Comprehensive survey of similarity measures for ranked based location fingerprinting algorithm

Comprehensive Survey of Similarity Measures for Ranked Based Location Fingerprinting Algorithm

Georgy Minaev
Tampere University of Technology
Tampere, Finland
Email: Georgy.Minaev@student.tut.fi

Ari Visa
Tampere University of Technology
Tampere, Finland
Email: Ari.Visa@tut.fi

Robert Piché
Tampere University of Technology
Tampere, Finland
Email: Robert.Piche@tut.fi

***Abstract*—Ranked Based Fingerprinting uses only ordering indices instead of actual Wi-Fi RSS values in order to make the algorithm insensitive to devices. A key component of the RBF algorithm is a similarity measure which is used to compare and find the closest ranked fingerprints. Previous papers study a few similarity measures; here we study 49 similarity measures in a test with a benchmark with publicly available indoor positioning database. For different similarity measures the positioning accuracy varies from 15.80 m to 55.22 m.**

The top 3 similarity measures are Lorentzian, Hamming and Jaccard. Hamming and Jaccard similarity measures have been studied in other papers while Lorenzian had not been studied with that kind of problems.

Keywords - Fingerprinting; Indoor positioning; localization

I. Introduction

The tradition Fingerprint method is described in [7]. The method uses offline and online stages. The Radio Map is built in offline phase. A floor plan is divided into points of interest (cells). The RSS values from Access Points (APs) are collected inside each cell and stored into a Radio Map, so every fingerprint of the Radio Map keeps information about the location in a cell and RSS values list.

In the online phase the location is estimated by comparing the Radio Map with the received vector of RSS values. In the k-NN method the nearest location is that one which gives the best agreement between the Radio Map vectors and the received vector of RSS values.

The weakness of the tradition method is that in the online phase devices are not calibrated, so the devices are capable to report different RSS values for the same Wi-Fi station [12]. Non calibration device means the devices return different RSS values. Even a radio map can be obtained by crowdsourcing with non calibrated devices. The position accuracy will also suffer in that case.

The reader should keep in mind that people in area or different temperatures and humidity [13] are capable to affect RSS values returned from an AP.

The Rank based fingerprint method (RBF) was published in 2011 [1]. It is based on the insight that the rank or ordering of detected AP ids that are detected at given location, sorted by RSS values, should remain the same between different devices. Because RBF method does not use RSS values directly, it should be more robust to diversity effects.

The RSS values are measured from different Wi-Fi stations. The registered vector is sorted from the strongest to the weakest RSS values in the Rank Based Fingerprinting algorithm [1]. The rank fingerprinting uses ordering of ranks but not RSS values, so the algorithm is invariant to bias and scaling. Finally, every detected Wi-Fi station obtains the new ranks based on its position in the sorted vector.

The new vectors are compared with vectors which are stored in the radio map. The radio map is a set of sorted vectors with Ranks. The sorted vectors are measured in the offline phase and stored into the radio map.

The paper [1] studies 5 similarity measures to quantify nearness between the rank vectors. The paper [8] introduces another similarity measure and finds a big improvement. The question, whether a better similarity measure, which can find nearness between ranked fingerprint even better, arises. The paper [5] studies 45 different distance/similarity measures. These are similarity measures between probability density functions, but they can also be applied to compare rank fingerprints. The purpose of our paper is to investigate the use of the methods which are described in [5] to find similarities between the rank vectors.

Section II describes different indoor positioning algorithms, which are based on the RBF. Section III describes the RBF method which is used in the paper. The similarity measures are described in section IV. The baseline traditional fingerprint positioning method is described in section V. Section VI describes the experimental results. Finally, section VII discusses the results and concludes the paper.

II. Related Work

The new Radio Map Reduction algorithm was introduced in 2012 [10]. The main goal of the algorithm is reducing computational complexity of the RBF localization algorithm. The paper [10] makes an assumption that the number of reference points in the database is the key factor of computational complexity. Some reference points are removed because they are too far from the position of a mobile device. The position of the doors between the rooms is also kept in the database. The algorithm consists of the range estimation and reference points extraction parts.

The new optimization of the RBF localization algorithm using Kalman Filter was introduced in 2013 [11]. The algorithm uses the Rank Based Fingerprinting algorithm and Kalman Filter together.

The ”improved Wi-Fi Indoor Positioning Method via Signal Strength Order Invariance” was introduced in 2014 [9]. The method improves the positioning accuracy and is implemented with a new similarity measure. The similarity measure described in the paper [9] cannot be implemented because of missing parameter a in explanation.

