




# Putting human behavior predictability in context

Wanyi Zhang<sup>1\*</sup> , Qiang Shen<sup>3</sup>, Stefano Teso<sup>1</sup>, Bruno Lepri<sup>4</sup>, Andrea Passerini<sup>1</sup>, Ivano Bison<sup>2</sup> and Fausto Giunchiglia<sup>1,3</sup>

\*Correspondence:

[wanyi.zhang@unitn.it](mailto:wanyi.zhang@unitn.it)

<sup>1</sup>Department of Information Engineering and Computer Science, University of Trento, Trento, Italy  
Full list of author information is available at the end of the article

## Abstract

Various studies have investigated the predictability of different aspects of human behavior such as mobility patterns, social interactions, and shopping and online behaviors. However, the existing researches have been often limited to a single or to the combination of few behavioral dimensions, and they have adopted the perspective of an outside observer who is unaware of the motivations behind the specific behaviors or activities of a given individual. The key assumption of this work is that human behavior is deliberated based on an individual's own perception of the situation that s/he is in, and that therefore it should also be studied under the same perspective. Taking inspiration from works in ubiquitous and context-aware computing, we investigate the role played by four contextual dimensions (or modalities), namely time, location, activity being carried out, and social ties, on the predictability of individuals' behaviors, using a month of collected mobile phone sensor readings and self-reported annotations about these contextual modalities from more than two hundred study participants. Our analysis shows that any target modality (e.g. location) becomes substantially more predictable when information about the other modalities (time, activity, social ties) is made available. Multi-modality turns out to be in some sense fundamental, as some values (e.g. specific activities like "shopping") are nearly impossible to guess correctly unless the other modalities are known. Subjectivity also has a substantial impact on predictability. A location recognition experiment suggests that subjective location annotations convey more information about activity and social ties than objective information derived from GPS measurements. We conclude the paper by analyzing how the identified contextual modalities allow to compute the diversity of personal behavior, where we show that individuals are more easily identified by rarer, rather than frequent, context annotations. These results offer support in favor of developing innovative computational models of human behaviors enriched by a characterization of the context of a given behavior.

**Keywords:** Human behavior; Personal context; Predictability; Diversity

## 1 Introduction

In the last decade, several works have investigated the role of randomness in human behavior and how predictable are various aspects of human activities such as mobility [1–8], social interactions [9–12], shopping [13, 14], and online [15] behaviors.

© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Research studies have also highlighted how similar mechanisms seem to govern different human activities. For example, people show a finite number of favourite places [6] and friends [11]. In a similar way, some individuals tend to explore and change favourite places [16] over time, as they do with friendships [11] and mobile phone apps [17], while others tend to maintain stable their behavior.

However, existing studies on human dynamics have been often limited to a single or to the combination of few behavioral dimensions (e.g. mobility and social interactions) [2, 3, 18–20]. Moreover, these studies have adopted the perspective of an outside observer who is unaware of the motivations behind the activities of a given individual.

In our work, we propose a different angle for analyzing the predictability of human behavior. In particular, our study revolves around the observation that, in typical circumstances, human behavior is deliberated based on an individual's own perception of the situation s/he is involved in, as captured by the notion of *personal context* [21–23]. For this reason, we analyze *regularity* and *diversity* in behavior through the joint interplay of four modalities of personal context (i.e. *time*, *location*, *activity*, and *social ties*) widely used in context-aware and ubiquitous computing communities [21–25].

In particular, we perform a rigorous statistical analysis of the effects of these four modalities of personal context on the predictability of human behavior using a month of collected mobile phone sensor readings and self-reported annotations about time, location, activity and social ties from more than 200 volunteers [26, 27]. Our analysis leverages information theoretic techniques introduced by studies on human mobility [3, 15, 28] to characterize the predictability of individual behavior for single modalities and extends them to study correlations across distinct modalities. In addition, we look at behavior diversity across individuals through the lens of the four identified contextual modalities.

Our analyses and findings offer several pieces of evidence in support of the role played by the investigated contextual modalities. As a first step, we have estimated the performance of an *ideal, optimal* classifier for independently predicting each modality (i.e. time, location, activity, and social ties) for each individual in the data set. This showed that an optimal classifier with access to the previous annotations *for the same user and contextual modality*, but not their chronological order, cannot do better than 45% to 65% accuracy. In other words, ignoring correlations across time and between contextual modalities entails a large *irreducible error* of 35% to 55%, depending on the target modality. Disclosing the order of past annotations (again, available for the target modality only) makes the optimal classifier performs much better and the irreducible error decreases to 10% to 15%. However, supplying the optimal classifier with information about the other modalities (e.g. providing time, activity, and social ties while predicting location) but not their order decreases the irreducible error even more, below 5%. This shows that taking inter-modality correlations into account makes a substantial difference in the predictability of an individual behavior and supports the idea that inter-modality correlations may be more important than short- and long-term correlations over time. These results, which hold for *optimal* classifiers, were shown to carry over to practical classifiers (namely, Random Forests) in a location recognition experiment. This experiment also shows that some locations that are hard or impossible to predict using sensor data suddenly become easy to predict when information from time, activity, and social ties is taken into account. This further highlights the fundamental importance of the jointly interplay of different contextual modalities in behavior analysis.

Then, the analysis was extended to determine the impact of subjectivity on predictability of behavior. Consistently with our finding that activity and location are strongly tied, we compared the impact of injecting objective versus subjective location information on the performance of optimal classifiers for activity and social ties. Here, subjective location was implemented using self-reported annotations, while objective location was derived from GPS measurements. Our results support the argument that subjective location is far more informative than objective location for predicting behavior.

In a final experiment, we investigated the role played by the identified four contextual modalities in studying behavior diversity across individuals. The goal of this experiment was to determine whether common or uncommon behaviors are what distinguishes different individuals. The results clearly show that, first of all, the context distribution is heavy tailed, and therefore that contextual modalities offer support for analyzing “rare” behaviors, and second that annotations in the tail of the context distribution are much more effective than those in the head at identifying individuals. This was verified in a practical identity recognition experiment.

## 2 Related work

In this section we review key works related to our paper from two distinct research areas: (i) capturing and modeling contextual information and subjective experiences using mobile sensing approaches, and (ii) modeling predictability, entropy and diversity of human behaviors.

### 2.1 Previous studies on capturing contextual information using mobile sensing

Numerous works adopt mobile sensing approaches on modeling and recognizing contextual information, which not only serve applications such as health and physical activity monitoring [29–32], mental health monitoring [33–35], and aging care [36, 37], but also benefit the research on understanding and predicting human individual behaviors and traits [19, 38–41].

A decade ago, Lane *et al.* [42] surveyed a number of existing studies on mobile phone sensing algorithms, applications, and systems and pointed out that how to characterize contextual information is one of the most challenging research problems in the mobile sensing community. More in general, ubiquitous computing and context-aware computing researchers have produced several works on modeling and characterizing the contextual dimensions of human behaviors and activities [18, 19, 21, 43].

For example, various studies have leveraged the Experience Sampling Method (ESM) [44] approach on mobile devices to capture self-reported contextual information on the daily activities and routines of people [45–47]. ESM is a methodology aiming at collecting information on behaviors and feelings of study participants throughout their daily activities [44]. As in traditional diary studies, ESM collects data by means of study participants’ self-reports; however, study participants, unlike in diary studies, are proactively triggered at various moments during the day. Along this line, a group of ubiquitous computing researchers has designed Aware [48, 49], a platform for context-aware mobile sensing that captures different contextual information such as time, location and proximity interactions.

In our current work, we took inspiration from these previous efforts and we focus on collecting information on four contextual modalities often investigated in the past, namely

*time, location, activity, and social ties*. However, we advance the state-of-the-art performing a rigorous and extensive analysis of the joint effects of these contextual modalities on the *regularity* and *diversity* of human behaviors. Doing this, we merge contributions from ubiquitous and context-aware computing communities with computational social science approaches characterizing the predictability of human behaviors by means of information theoretic measures [3, 15, 28].

## 2.2 Previous studies on human predictability

In addition to studies and methods focused on capturing and modeling contextual information relevant for understanding human behaviors, there is a large body of work on human predictability. Notably, these studies are canonically split into topics based on the modality and data being considered: researches on human mobility look at location data [2, 3, 6, 8, 15, 28], studies on behavioral routines analyze activities and recurrent patterns [9, 50], and finally work on social networks investigates the role played by social relations and interactions [24]. Only few works consider combinations of distinct modalities, such as location and social ties [6, 18–20], or look at the predictability of human behavior at a more general level [9]. Here, we make use of the statistical and information theoretic measures developed mainly in studies on modeling human mobility [2, 3, 15], while extending them to the analysis of multiple modalities. Our experiments show that the four contextual modalities identified are fundamental for determining, and thus for analyzing, predictability and diversity of human behaviors.

## 3 Materials and methods

Given the variety and complexity of individual experiences, formalizing context in its entirety is essentially impossible, and application-specific or study-specific solutions are necessary. In our paper, we focus on four modalities of context—*time, location, activity, and social ties*—widely used in ubiquitous computing communities for capturing and describing situations occurring in everyday life [21–25].

