March - April 2020
ISSN: 0193 - 4120 Page No. 3905 - 3912

Wireless Spatial Data Broadcast based on Hybrid Query Likelihood Prediction Model

Seokjin Im[1]
[1]Computer Engineering, Sungkyul Univ., Anyang, Republic of Korea
imseokjin@gmail.com[1]

Article Info
Volume 83
Page Number: 3905 - 3912
Publication Issue:
March - April 2020

Article History
***Article Received**: 24 July 2019*
***Revised**: 12 September 2019*
***Accepted**: 15 February 2020*
***Publication**: 23 March 2020*

Abstract:
Wireless data broadcast system can support data access services for any number of the clients. In the system, it is critical for the performance in order for the server to disseminate the preferred data items of the clients to the wireless channel. This paper proposes a query likelihood model that can predict data preference of the clients in near future. The proposed model predicts the query likelihood by considering the data preference of the clients and the cell preference in a hybrid manner. The cell preference means the probability for the clients to reside within the broadcasting service area. The data preference is modeled with an artificial neural network in linear regression. With the proposed query likelihood prediction model, the broadcast server enables to reflect actively the needs for data items of the clients in the near future. Through intensive simulations, the effectiveness of the proposed model is shown with respect to the access time and tuning time of the clients in the broadcasting system.

Keywords:*Query Likelihood, Data Preference, Cell Preference, Wireless Data Broadcast.*

I. Introduction

Wireless communication systems provide the mobile clients with unlimited information services at any time and any place where they are located. According to the explosive increase of the number of the clients, the systems face challenges to meet the requests from the clients. For the efficient services, the systems adopt high speed wireless communication technology with high data rate and bandwidth like 5G [1, 2]. However, the systems also have to expand the capability to respond the information requests from the clients, according to the increase of the number of the clients.

The wireless data broadcast system based on high speed communication technology like 5G enables the seamless information services for any number of the clients, regardless of increasing the number of the clients, without expanding the capacity of the system [3, 13, 14, 15]. It results from the reason the broadcast system can accommodate any number of the clients simultaneously. In the system, a server broadcasts data items to the wireless channel, and then each client tunes into the channel and downloads its desired items from the channel. Thus, each client using the system is not affected by the others in the process of downloading its queried data items [11, 12]. The system can be an alternative for efficient information services providing the scalability according to the increase of the clients. For example, Figure 1 shows a broadcasting system in which the broadcast server disseminates data items over a downlink channel that is a wireless channel with high bandwidth. The client tunes into the downlink channel and downloads its queried data item from the channel.

Also, the system supports location dependent information services by the server broadcasting spatial data, like the information on shopping mall or historical places, on the downlink channel.

In order for the clients to download data items from the channel, the clients stay on the channel until their queried data items appear, consuming energy in the active mode [6, 7, 10]. It deteriorates the energy-efficiency of the clients. For energy-efficient data access of the clients, the system applies an air index to the downlink channel. The air index keeps the information on broadcasting time of each data item. The server broadcasts data items with the index in the interleaved manner to the downlink channel. Using the index, the clients can obtain the broadcasting times of their queried data items. After tuning into the downlink channel, the clients access the index on the channel at the first. With the index, the clients search the broadcasting times of their queried items, and then they enter into the doze mode for saving their energy. The clients wake up into the active mode at the broadcasting times of the data items and download the items from the channel. Thus, the clients can download selectively only their queried data items.

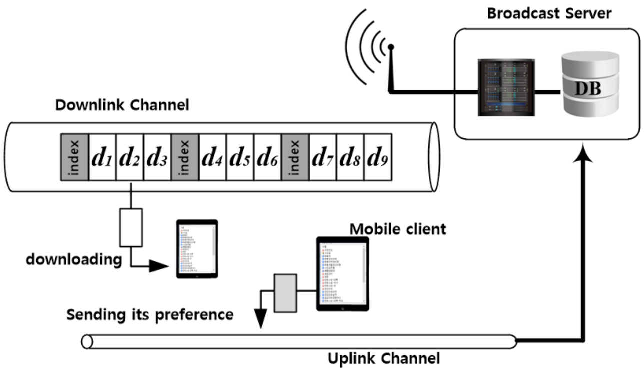

Figure 1. A Broadcast System

In this environment, it is critical for the server to select data items that the client try to download from the database and to disseminate them. For supporting the selection in the system, the clients send their preferences to data items to the broadcasting server through the uplink channel with low bandwidth. With the preferences from the clients, the server calculates the query likelihood of each data item in the database for broadcasting. The query likelihood means the possibility for the clients to download on the wireless channel. Then, the server sorts data items in the database according to the calculated query likelihood, selects top k items with high query likelihood, and disseminates the selected items.

