Risk Estimation in COVID-19 Contact Tracing Apps*

Juan C. Bellassai[2], Pablo G. Madoery[1,3], Ramiro Detke[1], Lucas Blanco[1], Sandro Comerci[1], María S. Marattin[1], Juan Fraire[1], Aldana González Montoro[2], Grisel Britos[2], Silvia Ojeda[2], and Jorge M. Finochietto[1,3]

[1] Facultad de Ciencias Exactas, Físicas y Naturales - Universidad Nacional de Córdoba
[2] Facultad de Matemática, Astronomía, Física y Computación
[3] Instituto de Estudios Avanzados en Ingenieria y Tecnologia (IDIT) - CONICET

Abstract. In the context of COVID-19, contact tracing has shown its value as a tool for contention of the pandemic. In addition to its paper based form, contact tracing can be carried out in a more scalable and faster way by using digital apps. Mobile phones can record digital signals emitted by communication and sensing technologies, enabling the identification of risky contacts between users. Factors such as proximity, encounter duration, environment, ventilation, and the use (or not) of protective measures contribute to the probability of contagion. Estimation of these factors from the data collected by phones remains a challenge. In this work in progress we describe some of the challenges of digital contact tracing, the type of data that can be collected with mobile phones and focus particularly on the problem of proximity estimation using Bluetooth Low Energy (BLE) signals. Specifically, we use machine learning models fed with different combinations of statistical features derived from the BLE signal and study how improvements in accuracy can be obtained with respect to reference models currently in use.

Keywords: COVID-19 · Bluetooth Low Energy· Contact tracing · Proximity estimation · Feature selection

1 Introduction

The COVID-19 pandemic has led to rethink how currently available technology can help fight it. In particular, contact tracing, which consists in identifying close contacts that can be at risk of being infected with the virus, can benefit from smartphones [1]. While contact tracing existed in public health long before this pandemic, it has never been automated by relying on data collected by smartphones rather than people [2]. The widespread use of smartphones around the world enables a unique opportunity to implement contact tracing apps as a supplementary mean to control the outbreak. By systematically collecting data from close contacts, these apps can achieve unprecedented speed and coverage and therefore reduce the spread of the infection by alerting potential infected individuals earlier than with traditional methods, even before symptoms onset [3–5]. The concept has been widely implemented by public and private sectors around the world [6–9]. Google and Apple have released tracing services in their operating systems, enabling a true global potential of the approach [10]. However, these apps have generated much discussion around their key attributes, including system architecture, data management, privacy, security, and the actual accuracy to estimate the risk of contacts [11–14]. In this context, two major challenges arise.

* Supported by Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación

On the one hand, it is not known with complete certainty to what extent various factors such as proximity, duration, environment, ventilation, and the use of protective measures contribute to the risk of a contact. On the other hand, there are limitations both in the process of collecting data via smartphones and in the potential capacity of these data to provide accurate estimates of the factors mentioned above. The **ContactAr** project is a research initiative funded by the 2020 COVID-19 Call from the National Ministry of Science and Technology in Argentina which aims at exploring ways in which smartphone data can be used to estimate as accurately as possible some of the factors that contribute to the risk of contagion. The rest of the work is organised as follows. Section 2 discusses the process of obtaining different data through the use of smartphones and their possible application for risk characterization of contacts. Section 3 describes the problem of proximity estimation with emphasis on feature selection, model training and the experiments. Finally, Section 4 describes the results obtained so far and the work currently in progress in the framework of the **ContactAr** project.

2 Data collection

Modern cell phones can capture data from different sources including communication technologies (cellular network (2G/3G/4G), WiFi, Bluetooth), location technologies (GPS) and hardware/software based sensors as described in Table 1. In order to collect these data in a systematic way we developed a custom Android application which is available online in a public repository [4]. This app allows us to develop experiment campaigns using different cell phone models, while varying distance between phones, their position (horizontal, vertical), the environment (indoor, outdoor), and the length of recording times.