Finally, RBF with similarity measure based on Spearman rank correlation was introduced in [14]. In tests reported in [8] this method was found to be better than other RBF similarity measures. The Spearman rank corrrelation coefficient is calculated according to

$$\rho = \frac{\sum_{n=1}^{N_c}[(V_F(n) - \bar{R}_F)(V_S(n) - \bar{R}_S)]}{\sqrt{\sum_{n=1}^{N_c}(V_F(n) - \bar{R}_F)^2 \sum_{n=1}^{N_c}(V_s(n) - \bar{R}_S)^2}} \quad (1)$$

where V_S and V_F are ranked fingerprints, N_c is the number of APs. $\bar{R}_F$ and $\bar{R}_S$ are calculated according to

$$\bar{R}_F = \frac{1}{N_c}\sum_{n=1}^{N_c} V_F(n) \quad (2)$$

$$\bar{R}_S = \frac{1}{N_c}\sum_{n=1}^{N_c} V_S(n) \quad (3)$$

The distance between two fingerprints is calculated according to

$$d = 1 - \rho \quad (4)$$

III. Rank Based Fingerprinting

Here are the details of our implementation of RBF and the testing procedure.

The RBF method, which is used, is based on the method described in the paper [1]. The radio map consists of the rank values and the coordinates of locations.

The algorithm first computes the similarities between the rank fingerprint of the unknown location and the rank fingerprints in the radio map. The k-NN (k-Nearest Neighbors) algorithm is then used to estimate the device position as follows. The k ranked fingerprints with the smallest distances is used to estimate the new location. The new location is obtained by using mean of the k most similar rank fingerprint locations. Finally, the error is calculated by finding Euclidean distance between the estimated position and the tested position.

IV. Similarity Measures

The RBF algorithm uses similarity measures to compare ranking vectors. All lists of IDs of heard APs, sorted by RSS are transformed to vectors of the same length. The rank vectors contain only IDs of APs with the highest RSS values, sorted in descending order by RSS.

The methods from 11 to 48 in table II are implemented directly with equations in the paper [5].

TABLE I
The best 5 RBF similarity measures

11	Lorentzian	$d = \sum_{i=1}^{n} \ln(1+ \mid P_i - Q_i \mid)$	[5]
8	Hamming	$d = (\#(P_i \neq Q_i)/n)$	Matlab
9	Jaccard	$d = \frac{\#[(P_i \neq Q_i) \cap ((P_i \neq 0) \cup (Q_i \neq 0))]}{\#[(P_i \neq 0) \cup (Q_i \neq 0)]}$	Matlab
24	Wave Hedges	$d = \sum_{i=1}^{n} \frac{\mid P_i - Q_i \mid}{\max(P_i, Q_i)}$	[5]
18	Canberra	$d = \sum_{i=1}^{n} \frac{\mid P_i - Q_i \mid}{P_i + Q_i}$	[5]

The methods from 1 to 9 in table II are implemented with Matlab Statistics and Machine Learning Toolbox 2017a function `pdist`.

The method 10 in table II is Cross-correlation. The method 10 is implemented with Matlab function `xcorr`, as it is shown in the code fragment:

```
[acor,lag] = xcorr(p,q);
distance = - max(acor);
```
(5)

where p and q are rank fingerprints.

The method 49 in table II is Spearman rank correlation coefficient [14].

The best 5 methods are shown in table I.

None of the methods has parameters except Minkowski distance. The default parameter is $P = 1$. The parameter $P = 2$ shows the same RMSE error as Euclidean distance. The Euclidean distance shows slightly better results, than Minkowski with $P = 3, 4, 5, 6$.

V. Comparison baseline

The Baseline fingerprinting method is needed to show which similarity measures are acceptable. A positioning algorithm should show a better result than the baseline method.

The Baseline method is chosen as described in the article [4]. All RSS values are first transformed to new ones according to

$$\text{NewValue} = \begin{cases} 0, \text{ if RSS value is missing} \\ \text{RSS} - \min + 1, \text{ otherwise} \end{cases} \quad (6)$$

where min is the minimum RSS value in the database.

The baseline positioning algorithm is the 1-NN as it is described in the paper [4]. The baseline algorithm is used to estimate the location of each test fingerprint.