The query likelihood model of existing wireless data broadcasting system using preferences of the clients considers only past preferences of the clients. It is because the existing systems calculate the query likelihood using the past preference information of the clients. It cannot reflect actively the changes of the need for data items of the clients and results to deteriorate the performances of the clients for accessing their queried data items.

In this paper, we propose a prediction model for calculating the query likelihood of each spatial data item for near future in order to cope with actively the need of the clients. We hybridize the preference to data items and it to the location of the clients. In the proposed hybrid scheme, the data preference is calculated for the near future. That means that the data preference to an item is the possibility for the clients to try to download it from the downlink channel. We calculate the data preference with an artificial neural network in linear regression. For the preference to the location of the client, we partition the data space, on which the clients reside, into $n*n$ grid. Using the grid, we calculate the cell preference that means the preference to the locations of the clients. Thus, the proposed query likelihood prediction model calculates the likelihood of data items to be queried with the data preference and the cell preference.

We organize the rest of the paper as follows. In Section II, we summarize the related works. Section IIIpresents the proposed hybrid query likelihood prediction model. In Section IV, we experiment and evaluate the performance metrics of proposed model through simulations. In section V, we conclude the paper.

TEST Engineering & Management
March - April 2020
ISSN: 0193 - 4120 Page No. 3905 - 3912

II. Related Works

2.1 Allocation of the Air Index on the Channel

The air index on the wireless channel supports the clients to download queried data items energy-efficiently through allowing them to selectively listen to their queried items on the channel. Using the index information on the channel, the clients process the given query in three steps as follows:

[Step 1] In this step, a client trying to process a given query carries out the initial probe. At first, the client gets the broadcasting time of the index information with the first bucket (the smallest logical unit delivering the broadcasting time information) on the channel after it tunes into the channel. The client decides the broadcasting time of the index, then waits for the index in the doze mode.

[Step 2] The client wakes up at the broadcasting time of the index and downloads it. The client searches the index in order to decide the broadcasting time of the queried data items. Then, the client switches into the doze mode until the broadcasting times for data items.

[Step 3] At the broadcasting times, the client wakes up into the active mode and downloads all queried data items.

The broadcast server allocates the index with data items in the interleaved manner on the wireless channel. The server uses one of two ways as the allocation scheme, (1, *m*) allocation scheme and distributed allocation scheme. With (1, *m*) allocation scheme, the server duplicates the index for all data items to be broadcast in front of every *m* fraction of the items. In the scheme, the clients can obtain the broadcasting time of all data items on the wireless channel. However, the scheme makes the length of the broadcast cycle because it duplicates the index for all data items *m* times in a broadcast cycle. Also, the scheme makes the waiting time of the clients for the index information after tuning into the channel [4, 5, 8, 9].

For the distributed allocation scheme, the server partitions the data items to be broadcast into *k* fractions and then organizes the indexes for data items in each fraction [7, 8]. The server disseminates each index in front of the fraction for the index. The distributed allocation scheme has strong points over the (1, *m*) indexing scheme in the aspect of the length of the broadcast cycle and the waiting time for index information on the wireless channel. The distributed allocation scheme makes shorter the length of the broadcast cycle and waiting time for index information than the (1, *m*) allocation scheme. In the distributed scheme, however, the link information from an index to other indexes on the channel can affect the performance of the clients. It is critical to keep a link information in a distributed index to other indexes.

2.2Data Preference Model

In order to consider the preference to data items of the client in the wireless data broadcast, the server broadcasts preferred data items more times in a broadcast cycle. In contrast, the server disseminates regular items once in a cycle. The server selects preferred data items using the information that the clients send to the server.

The authors in [8] use the probability of each data items for the selection. The clients send the interested area A_{int} to the server. Then, the server calculates the probability of a data item, as query likelihood, by the number of interested areas containing of the item. The server sorts the all the items by the probability and selects k items with high probability as preferred data items.

