k-NN Empowered LoRaWAN Localization for Surface and Underground Scenarios: Work-in-Progress Report

Ekaterina Svertoka[1,2,*], Alexandru Rusu-Casandra[1], Radim Burget[2], Jari Nurmi[3] and Aleksandr Ometov[3]

[1] *Department of Telecommunications, Faculty of Electronics, Telecommunications and Information Technology, University Politehnica of Bucharest, 061071 Bucharest, Romania*

[2] *Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, 616 00 Brno, Czech Republic*

[3] *Unit of Electrical Engineering, Faculty of Information Technology and Communication Sciences, Tampere University, 33720 Tampere, Finland*

Abstract

This short paper explores the usability of Long Range Wide Area Network (LoRaWAN) technology for localization within the context of modern Industry 5.0 wireless networks. Traditional localization methods have often fallen short in providing meaningful accuracy in this domain. Our research addresses this gap by investigating the potential of LoRaWAN for localization, synthesizing key findings and advancements. Two primary contributions are presented: the analysis of two underground LoRaWAN datasets, valuable resources for researchers and practitioners, and the proposal of two innovative k-nearest neighbors (k-NN) algorithms designed to enhance position estimation accuracy through optimized nearest neighbor selection. By integrating preprocessing strategies with these algorithms, an improvement in accuracy of up to 17% is achieved.

Keywords

LoRaWAN, localization, accuracy, Machine Learning, ML, dataset

1. Introduction

Improving safety protocols within industrial landscapes, especially in hazardous settings where humans interface with autonomous machinery, is a continual pursuit reshaping operational paradigms. This development is significantly augmented by the integration of cutting-edge technologies, notably within the realm of Industry 5.0 [1, 2]. A major aspect of this evolution is the amalgamation of the Internet of Things (IoT) [3] and wearable tech, offering transformative potential for improving safety measures in challenging industrial contexts [4].

Wearable devices have gained traction owing to their ability to enhance safety, particularly in rugged environments [5]. However, the quest for enhanced functionality, reduced energy consumption, and miniaturization poses intricate engineering challenges. Among these, localization emerges as a critical domain, especially when considering LoRaWAN technology as means for communication or localization [6].

LoRaWAN, a robust communication solution for IoT applications, including industrial wearables, is very heavily used in IoT deployments [7]. However, its viability for localization applications necessitated deeper scrutiny due to inherent bandwidth limitations, which may impinge upon accuracy. Nevertheless, LoRaWAN's attributes such as low-power operation, expansive coverage, scalability, market availability, and robust propagation capabilities render it an attractive contender for IoT-based localization traditional methodologies like Wi-Fi, BLE, or UWB [8].

WIPHAL 2024: Work-in-Progress in Hardware and Software for Location Computation, June 25-27, 2024, Antwerp, Belgium

*Corresponding author.

esvertoka@elcom.pub.ro (E. Svertoka); alexandru.rusu@upb.ro (A. Rusu-Casandra); burgetrm@vut.cz (R. Burget); jari.nurmi@tuni.fi (J. Nurmi); aleksandr.ometov@tuni.fi (A. Ometov)

0000-0001-6384-5113 (E. Svertoka); 0000-0003-3242-3561 (A. Rusu-Casandra); 0000-0003-1849-5390 (R. Burget); 0000-0003-2169-4606 (J. Nurmi); 0000-0003-3412-1639 (A. Ometov)

This paper delves into the feasibility of harnessing LoRaWAN for localization tasks based on real life datasets, leveraging both preprocessing and postprocessing methodologies to augment accuracy, particularly for wearable devices employed in safety-critical scenarios. Building upon antecedent research that furnished open fingerprinting datasets acquired by the authors in "wild" conditions, this study progresses by proffering two novel k-NN-based algorithm modifications customized to optimize nearest neighbor selection for refined position estimation. Additionally, it reflects on two underground LoRaWAN-based datasets, augmenting algorithmic refinement and serving as invaluable assets for the broader research community following work [8].

The paper assesses LoRaWAN-based localization methodologies, finding conventional techniques wanting, while ML, notably k-NN, substantially curtail errors. Diverse preprocessing strategies further impacting the accuracy, with indoor datasets showcasing superior precision, followed by outdoor and subterranean environments. The proposed k-NN algorithms and associated strategies yield a marked accuracy enhancement of up to 17%. While not rivaling leading-edge technologies in precision, LoRaWAN evinces promise for applications such as logistics and agriculture, demonstrating commendable accuracy under appropriate conditions.