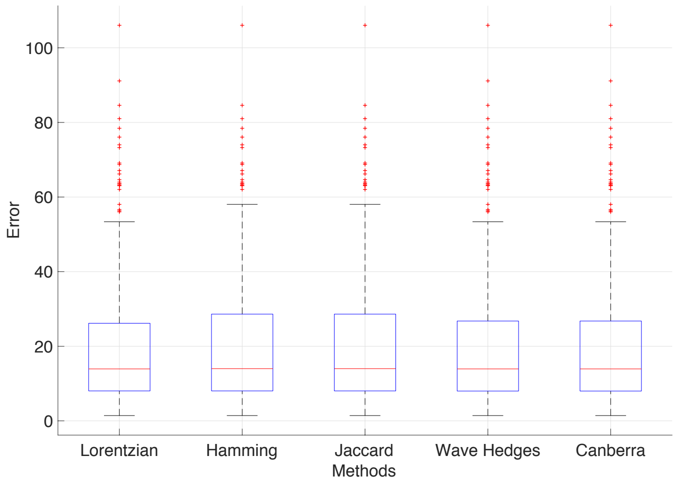

Fig. 1. Simulation results, where $k = 1$ and number of APs $N = 1$.

Here fingerprint is a vector of length of total number of Wi-Fi stations in the database. Values of the vector are RSS values, updated according to the equation (6).

The distance between two fingerprints p, q is calculated according to

$$d(p, q) = \sqrt{\sum_{i=1}^{N}(p_i - q_i)^2} \tag{7}$$

where N is the total number of APs in the database.

VI. Experimental Results

The database is open [15], and is described in the articles [2], [3]. The data was collected in two buildings in Tampere University of Technology with a Windows Tablet. We study the data obtained in Building 1 of size approximately 141 m to 54 m. It has 1479 fingerprints over 4 floors.

The paper studies 80% of randomly selected of database's fingerprints (a sorted list of heard APs with the measurement location) as Radio map. The remaining fingerprints are used to simulate the online measurements.

RMSE errors of different methods are shown in table II.

The best 5 results based on RMSE and $k = 5$ are Lorentzian, Hamming, Jaccard, Wave Hadges, Canberra distances.

The methods Kullback-Leibler, Cross-Corellation, Inner Product, K divergence and Bhattacharyya had the highest RMSE, they were beaten by the baseline method's RMSE (52.94 m).

The simulations show that Lorentzian distance similarity measure is slightly more accurate than the similarity measures used in [1], [8].

The tests were also made using componentwise median instead of mean. The simulations give almost the same results. The results might be different with more noisy data.

Figure 1 refers to the results with $k = 1$ and $N = 1$.

Figure 2 refers to the results with $k = 3$ and $N = 3$.

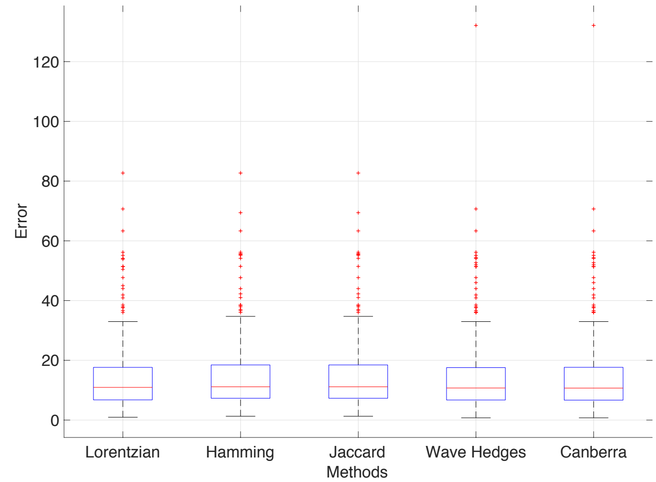

Fig. 2. Simulation results, where $k = 3$ and number of APs $N = 3$.

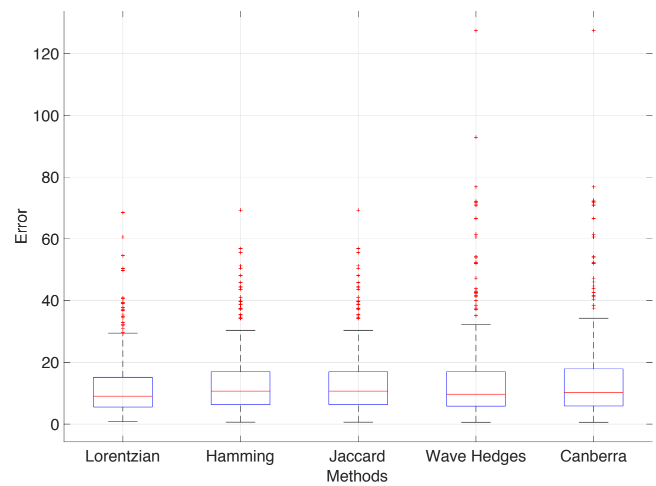

Fig. 3. Simulation results, where k=5 and number of APs $N = 5$.