The data obtained from our experiments allow us to estimate several of the factors that contribute to the risk of infection associated with a close contact. However, as a first step, we decided to put the focus on how *Bluetooth Low Energy (BLE)* data can be used to estimate proximity between two people (phones) and we leave as future work the incorporation of the sensor data from Table 1 to investigate its impact to achieve improvements in risk estimation. The next section explains the experiments that were performed and the specific BLE data that was used to feed the machine learning models.

Sensor	Common uses	Risk estimation uses
light	controlling screen brightness	environment awareness (indoor/outdoor)
accelerometer	motion detection (shake, tilt, etc)	proximity
linear acceleration	monitoring acceleration along a single axis	proximity
gyroscope	rotation detection (spin, turn, etc)	proximity
gravity	motion detection (shake, tilt, etc)	proximity
magnetic field	creating a compass	proximity
rotation vector	motion detection and rotation detection	proximity
proximity	phone position during a call	social situation, proximity
activity	start/end of activity (walking, in_vehicle, etc)	social situation, proximity, duration
step counter	count steps	social situation, proximity, duration

Table 1. Other sensors.

3 Proximity estimation

Distance is the most straightforward target variable used to estimate proximity between two objects. When two BLE devices communicate, at any particular moment, one of them takes the role

[4] Contactar public repository: https://lcd-unc.github.io/

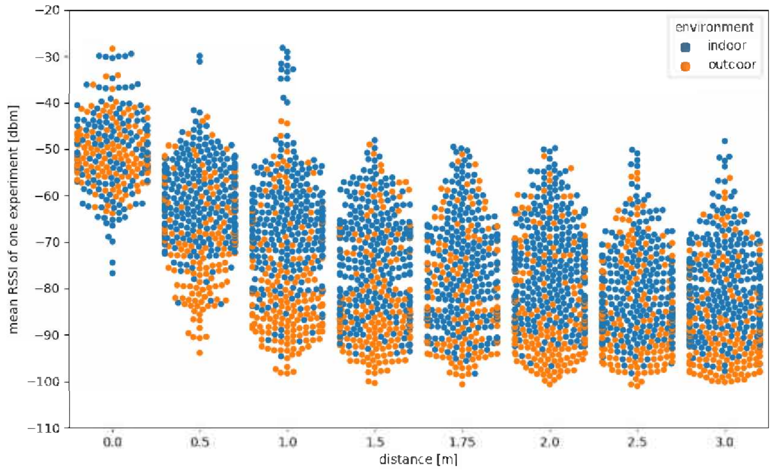

Fig. 1. Variability of mean RSSI over different distances, environments and smartphone models obtained in the second campaign.

of advertiser while the other assumes the role of listener. The former emits its signal at a certain power level, while the latter observes this power attenuated in terms of a Received Signal Strength Indicator (RSSI). Most current applications estimate distance by averaging RSSI values over a period of time. However, RSSI values typically fluctuate in time due to factors such as obstacles and reflections of particular environments, or due to different BLE chipsets and/or antenna configurations of specific device models. Figure 1 shows the variability of mean RSSI values (one dot per experiment) over different distances, environments and smartphone models obtained in our second campaign of experiments. In this work, we propose to incorporate other features, in addition to the mean RSSI, to feed machine learning models in order to assess whether it is possible to obtain accuracy improvements in distance estimation.

Specifically, 15 statistical features were derived from the normalized RSSI values: mean, trimmed mean, median, first and third quartiles, minimum, maximum, standard deviation, range, interquartile range, L1 distance to the mean ($\frac{1}{n}\sum_{i=1}^{n}|x_i - \overline{x}|$), L1 distance to the median ($\frac{1}{n}\sum_{i=1}^{n}|x_i - \tilde{x}|$), kurtosis, skewness, and the unique values count describing the series.

