How to Find Appropriate Automobile Exhibition Halls: Towards a Personalized Recommendation Service for Auto Show

Danhuai Guo
Computer Network Information Center
Chinese Academy of Sciences
Beijing, China
e-mail: guodanhuai@cnic.cn

Yingqiu Zhu
Computer Network Information Center
Chinese Academy of Sciences
Beijing, China
e-mail: inkzhu@ruc.edu.cn

Wei Xu
School of Information Renming
University of China
Beijing, China
e-mail: weixu@ruc.edu.cn

Yong Ge
College of Computing and Informatics
UNC Charlotte
Charlotte, NC, USA
e-mail: yge1@uncc.edu

Yuanchun Zhou
Computer Network Information Center
Chinese Academy of Sciences
Beijing, China
e-mail: zyc@cnic.cn

Jianhui Li
Computer Network Information Center
Chinese Academy of Sciences
Beijing, China
e-mail: lijh@cnic.cn

***Abstract*—This paper proposes a novel recommendation service to help visitors to find their proper automobile exhibition halls for auto show. In the proposed method, both temporal and spatial features of visitors are first considered to construct their profiling, and then extract their interests based on visitors' clustering. Finally, highly desired exhibition halls are personalized recommended to proper visitor. The proposed recommender system consists of three modules including relevance module, quality module and integration module. The relevance module is developed to measure the relationship of an automobile exhibition and a visitor, while the quality module is constructed to analyze the quality of each automobile exhibition. The integration module is to combine two modules above for appropriate automobile exhibition. The proposed approach is well validated using a real world dataset, and compared with several baseline models. Our experimental results indicate that in terms of the well-known evaluation metrics, the proposed method can achieves more useful and feasible recommendation results, and our finding highlights that the proposed method can help both visitors to find a more appropriate automobile exhibition halls, and manage officers to reduce more management cost.**

Keywords—Auto show, automobile exhibition halls, recommendation, profiling

I. Introduction

Service recommendation is critically important in both online and offline situations, such as movie recommendation, music recommendation, conference recommendation and auto recommendation. For online information recommendation, comparing with traditional content-based and collaborative filtering-based recommendation [1-4], hybrid models with social factors in big data contexts have been emerged in Web 2.0 age [5-7]. For example, Meng et al. [5] proposed a keyword-aware service recommendation model in big data context. Yin et al. [6] considered both location and content for travel recommendation, while Zhang et al. [7] offered a hybrid content and relation analysis for document recommendation. Recently, with the rapid development of Internet of Thing (IoT), offline situations, such as conference recommendation and auto recommendation, have been got researchers' attention. For instance, Asabere et al. [8] offered an improved session recommendation for a given conference with social information. Wongchokprasitti et al. [9] collected the wisdom of the user community, and recommend sessions for researchers.

Like attending conference sessions, watching auto show is a popular activity but with time constraint. So, how to choose appropriate automobile exhibition halls in right time is a significant but important task. The auto show organizer will provide visitors maps of location of auto brands and timetable of performance such as launch of new brand or new car and model show for general purpose. Due to vary of motivation of attending auto show, the general services provided cannot cover the personal demands.

Being restricted by the price or precision of in-door positioning technique, such as Time of Arrival (TOA) [10], Time Difference of Arrival (TDOA) [11], Angle of Arrival (AOA) [12], location fingerprinting [13-14], the current path commendation fail to provided accurate scheme both in spatial and temporal dimension. The WiFi-based indoor localization method, which uses Received Signal Strength Indicator (RSSI) measures to estimate the distance between receiver and access point, calculates the location of receiver by the RSSI from access points. Due to the low price, it has been widely used in many systems even some commercial applications, such as RADAR [15].

In this paper, we propose a novel personalized recommendation service for auto show recommendation by WiFi-based indoor localization technique. In the proposed method, both temporal and spatial features of visitors are

978-1-4673-8493-3/15 $31.00 © 2015 IEEE
DOI 10.1109/ICDMW.2015.171

first considered using RSSI to construct their profiling, and then extract their interests based on visitors' clustering. What is more, location-aware features are also considered for recommendation. Finally, highly desired exhibition halls are recommended to each visitor.