The rest of the paper is organized as follows. Section 2 provide a brief background information. Further, Section 3 outlines the main k-NN related modifications. Next, Section 4 depicts selected numerical results based on real-life datsets for both underground and traditional localization cases. The last section concludes the work.

2. Background

This section draws upon previously gathered open-access LoRaWAN datasets documented in [8] and accessible via [9]. To maintain coherence, we initiate this segment with a concise overview of their configuration, followed by an examination of baseline localization errors and a comparison of preprocessing methodologies.

Two distinct measurement campaigns were conducted at Brno University of Technology (BUT) and Politehnica București National University for Science and Technology (UNSTPB). Each campaign adhered to a consistent two-stage protocol: an "offline" phase encompassing map preparation and LoRaWAN LG308 gateway (GW) deployment, succeeded by an "online" phase involving direct data collection utilizing LoRa Field Test Devices (FTDs) [10, 11]. Detailed procedural information regarding campaign organization is available in [8].

This endeavor yielded seven distinct datasets [9], varying across several key parameters including environment, spread factor (SF), measurement area, number of measurement points (MP), spacing, LoRaWAN gateways (GW), as well as the mean random choice k-NN-obrained error. A summary of pivotal parameters, with comprehensive dataset details available in [9].

- **ds1**: BUT, Building, GW place – Building, Spacing – 1m, GWs – 7, Indoor; Mean error: 3.4m;
- **ds2**: BUT, Parking (fl. -1), GW place – Building, Spacing – 3m, GWs – 7, Underground; Mean error: 11.6m;
- **ds3**: BUT, Parking (fl. -2), GW place – Building, Spacing – 3m, GWs – 7, Underground; Mean error: 13.3m;
- **ds4**: BUT, Parking (fl. -1) and Parking (fl. -2), Spacing – 2.5m, GWs – 6, Underground; Mean error: 5.6m;
- **ds5**: BUT, Parking (fl. -2) and Parking (fl. -2), Spacing – 2.5m, GWs – 6, Underground; Mean error: 7.2m;
- **ds6**: UNSTPB, Building, GW place – Building, Spacing – 1m, GWs – 9, Indoor; Mean error: 3.3m;
- **ds7**: UNSTPB, Alley, GW place – Building, Spacing – 1m, GWs – 8, Outdoor Mean error: 5.3m.

Initially, localization accuracy was approximated using conventional methodologies such as trilateration, Weighted Centroid Algorithm (WCA), and k-Nearest Neighbors (k-NN). Trilateration and WCA, being reliant on receiver positions, were directly applicable without significant data manipulation. Conversely, k-NN necessitated redundancy exclusion stemming from multiple fingerprints per location. In our baseline k-NN implementation, the "random choice" selection strategy was adopted, wherein one fingerprint per location was randomly chosen for subsequent calculations. Following redundancy removal, the remaining data was partitioned into training

(20%) and testing (80%) subsets. Additional insights into the application of these localization algorithms can be found in the seminal work [8].

Localization error estimation employed the Root Mean Square Error (RMSE) metric, widely utilized for accuracy assessment: $RMSE = \sqrt{\frac{\sum^{K}((x_i - \hat{x}i)^2 + (yi - \hat{y}_i)^2)}{K}}$, where x and y denote real coordinates, $\hat{x}$ and $\hat{y}$ represent predicted object coordinates (utilizing Trilateration, WCA, k-NN), and K denotes the number of MPs to be estimated.

Overall, our previous observation proved that trilateration and WCA exhibit notably low accuracy within a few tens of meters, rendering them relatively ineffectual for localization purposes. Conversely, k-NN yields markedly superior and generally promising results across most scenarios, particularly for indoor scenarios (ds1 and ds6), boasting an accuracy up to 10 times greater than methods reliant on GW location.

Consequently, k-NN with random choice was designated as the baseline, with trilateration and WCA omitted from further comparisons to streamline visualization. The baseline findings revealed an average error of 3.4m for the indoor environment and 5.3m for outdoor scenarios.

In pursuit of heightened localization precision, we conducted experiments exploring various redundancy reduction techniques essential for k-Nearest Neighbors (k-NN). In addition to the "random choice" strategy, we investigated "averaging" (based on averaged fingerprints per location) and "maximum" (relying on fingerprints with the highest average Received Signal Strength Indicator (RSSI) level) methodologies. As noted in [8], while discernible disparity between the two strategies is minimal across the datasets, a marginal superiority of the "averaging" strategy. Notably, the "averaging" strategy exhibits preeminence in underground datasets characterized by signal attenuation and general instability. The calculations were conducted across five datasets, excluding underground ds2 and ds3, thus exhibiting slight favoritism towards the "maximum" strategy.