This scheme does not reflect actively the changes of needs of the clients for data items because the interested area is the information of the past. In this paper, we propose hybrid query likelihood model that enables to cope actively with the change of the needs of the client for data items by predicting.

III. Query Likelihood Prediction Model

For the proposed query likelihood prediction model, we adopt a broadcasting system that has downlink channel for broadcasting index information and data items and uplink channel for the server to collect the preference information from

TEST Engineering & Management
March - April 2020
ISSN: 0193 - 4120 Page No. 3905 - 3912

the clients, as shown Figure 1. In the system, the server holds a spatial database for providing location dependent information services to the clients. Figure 2 shows the server maintaining a spatial database. The database keeps data record R_i for spatial data item d_i as below:

$$R_i = < (d_x, d_y), CTG, Value > \quad (1)$$

Here, (d_x, d_y) is the longitudinal and latitudinal location of d_i. CTG means the data category of d_i and Value means the field of values of d_i.

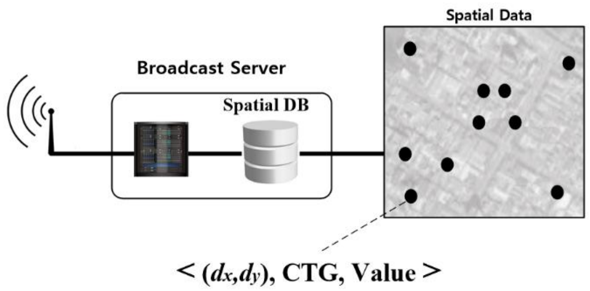

Figure 1. Spatial Database and Data Record

The proposed query likelihood model, QLike, is organized with two terms of data preference model DPref and cell preference CPref as shown in Figure3.

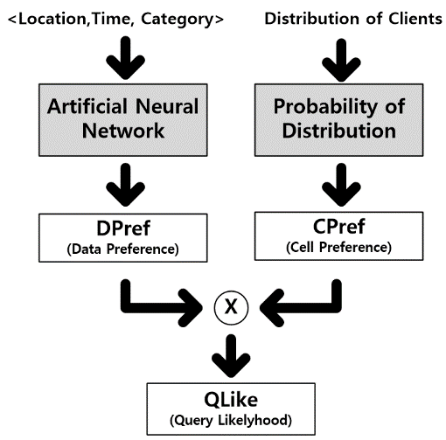

Figure 2. Proposed QLike Model

3.1 Data Preference

In order for the server to calculate the data preference *DPref*, the clients send client information *CI* to the server as below.

$$CI = < (X, Y), T, W_{CTG1}, W_{CTG2}, \dots W_{CTGk}> \quad (2)$$

Here, Here, (X, Y) is the current location of a client. T means the current time when the client sends CI and W_{CTGi} means the weight for the data category of spatial data items.

With CI, the server models data preference DPref in linear regression as below.

$$DPref = w_0 + \sum w_k x_k \quad (3)$$

Here, w_k means the weight value for the DPref and $x_k$means each value in CI from the client. The value of w_k is modeled with the artificial neural network. The location (X, Y) in CI enables that DPref reflects the effect of the location to the preference. Also, the time information T in CI enables that DPref predicts the preference to the data items of the clients for the near future because DPref is the linear model. The weight values for data categories in CI reflect the interests of the clients in the DPref.

3.2Cell Preference

In order to hybridize the preference to the area of the clients to reside in data space with data preference DPref, the server partitions the data space that the spatial. The server partitions the data space into n*n cell grid and consider the population of the clients over the cell grid. Figure 4 shows an example of 4*4 cell partition for the data space in Figure 2.

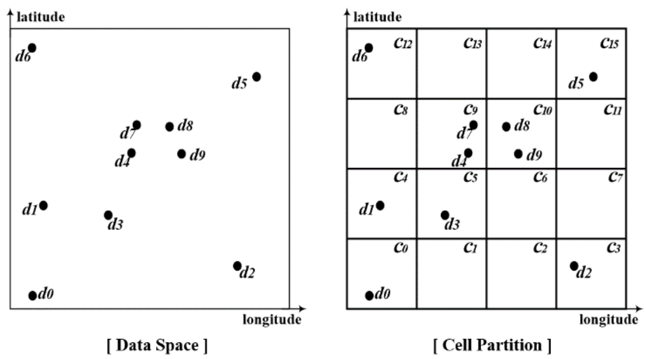

Figure 3. Data Space and its Cell Partition

The server calculates cell preference CPref with (X, Y) in CI, that means the preference to the cells in the grid partition, as below.

