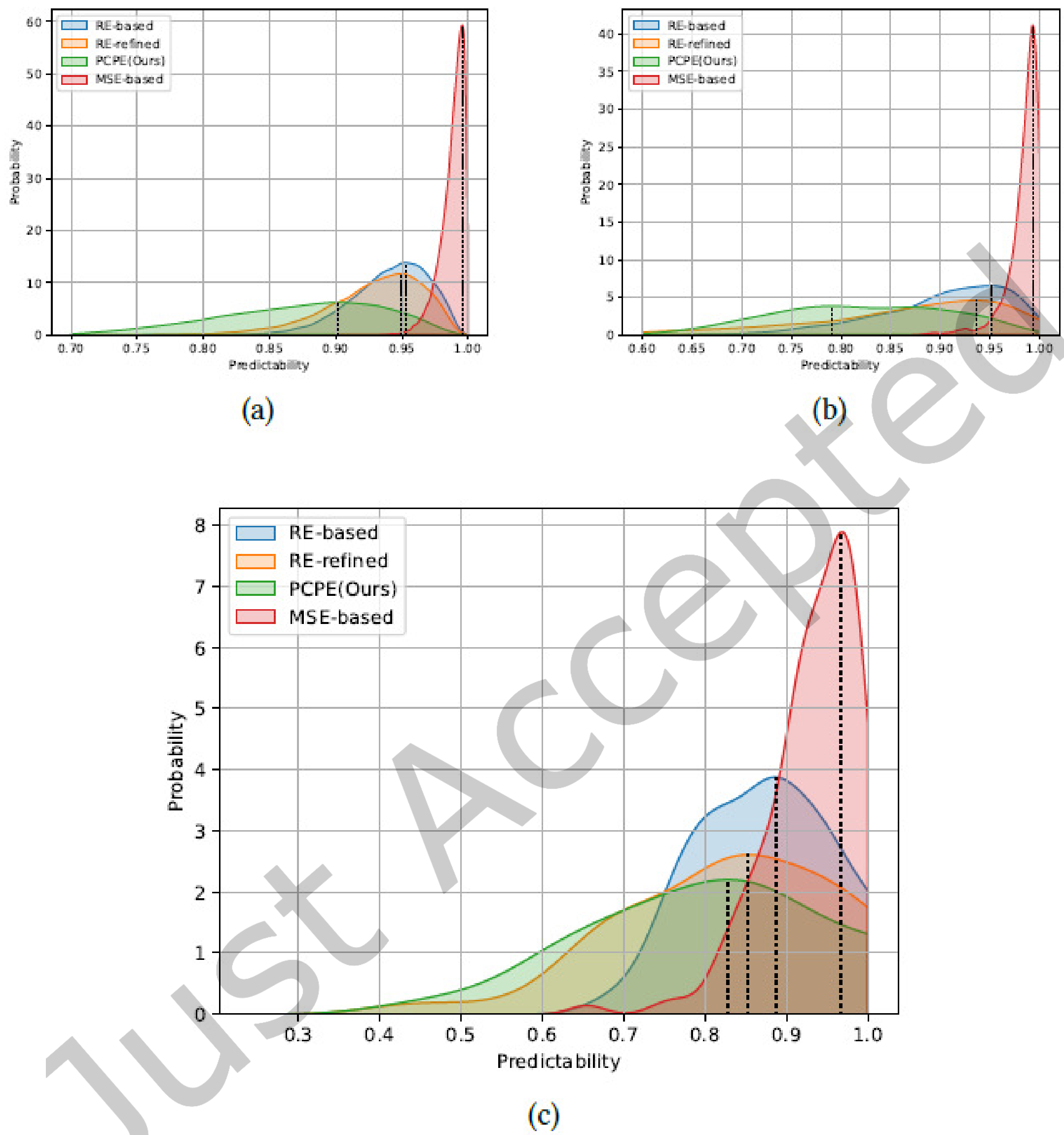

Fig. 3. Fig(a) shows the result of the probability distributions of predictability measured on the D4D dataset. The predictability given by our method is 90.2% while 95.3%, 94.9%, and 99.6% are given by the other three methods. Fig(b) shows the result of the probability distributions of predictability measured on Tsinghua dataset. Fig(c) shows the result of the predictability given by our method is 79.1% while 95.1%, 93.6%, and 99.3% are given by the other three methods. The probability distributions of predictability were measured on Geolife dataset. The predictability given by our method is 82.8% while 88.5%, 85.1%, and 96.7% are given by the other three methods.

holiday, and temperature. The predictability is improved by 1.5% with all external factors, which is also indicated by Table II.