Despite surpassing initial expectations, the current localization accuracy estimates by LoRaWAN still fall short of leading technologies such as Ultra-Wideband (UWB) or Global Positioning System (GPS). Consequently, we continued our quest for accuracy refinement. The noted in [8] experiments involved adjustments to measurement campaign parameters (Spread Factor, number of Gateways, spacing) and machine learning algorithms. Our analysis corroborated k-NN as the most accurate method for our specific problem, prompting further research extensions detailed in the subsequent sections.

3. Description of the ML adjustments

Upon scrutinizing the dataset using k-Nearest Neighbors (k-NN), a discernible observation emerges: the shortest distances in the distance matrix may correspond to measurement points (MPs) quite distant from the point under estimation. This phenomenon, extensively elaborated in our prior conference paper dedicated to LoRaWAN-based outdoor localization [12], primarily signifies the low stability of the signal map, resulting in erroneous neighbor selection. To mitigate this drawback, an enhancement strategy can be integrated into the k-NN algorithm, aimed at reducing its impact.

The *Modification 1* strategy proposes to reassess the relevance of selected nearest neighbors based on the Euclidean distance between them and the origin points. Here, one of the corners of the measurement area perimeter serves as the origin point. The rationale behind this approach posits that the nearest neighbor group should ideally reside within the same vicinity. Consequently, the proposition involves computing Euclidean distances from the origin to each neighbor, delineating an approximate localization zone based on mean distance, and discarding outliers, i.e., neighboring points outside this zone.

The *Modification 2* follows a similar trajectory, albeit with a slight variation. Instead of computing Euclidean distance to the origin point, it's calculated from each neighbor to the preliminary estimated value provided by k-NN. Consequently, the localization zone is established by one border, extending from 0 to the mean value of Euclidean distances plus a designated allowance.

However, *Modification 1* acknowledges certain limitations, particularly its efficacy at low k values, predicated on the assumption that the outlier count in the group is fewer than relevant neighbors. Nevertheless, a potential issue arises: the likelihood of a larger outlier group compared to 'valid' neighbors, i.e., those within the localization zone. To address this, a condition is proposed wherein irrelevant neighbors are eliminated only if their count n is three times less than the total neighbor count k: $n \leq k/3$.

For the *Modification 2*, this work investigates two options: *Modification 2.1* employs unstable k when the number of the nearest neighbors, based on which the estimated position is calculated, is not replenished after the elimination of outliers. *Modification 2.2* uses stable k when the number of removed outliers n is replenished by the next neighbors from the distance matrix d.

4. Numerical results

To estimate the general gain from the modifications proposed in this work, it is necessary to compare the strategies against the baseline using preprocessing, i.e., averaging redundancy reduction, which has proven to be the most effective. The results for this setup are presented in Figure 1.

The results are summarized in Table 1 show that *Modification 2* remains the most successful, with which it was possible to achieve an improvement in localization quality by 17.2% on average across all datasets, while for ds1 and ds5 this number exceeds 25%. Among the datasets, the accuracy improvement exceeds 13% for all of them except ds2 and ds3 – two underground datasets with non-local placement of the GWs.

Table 1
Relative Advantages of proposed strategies with preprocessing over the baseline

Strategy	ds1	ds2	ds3	ds4	ds5	ds6	ds7	$\bar{x}$
1	+25.2%	+7.3%	+6.0%	+19.0%	+23.5%	+15.1%	+21.6%	**+16.8%**
2.1	+25.9%	+7.3%	+10.3%	+19.0%	+25.6%	+13.4%	+19.1%	**+17.2%**
2.2	+15.9%	+4.2%	+5.3%	+19.0%	+25.6%	+13.4%	+21.6%	**+15.0%**

The charts indicate significant improvements in LoRaWAN localization accuracy, up to 17.2% on average, with the implementation of a simple redundancy reduction technique and proposed algorithms. While underground datasets posed challenges, indoor and outdoor scenarios showed results comparable to Wi-Fi and GPS, respectively, albeit within a few meters of accuracy. This level of precision may suffice for many industrial applications where extreme precision is not critical, including logistics, agriculture, and smart factories.

Basic k-NN emerged as the most precise among the tested ML algorithms. Among the proposed accuracy-enhancing strategies, reassessment from the estimated value with stable k (*Modification 2.2*) yielded the most significant increase in accuracy, averaging 17.2% with preprocessing. Other strategies also performed well, with an average performance exceeding 15%.

Notable error reductions were observed in indoor (ds1), outdoor (ds7), and certain underground datasets (ds4 and ds5), with over 6% reductions seen in the remaining datasets.