Due to the high dependency among many of these explanatory variables, it is not worth considering all possible combinations but only a subset of them. Taking into account the Pearson correlation coefficient between every pair of these features plotted in the Fig. 2 heatmap, we defined 3 different groups with strong within-group correlations:

- **Measures of position (Group 1)**; mean, trimmed mean (tmean), median, first quartile (q1), third quartile (q3), minimum (min) and maximum (max).
- **Measures of dispersion (Group 2)**; standard deviation (std), range interquartile range (iqr), mean L1 distance to the mean (dis1) and mean L1 distances to the median (dis2).
- **Measures of shape (Group 3)**; skewness (skew), kurtosis (kur) and count (cnt).

Within-groups correlations for groups 1 and 2 are larger than between-groups correlations. For instance, all within-group correlations for Group 1 are larger than 0.9 (except for $corr(min - q3) = 0.88$). Features in Group 3 are the less correlated. We evaluated combinations of 1, 2 and 3 features

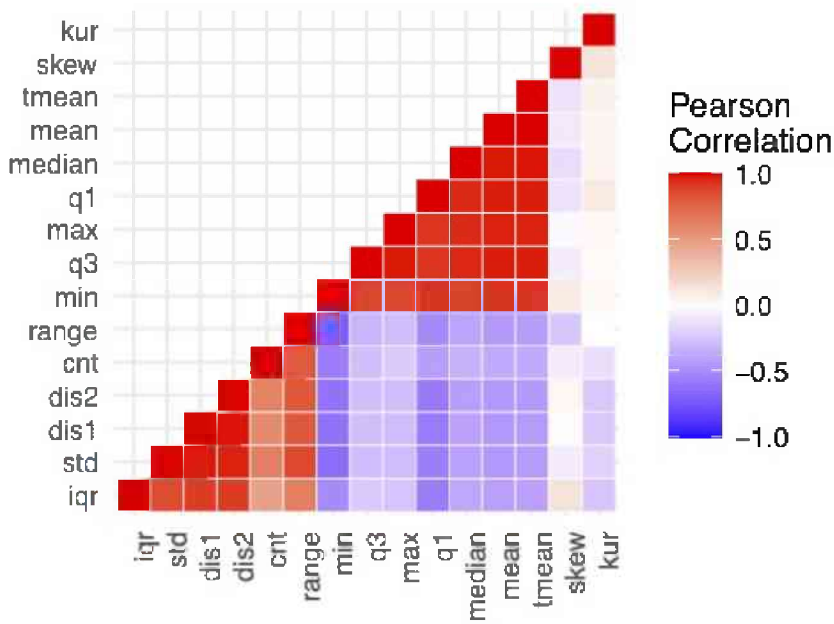

Fig. 2. Correlation between all different features.

in the following way: when using a unique feature it was possible to choose any feature of the position group. When 2 features were used, it was possible to choose 1 position feature and 1 dispersion feature or 1 position feature and 1 shape feature. When 3 features were evaluated a feature from each group was selected.

Training data were drawn from 2 campaigns of experiments. The first campaign considered 2 MOTO G(8) PLAY and 2 MOTOROLA ONE ZOOM mobile phones placed in horizontal positions at different distances (in ranges from 0 to 4 meters) and in different environments (indoor, outdoor) that mimic social gatherings where people leave their devices on a table. On the other hand, the second campaign incorporated a greater diversity of mobile phone models by adding 4 SAMSUNG, 4 XIAOMI and 2 MOTOROLA devices in other environments and distances, and with different horizontal and vertical orientations. Each experiment consisted of at least 2 phones on the same environment which periodically emitted and scanned BLE beacons during an observation window of 5 minutes in which the described features were computed. A total of 517 experiments were carried out in the first campaign (248 indoor and 269 outdoor) at distances of 0.0, 0.5, 0.8, 1.0, 1.5, 1.8, 2.0, 3.0 and 4.0 meters between phones, and a total of 2844 experiments were carried out in the second campaign (1300 indoor and 1544 outdoor) at distances of 0.0, 0.5, 1.0, 1.5, 1.75, 2.0, 2.5 and 3.0 meters as shown in Fig. 1.