The proposed recommender system consists of three modules including relevance module, quality module and integration module. The relevance module is developed to measure the relationship of an automobile exhibition and a visitor, while the quality module is constructed to analyze the quality of each automobile exhibition. The integration module is to combine two modules above for appropriate automobile exhibition.

The rest of paper is organized as follows. Section 2 proposes a novel temporal-spatial recommendation method by user clustering. Then the proposed demo is implemented in Section 3, and the proposed method is well verified in Section 4. Section 5 summarized the whole paper and provides the drawbacks and future work of this research.

II. PRELIMINARIES

A. Mechanism for Positioning

The Wi-Fi received signal strength (RSS) which we used in indoor localization problem is impressionable by environmental change which makes the indoor positioning methods an extremely challenging problem. For the plurality of RSSI values received from the same node, due to various disturbances, there must be an error caused by a small probability event. We select the RSSI value with high probability interval by the Gaussian model as valid value, and then get the geometric mean value. This method can effectively reduce the effect of the data of small probability and large disturbances on the overall measurement data and improve the accuracy of positioning.

$$\mu=\frac{1}{n}\cdot\sum_{k=1}^{n}\mathrm{RSSI}_{(k)} \quad (1)$$

$$\sigma=\sqrt{\frac{1}{n-1}\cdot\sum_{k=1}^{n}(\mathrm{RSSI}_{(k)}-\mu)^2} \quad (2)$$

We use Gaussian filter to choice the RSSI values within the interval $(\mu-\sigma\leq \mathrm{RSSI}\leq\mu+\sigma)$

$$\overline{\mathrm{RSSI}}=\frac{1}{N}\cdot\sum_{k=1}^{n}\mathrm{RSSI}_{(k)},\ (\mu-\sigma\leq \mathrm{RSSI}_{(k)}\leq\mu+\sigma) \quad (3)$$

Gaussian model solves the RSSI which has been subject to interference and poor stability in the actual test and improves the positioning effect.

Besides, we use the proposed Linear Coding Scheme [19] to remove the RSS difference due to the device heterogeneity which is proved to achieve better localization accuracy than methods which do not used this algorithm.

After that, the moving of visitor was then transferred to spatio-temporal series of position with location and time. {ID, X1,Y1,T1,X2,Y2,T2⋯}

B. Definition of visiting of interest

By spatial overlay with location the exhibition stands, the spatio-temporal position series can be transferred to temporal series of exhibition stands of each visitors. The stay in specific auto band exhibition stand will be regarded by valid visiting of interest auto band if the stay time greater than threshold (3 minutes was adopted)

ID, T1, Band*, T2, Band**⋯

III. RELEVANCE ANALYSIS MODULE

Relevance module is widely applied in recommendation. In the relevance analysis, we consider similarities among visitors' behaviors using clustering. First of all, visitors with similar behaviors are clustered using K-Means. Second, collaborative filtering technique is employed to recommend proper auto show.

A. K-means clustering

For each visitor, his moving paths and stay time in exhibition stands have been recorded since he entered the hall until the end of his visit. To identity a visitor's interests, the durations of his stays in exhibition stands are considered. Then a visitor-stand matrix is built to determine the relationships between visitors and brand stands:

$$\begin{bmatrix} vt_{11} & \cdots & vt_{1j} & \cdots & vt_{1p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ vt_{i1} & \cdots & vt_{ij} & \cdots & vt_{ip} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ vt_{n1} & \cdots & vt_{nj} & \cdots & vt_{np} \end{bmatrix}$$

where vt_{ij} denotes the i th visitor's stat time in the j th stand.

Based on the visitor-stand matrix, visitors are clustered using K-means. For a set of visitors with their stay time series $\{T_1, T_2, \cdots, T_n\}$, the K-means clustering equals to a minimization problem as follows:

$$\min\sum_{j=1}^{k}\sum_{i=1}^{n}\left\|T_i-m_j\right\| \quad (4.1)$$

$$m_j=\frac{\sum_{i=1}^{n}y_{ij}T_i}{\sum_{i=1}^{n}y_{ij}} \quad (4.2)$$

where T_i denotes the i th visitor's stay time. m_i denotes the mean vector of the j th cluster. And y_{ij} would be 1 if the i th visitor belongs to the j th cluster. Otherwise y_{ij} would be 0.