Here, we illustrate these contextual modalities using a simple university life scenario, in which a student is attending a lecture at the University of Trento, at 11:00 AM, together with a friend named “Shen”.

Formally, the context can be represented as a tuple:

$$\text{Context} = \langle \text{TIME}, \text{WE}, \text{WA}, \text{WO} \rangle,$$

where:

- TIME is the *temporal context*: It answers “What **T**IME is it?” and encodes the time in which that context was observed, e.g. “morning”;
- WE is the *endurant context* and answers “**W**hEre are you?” It indicates the relevant location that a person is at, e.g. “classroom”;
- WA is the *perdurant context* and answers “**W**hAt are you doing?” It refers to the main activity taking place, e.g. “lesson”;
- WO is the *social context* and answers “**W**hO are you with?” It covers all the relevant people in the current context, e.g. “teacher”, “classmates”, and “Shen”.

### 3.1 Data collection

The data was collected as part of the Smart UniTn 2 project, which lasted from the 7th of May to the 7th of June 2018, for a total of four weeks (32 days) [27]. The research protocol was designed on top of a prior analogous, but slightly smaller, study [26]. More precisely, following the research protocol an e-mail inviting participation in the data collection was sent to all 12,000 regularly enrolled students at the University of Trento. The e-mail clearly explained that students could choose to participate in the study for two or four weeks, and that in the first two weeks they would receive a notification every half hour, while in the second two weeks every two hours. Moreover, as stated by Keusch *et al.* [51], the willingness to participate in mobile data collection is strongly influenced by the incentive promised for study participation. Thus, a monetary incentive was introduced to encourage prompt and truthful reporting. A reward of 20 euros was promised to each participant every two weeks. In addition, each participant was informed that at the end of the survey there would be a lottery among those who responded to more than 75% of the notifications consisting of 3 prizes of 100 euros for the first two weeks and 3 prizes of 150 euros for the second two weeks. From the 1089 volunteers, a stratified random sample of 237 students from 10 different departments at the University of Trento, Italy, was invited to participate in the survey.

Following the Italian regulations, all participants were asked to sign informed consent forms and the study was conducted in accordance to them. The research protocol and the informed consent forms were also approved by the Ethical Committee of the University of Trento.

The data was logged using the i-Log app [52], which all volunteers were required to install on their mobile phones. The app records measurements from several sensors, both hardware (e.g. GPS, accelerometer, gyroscope) and software (e.g. applications running on the device). Table 1 lists all sensors with their frequencies and units of measurement. The app was also used to track the personal context of each study participant (namely their current activity, location, and social context) by periodically administering questionnaires. Figure 1 reports the questions appearing in each questionnaire and the set of possible answers. The participants had 150 minutes (2.5 h) since submission to answer a questionnaire. If a study participant failed to timely reply to five consecutive questionnaires, the oldest one was dropped and the answer treated as a missing value.

As previously said, the data collection was split in two phases, each two weeks long. In the first phase (7th to 24th of May) questionnaires were submitted to the volunteers every 30 minutes, while in the second one (25th of May to 7th of June) every 2 hours, to lessen the cognitive load. In this second stage, the volunteers were also specifically requested to leave the app running at all times.

### 3.2 Data preprocessing

Despite these precautions, the self-reported annotations are not unlikely to be noisy and biased. This is compatible with earlier observations about a similar collection experiment [53]. In order to minimize the remaining bias, the raw annotations were cleaned as follows. In a first step, a simple criterion was used to identify valid (that is, “trustworthy”) study participants. A participant was deemed valid if s/he failed to reply no more than 7 times within any 10-hour window, completed all questionnaires for at least 13 days, and provided at least 300 valid answers. All of these conditions must hold for a study participant to be

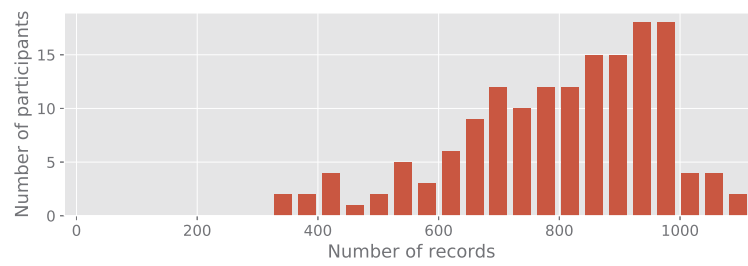
**Table 1** List of sensors. Proximity triggers when the phone detects close objects, like the subject's hand or head

Sensor	Frequency	Unit
Acceleration	20 Hz	m/s <sup>2</sup>
Linear Acceleration	20 Hz	m/s <sup>2</sup>
Gyroscope	20 Hz	rad/s
Gravity	20 Hz	m/s <sup>2</sup>
Rotation Vector	20 Hz	Unitless
Magnetic Field	20 Hz	μT
Orientation	20 Hz	Degrees
Temperature	20 Hz	TC
Atmospheric Pressure	20 Hz	hPa
Humidity	20 Hz	%
Proximity	On change	0/1
Position	Every minute	Lat./Lon.
WIFI Network Connected	On change	Unitless
WIFI Networks Available	Every minute	Unitless
Running Application	Every 5 seconds	Unitless
Screen Status	On change	0/1
Flight Mode	On change	0/1
Battery Charge	On change	0/1
Battery Level	On change	%
Doze Modality	On change	0/1
Headset plugged in	On change	0/1
Audio mode	On change	Unitless
Music Playback	On change	0/1
Audio from the internal mic	10 seconds per minute	Unitless
Notifications received	On change	Unitless
Touch event	On change	0/1
Cellular network info	Once every minute	Unitless
Activity Performed (Google data)	Once every minute	Unitless

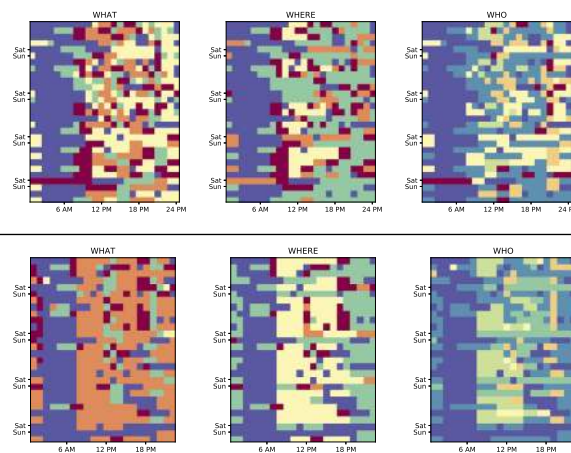
What are you doing?	Where are you?	With whom are you?	What is your mood?
Sleeping	Home Apartment Room	Alone	1. 😊
Self-care	Relatives Home	Friend(s)	2. 😊
Eating	House (friends others)	Relative(s)	3. 😊
Study	Classroom / Laboratory	Classmate(s)	4. 😊
Lesson	Classroom / Study hall	Roommate(s)	5. 😊
Social life	University Library	Colleague(s)	
Watching YouTube Tv-shows etc.	Other university place	Partner	
Social media (Facebook Instagram etc.)	Canteen	Other	
Travelling (*)	Other Library		
Coffee break cigarette beer etc.	Gym	(*) How are you moving?	
Phone calling; in chat WhatsApp	Shop supermarket	By subway	
Reading a book; listening to music	Pizzeria pub bar restaurant	By car	
Movie Theatre Concert Exhibit ...	Movie Theatre Museum	By foot	
Housework	Workplace	By bike	
Shopping	Other place	By bus	
Sport	Outdoors	By train	
Rest/nap		By motorbike	
Hobbies		Other	
Work			

**Figure 1** The questionnaire used in the Smart UniTn 2 project

deemed valid. A total of 184 study participants were marked as valid. The records of all invalid study participants were discarded. The next step was to delete events with invalid or missing values (like empty string labels) and records spuriously occurring before the



**Figure 2** Number of annotations per participant in the final data set. Most study participants have between 400 and 1000 fully annotated contexts



**Figure 3** Annotations of an irregular (top) and a regular (bottom) study participant. For each participant, left to right: annotations for What, Where, and Who. Each row is a day, columns are hours. Colors indicate different values. The values were grouped by similarity, for interpretability. Missing values are in blue. Weekends are indicated on the y axis

8th of May or after the 5th of June. Finally, in order for all statistics of the data in the two phases to be directly comparable, the records obtained from the second phase were replicated four times.