Since the actual goal of contact tracing apps is to determine close contact rather than the exact distance among devices, proximity can instead be treated as a binary classification problem. Given the World Health Organization (WHO) defines a close contact as a distance within almost 2 meters, the categories were defined by *close contact* (for distances less than two meters) and *non-close contact* (distances greater or equal than two meters). Then, we trained and evaluated three different machine learning models: Logistic Regression (LR), Support Vector Machine (SVM) and Random Forest (RF) using the feature selection method described above. In order to obtain correct estimates of the models' accuracy, classes were balanced, erroneous values (generated by the hardware in the data collection process) were removed and each model was trained and evaluated using grid search with cross validation with KFold=5 and with 5 repetitions. Table 2 shows the hyperparameters values used in the grid search with cross validation process. Therefore, each accuracy score was computed in the validation sets as the average of 25 values by using the Python scikit-learn library. Furthermore, 20% of the data was kept out as test set to evaluate final model metrics.

	Hyperparameters
LR	[{'penalty' : ['l2'], 'C' : [1, 5, 10, 100]}]
SVM	[{'kernel' : ['rbf'], 'C' : [1, 10, 100]}, {'kernel' : ['poly'], 'degree' : [2], 'C' : [1, 10, 100]}]
RF	[{'n_estimators' : [50, 100], 'max_depth' : [4, 6, 10, 14], 'criterion' : ['entropy']}]

Table 2. Hyperparameters optimization

4 Preliminary results and current work

First campaign results showed that LR models have the lowest performance, while RF ones slightly outperform SVM models. All machine learning models outperform the benchmark used as reference, even when a single feature is considered. Furthermore, a better accuracy can be obtained in outdoor locations with respect to indoor ones. Indoor proximity estimation can benefit more from the introduction of more features (up to 3 features) with respect to the outdoor estimation case (up to 2 features). Accuracy can be increased about 10% when multiple features are considered if the device is aware of its environment, reaching a performance of up to 83% in indoor spaces and up to 91% in outdoor ones. These results encourages future contact tracing apps to integrate this awareness not only to better assess the associated risk of a given environment but also to improve the proximity accuracy. Figure 3 shows the confusion matrices obtained in the test set using 1, 2 and 3 features with the SVM model fit in indoor environments. The benefit of incorporating more features both to increase accuracy and to achieve a better balance between false positives and false negatives can be appreciated.

Second campaign results showed accuracy values around 71% for indoor environments and 66% for outdoor environments. Furthermore, no significant improvements in accuracy were observed as more features are added. Additionally, a feature importance analysis was performed to quantify how models improve accuracy by incorporating information about the environment, the

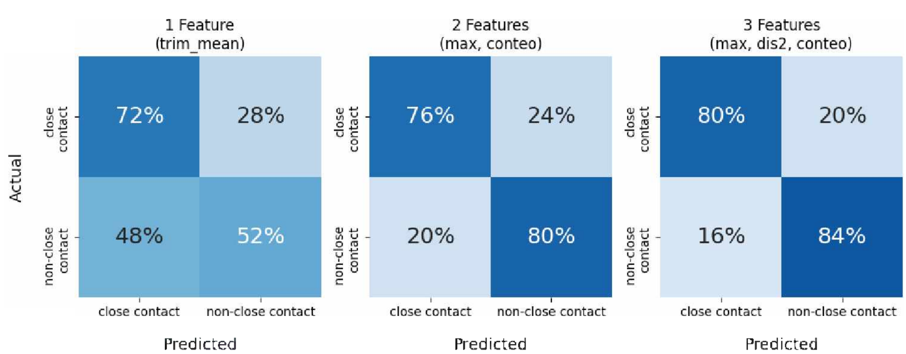

Fig. 3. Confusion matrices obtained in the test set using 1, 2 and 3 features with the SVM model fit in indoor environments.

position of the phones and phone models. This analysis showed that the most important context feature was the phone position with around 2.5% of relevance, the second one was the environment with around 2.2% of relevance and the last one was the phone model with less than 0.05%.