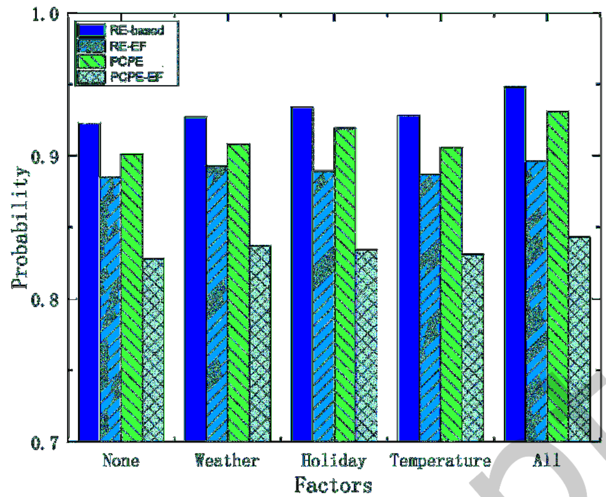

Fig. 4. The predictability of integrating external factors. We integrate our MI-based method to consider external factors into the RE-based method and compare it with our proposed method. We can intuitively see the influence of external factors on predictability in the figure.

Table 2. Predictor Accuracy and Behavior Predictability

	Accuracy				Predictability(P_A)			
	ARIMA	*MRF*	*RNN*	*LSTM*	*MSE-based*	*RE-refined*	*RE-based*	*PCPE(Ours)*
No Factors	0.641	0.713	0.746	0.753	0.967	0.851	0.885	0.828
With Factors	0.701	0.819	0.795	0.806	0.975	0.863	0.896	0.843

5.3.3 Evaluate Predictability with Predictors. Besides the comparison of predictability methods, we also evaluate our method by comparing the predictability to the accuracy of other human mobility predictors, i.e. *ARIMA*, *MRF*, *RNN* and *LSTM*. We have trained the model to obtain parameters that make the algorithm achieve the best performance.

Table II shows the accuracy and predictability measured by different methods. We can see that Whether external factors are considered or not, the predictability calculated by the method we proposed is higher than the accuracy of the state-of-the-art predictors, which illustrates the validity and reasonableness of predictability proposed for one side.

6 CONCLUSION

The predictability of human mobility is a fundamental characteristic of the movement sequence. In this paper, we have proposed a model based on additive Markov chains to measure the probability of exploration and developed a method for quantifying mobility predictability. We use mutual information to measure the reduction in uncertainty caused by external factors, and further propose a calculation method for predictability considering external factors. Experiments on simulation data and three real-world datasets proved that our result is a tighter and more accurate upper bound of predictability of human mobility. On the other hand, the incorporation of external factors increased predictability by reducing the uncertainty of behavior (caused a 1.5% raise in predictability). In future, we will consider the influence of other characteristics of human mobility on predictability. How to apply our methods to other time series and how to use predictability to guide the design of prediction algorithms are open problems.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (No. 61960206008, 62032020), and the National Science Fund for Distinguished Young Scholars (No. 62025205).

REFERENCES

[1] Daniel Austin, Robin M Cross, Tamara Hayes, and Jeffrey Kaye. 2014. Regularity and predictability of human mobility in personal space. *PloS one* 9, 2 (2014), e90256.

[2] F. Bartoli, G. Lisanti, L. Ballan, and A. Del Bimbo. 2018. Context-Aware Trajectory Prediction. In *2018 24th International Conference on Pattern Recognition (ICPR)*. 1941–1946. https://doi.org/10.1109/ICPR.2018.8545447

[3] Vincent D Blondel, Markus Esch, Connie Chan, Fabrice Clérot, Pierre Deville, Etienne Huens, Frédéric Morlot, Zbigniew Smoreda, and Cezary Ziemlicki. 2012. Data for development: the d4d challenge on mobile phone data. *arXiv preprint arXiv:1210.0137* (2012).

[4] Eunjoon Cho, Seth A Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1082–1090.

[5] Andrea Cuttone, Sune Lehmann, and Marta C. González. 2018. Understanding predictability and exploration in human mobility. *Epj Data Science* 7, 1 (2018), 2.

[6] Weizhen Dang, Haibo Wang, Shirui Pan, Pei Zhang, Chuan Zhou, Xin Chen, and Jilong Wang. 2022. Predicting Human Mobility via Graph Convolutional Dual-attentive Networks. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 192–200.

[7] Timothy DelSole and Michael K Tippett. 2007. Predictability: Recent insights from information theory. *Reviews of Geophysics* 45, 4 (2007).

[8] Francis X Diebold and Lutz Kilian. 2001. Measuring predictability: theory and macroeconomic applications. *Journal of Applied Econometrics* 16, 6 (2001), 657–669.

