Missing Data Completion for Telco Localization via Adversarial Learning

1st Jinhua Lv, Weixiong Rao
Dept. of Software Engineering
Tongji University
Shanghai, China
jhlv@tongji.edu.cn

***Abstract*—This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX. This document is a model and instructions for LaTeX.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Recent years witnessed the rapid spreading of cellular networks and pervasive mobile devices (MDs). The telecommunication (Telco) data, as trace of MDs in cellular networks, has many important applications for Telco operators, e.g., city-scale Telco localization [1], churn prediction of subscribers [2] and user experience assessment [3]. In particular, as an important complementary technique of GPS, Telco localization is aimed at recovering mobility patterns of MDs at fine grained level (e.g., 20 meters). Unlike the call detail records (CDRs), Telco localization techniques mainly focus on measurement records (MRs), which measures radio signal strengths (RSSIs) between MD and its connected cells in Telco networks. In the past years, a plenty of Telco localization methods have been proposed to improve the performance under challenging city environment [4]–[8].

However, these localization models are suffering from missing signal strength (RSSI) values with varying degrees. Zhu et al. [1] found that nearly 50% of real world MR records have RSSI with only 1 or 2 cells. Ray et al. [7] proposed a localization model based on that RSSI values from neighboring cells are all missing. In the worst case, there is no RSSI at all, left only the serving cells in MR records [9]. The missing data problem significantly deteriorates the performance of Telco localization. There are two main reasons of missing values in MR records. One is that the mobile phones do not provide API to access neighboring cells. The other is that the RSSI is lost, due to communication failure or data corruption. Consequently, it is desired to develop a methodology with high completion performance to estimate the missing data.

The most frequently used methods for data completion are interpolation, statistics and nearest neighbors [10]. These methods simply fill missing values by part of the data set, do not generate high quality complete data. . The recent proposed algorithms built on deep learning (e.g. denoising autoencoders (DAE)) [11] learn the conditional probability from complete parts to missing parts, based on the whole data set. Given existing data completion methods, there are two major challenges in applying them directly for MR data.

(1) The existing methods cannot capture the correlation of cell locations and associated RSSI, which is most important in completing missing RSSI.

(2) Due to complex RF propagation phenomena (e.g., multi-path and non-line-of-sight), the signal strengths often fluctuate in a wide range. Noise caused by such fluctuation can hurt the inference of missing RSSI.

To this end, we propose TelcoGAN that generalizes Generative Adversarial Nets [12] with to generate complete data set.

Besides, there is a major difference in our problem. The missing MR data to complete is collected without location labels (e.g., GPS coordinates) due to GPS issues [13]. However, by retrieving locations from location-based services, is is available in historical collected MR data [14]. How to exploit these labels to help the generation of missing RSSI remains challenging.

II. Preliminaries

In this section, we first give a detailed description of MR data and some basic notations, then followed by the problem definition and overview of GANs [12] with its variants.

TABLE I
AN EXAMPLE OF 4G LTE MR RECORD

MRTime	**2017/5/31 14:12:06**	IMSI	*****012**
Serving_eNodeBID	**99129**	Serving_CellID	**1**
eNodeBID_1	**99129**	CellID_1	**1**
RSRP_1	**-93.26**	RSSI_1	**-67.18**
...	...	...	...
eNodeBID_6	**99145**	CellID_6	**5**
RSRP_6	**-90.02**	RSSI_6	**-50.92**

A. Data Description and Notations

Telco localization techniques mostly focus on MR data, which is generated when MDs connect to nearby cells in Telco networks. Generally, the MR data can be categorized into two aspects: (1) connected cells data including cell ids, GPS locations; (2) continuous signal strength data, such as RSRP, RSSI. Table I shows an example of MR record in 4G LTE network. The IMSI records a unique card ID and the MRTime preserves time stamp of once connection. Besides, A piece of MR records up to 6 nearby cells (eNodeBID and CellID) and radio signal strength indicators (RSSI) for each.

REFERENCES

[1] F. Zhu, C. Luo, M. Yuan, Y. Zhu, Z. Zhang, T. Gu, K. Deng, W. Rao, and J. Zeng, "City-scale localization with telco big data," in *Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM 2016, Indianapolis, IN, USA, October 24-28, 2016*, 2016, pp. 439–448.

[2] Y. Huang, F. Zhu, M. Yuan, K. Deng, Y. Li, B. Ni, W. Dai, Q. Yang, and J. Zeng, "Telco churn prediction with big data," in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia, May 31 - June 4, 2015*, 2015, pp. 607–618.

[3] C. Luo, J. Zeng, M. Yuan, W. Dai, and Q. Yang, "Telco user activity level prediction with massive mobile broadband data," *ACM TIST*, vol. 7, no. 4, pp. 63:1–63:30, 2016.

[4] M. Ibrahim and M. Youssef, "A hidden markov model for localization using low-end GSM cell phones," in *Proceedings of IEEE International Conference on Communications, ICC 2011, Kyoto, Japan, 5-9 June, 2011*, 2011, pp. 1–5.

[5] S. Hara, D. Anzai, T. Yabu, T. Derham, and R. Zemek, "Analysis on TOA and TDOA location estimation performances in a cellular system," in *Proceedings of IEEE International Conference on Communications, ICC 2011, Kyoto, Japan, 5-9 June, 2011*, 2011, pp. 1–5.

[6] M. Ibrahim and M. Youssef, "Cellsense: An accurate energy-efficient GSM positioning system," *IEEE Trans. Vehicular Technology*, vol. 61, no. 1, pp. 286–296, 2012.

[7] A. Ray, S. Deb, and P. Monogioudis, "Localization of LTE measurement records with missing information," in *35th Annual IEEE International Conference on Computer Communications, INFOCOM 2016, San Francisco, CA, USA, April 10-14, 2016*, 2016, pp. 1–9.

[8] R. Margolies, R. A. Becker, S. D. Byers, S. Deb, R. Jana, S. Urbanek, and C. Volinsky, "Can you find me now? evaluation of network-based localization in a 4g LTE network," in *2017 IEEE Conference on Computer Communications, INFOCOM 2017, Atlanta, GA, USA, May 1-4, 2017*, 2017, pp. 1–9.

[9] K. Perera, T. Bhattacharya, L. Kulik, and J. Bailey, "Trajectory inference for mobile devices using connected cell towers," in *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, Bellevue, WA, USA, November 3-6, 2015*, 2015, pp. 23:1–23:10.

[10] O. G. Troyanskaya, M. N. Cantor, G. Sherlock, P. O. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman, "Missing value estimation methods for DNA microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2001. [Online]. Available: https://doi.org/10.1093/bioinformatics/17.6.520

[11] L. Gondara and K. Wang, "MIDA: multiple imputation using denoising autoencoders," in *Advances in Knowledge Discovery and Data Mining - 22nd Pacific-Asia Conference, PAKDD 2018, Melbourne, VIC, Australia, June 3-6, 2018, Proceedings, Part III*, 2018, pp. 260–272.

[12] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, 2014, pp. 2672–2680.

[13] Y. Zhang, W. Rao, M. Yuan, J. Zeng, and H. Yang, "Confidence model-based data repair for telco localization," in *18th IEEE International Conference on Mobile Data Management, MDM 2017, Daejeon, South Korea, May 29 - June 1, 2017*, 2017, pp. 186–195.

[14] X. Huang, Y. Li, Y. Wang, X. Chen, Y. Xiao, and L. Zhang, "CTS: A cellular-based trajectory tracking system with gps-level accuracy," *IMWUT*, vol. 1, no. 4, pp. 140:1–140:29, 2017.