TEST Engineering & Management
March - April 2020
ISSN: 0193 - 4120 Page No. 3905 - 3912

$$CPref(C_i) = n(C_i) / N_{client} \quad (4)$$

Here, $n(C_i)$ means the number of the clients staying currently within cell C_i. Also, N_{client} means the number of all the clients in data space. The cell preference CPref means the probability of the distribution of the clients over the grid cell.

3.3Query Likelihood

As shown in Figure 3, the proposed query likelihood QLike of data items is calculated as below.

$$QLike = DPref \cdot CPref \quad (5)$$

Thus, the query likelihood QLike hybridizes data preference and area preference, and allows to predict the preference for near future. Using QLike, the server copes with the changes of needs for data items.

IV. Expriment and Performance Evaluation

We experiment to evaluate the accuracy of the DPref and to compare the performance of the broadcast system adopting the query likelihood model QLike with other broadcasting systems by intensive simulations.

4.1 SimulationEnvironment

For evaluating a broadcasting system adopting the proposed query likelihood, we have implemented a simulation testbed with SimJava, a simulation package in discrete time. The testbed is organized with one broadcast server, a downlink wireless channel, an uplink wireless channel, and 150 mobile clients. The clients send client information CI to the server. With CI received from the clients, the server calculates QLike for the data items. In order to broadcast, the server uses a real skewed spatial data set of 8253 places in California with five categories. In the simulation testbed, the clients process a window query, that find all data items within a given query window.

We use simulation parameters as follows. We set the size of a spatial data item to 2048 bytes, the bucket size to 512 bytes, n for partitioning the data space into cell grid to 16. Also, the size of a query window to 0.09 that means the ratio of query window to the data space.

For the comparison of the performances, we adopt the query likelihood QLike to the testbed where the server uses GDIN as its index. We compare the testbed with broadcasting systems without adopting QLike and with the use of GDIN, DSI, respectively. We use the access time and tuning time as performance metrics. The access time means the time from the moment tuning into a downlink channel to the moment downloading all the queried data items from the channel. The tuning time means the duration of time in active mode for the access time.

4.2Prediction Model Evaluation

In order to evaluate the accuracy of the DPref, we set up an artificial neural network for DPref. The neural network is organized with an input layer, a hidden layer, and an output layer. For the network setup, we generate a set of CI information with 4000 CI records and use 3600 CIs for the study of the network and use 400 CIs for the test.

In the evaluation, we change the number of the nodes in the hidden layer from 3 to 9. For each number of the nodes, we setup the neural network for DPref and test the organized DPref model.

Figure 5 and Figure 6 show the neural network with 7 nodes and 9 nodes in the hidden layer. Figure 7 shows the sum of squared error of the organized neural network, according to the number of nodes in the hidden layer. The figure shows SSE between 1 and 2 %. That means that the organized neural network for DPref with 7 nodes or 9 nodes in the hidden layer shows more accurate prediction of data preference of the clients.