Figure 3 refers to the results with $k = 5$ and $N = 5$.

Measurements show that the number of ranks significantly affects the RMSE errors. While, RMSE error is not changed significantly with increasing k-NN value from 3 to 5.

Ten replications show that the Lorentzian measure is the best in all simulations with 80% of training data, $N = 5$. Figure 4 refers to the 10 results of Lorentzian measure with number of APs $N = 5$.

VII. Discussion and Conclusion

The effect of different similarity measures on RBF algorithm for indoor navigation is shown. Some similarity measures show better position prediction accuracy, while some measures do not pass the baseline criteria.

The RMSE results of Ranked Fingerprinting are compared with traditional Fingerprinting algorithm, described in Comparison baseline section V.

The best similarity measure is Lorentzian.

The experiments show that the number of APs which are used in similarity calculation affects the position accuracy

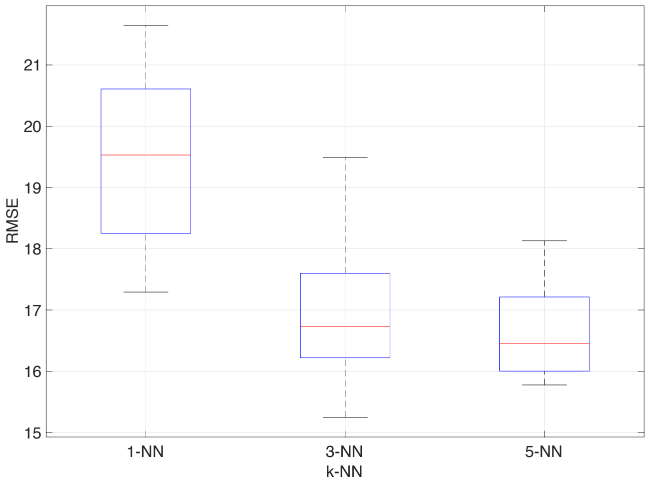

Fig. 4. 10 simulation results of Lorentzian measure, with number of APs $N = 5$.

significantly. While, increasing the k-NN number more than 3 does not affect significantly the position accuracy.

There are several places for improvement. The list of similarity measures could be extended. The similarity measures, which showed the best position prediction accuracy results, should be tested with different methods, for instance, with Kalman Filter as it is shown in the paper [11].

TABLE II
RMSE ERROR OF DIFFERENT METHODS, WITH THE MINIMUM NUMBER OF APS EQUAL TO 5 AND K = 5

id	Measure	1-NN (m)	mean 5-NN (m)	median 5-NN (m)
11	Lorentzian	**18.18**	**15.80**	**17.51**
8	Hamming	23.14	17.571	17.89
9	Jaccard	23.14	17.57	17.89
24	Wave Hedges	21.09	21.59	26.53
18	Canberra	21.30	21.98	26.91
38	Clark	21.30	21.98	26.91
2	Cityblock	20.44	22.43	26.80
3	Minkowski	20.44	22.43	26.80
15	Gower	20.44	22.43	26.80
19	Intersection	20.44	22.43	26.80
12	Sørensen	20.39	22.43	27.00
16	Soergel	20.39	22.43	27.00
17	Kulczynski	20.39	22.43	27.00
20	Czekanowski	20.39	22.43	27.00
21	Motyka	20.39	22.43	27.00
22	Ruzicka	20.39	22.43	27.00
23	Tanimoto	20.39	22.43	27.00
37	Divergence	23.84	24.04	28.47
34	Neyman χ^2	22.73	24.06	28.99
36	Probabilistic Symmetric χ^2	23.73	24.51	29.21
43	Jensen-Shannon	23.73	24.56	29.28
42	Topsøe	23.73	24.56	29.28
44	Jensen difference	23.73	24.57	29.28
40	Jeffreys	23.73	24.63	29.30
45	Taneja	23.84	24.65	29.32
39	Additive Symmetric χ^2	24.15	24.78	29.46
33	Pearson χ^2	25.85	24.86	30.04
46	KumarJohnson	24.48	24.90	29.50
13	Dice	23.69	25.00	30.55
1	Euclidean	22.53	25.50	30.43
29	Hellinger	27.87	25.64	27.96
30	Matusita	27.87	25.64	27.96
31	Squared-chord	27.87	25.64	27.96
35	Squared χ^2	22.86	25.74	30.16
47	Avg	23.20	25.76	29.39
4	Chebychev	23.91	26.77	31.48
32	Squared-Euclidean	36.72	29.04	30.04
5	Cosine	34.43	31.61	34.21
49	Spearman Rank Correlation Coefficient	46.60	38.23	41.95
6	Correlation	46.49	38.27	41.91
7	Spearman	49.58	39.04	43.80
26	Harmonic Mean	55.19	49.38	48.46
27	Fidelity	50.60	50.39	50.84
14	Kumar-Hassebrook (PCE)	53.66	52.98	53.57
48	Kullback-Leibler	56.88	55.16	56.75
10	Cross-correlation	57.04	55.17	56.75
25	Inner Product	57.05	55.17	56.75
41	K divergence	56.93	55.21	56.77
28	Bhattacharyya	56.97	55.22	56.77