[9] Douglas do Couto Teixeira, Jussara M Almeida, and Aline Carneiro Viana. 2021. On estimating the predictability of human mobility: the role of routine. *EPJ Data Science* 10, 1 (2021), 49.

[10] M. Fouladgar, M. Parchami, R. Elmasri, and A. Ghaderi. 2017. Scalable deep traffic flow neural networks for urban traffic congestion prediction. In *2017 International Joint Conference on Neural Networks (IJCNN)*. 2251–2258. https://doi.org/10.1109/IJCNN.2017.7966128

[11] R. Fu, Z. Zhang, and L. Li. 2016. Using LSTM and GRU neural network methods for traffic flow prediction. In *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. 324–328. https://doi.org/10.1109/YAC.2016.7804912

[12] Huiji Gao, Jiliang Tang, and Huan Liu. 2012. Mobile location prediction in spatio-temporal context. In *Nokia mobile data challenge workshop*, Vol. 41. 1–4.

[13] Joshua Garland, Ryan James, and Elizabeth Bradley. 2014. Quantifying time-series predictability through structural complexity. *CoRR, abs/1404.6823* (2014).

[14] Cathy Hohenegger and Christoph Schär. 2007. Predictability and error growth dynamics in cloud-resolving models. *Journal of the Atmospheric Sciences* 64, 12 (2007), 4467–4478.

[15] Tanvi Jindal, Prasanna Giridhar, Lu-An Tang, Jun Li, and Jiawei Han. 2013. Spatiotemporal periodical pattern mining in traffic data. In *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*. 1–8.

[16] Xiangjie Kong, Kailai Wang, Mingliang Hou, Feng Xia, Gour Karmakar, and Jianxin Li. 2022. Exploring human mobility for multi-pattern passenger prediction: a graph learning framework. *IEEE Transactions on Intelligent Transportation Systems* 23, 9 (2022), 16148–16160.

[17] V Krishnamurthy. 2019. Predictability of weather and climate. *Earth and Space Science* 6, 7 (2019), 1043–1056.

[18] Juha K Laurila, Daniel Gatica-Perez, Imad Aad, Olivier Bornet, Trinh-Minh-Tri Do, Olivier Dousse, Julien Eberle, Markus Miettinen, et al. 2012. *The mobile data challenge: Big data for mobile computing research.* Technical Report.

[19] Yan Leng, Dominiquo Santistevan, and Alex Pentland. 2021. Understanding collective regularity in human mobility as a familiar stranger phenomenon. *Scientific Reports* 11, 1 (2021), 19444.

[20] Xiaolong Li, Gang Pan, Zhaohui Wu, Guande Qi, Shijian Li, Daqing Zhang, Wangsheng Zhang, and Zonghui Wang. 2012. Prediction of urban human mobility using large-scale taxi traces and its applications. *Frontiers of Computer Science* 6, 1 (2012), 111–121.

[21] Xin Lu, Erik Wetter, Nita Bharti, Andrew J Tatem, and Linus Bengtsson. 2013. Approaching the limit of predictability in human mobility. *Scientific reports* 3 (2013), 2923.

[22] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Wang. 2015. Traffic Flow Prediction With Big Data: A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems* 16, 2 (2015), 865–873. https://doi.org/10.1109/TITS.2014.2345663

[23] Eduardo Mucelli Rezende Oliveira, Aline Carneiro Viana, Carlos Sarraute, Jorge Brea, and Ignacio Alvarez-Hamelin. 2016. On the regularity of human mobility. *Pervasive and Mobile Computing* 33 (2016), 73–90.

[24] Tim N Palmer. 2000. Predicting uncertainty in forecasts of weather and climate. *Reports on progress in Physics* 63, 2 (2000), 71.

[25] Zhao Pei, Xiaoning Qi, Yanning Zhang, Miao Ma, and Yee-Hong Yang. 2019. Human trajectory prediction in crowded scene using social-affinity Long Short-Term Memory. *Pattern Recognition* 93 (2019), 273–282.

[26] Frank Pennekamp, Alison C Iles, Joshua Garland, Georgina Brennan, Ulrich Brose, Ursula Gaedke, Ute Jacob, Pavel Kratina, Blake Matthews, Stephan Munch, et al. 2019. The intrinsic predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs* 89, 2 (2019), e01359.

[27] Adam Sadilek and John Krumm. 2012. Far out: Predicting long-term human mobility. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 26.

[28] Afees A Salisu, Kazeem Isah, and Lateef O Akanni. 2019. Improving the predictability of stock returns with Bitcoin prices. *The North American Journal of Economics and Finance* 48 (2019), 857–867.