While certain conditions, like a concentration of 7 GWs per 50,000 square meters with uniform distribution, seem necessary for indoor and outdoor scenarios, the implications for underground environments remain uncertain. Although the presented underground datasets exhibited low localization accuracy, it's unclear if this is representative of the broader environment or specific to the measurement campaign.

Despite proposed accuracy improvements, LoRaWAN-based localization still lags behind leading technologies. Nonetheless, it may find utility in sectors where extreme precision is not essential, leveraging its existing infrastructure for communication in logistics, agriculture, or smart factories.

5. Conclusions

To draw the intermediate conclusion, we start with the trilateration and WCA methodologies that exhibit inefficiency, yielding an approximate error of 30 meters. Conversely, Machine Learning (ML) algorithms showcase significant error reduction, often by several folds. Notably, k-Nearest Neighbors (k-NN) emerges as the most precise ML technique, reducing mean localization error by up to a factor of 10 compared to trilateration. Among the k-NN extensions proposed in this paper, leveraging reassessment from the estimated position with unstable k shows promise.

Baseline k-NN, employing random fingerprint selection, is eclipsed by "averaging" and "maximum" preprocessing strategies in spite of redundancy reduction strategy. Of the two, the "averaging" strategy slightly outperforms the "maximum" approach. Optimal strategy selection hinges on dataset signal levels: "averaging" proves more effective for low signal levels, while "maximum" is preferable otherwise.

Notably, superior accuracy is observed in indoor datasets ($\sim$ 2.6 m), followed by outdoor ($\sim$ 4 m) and underground datasets ($\sim 5 - 12$ m).

In conclusion, this paper represents a culmination of efforts in exploring LoRaWAN's localization potential. It introduces new underground datasets, contributing to algorithm refinement, and proposes novel k-NN-based algorithms aimed at enhancing position estimation accuracy. The integration of these algorithms with redundancy reduction strategies yields a notable accuracy improvement of up to 17%. While not competing with leading precision technologies, LoRaWAN-based localization remains viable for specific applications such as logistics, agriculture, or smart factories, where extreme precision is not paramount.