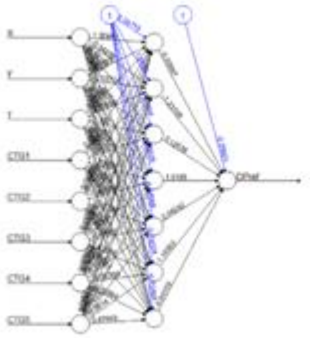

Figure 4. DPref with 7 nodes in the Hidden layer

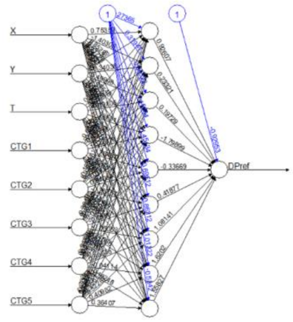

Figure 5. DPref with 9 nodes in the Hidden layer

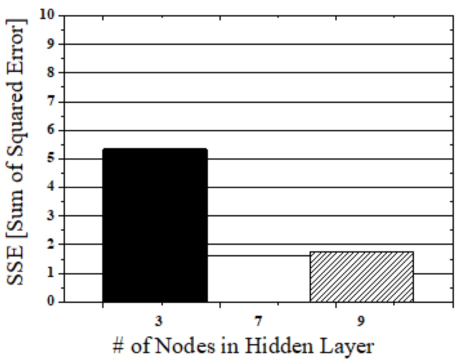

Figure 6. SSE of the DPref Model

4.3Broadcasting Performance Evaluation

To evaluate the proposed QLike model in the broadcasting system, we compare a broadcasting system adopting QLike with two systems without applying QLike. The system applied with QLike uses GDIN as an index scheme, called BS_QLike [8]. One of the two without QLike uses also GDIN and the other uses DSI as its index scheme, called BS_GDIN and BS_DSI respectively [6].

At first, we compare the access time of the three schemes, BS_QLike, BS_GDIN, and BS_DSI. Through comparing the access time of the three schemes, we evaluate how quickly the clients can obtain their queried data items at each broadcasting system.

Figure 8 depicts the access time of the three schemes. That figure describes that the clients in BS_QLike can obtain their queried data items most quickly than in BS_GDIN and BS_DSI. This is because BS_QLike can disseminate data items based on the prediction of data items that the clients is possible to query, while BS_GDIN and BS_DSI cannot change the data items to be broadcast on the channel. The two broadcasting systems broadcast the same data set on the channel. BS_QLike lessens the number of data items to be broadcast by the system broadcast data items with high possibility to be queried. Thus, BS_QLike actively copes with the changes of needs for data items of the clients.

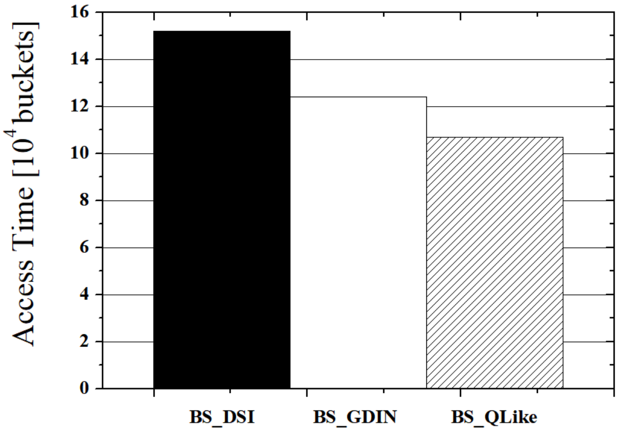

Figure 7. The Access Time Comparison

Next, we compare the tuning time of the three schemes. The tuning time shows that how much the clients consume the energy for processing given queries on the wireless channel, because the time measures time duration in active mode in which they listen to the channel consuming energy.

Figure9 describes that BS_QLike and BS_GDIN show almost same tuning time and they outperform BS_DSI. The almost same tuning time between BS_QLike and BS_GDIN results from they use the same indexing scheme GDIN. The tuning time depends on the indexing scheme in the same conditions of other factors in the system. BS_DSI shows longer tuning time than BS_GDIN and BS_QLike because the indexing scheme DSI makes the clients listen to more buckets on the channel than the indexing scheme GDIN.

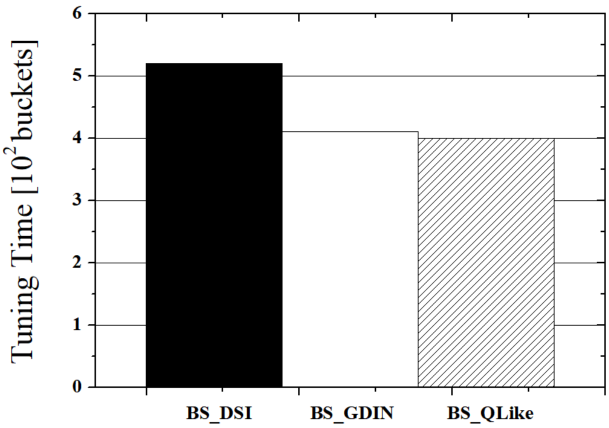

Figure8. The Tuning Time Comparison

V. Conclusion

In this paper, we proposed the hybrid query likelihood model that is capable to predict the possibility of data items to be queried in near future. The proposed model hybridizes the preference to data items of the clients and the probability for the clients to reside within the broadcasting service space.