To access the similarity between any two visitors, Euclidean distance is used to measure the relevance between them. The distance can be expressed as follows:

$$d(T_i,T_j) = \|T_i - T_j\| = \sqrt{\sum_{k=1}^{p} \left| vt_{ik} - vt_{jk} \right|^2} \quad (5)$$

where $d(T_i,T_j)$ denotes the Euclidean distance between the the i th visitor and the j th visitor.

However, there remains a problem, namely the comparison between visitors who have already finished their visits and visitors who have not. Obviously, the distance determined above would fail to describe the relevance correctly when two visitors' stay time are not in the same dimensions. Therefore, the Euclidean distance is associated with special parameters as follows:

$$d(T_i,T_j) = \sqrt{\sum_{k=1}^{p} b_{ijk} \left| vt_{ik} - vt_{jk} \right|^2} \quad (6)$$

b_{ijk} would be 1 if both the i th visitor and the j th visitor have visited the k th stand. Otherwise b_{ijk} would be 0.

K-means clustering is implemented as shown in algorithm 1. The clustering work is repeated every 5 minutes in order to approach real-time results.

Input:

T : all visitor samples with their stay time in exhibition stands

k : the number of clusters

Output:

C: k clusters.

1. select k samples randomly and build an initial set of K-means;

Begin iteration

1. assign each visitor sample to the cluster whose mean yields the least distance. The i th cluster can be expressed as follows:

$Cluster_i = \{T_p : d(T_p, m_i) \le d(T_p, m_j), \forall j, 1 \le j \le k\}$

where m_i denotes mean of the i th cluster. Each Tp is assigned to exactly one cluster;

2. update all clusters by calculating the new means.

Until no mean changes

C is the set of k clusters. Each cluster contains numbers of visitor samples.

Figure 1. K-means clustering.

B. collaborative filtering recommendation

Those who appear in the same cluster may have similar behavior patterns or share the same interests. Therefore, we can recommend possible stands to a visitor according to records within the cluster that he belongs to.

After clustering, each cluster is formed as follows:

$$cluster_i = \{(vt_1, vt_2, \cdots, vt_p)\} \quad (7.1)$$

$$m_i = \left(\overline{vt_1}, \overline{vt_2}, \cdots, \overline{vt_p}\right) \quad (7.2)$$

where $cluster_i$ stands for the i th cluster. m_i denotes the mean of the i th cluster. $\overline{vt_j}$ denotes the mean stay time in the j th stand of all visitors who belong to the cluster.

Finally, combining collaborative filtering technique with location information, the relevance between a visitor and a stand can be expressed as follows:

$$RS_{ijt} = \frac{\overline{vt_j}}{D_{ijt}} \quad (8)$$

where RS_{ij} denotes the relevance between the i th visitor and the j th exhibition stand at time t. $\overline{vt_j}$ denotes the mean stay time in the j th stand among the cluster that contains the visitor. D_{ijt} denotes the distance between the visitor and the stand at time t.

IV. Quality Analysis Module

Relevance module can gain an increasingly accurate representation of visitor preferences over time. However, relevance module still faces challenges like the cold start problem. Therefore, quality module is applied to support the recommendation. Besides, the quality module provides information based on contents of the exhibition stands, which counts greatly in visitors' decision-making.

In the quality module, several aspects have been suggested including auto brand, popularity and convenience in terms of spatial features. First of all, we compute the popularity according to the amount of visitors in whole; while measure the convenience in terms of visitors in short time. Then, we give the score auto brand according to brand ranking. Finally, we adopt a weighted scheme to generate a quality measure of overall quality of auto show.

To measure the popularity of an exhibition stand, both accumulative and real-time amounts of visitors are considered. The accumulate amount serves as an indicator to describe the degree how popular a stand has been since the start of the exhibition. However the real-time amount, which records the current number of visitors in an exhibition stand, is used to measure the real-time popularity. Besides, whether a stand is giving a special program also influences visitors' attention. According to all above, the quality of an exhibition stand can be expressed as follows:

$$QS_{it} = \alpha_a \cdot SVA_{it} + \alpha_r \cdot SVR_{it} + IFP_{it} \quad (9)$$

where QS_{it} denotes the i th stand's quality at time t. SVA_{it} denotes the accumulate amount of visitors while SVR_{it} denotes the temporal amount. IFP_{it} can be an additional score if there is a special performance in i th stand at time t. α_a and α_r are parameters used to integrate the accumulate and temporal amounts.