Overall, our analyses show that it is possible to use this technology to improve contact tracing techniques and thus help containing the spread of the pandemic, provided that suitable accuracy values are achieved. In our current work, we are addressing new challenges to further improve its performance. Among others, these challenges include:

Environment awareness Given that awareness of the environment has shown its value in estimating both distance and risk in general, we are investigating how other phone sensors can be used to discriminate the environment. In particular, we were able to predict the environment (indoor or outdoor) by using the light sensor with an accuracy of 87%. Furthermore, we claim that this result could be further improved by the incorporation of other data potentially available on the phone such as the number of GPS satellites at line of sight. This value is generally larger in outdoor environments compared to indoors. Besides, voting schemes could be implemented if phones advertise in their BLE messages (using a flag field) the type of environment they locally estimate.

Differences due to devices With the second campaign, accuracy values decreased and environmental awareness contribution was lower mainly due to the presence of different cell phone brands. We are interested in the nature of the effect of these factor on BLE signals and therefore on model performances.

References

1. N. Ahmed, R. A. Michelin, W. Xue, S. Ruj, R. Malaney, S. S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, and S. K. Jha. A Survey of COVID-19 Contact Tracing Apps. *IEEE Access*, 8:134577–134601, 2020.
2. Isobel Braithwaite, Thomas Callender, Miriam Bullock, and Robert W Aldridge. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. *The Lancet Digital Health*, 2020.
3. Mirjam E Kretzschmar, Ganna Rozhnova, Martin CJ Bootsma, Michiel van Boven, Janneke HHM van de Wijgert, and Marc JM Bonten. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *The Lancet Public Health*, 5(8):e452–e459, 2020.
4. Joel Hellewell, Sam Abbott, Amy Gimma, Nikos I Bosse, Christopher I Jarvis, Timothy W Russell, James D Munday, Adam J Kucharski, W John Edmunds, Fiona Sun, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, 2020.
5. Matt J Keeling, T Deirdre Hollingsworth, and Jonathan M Read. The Efficacy of Contact Tracing for the Containment of the 2019 Novel Coronavirus (COVID-19). *medRxiv*, 2020.
6. Vernon J Lee, Calvin J Chiew, and Wei Xin Khong. Interrupting transmission of COVID-19: lessons from containment efforts in Singapore. *Journal of Travel Medicine*, 27(3):taaa039, 2020.
7. Cheryl Lin, Wendy E Braund, John Auerbach, Jih-Haw Chou, Ju-Hsiu Teng, Pikuei Tu, and Jewel Mullen. Policy Decisions and Use of Information Technology to Fight Coronavirus Disease, Taiwan. *Emerging infectious diseases*, 26(7):1506, 2020.
8. Suneela Garg, Nidhi Bhatnagar, and Navya Gangadharan. A case for participatory disease surveillance of the COVID-19 pandemic in India. *JMIR Public Health and Surveillance*, 6(2):e18795, 2020.
9. Paul Mozur, Raymond Zhong, and Aaron Krolik. In coronavirus fight, China gives citizens a color code, with red flags. *New York Times*, 1, 2020.
10. Douglas J Leith and Stephen Farrell. GAEN Due Diligence: Verifying the Google/Apple Exposure Notification API. *CoronaDef21, Proceedings of NDSS '21*, 2021, 2020.
11. Leonardo Maccari and Valeria Cagno. Do we need a contact tracing app? *Computer Communications*, 2020.

12. M Burgess. Why the NHS Covid-19 contact tracing app failed, 2020.
13. Michael Klenk, Hein Duijf, and Christian Engels. Ethics of Digital Contact Tracing and COVID-19: Who Is (Not) Free to Go? *Available at SSRN 3595394*, 2020.
14. Hyunghoon Cho, Daphne Ippolito, and Yun William Yu. Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. *arXiv preprint arXiv:2003.11511*, 2020.