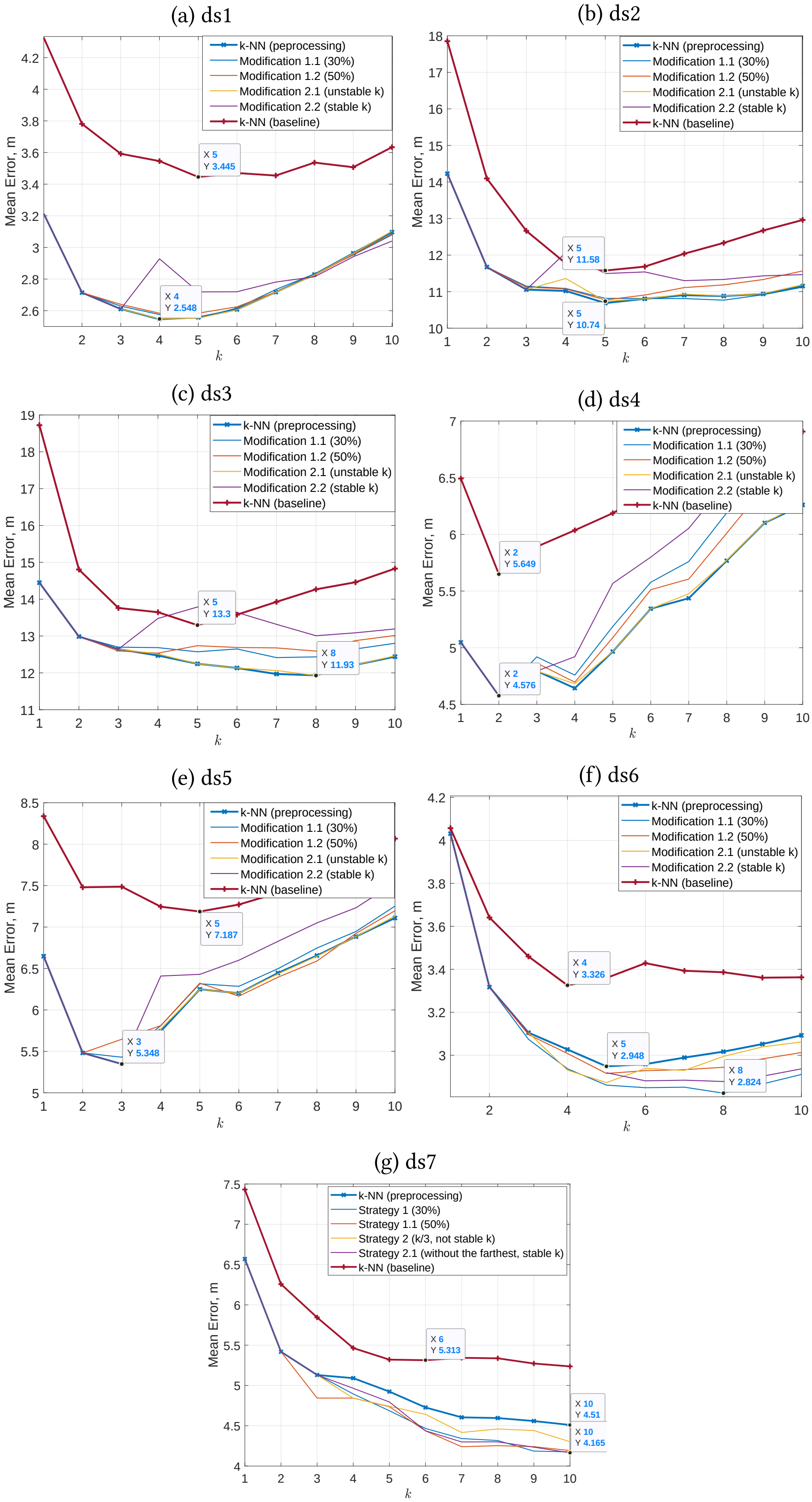

Figure 1: Result of applying modifications (k-NN with preprocessing)

Acknowledgments

The authors gratefully acknowledge funding from European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska Curie grant agreements No. 813278 (A-WEAR), 956090 (APROPOS), and support grant PN-III-P3-3.6-H2020-2020-0124 from the Romanian Ministry of Research.

References

[1] N. Rane, Transformers in Industry 4.0, Industry 5.0, and Society 5.0: Roles and Challenges (2023).

[2] H. R. Chi, C. K. Wu, N.-F. Huang, K.-F. Tsang, A. Radwan, A Survey of Network Automation for Industrial Internet-of-Things toward Industry 5.0, IEEE Transactions on Industrial Informatics 19 (2022) 2065–2077.

[3] J. Hosek, P. Masek, S. Andreev, O. Galinina, A. Ometov, F. Kropfl, W. Wiedermann, Y. Koucheryavy, A SyMPHOnY of Integrated IoT Businesses: Closing the Gap between Availability and Adoption, IEEE Communications Magazine 55 (2017) 156–164.

[4] P. Coelho, C. Bessa, J. Landeck, C. Silva, Industry 5.0: The Arising of a Concept, Procedia Computer Science 217 (2023) 1137–1144.

[5] A. Moradbeikie, A. Keshavarz, H. Rostami, S. Paiva, S. I. Lopes, Improving LoRaWAN RSSI-based Localization in Harsh Environments: The Harbor Use Case, in: World Conference on Information Systems and Technologies, Springer, 2023, pp. 14–25.

[6] H. Fugate, H. Alzraiee, Quantitative Analysis of Construction Labor Acceptance of Wearable Sensing Devices to Enhance Workers' Safety, Results in Engineering 17 (2023) 100841.

[7] I. Aqeel, E. Iorkyase, H. Zangoti, C. Tachtatzis, R. Atkinson, I. Andonovic, LoRaWAN-implemented Node Localisation based on Received Signal Strength Indicator, IET Wireless Sensor Systems 13 (2023) 117–132.

[8] E. Svertoka, A. Rusu-Casandra, R. Burget, I. Marghescu, J. Hosek, A. Ometov, LoRaWAN: Lost for localization?, IEEE Sensors Journal 22 (2022) 23307–23319.

[9] Collected Open Access Datasets, 2024. Available online: https://drive.google.com/drive/folders/15j0itClvJuGviUhXWimZutIC980rehSx (accessed on 15.04.2024).

[10] Dragino, LG308 LoRaWAN Gateway User Manual.Document Version: 1.4.0 , 2022. Available online: https://www.dragino.com/downloads/downloads/LoRa_Gateway/LG308-LG301/ (accessed on 15.04.2024).

[11] Adeunis, FTD: network tester, 2024. Available online: https://www.adeunis.com/en/produit/ftd-network-tester/ (accessed on 15.04.2024).

[12] E. Svertoka, I. Marghescu, A. Rusu-Casandra, R. Burget, J. Hosek, P. Masek, A. Ometov, Evaluation of Real-Life LoRaWAN Localization: Accuracy Dependencies Analysis Based on Outdoor Measurement Datasets, in: Proc. of 14th International Conference on Communications (COMM), IEEE, 2022, pp. 1–7.