After the processing, there are 156 study participants with 124,963 records in total in the data set. The processed dataset consists of several time series  $\mathbf{x}^{u,m}$ , one time series for each valid study participant  $u \in \{1, \dots, 156\}$  and contextual modality  $m \in \{\text{TIME, WE, WA, WO}\}$ . Each time series can be viewed as a vector  $\mathbf{x}^{u,m} = (x_1^{u,m}, \dots, x_T^{u,m})$ , where  $T$  is the number of questionnaires administered to a study participant during the collection procedure and every  $x_t^{u,m}$ , with  $t = 1, \dots, T$ , indicates the annotation for modality  $m$  reported by study participant  $u$  at time  $t$ . The number of annotations per study participant, reported in Fig. 2, shows that most participants have in-between 400 and 1000 records. To get an intuition of the regularity in the behavior of volunteers, we selected two participants with the highest and lowest annotation diversity and visualize their annotations in Fig. 3. The most regular study participant has a distinctly simpler behavior than the other one, as expected. The figure also shows that even the behavior of the more regular volunteer is still quite irregular and displays substantial variability across days and across weeks.

### 4 Results

We organize this section in several subsections. First of all (Sects. 4.1, 4.2, and 4.3), we analyze each contextual modality in isolation. Second (Sects. 4.4 and 4.5), we study the influence of the four modalities on one another. Then, empowered by the positive results of the previous sections we proceed the analysis by studying the impact of subjectivity on predictability (Sect. 4.6). We conclude (see Sects. 4.7 and 4.8) by providing evidence that the investigated contextual modalities are useful in computing the diversity of personal behavior across individuals.

#### 4.1 Intra-modal predictability

Figure 4 reports the distribution of annotations in the data. The plot shows that, for all contextual modalities, few values take up most of the mass. Roughly speaking, this means that study participants spend most of their time performing four basic activities (namely studying, sleeping, eating, and moving between locations, which account for about 55% of the records), mostly stay at home (either their home or their relatives, more than 50%), and mostly by themselves or with their friends (almost 50% and 16%, respectively). TIME is special in that its annotations are extremely regular and mostly determined by the experimental setup rather than by individual preferences. This is especially true for nocturnal annotations, as the user can set i-Log to “sleep mode” so that it will automatically reply to the questionnaires accordingly during the night. For this reason, TIME is omitted from the

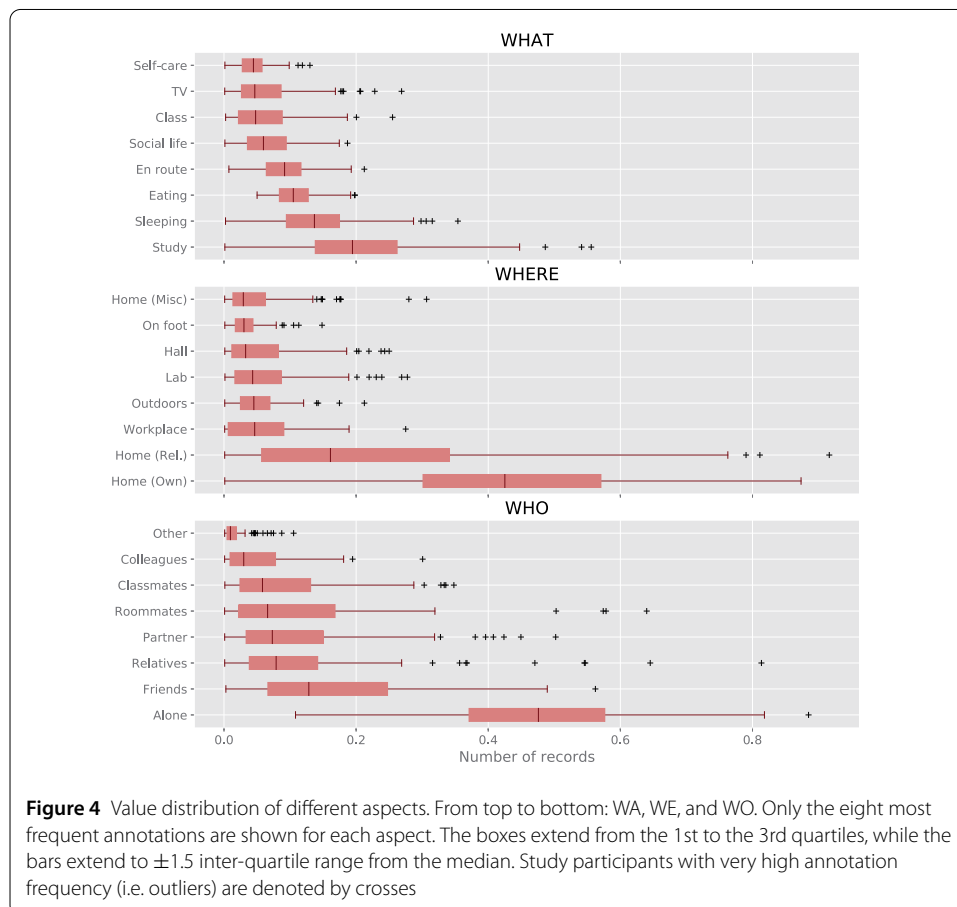




figure. The profile transpiring from the data reflects the source demographics.<sup>1</sup> The concentration of mass on few preferred values is consistent with previous studies on mobility [3, 6].

We are interested in understanding to what degree individual modalities are predictable and whether some modalities are intrinsically more predictable than others. In line with previous work [2, 3, 15, 28], we answer these questions using *entropy* and *predictability*. We introduce these notions in turn.

### 4.2 Entropy and predictability

*Entropy* measures the number of bits necessary to encode a random source: an entropy of  $b$  bits indicates that, on average, an individual who chooses her/his next value (i.e. location, activity, or social tie) randomly according to the ground-truth distribution will be found in  $2^b$  distinct states with high probability [54]. Hence, higher entropy implies higher uncertainty. In order to evaluate the contribution of different factors, consistently with previous studies [3, 15], we estimated three forms of entropy:

- (1) The *random entropy*, defined as:

$$H_{\text{rand}}(X^{u,m}) = \log_2 N^{u,m},$$

where  $X^{u,m}$  is a random variable that represents the value of modality  $m$  for individual  $u$  and  $N^{u,m}$  is the number of distinct values observed for that modality and individual in the full data set. The random entropy assumes that the study participant is equally likely to choose any of the values that s/he has annotated.

- (2) The *time-uncorrelated* or *flat entropy*, defined as:

$$H_{\text{flat}}(X^{u,m}) = - \sum_x \Pr(X^{u,m} = x) \log_2 \Pr(X^{u,m} = x),$$

where the sum runs over all the possible values for modality  $m$  and  $\Pr(X^{u,m} = x)$  denotes the empirical probability that individual  $u$  reported value  $x$  for modality  $m$ , as estimated from the data. The flat entropy is more informed than the random entropy as it takes the full value distribution into account.

- (3) The *true entropy*, defined as the limit of the joint entropy:

$$H_{\text{time}}(X^{u,m}) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T H(X_1^{u,m}, \dots, X_t^{u,m}).$$

Here  $X_t^{u,m}$  is a random variable that captures the value provided by individual  $u$  for modality  $m$  at time  $t$ , and the joint entropy  $H(X_1^{u,m}, \dots, X_t^{u,m})$  measures the disorder of  $t$  random variables:

$$- \sum_{x_1, \dots, x_t} \Pr(X_1^{u,m} = x_1, \dots, X_t^{u,m} = x_t) \times \log_2 \Pr(X_1^{u,m} = x_1, \dots, X_t^{u,m} = x_t).$$

---

<sup>1</sup>Considering that the volunteers are university students, the self-reported amount of studying is likely to be a (slight) over-estimate.

Compared to the flat entropy, the true entropy takes correlations over time, including short- and long-range correlations, into account. The true entropy is estimated from the data using the Lempel–Ziv estimator [55].

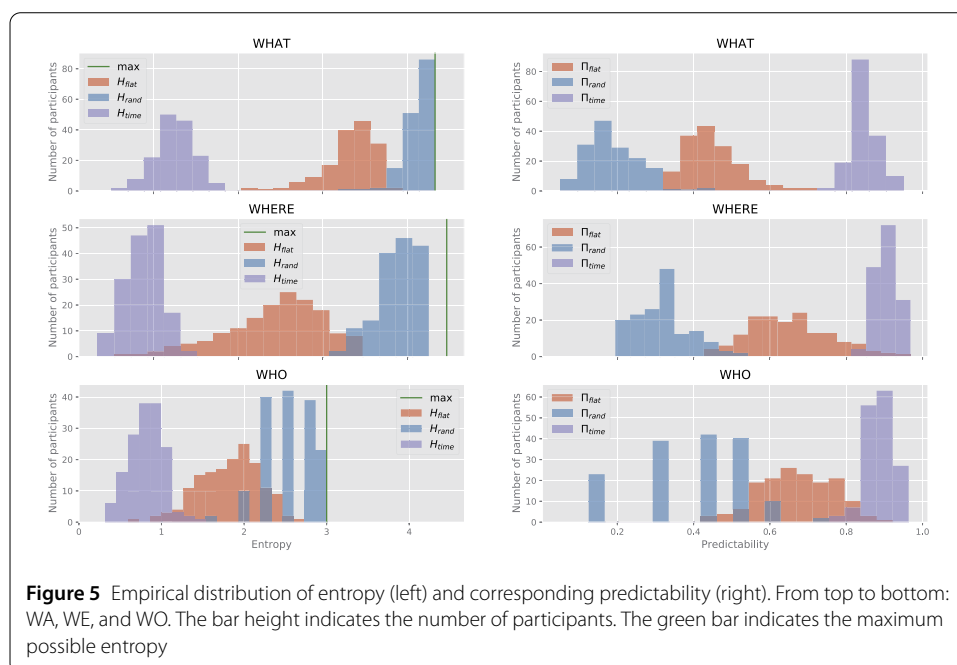
While entropy measures uncertainty, it only gives indirect information about how “easy to guess” a random source is. This is better captured by the notion of *predictability*, which was introduced to assess regularity of human mobility [3]. Formally, the predictability  $\Pi(X) \in [0, 1]$  of a random variable  $X$  is the accuracy of an *optimal* classifier for  $X$ , that is, the probability that this classifier outputs the correct value. As a consequence, if the predictability of a random variable is 0.8, then no classifier can have an accuracy higher than 80%—or, in other words, all classifiers must be mistaken 20% of the time. This means that predictability measures the *irreducible error* intrinsic in a random source. A notable property of the predictability  $\Pi$  is that, thanks to Fano’s inequality [54], it can be derived directly from the entropy  $H$  by solving the equation:

$$H = -(\Pi \log_2 \Pi + (1 - \Pi) \log_2(1 - \Pi)) + (1 - \Pi) \log_2(N - 1). \tag{1}$$

Here  $N$  is the number of distinct values that  $X$  can take. Please see [3] for a detailed derivation. For our goals, it suffices to know that, very intuitively, lower entropy entails higher predictability. In order to measure the effect of annotation distribution and correlations over time, the predictability of each individual  $u$  and modality  $m$  was obtained by solving Eq. (1) using the random, flat, and true entropy. The resulting values are indicated as  $\Pi_{\text{rand}}^{u,m}$ ,  $\Pi_{\text{flat}}^{u,m}$ , and  $\Pi_{\text{time}}^{u,m}$ , respectively.

### 4.3 Results for intra-modal entropy and predictability

Figure 5 illustrates the distribution of entropy (left) and predictability (right) for each modality. The histograms show that while all modalities are to some extent regular, some are more regular than others. This is partly due to the fact that the theoretical maximum



**Table 2** Empirical entropy (left) and predictability (right) averaged over all study participants and standard deviation thereof

Modality	$H_{\text{rand}}$	$H_{\text{flat}}$	$H_{\text{time}}$	$\Pi_{\text{rand}}$	$\Pi_{\text{flat}}$	$\Pi_{\text{time}}$
WHAT	$4.11 \pm 0.15$	$3.32 \pm 0.31$	$1.25 \pm 0.23$	$0.20 \pm 0.07$	$0.45 \pm 0.07$	$0.85 \pm 0.03$
WHERE	$3.88 \pm 0.24$	$2.43 \pm 0.57$	$0.87 \pm 0.22$	$0.32 \pm 0.07$	$0.65 \pm 0.10$	$0.90 \pm 0.03$
WHO	$2.58 \pm 0.30$	$1.82 \pm 0.37$	$0.82 \pm 0.20$	$0.40 \pm 0.15$	$0.67 \pm 0.10$	$0.89 \pm 0.04$

of the entropy is  $\log_2 N^m$ , and it is controlled by the number of possible values  $N^m$  for modality  $m$ . Hence, modalities with more states, like activity and location, are intrinsically more uncertain and less predictable than modalities with fewer states. In our setting, the theoretical maximum of the entropy (represented in the entropy plots by a green line) is about 4.4 for location, 4.3 for activity, and 3 for social tie. The plots show that entropy is largely determined by distributional information and short- and long-range correlations always impact the measured entropy: random entropy (blue) is always much higher than flat entropy (red), which is itself much higher than true entropy (purple). These changes in uncertainty demonstrate that taking annotation distribution and time correlations into account can substantially lower uncertainty and increase predictability. The same effect can be observed for all modalities, with some differences. For all entropy measures, the WA modality has the highest entropy, followed by WE and WO. However, the difference between modalities is more pronounced for the random and flat entropy, while it is limited for the true entropy, confirming the usefulness of taking time correlations into account.

Figure 5 (right) shows predictability of each modality for the different types of entropy. Comparing these histograms with those on the left makes it clear that increasing the amount of information dramatically increases predictability, as expected. Table 2 reports means and standard deviations of empirical entropy and predictability of each modality and type of entropy. The predictability for the true entropy  $\Pi_{\text{time}}$  (and hence maximal prediction accuracy) is 85% for activity, 89% for social tie, and 90% for location. This entails that irreducible error, even when taking all the available information into account, is about 10%–15% across modalities. The irreducible error for the flat entropy is even larger, 35%–55%.

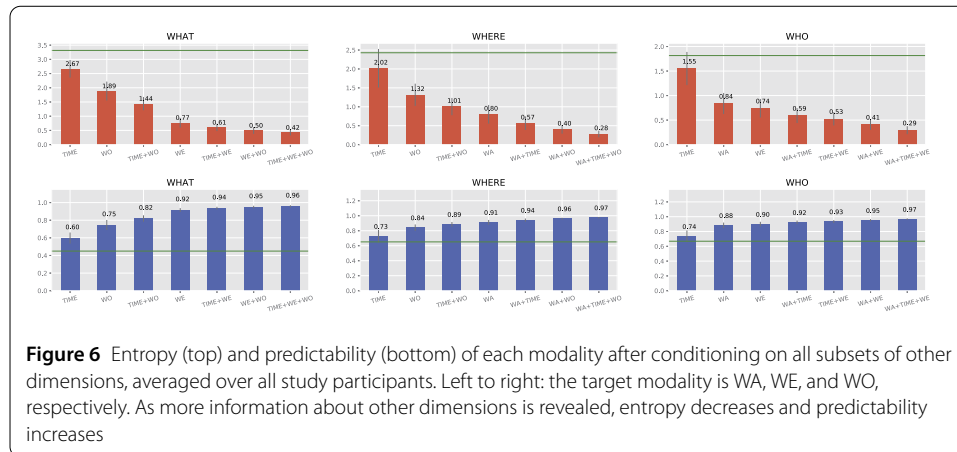
The standard deviation of predictability—that is, the spread of the histogram—does considerably shrink as more information is taken into consideration. This points at the fact that, as more information is considered, all participants appear to act more predictably. It is worth noting that, however, the standard deviation of the true entropy is non-zero, hinting at the fact that some participants are *intrinsically* less predictable than others. This partially motivates our study of behavior diversity across individuals, presented later on.

#### 4.4 Inter-modal predictability

So far, we have studied individual modalities taken in isolation. This approach is simplistic in that it neglects correlations between modalities, which we hypothesize to be very significant. In the following, we study the effect of inter-modal correlations on predictability.

This is achieved by estimating the conditional entropy  $H(X^{u,m}|X^{u,m'})$ , which quantifies the number of bits  $b$  needed to encode a random source  $X^{u,m}$  assuming that  $X^{u,m'}$  is known (with  $m' \neq m$ ). Intuitively, the more  $X^{u,m'}$  influences or determines  $X^{u,m}$ , the lower the conditional entropy [54]. The conditional entropy is defined as:

$$H(X^{u,m}|X^{u,m'}) = \sum_{x'} \Pr(X^{u,m'} = x') H(X^{u,m}|X^{u,m'} = x'),$$



**Figure 6** Entropy (top) and predictability (bottom) of each modality after conditioning on all subsets of other dimensions, averaged over all study participants. Left to right: the target modality is WA, WE, and WO, respectively. As more information about other dimensions is revealed, entropy decreases and predictability increases

where  $H(X^{u,m}|X^{u,m'} = x')$  is the entropy of  $X^{u,m}$  estimated only on those records that satisfy  $X^{u,m'} = x'$ . An issue with conditioning is that it is incompatible with the full entropy  $H_{\text{time}}$ , as it breaks time correlations: two non-consecutive records may appear to be consecutive in the conditional data set simply because they satisfy the same condition  $X^{u,m'} = x'$  and none of the records in-between them does. This means that the conditional and unconditional entropy cannot be compared directly.<sup>2</sup> For this reason, in the following we use the flat, time-uncorrelated entropy  $H_{\text{flat}}$  in all computations.

The reduction in flat entropy due to conditioning, averaged over all study participants, is illustrated in Fig. 6 (top). The green line represents the entropy prior to conditioning (as reported in Table 2), while the red bars represent the conditional entropy. The change in predictability is reported below in the same figure. The plots show very clearly that *in all cases, inter-modal information substantially reduces uncertainty and improves predictability.*<sup>3</sup> Indeed, conditioning any modality on the rest of the context (including TIME, rightmost bar in the plots) reduces entropy by more than 80% and increases predictability by at least 30%. More in detail, upon conditioning on the full context model, the entropy drops from 3.32 to 0.42 for WA, from 2.43 to 0.28 for WE, and from 1.82 to 0.29 for WO. At the same time, the predictability goes from 0.45 to 0.96 for WA, from 0.65 to 0.97 for WE, and from 0.67 to 0.97 for WO, cf. Table 2. This shows that the potential gain in accuracy from using multi-modal contextual dimensions is extremely large for all the modalities. The results for predictability make this point even clearer, as conditioning gives an impressive reduction of the *irreducible error* (that is,  $1 - \Pi$ ). In particular, the irreducible error of WA sees a huge drop from 55% to 4%, that of WE from 35% to 3%, and that of WO from 33% to 3%. This is consistent with our argument that time, location, activity, and social ties strongly influence each other, and provides empirical evidence in favor of our approach of taking into consideration all the four contextual dimensions.