Furthermore, we match the content of exhibitions with visitors' interests. When a visitor is first identified, auto shows are recommended according to the rank result of their quality. After he has visited several stands, we can recommend auto shows considering his possible interests, which can be revealed by his behaviors.

Here we use stay time to reflect visitors' interests. According to the definition of visiting of interest, the i th visitor's stay time in brand stands is recorded as $\{vt_{i1}, vt_{i2}, \cdots, vt_{ip}\}$. vt_{ij} denotes his stay time in the j th stand. Each exhibition sta p nd have detail information, including the brand name and a series of keywords $\{key_1, key_2, \cdots, key_n\}$, which highlight its attractions. The keywords are seemed as attributes that describe the corresponding exhibition. The stay time series linked to brands can be transferred to a new series as $\{vtk_{i1}, vtk_{i2}, \cdots, vtk_{im}\}$, where rtk_i denotes the stay time linked to the i th keyword. With a threshold τ set, interests of the visitor can be expressed as $\{key_{ij} : vtk_{ij} \geq \tau\}$.

To match a visitor's interests and a stand's attributes, Pearson correlation coefficient is introduced. Both the interest series and attribute series are transferred to boolean vectors. Then the Pearson correlation coefficient between them is determined as follows:

$$p_{ij} = \frac{\sum_{k=1} (v_{ki} - \overline{v_i})(s_{kj} - \overline{s_j})}{\sqrt{\sum_{k=1} (v_{ki} - \overline{v_i})^2 \sum_{k=1} (s_{kj} - \overline{s_j})^2}} \tag{10}$$

where p_{ij} denotes the Pearson correlation coefficient between the i th visitor and the j th stand. v_{ki} would be 1 if the k th keyword exists in the interest series of the i th visitor. Otherwise it would be 0. s_{kj} would be 1 if the k th keyword exists in the attribute series of the j th stand. Otherwise it would be 0.

Finally, with location information taken into consideration, the quality of an exhibition stand to a visitor is expressed as follows:

$$MQS_{ijt} = \frac{QS_{jt} \cdot p_{ij}}{D_{ij}} \tag{11}$$

where MQS_{ijt} denotes the integration of the quality of the j th stand and the Pearson correlation coefficient between the i th visitor and the stand at time t. D_{ijt} denotes the distance between the i th stand and the j th visitor at time t.

V. AGGREGATION

When some visitor wants to watch an auto show when he just enters the automobile exhibition hall, we can recommend auto show according to the quality module. After someone has watched some auto shows, we can recommend other auto show by combining both quality module and relevance module. Therefore the situations can be classified to two stages. In the first stage, there is not enough data for relevance analysis module and the recommendation depends on the quality analysis. In the second stage, the relevance analysis module is introduced .

In the later scenario, the quality score, and relevance score, which are derived from the former analysis modules, need to be further aggregated with appropriate weighting strategy.

Another factor that can not be ignored is people density. For each stand, its density can be expressed as follows:

$$DS_{it} = \frac{SVR_{it}}{Vol_i} \tag{12}$$

where DS_{it} denotes the people density of the i th exhibition stand at time t. Vol_i denotes the total capacity of the i th stand. When people density achieve a relatively high level, the stand is not convenient for visitors.

When quality, relevance and people density are measured, the matching of a visitor and an exhibition stand is carried out. We use a formula to combine these objects as follows:

$$A_{ijt} = \frac{(\lambda \cdot \phi(rt) \cdot MQS_{ijt} + \gamma \cdot \varphi(rt) \cdot RS_{ijt})}{DS_{jt}} \tag{13}$$

where A_{ijt} denotes the aggregation of the quality of the j th stand for the i th visitor, the relevance between them and people density of the stand at time t. λ and γ are weights of the quality and the relevance. To confirm the weights of parameters involved is a challenging but difficult task. Among aggregation models, analytic hierarchy process (AHP) is one well known method to solve multi-criteria decision-making (MCDM) problems, which determinate the relative importance or weight of criteria by mathematical pair-wise comparison. It has been applied extensively in many research fields, such as social network analysis and recommendation systems. In this research, we choose AHP approach as our aggregation model.