We have set up an artificial neural network in linear regression for predicting data preference of the clients in the near future. We showed the accuracy of the proposed model for data preference by showing the sum of squared of the model. In order to evaluate the proposed model, we have adopted the model to a broadcast system, and compared the access time and tuning time with other broadcasting systems. Through the performance evaluation for broadcasting, we have shown that the proposed model enables the mobile clients to obtain their queried data items more quickly than other systems without the model. With the proposed model, the broadcast server can reflect actively changes of the needs for data items in the near future.

VI. Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2018R1D1A1B07050748).

VII. References

[1] M. A. M. Albreem. (2015). 5G wireless communication systems: Vision and challenges. -Proc. of 2015 International Conference on Computer, Communications, and Control Technology, Malaysia, pp. 493–497.

[2] X. Liu. (2018). Initial Study on the Architecture of Field Observation in 5G Era.-2018 IEEE 5G World Forum, CA, USA, pp. 524–527.

[3] Y. Yao, X. Tang, E. P. Lim and A. Sun. (2006). An energy-efficient and access latency optimized indexing scheme for wireless data broadcast. - IEEE Trans. Knowl. Data Eng., vol. 18, no.8, pp 1111-1124, Aug.

[4] B. Zheng, W. C. Lee, and D. L. Lee. (2004). Spatial Queries in Wireless Broadcast systems. - Wireless Netw., vol 10, no 6. Pp 723-736, Dec.

[5] W. C. Lee and B. Zheng. (2005). DSI: A fully distributed spatial index for location-based wireless broadcast services. - Proc. 25th IEEE International Conference on Distributed Computing Systems.

[6] B. Zheng, W. C. Lee, K.C.K Lee, D. L. Lee, and M. Shao. (2009). A distributed spatial index for error-prone wireless data broadcast. - VLDB J., vol. 18. No. 4, pp. 959-986.

[7] S. Im, M. Song, J. Kim, S. Kang, and C. Hwang. (2006). An error-resilient cell-based distributed index for location-based wireless broadcast services. - Proc. 5th ACM Workshop MobiDE 06, pp. 59-66, Chicago. USA, June.

[8] S. Im, H. Youn, J. Choi, and J. Ouyang. (2011). A novel air indexing scheme for window query in non-flat wireless spatial data broadcast. - Journal of Comm. And Networks, vol. 13, no. 4, pp. 400-407.

[9] S. Dhanked and V.Goel. (2014).Hshing techniques for broadcasting in wireless data

March - April 2020
ISSN: 0193 - 4120 Page No. 3905 - 3912

environment.-2014 International Conference on Issues and Challenges in Intelligent Computing Techniques, India.

[10] S. Mavridopoulos, P. Nicopolitidis, G. Papadimitriou, and P. Sarigiannidis. (2015). Broadcast Levels: Efficient and Lightweight Schedule Construction for Push-based Data Broadcasting System. IEEE Trans. on Broadcasting, vol. 61, no. 3, pp. 470-481.

[11] J. Kuang, and S. Yu. (2017). Broadcast-based Content Delivery in Information-Centric Hybrid Multihop Wireless Networks. IEEE Comm. Letters, vol. 21, no. 4, pp. 889-892.

[12] A. B. Waluyo, F. Zhu, D. Taniar, and B. Srinivasan. (2014) Design and Implementation of a Mobile Broadcast System. 2014 IEEE 28th International Conference on Advanced Information Networking and Applications, Vitoria, Canada.

[13] K. Tao, Y. Kou, X. Lu, and R.Song. (2015). Design of data distribution protocols for next generation broadcast wireless systems.-2015 10th International Conference on Communication and Networking in China, Shanghai.

[14] M. Veeresha, and M. Sugumaran. (2016). k-Nearest neighbor queries in wireless broadcast environments. International Conference on Communication and Electronics Systems, Coimbatore, India.

[15] S. Luan, and P. He. (2014). A novel on-demand data broadcast scheme in wireless networks. -6th International Symposium on parallel Architectures, Algorithms and Programming, Beijing, China.