The magnitude of entropy reduction is largely independent of the target modality: conditioning reduces entropy of WA by 84%, of WE by 86%, and of WO by 81%, and increases predictability by 160%, 133%, and 131%, respectively. At the same time, some modalities

<sup>2</sup>A naïve comparison shows that the conditional full entropy appears to be larger than the unconditional full entropy, which is clearly impossible.

<sup>3</sup>The conditional entropy is—by definition—never larger than the unconditional entropy, that is,  $H(X|X') \leq H(X)$ , regardless of the relation between  $X$  and  $X'$ . Still, if  $X'$  is independent of  $X$ , then conditioning has no effect on entropy. This is clearly not the case in our plots.

ties appear to carry more information than others: while conditioning on TIME shrinks entropy by only 15–20%, conditioning on WO, WA, and WE reduces entropy by 45%, 54–67%, and 59–77%, respectively. The four modalities can be ordered by average impact as  $TIME < WO < WA < WE$ , meaning that *TIME* is the least informative modality and location the most informative one in the setting under investigation in this study. The largest impact is observed when conditioning activity on location or vice versa, although conditioning on multiple modalities makes this effect more noticeable.

Comparing these results, which refer to flat entropy and predictability and that therefore ignore correlations over time, with those for full entropy supports the idea that inter-modal correlations are more influential than pure temporal correlations. Indeed, the full entropy of WA, WE, and WO reported in Table 2 are 1.25, 0.87, and 0.82, respectively, while the flat entropy obtained upon conditioning on the rest of the context is much lower, namely 0.42, 0.28, and 0.29, respectively.

#### 4.5 Location prediction in practice

The above analysis shows that taking multiple contextual modalities into account helps to identify regularities in the behavior of individuals. Along this line, we also expect that some activities, locations, or social relationships cannot be predicted unless information from other modalities is available. Furthermore, while predictability measures the performance of an *optimal* classifier, it is important to study whether improvements in predictability due to conditioning affect the performance of real classifiers *in practice*.

To investigate this issue, we carried out a practical location prediction experiment. Specifically, we measured the difference in prediction performance between a prototypical statistical classifier [56] that predicts location from sensor measurements and that of analogous classifiers that were additionally given annotations about activity and/or social ties. As for the classifier, we opted for Random Forests due to their performance and reliability [57].

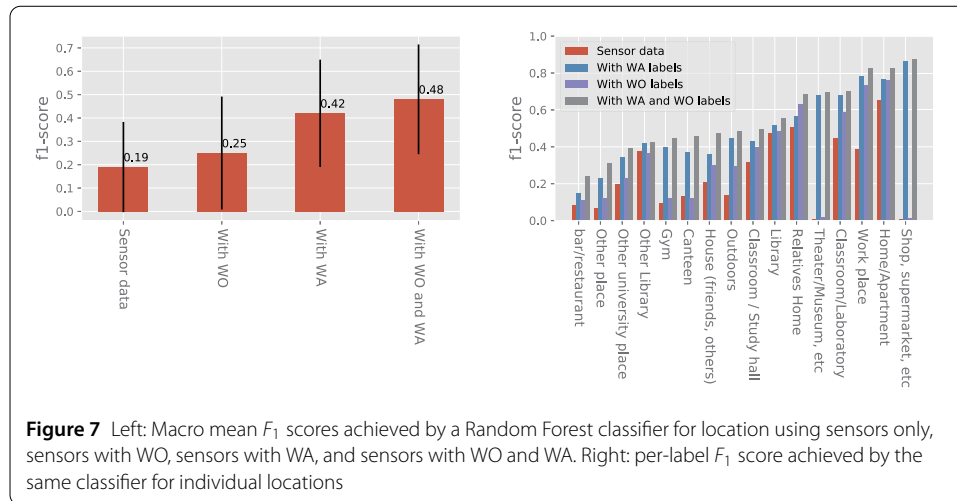
We trained one Random Forest classifier for each participant  $u$ . Each Random Forest takes as inputs the sensor measurements  $s_t^u$  of user  $u$  at time  $t$ —and optionally the annotations for the activity  $x_t^{u,WA}$  and social ties  $x_t^{u,WO}$ —and predicts the corresponding location  $x_t^{u,WE}$ . For simplicity, the sensor measurements  $s_t^u$  were restricted to features derived from GPS information, and in particular to longitude, latitude, and total distance traveled by the subject since the last questionnaire. This simple setup is sufficient for location prediction, and readings from the other sensors were found empirically to not be very relevant for the task at hand.

Prediction performance was evaluated using a 5-fold cross validation procedure. Namely, for each study participant, her/his records were randomly partitioned into 5 folds: one fold was used for performance evaluation while the remaining four were used for training the classifier. This step was repeated five times by varying the test fold. The performance of the Random Forest was taken to be the average over the five repeats.

For each user, we evaluated the impact of inter-modal annotations by comparing the performance of four classifiers: a baseline Random Forest that uses only GPS-derived inputs  $s_t^u$  and three Random Forests—with the very same depth—that were given also annotations for WA and/or WO as inputs. All hyper-parameters were kept to their default values.<sup>4</sup> except for forest depth, which was selected on a separate validation set to optimize

---

<sup>4</sup>As given in the `scikit-learn` package, version 0.24 [58]



the performance of the *baseline* Random Forest. In order to account for annotations skew (i.e. some locations are naturally more frequent than others), performance was measured using the macro  $F_1$  score. The latter is simply the  $F_1$  score of individual locations averaged over all locations.

The overall macro  $F_1$  scores averaged across study participants, as well as a breakdown of the  $F_1$  scores for individual locations, are reported in Fig. 7. The plots show that GPS information can predict reasonably well several locations (red bars), like “Home”, “Relative’s home”, and “Library”, among others, on which the baseline Random Forest achieves 40%  $F_1$  score. We conjecture this to be partially due to the fact these locations are very specific—in our data, the home of most users is unique and often easily identified from even few examples—and partially due to the abundance of annotations for these locations, cf. Fig. 4. GPS information, however, is clearly insufficient for locations like “Shop/Supermarket/etc.,” “Theater/Museum/etc.,” “Gym”, which are far more generic. Here the baseline Random Forest performs very poorly. This can be explained by two facts. First, these locations are composed of multiple objective locations (e.g. different shops, some of which possibly never observed during training), and therefore they are harder to predict based on GPS data alone. Second, the number of annotations for these locations is much lower.

Performance dramatically improves once WA and WO are supplied as inputs. In particular, the overall  $F_1$  score increases by about 30%. Moreover, while knowledge of either WO or WA always helps recognition performance, supplying both improves performance even more, as expected. We also note that WA is more useful than WO in general. These observations are consistent with the results for the optimal classifier.

One question is whether these results are influenced by the performance of particularly easy to predict classes. We assessed this possibility by computing a variant of the macro  $F_1$  that considers the *median* (rather than the mean) performance over classes, and as a result is naturally insensitive to classes that perform exceptionally well or exceptionally badly. The results are as follows: the macro mean  $F_1$  for the four cases (sensors only, sensors with WO, sensors with WA, and sensors with WO and WA) is 0.19, 0.25, 0.42 and 0.47 respectively, whereas the macro median  $F_1$  is 0.09, 0.13, 0.43 and 0.47. The more significant difference between macro mean and median  $F_1$  appears when no activity information is present: the *baseline* drops by about 10% and the “with WO” Random Forest

by 13%. However, the latter can be almost entirely explained by the former: adding social information contributes roughly +5% to both macro mean and median  $F_1$  (from 0.19 to 0.25 and from 0.09 to 0.13, respectively). Summarizing, this shows that the macro mean  $F_1$  overestimates the quality of the *sensor-only baseline* by about 10%. This probably occurs because all on-the-way locations like driving and walking are very hard to predict from sensors only (they individually achieve less than 8%  $F_1$ ), meaning that the macro median  $F_1$  tends to consider the higher-performing classes as outliers and ignores them. Most importantly, the contribution of inter-modal information to predictive performance is confirmed even by this more strict metric.