$\phi(rt)$ and $\varphi(rt)$ are dynamic weights that change according to the residence time of a visitor. When a visitor just enters the exhibition hall for a while , accessible data is not so adequate to support relevance analysis. Therefore the quality of a stand serves as a more significant role in aggregation then. However, when a visitor spends considerable time in the exhibition hall, his behaviours can be well extracted and provide instructive references. Thus, the relevance contributes more when a visitor stays for a long time. As the residence time increases, $\varphi(rt)$ increases while $\phi(rt)$ decreases.

After aggregation, the final auto show ranking list can be more appropriate and accurate. For each visitor, the aggregation is implemented every 5 minutes, in order to fit the real-time situation.

VI. Quality Analysis Module

On account of finding a verification of our algorithm, we choose the data collected one day during the China Chongqing International Auto Industry Fair. There are eight exhibition rooms and each room is arranged to different auto manufacturer. In each room, there are ten APs located in the permanent position and consistently collect the RSS from the mobile devices which can receive the signal of these transmitters. We choose one of the eight exhibition rooms to test our algorithm.

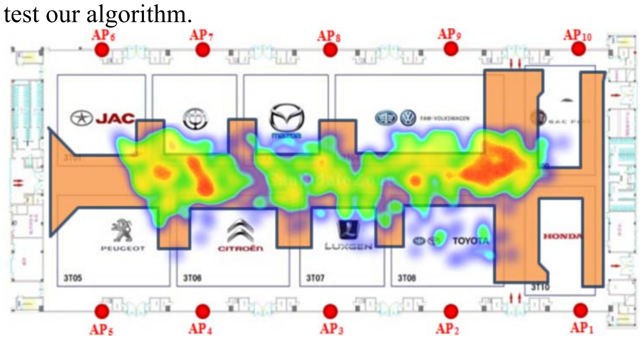

Figure 2. HeatMap of location of visitors in one exhibition hall

VII. Empirical Analysis

As mentioned above, the 2014'Chongqing car exhibition was hold in inter-connected eight exhibition halls. Each exhibition hall has set up 16 WiFi hotspot access points. All the turned on access devices, such as mobile, pad, computer, would be scanned by near hotspots devices. The Mac address, which is unique to each internet access devices, of the devices, as well as the signal attenuation ratio was recorded. Based on the resection method of surveying, if one internet access device can be scanned by three or above WiFi hotspots, then the related location of devices can be calculated. The raw dataset, record all the being turned-on internet devices in two days, exceed the volume of 100GByte.

Similar to traditional recommendation and search system, we recommend a list of automobile exhibition halls to visitors based on their profiling and ask them to rate recommendations.

To verify the accuracy of the proposed recommendation method, the Average Rating score (AR) and Normalized Discounted Cumulative Gain (NDCG) are selected as the performance metrics(Adomavicius and Tuzhilin, 2005). These metrics are computed over the top 1 and 5 recommended experts. AR is computed among the ratings from all the visitors and it indicates the average rating of all the recommendations. NDCG is a commonly adopted metric for evaluating a search engine's performance and it is for gradual judgments (i.e., automobile exhibition halls are non-relevant or more or less relevant to the query).

In this section, we present the accuracy and scalability results from experiments. In accuracy, the performance comparison of our proposed method (PM) and the relevance-based method (RM) is listed in Table 1.

TABLE I. Experiment results in terms of AR and NDCG

		PM	RM
AR	***Top 5***	4.23	3.87
	Top 10	4.08	3.65
NDCG	***Top 5***	0.73	0.63
	Top 10	0.68	0.60

As can be seen from Table 1, we can see that our proposed method achieves the best performance in terms of both the AR metric and the NDCG metric. The AR score obtained by using the relevance-based method is 3.87 in terms of returning top five halls and these results are acceptable. However, our proposed method achieves more than 9.3% improvements over baselines. The improvements of AR at top 10 returned halls also indicate that our method can find more relevant halls than the another one. We further evaluate the rank performance of two methods. NDCG scores reflect the browsing efforts of the visitors before locating the relevant halls. In terms of NDCG values, the proposed method achieves at 4.08, and over 11.8% improvements when using the baseline. The improvements on the NDCG value clearly show that our method is more effective than the relevance-based method, which gives a higher ranking for finding proper automobile exhibition halls.