[29] Samuel V Scarpino and Giovanni Petri. 2019. On the predictability of infectious disease outbreaks. *Nature communications* 10, 1 (2019), 1–8.

[30] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. 2018. A Comparison of ARIMA and LSTM in Forecasting Time Series. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*.

[31] Gavin Smith, Romain Wieser, James Goulding, and Duncan Barrack. 2014. A refined limit on the predictability of human mobility. In *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 88–94.

[32] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010. Limits of predictability in human mobility. *Science* 327, 5968 (2010), 1018–1021.

[33] Douglas Do Couto Teixeira, Aline Carneiro Viana, Jussara M. Almeida, and Mrio S. Alvim. 2021. The Impact of Stationarity, Regularity, and Context on the Predictability of Individual Human Mobility. *ACM Trans. Spatial Algorithms Syst.* 7, 4, Article 19 (jun 2021), 24 pages. https://doi.org/10.1145/3459625

[34] Vadym E. Vekslerchik, Sergiy S. Melnik, Galyna. M. Pritula, and Oleg V. Usatenko. 2018. Correlation Properties of Additive Linear High-Order Markov Chains. In *2018 9th International Conference on Ultrawideband and Ultrashort Impulse Signals (UWBUSIS)*.

[35] Yingzi Wang, Nicholas Jing Yuan, Defu Lian, Linli Xu, Xing Xie, Enhong Chen, and Yong Rui. 2015. Regularity and conformity: Location prediction using heterogeneous mobility data. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 1275–1284.

[36] Paiheng Xu, Likang Yin, Zhongtao Yue, and Tao Zhou. 2019. On predictability of time series. *Physica A: Statistical Mechanics and its Applications* 523 (2019), 345–351.

[37] Tao Xu, Xianrui Xu, Yujie Hu, and Xiang Li. 2017. An entropy-based approach for evaluating travel time predictability based on vehicle trajectory data. *Entropy* 19, 4 (2017), 165.

[38] Ming Yan, Shuijing Li, Chien Aun Chan, Yinghua Shen, and Ying Yu. 2021. Mobility prediction using a weighted Markov model based on mobile user classification. *Sensors* 21, 5 (2021), 1740.

[39] Fei Yi, Zhiwen Yu, Fuzhen Zhuang, and Bin Guo. 2019. Neural Network based Continuous Conditional Random Field for Fine-grained Crime Prediction.. In *IJCAI*. 4157–4163.

[40] Fei Yi, Zhiwen Yu, Fuzhen Zhuang, Xiao Zhang, and Hui Xiong. 2018. An integrated model for crime prediction using temporal and spatial factors. In *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 1386–1391.

[41] Donghan Yu, Yong Li, Fengli Xu, Pengyu Zhang, and Vassilis Kostakos. 2018. Smartphone app usage prediction using points of interest. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–21.

[42] Zhiwen Yu, Huang Xu, Zhe Yang, and Bin Guo. 2015. Personalized travel package with multi-point-of-interest recommendation based on crowdsourced user footprints. *IEEE Transactions on Human-Machine Systems* 46, 1 (2015), 151–158.

[43] Zhiwen Yu, Fei Yi, Qin Lv, and Bin Guo. 2018. Identifying on-site users for social events: Mobility, content, and social relationship. *IEEE Transactions on Mobile Computing* 17, 9 (2018), 2055–2068.

[44] Quan Yuan, Wei Zhang, Chao Zhang, Xinhe Geng, Gao Cong, and Jiawei Han. 2017. PRED: Periodic region detection for mobility modeling of social media users. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. 263–272.

[45] G. Udny Yule. 1927. On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers. *Philosophical Transactions of The Royal Society B Biological Sciences* 226, 636–646 (1927), 267–298.

[46] Chao Zhang, Keyang Zhang, Quan Yuan, Haoruo Peng, Yu Zheng, Tim Hanratty, Shaowen Wang, and Jiawei Han. 2017. Regions, periods, activities: Uncovering urban dynamics via cross-modal representation learning. In *Proceedings of the 26th International Conference on World Wide Web*. 361–370.

[47] Chao Zhang, Kai Zhao, and Meng Chen. 2022. Beyond The Limits of Predictability in Human Mobility Prediction: Context-Transition Predictability. *IEEE Transactions on Knowledge and Data Engineering* (2022), 1–1. https://doi.org/10.1109/TKDE.2022.3148300

[48] Yu Zheng, Xing Xie, Wei-Ying Ma, et al. 2010. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.* 33, 2 (2010), 32–39.

[49] Eric Zivot and Jiahui Wang. 2006. Vector autoregressive models for multivariate time series. *Modeling financial time series with S-PLUS®* (2006), 385–429.