REFERENCES

[1] J.Machaj, P. Brida and R.Piché *Rank based fingerprinting algorithm for indoor positioning*, in International Conference on Indoor Positioning and Indoor Navigation (IPIN) 2011 pp.1-6.

[2] S. Shrestha, J. Talvitie, and E.S. Lohan, *Deconvolution-based indoor localization with WLAN signals and unknown access point locations*, in Proc. of IEEE ICL-GNSS, Jun2013, Italy

[3] J.Talvitie,E.S.Lohan,and M. Renfors, *Effect of Coverage Gaps and Measurement Inaccuracies in Fingerprinting based Indoor Localization*, in Proc. of IEEE ICL-GNSS conference, Jun 2014, Helsinki, Finland

[4] Joaquin Torres-Sospedra, Raul Montoliu, German M. Mendoza-Silva, Oscar Belmonte, David Rambla, and Joaquin Huerta *Providing Databases for Different Indoor Positioning Technologies: Pros and Cons of Magnetic Field and Wi-Fi Based Positioning*, Hindawi Publishing Corporation, Mobile Information Systems, Volume 2016, Article ID 6092618, 22 pages, http://dx.doi.org/10.1155/2016/6092618

[5] Cha, S.H., 2007. *Comprehensive survey on distance/similarity measures between probability density functions*.

[6] T. M. Cover and P. E. Hart, *Nearest Neighbor Pattern Classification*,IEEE Transactions on Information theory, vol. 13, no. 1, pp. 2127, 1967.

[7] Ville Honkavirta, Tommi Perala, Simo Ali-Loytty, and Robert Piché. *A Comparative Survey of WLAN Location Fingerprinting Methods*, In WPNC, volume 2009, 2009

[8] Xie, Yaqin, Wang, Yan; Nallanathan, Arumugam; Wang, Lina.*An Improved K-Nearest-Neighbor Indoor Localization Method Based on Spearman Distance*, In: IEEE Signal Processing Letters, Vol. 23, No. 3, 03.2016, p. 351-355.

[9] Jie Zhuang, Jiadong Zhang, Demin Zhou, Hong Pang, and Wei Huang *An Improved Wi-Fi Indoor Positioning Method via Signal Strength Order Invariance*, In IEEE International Conference on Computer and Information Technology, 2014

[10] J. Machaj, P. Brida *Optimization of Rank Based Fingerprinting Localization Algorithm*, 2012 International Conference on Indoor Positioning and Indoor Navigation

[11] J. Machaj, P. Brida and J. Benikovsky *Optimization of the RBF Localization Algorithm Using Kalman Filter*, 2013 Telecommunications and Signal Processing (TSP), 2013 36th International Conference on

[12] Christos Laoudias , Robert Piché and Christos G. Panayiotou (2013) Device self-calibration in location systems using signal strength histograms, Journal of Location Based Services, 7:3, 165-181, DOI: 10.1080/17489725.2013.816792

[13] ITU-R P.676-3 *Attenuation by Atmospheric Gases*

[14] R. Saraiva Campos, L. Lovisolo, *RF Fingerprinting Location Techniques*, in S.A.Zekavat, R.M.Buehrer, eds., *Handbook of Position Location: Theory, Practice and Advances*, Hoboken, NJ, USA: Wiley, 2011.

[15] E.S.Lohan, J.Talvitie, *WLAN RSS indoor measurement data*, Tampere University of Technology, 2015 `www.cs.tut.fi/tlt/pos/MEASUREMENTS_WLAN_FOR_WEB.zip`