VIII. Conclusions and Future Work

In this paper, a novel recommendation service is proposed to help visitors to find their proper automobile exhibition halls for auto show. In the proposed method, both temporal and spatial features of visitors are first considered to construct their profiling, and then extract their interests based on visitors' clustering. Finally, highly desired exhibition halls are recommended to each visitor. The proposed approach is well validated using a real world dataset, and compared with several baseline models. Our experimental results indicate that in terms of the well-known evaluation metrics, the proposed method can achieves more useful and feasible recommendation results, and our finding highlights that the proposed method can help both visitors to find a more appropriate automobile exhibition halls, and manage officers to reduce more management cost.

Several research questions can be further investigated in future. Firstly, this paper adopts AHP technique as the rank aggregation method. Possibly, other data fusion techniques can be considered and other techniques that model score distribution. Second, it is better to consider other social factors when recommending auto show. Therefore, we will extend this work by considering these factors in the future.

ACKNOWLEDGMENT

This work was supported in part by 973 Project (No. 2012CB316205), National Natural Science Foundation of China (No. 41371386, 91324019，91224006), Humanities and Social Sciences Foundation of the Ministry of Education (No. 14YJA630075), Beijing Nova Program (No. Z131101000413058).

REFERENCES

[1] Jannach D, Zanker M, Felfernig A, Friedrich G. Recommender systems: an introduction: Cambridge University Press; 2010.

[2] Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering. Internet Computing, IEEE. 2003; 7: 76-80.

[3] Zhang Y, Zhang B, Gao K, Guo P, Sun D. Combining content and relation analysis for recommendation in social tagging systems. Physica A: Statistical Mechanics and its Applications. 2012; 391:5759-68.

[4] Zheng N, Li Q, Liao S, Zhang L. Which photo groups should I choose? A comparative study of recommendation algorithms in Flickr. J Inf Sci. 2010; 36: 733-50.

[5] Meng, S., Dou, W., Zhang, X., Chen, J. KASR: akeyword-aware service recom-mendation method on MapReduce for big data applications. IEEE Transaction on Parallel and Distributed Systems, 2014, 1-11.

[6] Yin H., Sun Y., Cui, B., Hu Z., Chen, L. (2013). LCARS: A location-content-aware recommender system. KDD2013, 221-229.

[7] Zhang, Y., Zhang, B., Gao, K., Guo, P., Sun, D. Combining content and relation analysis for recommendation in socialtagging systems. Physica A., 391, 5759-5768.

[8] Asabere, N.Y. et al. (2014). Improving smart conference participation through socially aware recommendation. IEEE Transactions on Human-Machine Systems, Online, 1-12.

[9] Wongchokprasitti, C., Brusilovsky, P., Parra, D. Conference navigator 2.0: community-based recommendation for academic conferences. ACM SRS, 1-5.

[10] M. Kanaan, and K. Pahlavan, “Algorithm for TOA-based indoor geolocation,” Electronics Letters, vol. 40(22), 2004, pp. 1421-1422.

[11] H. Elkamchouchi, and M. A. E. Mofeed, “Direction-of-arrival methods (DOA) and time difference of arrival (TDOA) position location technique,” Proceedings of the Twenty-Second National Radio Science Conference (IEEE Cat. No. 05EX1047), 2005, 2005, pp. B12/1-10.

[12] D. Niculescu, B. Nath, and I. Ieee, "Ad hoc positioning system (APS) using AOA," Ieee Infocom 2003: The Conference on Computer Communications, Vols 1-3, Proceedings, Ieee Infocom Series, pp. 1734-1743, 2003.

[13] K. Kaemarungsi, P. Krishnamurthy, and Ieee, "Modeling of indoor positioning systems based on location fingerprinting," Ieee Infocom 2004: The Conference on Computer Communications, Vols 1-4, Proceedings, Ieee Infocom Series, pp. 1012-1022, 2004.

[14] K. Kaemarungsi, and P. Krishnamurthy, "Properties of indoor received signal strength for WLAN location fingerprinting." Proc. Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004. The First Annual International Conference on, pp. 14-23.

[15] S. Mazuelas, A. Bahillo, R. M. Lorenzo et al., “Robust Indoor Positioning Provided by Real-Time RSSI Values in Unmodified WLAN Networks,” Ieee Journal Of Selected Topics In Signal Processing, vol. 3(5), Oct, 2009, pp. 821-831.