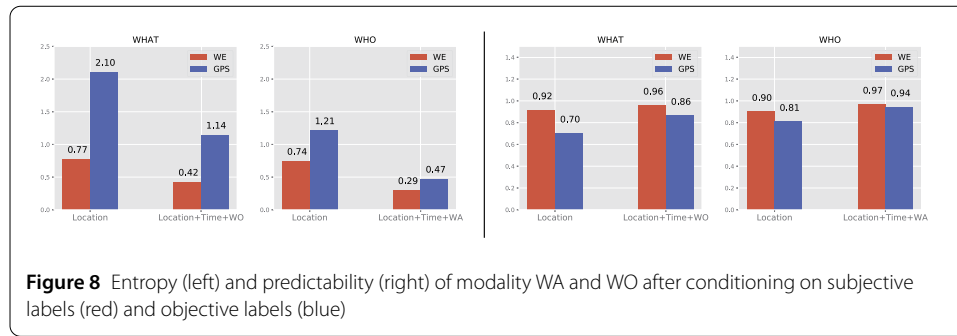
An important finding of this experiment is that some locations that were completely unpredictable from GPS data alone, are much easier to recognize if WA and WO annotations are supplied as inputs. The two most impressive examples are “Shop/Supermarket/etc.” and “Theater/Museum/etc.”, in which the correlation between location and activity boosts the  $F_1$  score from less than 5% to more than 70%. This very encouraging result offers further support for the jointly leverage of different contextual modalities, as some locations that are essentially impossible to recognize suddenly become essentially trivial to recognize when rich contextual information is provided.

#### 4.6 Subjectivity and predictability

Here, we investigate whether subjective annotations are more relevant than objective ones for determining predictability of behavior.

In particular, we compared the reduction in entropy due to conditioning on subjective location (namely, the WE annotations) to that due to conditioning on objective location, interpreted here in terms of GPS coordinates and related information. As in the location recognition experiment, we defined objective location using longitude, latitude, and total distance travelled since the last questionnaire. Computing the conditional entropy for continuous variables—in our case, the GPS coordinates—is not statistically straightforward. In order to avoid issues, we discretized the GPS information using a simple binning procedure. In particular, we allocated  $k = 3$  equal size bins for each of the three dimensions (longitude, latitude, amount travelled), for a total of 27 values for the objective data. This is done by using the `KBinsDiscretizer` class provided by `scikit-learn` [58] using the “quantile” strategy, which ensures that all bins contain roughly the same number of points. The number of bins roughly matches the number of subjective values (i.e. locations), which are 22. Since the variance of the conditional entropy estimator depends strongly on the number of alternative values, our choice of having roughly the same number of values for both subjective and objective data discourages the estimator from having dramatically different variances for the two cases.

A comparison of conditional entropy of WA and WO obtained by conditioning using subjective (red) versus objective (blue) location is reported in Fig. 8. The two left bars in each plot refer to conditioning the target modality using location only, while the two right bars indicate conditioning on all other modalities. There is a very clear difference between self-reported locations (WE) and GPS data: while knowing the GPS coordinates and traveled distance of the study participant reduces entropy in all cases, the reduction is far more modest than that obtained by conditioning on subjective location. The impact on predictability is analogous: GPS information provide a substantial boost to predictability (cf. Table 2), from 45% to 70% for WA and from 67% to 81% for WO. This is compatible



with the results obtained above for inter-modal correlations. The improvement is however always inferior to the one induced by subjective location: for WA, predictability is 70% when supplying objective location but goes up to 92% when supplying subjective annotations. For WO, the difference is less pronounced: 81% (objective location) against 90% (subjective). This is, again, likely due to the strong connection between activity and location. The situation is roughly unchanged if we condition the target modality on the rest of the context, namely location (either subjective or objective), time, and the remaining modality. These results show that subjectivity, besides being necessary for framing behavior from the subject's perspective, has a substantial effect on predictability and regularity of behavior in practice.

#### 4.7 Diversity: motivation

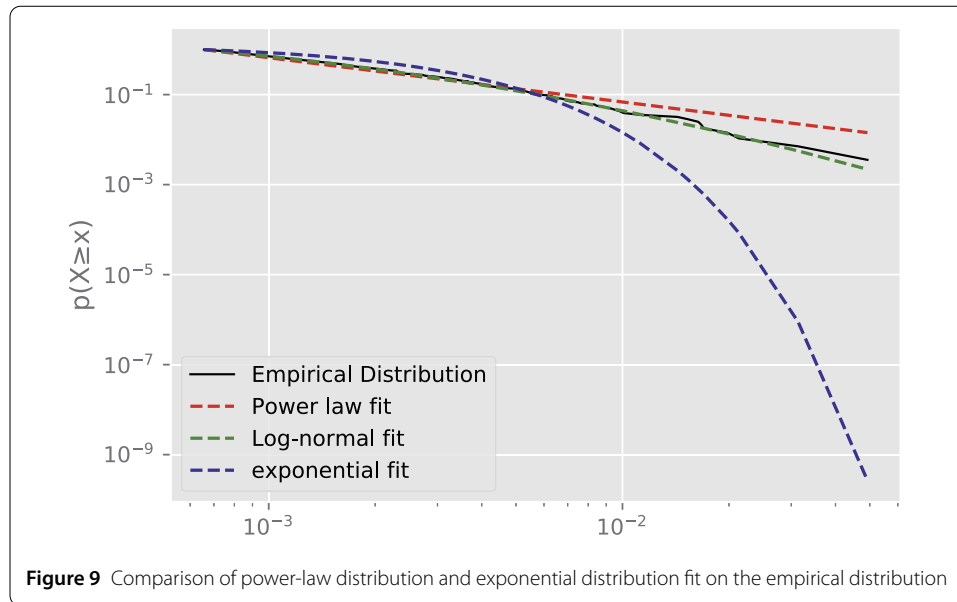
In the last experiment we studied the diversity of personal behavior. The motivation underlying this experiment is to provide some evidence of the intrinsic diversity, both objective and subjective, of the personal context of an individual. It is a widespread intuition that most of the time people behave similarly to each other. Indeed, everybody sleeps, eats, works, and socializes, and these activities take up most of our time. So, at a high level, everybody behaves the same during these high-frequency (subjective) activities. Our intuition is that individual differences manifest themselves in infrequent behaviors—for instance, while most people only go to the cinema in the evening, a cinephile has no issue going to a matinée.

A prerequisite to this argument is that rare behaviors occur often enough to be statistically meaningful. To determine whether this is the case, we checked whether the empirical distribution of context annotations is heavy-tailed. This was achieved by fitting three candidate distributions, a power law distribution, a log-normal distribution, and an exponential distribution to the data.<sup>5</sup> It is apparent from the plot shown in Fig. 9 that the log-normal distribution (with  $\mu = -8.2$ ,  $\sigma = 1.6$ ) offers a much better fit of the behavior of individuals than the exponential model, which is not heavy-tailed. This supports the idea that individual behavior described using the four identified contextual modalities is heavy-tailed, as expected.

Inspired by some studies on the uniqueness of mobility [60, 61] and apps usage [62] behaviors, we investigate whether annotations in the tail of the context distribution are indicative of personal identity, that is, whether it is easier to identify individuals using

<sup>5</sup>Using the `powerlaw` package [59].



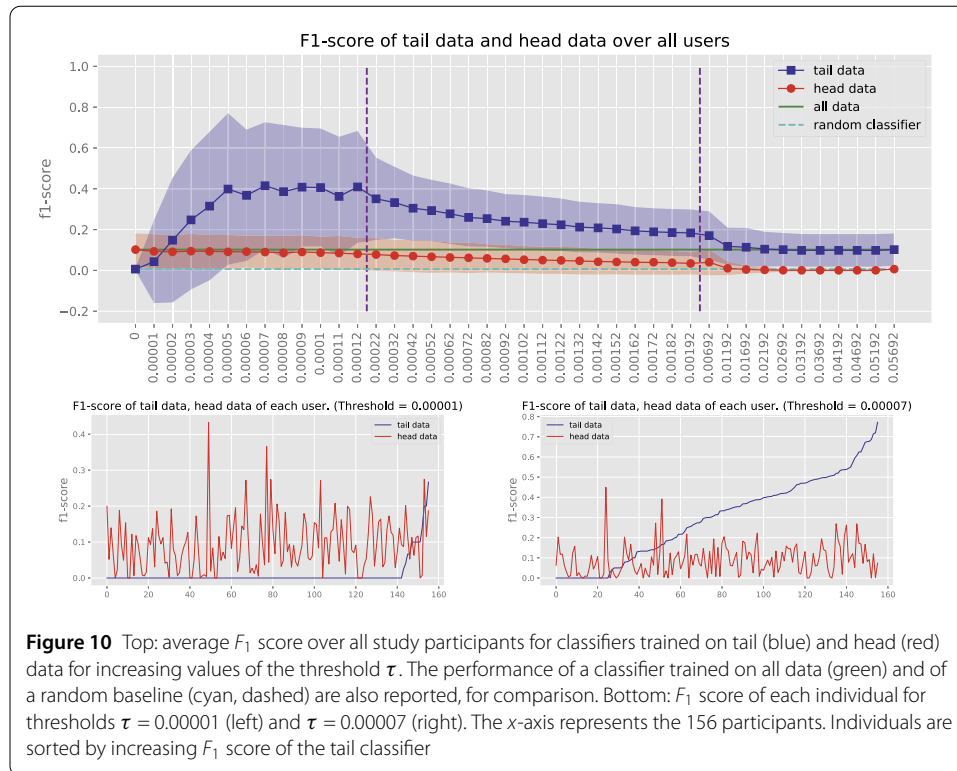


annotations from the tail or from the “head” of the distribution. For instance, in our university setting we expect common (head) annotations like (morning, classroom, lesson, classmates) to convey very little information about individual identity, as most university students attend lectures in the morning, and rarer (tail) annotations like (morning, workplace, work, alone) to be far more informative.

#### 4.8 Diversity: experiment and results

We designed a classification task in which the goal was to predict the identity of individuals based on context annotations only. All records in our data set were annotated with the ID of the subject they were generated by. The head and tail of the distribution were then defined using an arbitrary threshold  $\tau \geq 0$ : annotations that appear with frequency below  $\tau$  were taken to fall in the tail and the others in the head. Next, we trained two Support Vector Machine (SVM) classifiers [63] separately on the tail data and on the head data, and compared their performance. Both models received annotations for all modalities as inputs. As above, performance was measured in terms of  $F_1$  score (the higher the better) in a 10-fold cross validation setup. Notice that the number of personal IDs is 156, which is fairly large and renders the classification task highly non-trivial. For reference, the expected  $F_1$  score of a random classifier is  $1/156$  (indicated in cyan in the plots).

The results can be viewed in Fig. 10. The top plot shows the  $F_1$  score of the two classifiers as the threshold  $\tau$  is increased. Recall that a lower threshold entails that fewer annotations fall in the tail and more in the head. The threshold ranges from 0 (left of the plot), in which case no annotation falls in the tail, to the smallest value for which all data fall in the tail, which is  $\approx 0.57$  (right of the plot). Broadly speaking, the tail classifier always outperforms the head classifier by a large margin, while the head classifier never performs better than a classifier trained on both head and tail annotations (the green line in the figure). In order to better analyze the plot, we split it into three regions, highlighted by the purple lines (notice that the sticks on the  $x$ -axis are non-uniform.) In the leftmost region, the tail classifier does outperform the head classifier as soon as there are enough annotations in the tail, and it stabilizes at around 40%  $F_1$  score for  $\tau$  from about 0.00005 to 0.00012. Here the tail is



maximally informative, presumably because it only contains rare and informative context annotations. As the threshold increases and less “rare” annotations fall in the tail (middle region), the tail classifier drops off in performance but it still outperforms the “all” and the head classifiers. The head classifier also performs worse and worse, as more annotations move from the head to the tail. In the rightmost region, the tail converges to the full data set and hence the tail classifier converges to the performance of the “all” classifier.

A break-down of performance for different study participants is reported in Fig. 10 (bottom) for the two thresholds corresponding to the minimum ( $\tau = 0.00001$ ) and maximum ( $\tau = 0.00007$ ) of  $F_1$  respectively. Individuals are sorted on the x-axis according to the  $F_1$  score of the tail classifier, for readability. In the left figure, when  $\tau = 0.00001$ , the size of tail data is extremely small and only less than 20 users have annotations. This explains clearly why the performance of the tail classifier drops when the threshold is too small. On the other hand, for  $\tau = 0.00007$  (right figure) the overwhelming majority of individuals is more likely to be identified correctly by looking at their infrequent behaviors—with less than 10 exceptions. This provides evidence in support of the fact that the tail of the distribution conveys much more information about personal identity than the head. The “exceptional” participants themselves can also be easily explained. These individuals are hard to classify because their behavior is slightly more regular than that of the other volunteers, meaning that their most of their annotations occur more frequently and therefore are more likely to fall in the head of the distribution. Indeed, we verified that this issue disappears once the threshold is increased slightly (data not shown). A proper solution for this issue would be to choose the threshold  $\tau$  on a subject-by-subject basis. This is however orthogonal to our goals, and beyond the scope of this paper.

## 5 Conclusion

In this work, we have studied the predictability of human behavior through the notion of personal context. Our study captures a rich, multi-faceted picture of individual behavior by looking at four orthogonal but interrelated dimensions—namely time, location, activity, and social ties—viewed from the subject's own perspective. An empirical analysis on a large data set of daily behaviors shows the benefit of this choice: the different contextual modalities and their subjective description are shown to provide important cues about the predictability of individual behavior. Motivated by this, we also applied our contextual modalities to study behavioral diversity. The obtained results highlight that individuals are more easily identified from rarer, rather than more frequent, subjective context annotations.

This work can be extended in several directions. First and foremost, while our results are promising, we plan to further validate them in more settings and in specific applications. To this end, we are currently working on collecting a much larger data set, with students from several universities in four different countries, which will serve as a basis for a thorough investigation of the results presented here.

This work also highlights an interesting conundrum. Our results suggest that subjective annotations are very useful for predicting certain contextual modalities. However, these subjective annotations, obtained by filling questionnaires, have some degree of error related to, for example, the list of alternatives that are allowed to the respondent, e.g. the list of activities, places, or people; the memory effect of the respondent when s/he does not respond immediately; the social desirability effect that may prevent the study participant from reporting certain (socially disapproved) activities; and unreported activities when the participant perceives this as an intrusion into her/his privacy. Moreover, in practical applications, collecting self-reported annotations is not always an option. This means that in some settings and scenarios one has to compute predictions from sensor measurements only, which is likely to incur a substantial performance penalty. Going forward, one option is to replace the ground-truth self-reported annotations with predictions. This makes especially sense in a multi-task prediction pipeline in which all contextual modalities are predicted *jointly* from sensor measurements. This way, the predictor can leverage inter-modal correlations, which are key for inferring some locations and activities and for avoiding inconsistencies. This prediction pipeline would be fully operationalizable even in the absence of subjective annotations, so long as an initial training set is available. The downside is that replacing annotations with predictions does introduce noise into the system. Finding a complete solution to this problem is an interesting avenue for future work.

### Acknowledgements

Not applicable.

### Funding

This research has received funding from the European Union's Horizon 2020 FET Proactive project "WeNet—The Internet of us", grant agreement No. 823783, and from the "DELPhi—DiscovEring Life Patterns" project funded by the MIUR Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) 2017—DD n. 1062 del 31.05.2019.

### Availability of data and materials

The datasets generated and analysed during the current study are not publicly available due to pending full approval from UniTN DPO but may be available from Fausto Giunchiglia ([fausto.giunchiglia@unitn.it](mailto:fausto.giunchiglia@unitn.it)) on reasonable request.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

Conceived the study and data collection: IB and FG. Designed and performed the experiments: WZ, QS, ST, AP and FG. Analyzed and evaluated the results: WZ, QS, ST, AP and FG. Wrote the paper: WZ, QS, ST, BL, AP, IB and FG. All the authors read, reviewed and approved the final manuscript.

### Author details

<sup>1</sup>Department of Information Engineering and Computer Science, University of Trento, Trento, Italy. <sup>2</sup>Department of Sociology, University of Trento, Trento, Italy. <sup>3</sup>College of Computer Science and Technology, Jilin University, Changchun, China. <sup>4</sup>Center for Information and Communication Technology, Fondazione Bruno Kessler, Trento, Italy.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 18 February 2021 Accepted: 27 July 2021 Published online: 13 August 2021

### References

1. Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. *Nature* 439(7075):462–465
2. Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. *Nature* 453(7196):779
3. Song C, Qu Z, Blumm N, Barabási A-L (2010) Limits of predictability in human mobility. *Science* 327(5968):1018–1021
4. Lin M, Hsu W-J, Lee ZQ (2012) Predictability of individuals' mobility with high-resolution positioning data. In: Proceedings of the 2012 ACM conference on ubiquitous computing, pp 381–390
5. Smith G, Wieser R, Goulding J, Barrack D (2014) A refined limit on the predictability of human mobility. In: Proceedings of the 2014 IEEE international conference on pervasive computing and communications (PerCom), pp 88–94
6. Alessandretti L, Sapiezynski P, Sekara V, Lehmann S, Baronchelli A (2018) Evidence for a conserved quantity in human mobility. *Nat Hum Behav* 2(7):485–491
7. Cuttone A, Lehmann S, Gonzalez MC (2018) Understanding predictability and exploration in human mobility. *EPJ Data Sci* 7:2
8. Alessandretti L, Aslak U, Lehmann S (2020) The scales of human mobility. *Nature* 587(7834):402–407
9. Eagle N, Pentland AS (2009) Eigenbehaviors: identifying structure in routine. *Behav Ecol Sociobiol* 63(7):1057–1066
10. Eagle N, Pentland A, Lazer D (2009) Inferring friendship network structure by using mobile phone data. *Proc Natl Acad Sci* 106(36):15274–15278
11. Miritello G, Lara R, Cebrian M, Moro E (2013) Limited communication capacity unveils strategies for human interaction. *Sci Rep* 3(1):1–7
12. Saramäki J, Leicht EA, Lopéz E, Roberts SGB, Reed-Tsochas F, Dunbar RIM (2014) Persistence of social signatures in human communication. *Proc Natl Acad Sci* 11(3):942–947
13. Krumme C, Lorente A, Cebrian M, Pentland A, Moro E (2013) The predictability of consumer visitation patterns. *Sci Rep* 3:1645
14. de Montjoye Y-A, Radaelli L, Singh VK, Pentland A (2015) Unique in the shopping mall: on the re-identifiability of credit card metadata. *Science* 347(6221):536–539
15. Sinatra R, Szell M (2014) Entropy and the predictability of online life. *Entropy* 16(1):543–556
16. Pappalardo L, Simini F, Rinzivillo S, Pedreschi D, Giannotti F, Barabasi A-L (2015) Returners and explorers dichotomy in human mobility. *Nat Commun* 6(1):1–8
17. De Nadai M, Cardoso A, Lima A, Lepri B, Oliver N (2019) Strategies and limitations in app usage and human mobility. *Sci Rep* 9(1):1–9
18. Cho E, Myers SA, Leskovec J (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, pp 1082–1090
19. Do TMT, Gatica-Perez D (2012) Contextual conditional models for smartphone-based human mobility prediction. In: Proceedings of the 2012 ACM conference on ubiquitous computing, pp 163–172
20. Toole JL, Herrera-Yaque C, Schneider CM, González MC (2015) Coupling human mobility and social ties. *J R Soc Interface* 12(105):20141128
21. Dey AK (2001) Understanding and using context. *Pers Ubiquitous Comput* 5(1):4–7
22. Giunchiglia F, Bignotti E, Zeni M (2017) Personal context modelling and annotation. In: 2017 IEEE international conference on pervasive computing and communications workshops (PerCom workshops). IEEE, pp 117–122
23. Vaizman Y, Ellis K, Lanckriet G (2017) Recognizing detailed human context in the wild from smartphones and smartwatches. *IEEE Pervasive Comput* 16(4):62–74
24. Tran MH, Han J, Colman A (2009) Social context: supporting interaction awareness in ubiquitous environments. In: Proceedings of the 6th international conference mobile and ubiquitous systems: networking & services, MobiQuitous 2009. IEEE, pp 1–10
25. Bettini C, Brdiczka O, Henriksen K, Indulska J, Nicklas D, Ranganathan A, Riboni D (2010) A survey of context modelling and reasoning techniques. *Pervasive Mob Comput* 6(2):161–180
26. Giunchiglia F, Bison I, Bignotti E, Zeni M, Song D (2021) Trento 2016—a pilot on the daily routines of university students. Technical report—DataScientia dataset descriptors, University of Trento. Dataset soon to be available at <https://ri.internetofus.eu>
27. Bison I, Giunchiglia F, Zeni M, Bignotti E, Busso M, Chenu-Abente R (2021) Trento 2018—an extended pilot on the daily routines of university students. Technical report—DataScientia dataset descriptors, University of Trento. Dataset soon to be available at <https://ri.internetofus.eu>
28. Qin S-M, Verkasalo H, Mohtaschemi M, Hartonen T, Alava M (2012) Patterns, entropy, and predictability of human mobility and life. *PLoS ONE* 7(12):e51353

29. Dunton GF, Liao Y, Kawabata K, Intille SS (2012) Momentary assessment of adults' physical activity and sedentary behavior: feasibility and validity. *Front Psychol* 3:260
30. Liao Y, Intille SS, Dunton GF (2015) Using ecological momentary assessment to understand where and with whom adults' physical and sedentary activity occur. *Int J Behav Med* 22(1):51–61
31. Rabbi M, Aung MH, Zhang M, Choudhury T (2015) Mybehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. In: *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pp 707–718
32. Intille S (2016) The precision medicine initiative and pervasive health research. *IEEE Pervasive Comput* 15(1):88–91
33. Bogomolov A, Lepri B, Ferron M, Pianesi F, Pentland A (2014) Daily stress recognition from mobile phone data, weather conditions and individual traits. In: *Proceedings of the 22nd ACM international conference on multimedia*, pp 476–486
34. Wang R, Wang W, DaSilva A, Huckins JF, Kelley WM, Heatherton TF, Campbell AT (2018) Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 2(1):1–26
35. Wang W, Mirjafari S, Harari GM, Ben-Zeev D, Brian R, Choudhury T, Hauser M, Kane J, Masaba K, Nepal S, Sano A, Scherer E, Tseng V, Wang R, Wen H, Wu J, Campbell AT (2020) Social sensing: assessing social functioning of patients living with schizophrenia using mobile phone sensing. In: *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp 1–15
36. Lee ML, Dey AK (2015) Sensor-based observations of daily living for aging in place. *Pers Ubiquitous Comput* 19(1):27–43
37. Berke EM, Choudhury T, Ali S, Rabbi M (2011) Objective measurement of sociability and activity: mobile sensing in the community. *Ann Fam Med* 9(4):344–350
38. Farrahi K, Gatica-Perez D (2011) Discovering routines from large-scale human locations using probabilistic topic models. *ACM Trans Intell Syst Technol* 2(1):1–27
39. Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD (2016) Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. *Perspect. Psychol Sci* 11(6):838–854
40. Wang W, Harari GM, Wang R, Müller SR, Mirjafari S, Masaba K, Campbell AT (2018) Sensing behavioral change over time: using within-person variability features from mobile sensing to predict personality traits. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 2(3):141. <https://doi.org/10.1145/3264951>
41. Peltonen E, Sharmila P, Opoku Asare K, Visuri A, Lagerspetz E, Ferreira D (2020) When phones get personal: predicting big five personality traits from application usage. *Pervasive Mob Comput* 69:101269. <https://doi.org/10.1016/j.pmcj.2020.101269>
42. Lane ND, Miluzzo E, Hong L, Peebles D, Choudhury T, Campbell AT (2010) A survey of mobile phone sensing. *IEEE Commun Mag* 48(9):140–150
43. Dey AK, Abowd G (2000) Towards a better understanding of context and context-awareness. In: *Proceedings of the CHI 2000 workshop on the what, who, where, when, and how of context-awareness*
44. Larson R, Csikszentmihalyi M (1983) The experience sampling method. In: Csikszentmihalyi M (ed) *Flow and the foundations of positive psychology*. Jossey-Bass, New York
45. Raento M, Oulasvirta A, Eagle N (2009) Smartphones: an emerging tool for social scientists. *Social Methods Res* 37(3):426–454
46. Pejovic V, Lathia N, Mascolo C, Musolesi M (2016) Mobile-based experience sampling for behaviour research. In: Tkalcic M, De Carolis B, de Gemmis M, Odič A, Košir A (eds) *Emotions and personality in personalized services: models, evaluation and applications*. Springer, Cham
47. van Berkel N, Ferreira D, Kostakos V (2017) The experience sampling method on mobile devices. *ACM Comput Surv* 50(6):1–40
48. Dey AK, Wac K, Ferreira D, Tassini K, Hong J-H, Ramos J (2011) Getting closer: an empirical investigation of the proximity of user to their smart phones. In: *Proceedings of the 13th international conference on ubiquitous computing (UbiComp 2011)*
49. Ferreira D, Kostakos V, Dey AK (2015) Aware: mobile context instrumentation framework. *Front ICT* 2:6. <https://doi.org/10.3389/fict.2015.00006>
50. Lara OD, Labrador MA (2012) A survey on human activity recognition using wearable sensors. *IEEE Commun Surv Tutor* 15(3):1192–1209
51. Keusch F, Struminskaya B, Antoun C, Couper MP, Kreuter F (2019) Willingness to participate in passive mobile data collection. *Public Opin Q* 83:210–235
52. Zeni M, Zaihrayeu I, Giunchiglia F (2014) Multi-device activity logging. In: *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: adjunct publication*, pp 299–302
53. Zeni M, Zhang W, Bignotti E, Passerini A, Giunchiglia F (2019) Fixing mislabeling by human annotators leveraging conflict resolution and prior knowledge. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 3(1):1–23
54. Cover TM, Thomas JA (2012) *Elements of information theory*. Wiley, New York
55. Kontoyiannis I, Algoet PH, Suhov YM, Wyner AJ (1998) Nonparametric entropy estimation for stationary processes and random fields, with applications to English text. *IEEE Trans Inf Theory* 44(3):1319–1327
56. Ho TK (1998) The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 20(8):832–844
57. Breiman L (2001) Random forests. *Mach Learn* 45:5–32
58. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
59. Alstott J, Bullmore DP (2014) powerlaw: a python package for analysis of heavy-tailed distributions. *PLoS ONE* 9(1):e95816
60. De Montjoye Y-A, Hidalgo CA, Verleysen M, Blondel VD (2013) Unique in the crowd: the privacy bounds of human mobility. *Sci Rep* 3:1376

61. Rossi L, Walker J, Musolesi M (2015) Spatio-temporal techniques for user identification by means of gps mobility data. *EPJ Data Sci* 4:11
62. Sekara V, Alessandretti L, Mones E, Jonsson H (2021) Temporal and cultural limits of privacy in smartphone app usage. *Sci Rep* 11:3861
63. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---