

Progressive Content-Sensitive Data Retrieval in Sensor Networks

Bo Yang and Manohar Mareboyana

Department of Computer Science, Bowie State University, Maryland, United States

Abstract: Problem statement: For a sensor network comprising autonomous and self-organizing data sources, efficient similarity-based search for semantic-rich resources (such as video data) has been considered as a challenging task due to the lack of infrastructures and the multiple limitations (such as band-width, storage and energy). While the past research discussed much on routing protocols for sensor networks, few works have been reported on effective data retrieval with respect to optimized data search cost and fairness across various environment setups. This study presented the design of progressive content prediction approaches to facilitate efficient similarity-based search in sensor networks. **Approach:** The study proposed fully dynamic, hierarchy-free and non-flooding approaches. Association rules and Bayesian probabilities were generated to indicate the content distribution in the sensor network. The proposed algorithms generated the interest node set for a node based on its query history and the association rules and Bayesian rule. Because in most cases the data content of a node was semantically related with its interest of queries, the sensor network was therefore partitioned into small groups of common interest nodes and most of the queries can be resolved within these groups. Consequently, blind search approach based on flooding could be replaced by the heuristic-based unicasting or multicasting schemes, which drastically reduced the system cost of storage space, network bandwidth and computation power. **Results:** We verified the performance with experimental analysis. The simulation result showed that both Bayesian scheme and association scheme require much less message complexity than flooding, which drastically reduced the consumption of system resources. **Conclusion:** Content distribution knowledge could be used to improve the system performance of content-based data retrieval in sensor networks.

Key words: Information retrieval, sensor network, video processing

INTRODUCTION

Wireless sensor networks are shared-medium multi-hop systems consisting of radio-equipped sensor nodes. These types of networks are useful any situation where temporary network connectivity and communication are needed. The sensor nodes in sensor networks are capable of not only storing and processing data, but also performing complex operations through their communications, such as similarity-based retrieval of data objects. Data management provides an essential application for sensor networks, as is witnessed by the wealth of research literature^[1,2,11]. However, efficient retrieval of semantic-rich data in sensor networks is challenging due to the multiple constraints such as node mobility, computation capability, memory space and bandwidth.

To address the problem of data retrieval in sensor networks, a variety of approaches have been presented in the literature. Researchers have proposed methods based on centralized client-server architecture. Some examples of such approaches are presented in^[3,4]. One

common characteristic of these models is the reliance on a centralized storage (or head node) that would handle the queries from clients and forward back the results. This assumption violates the requirements of sensor networks where sensor nodes should be considered as equal peers and none of them should be given extra capability or responsibility. Moreover, the centralized models will cause single points of failure and therefore reduce the robustness and scalability.

Decentralized approaches overcome the reliance on a central directory server and employ unstructured peer-to-peer connections to resolve the queries. However, these approaches do not take link cost in the consideration when computing routes^[5,12]. This shortcoming causes drastic waste of system resources and makes the decentralized approaches unsuitable for practical applications of sensor networks. Some protocols have been proposed to reduce the search cost in the mobile environments, but these protocols still rely heavily on the flooding of query messages and may be too expensive to operate.

Corresponding Author: Bo Yang, Department of Computer Science, Bowie State University, Maryland, United States

As an alternative, a great deal of research has been done on organizing sensor networks based on content distribution. Search algorithms based on landmark hierarchies were proposed in^[6]. The hierarchy consists of data nodes and landmark nodes that behave as indices. Improvements to better organize the node contents and distance information by clustering the content-similar nodes and mapping them into semantic categories in the semantic space-such as semantic-aware hierarchy^[2]. However, topology/content changes may require costs to update such indices. Further improvements to search efficiency have led to continuing overlaid infrastructures (e.g., Chord^[7]) that use content/location mapping to direct the searches to nodes holding the requested data. This approach causes higher maintenance overhead in the process of updating the mapping relationships in accordance with content changes.

Another solution is the hash-table-derived approach. The Content Addressable Networks (CANs) provide a distributed hash table abstraction over the Cartesian space of the data objects. They employ (object, key) pairs to allow efficient storage exploitation and query processing^[9]. The implementations include logical overlay networks (e.g., pSearch^[5]) that use dimension-reduction techniques to reduce search cost. The content-navigated retrieval models rely on proactive hash tables to guarantee their efficiency in the retrieval process, which also present high challenges to data location restriction and hash table maintenance in practical applications.

Motivated by the strengths and weaknesses of centralized and decentralized approaches, we propose to utilize the cooperation among sensor nodes to improve the system performance. In this study, we present two progressive content prediction schemes-Association-based Content Prediction (ACP) and Bayesian-based Content Prediction (BCP)-to facilitate content-based retrieval. The fundamental idea is to predict the location of data sources based on query history and probability rules. Based on the content distribution knowledge, the data retrieval is performed with reduced network traffic and response time due to the employment of heuristic-based unicasting instead of blind broadcasting.

MATERIALS AND METHODS

Proposed schemes: Our schemes are motivated by three observations: First, a sensor node may have specific interest and query some data objects frequently. Second, the data contents of a node are often semantically similar to the queries it has issued.

Therefore, by analyzing the earlier queries delivered among the nodes, the system may forward later queries to a small collection of nodes which requested or resolved the earlier queries with similar semantics. Third, some nodes in the network may share similar interest and generate the same queries. These nodes can be grouped into common-interest clusters, in which any query can be cached and analyzed to facilitate later queries issued by nodes in the cluster. Based on the three observations, we propose to exploit the query history of sensor nodes to find the relationships among multimedia data objects and use this knowledge to determine the data contents of sensor nodes and ultimately improve the performance of content-based multimedia information retrieval.

Problem formulation: We introduce the formalized problem description of content-based retrieval in sensor networks. We consider an sensor network consisting of resource-constrained nodes moving within a specified geographic region. The hardware capacity of the nodes may vary from each other and each node is equipped with a cache space for the query history recording and data perfecting purpose. We assume reliable pair-wise message communications between sensor nodes in the order of their generation, using some existing routing protocol, such as AODV or DSR.

We consider the type of applications in which all nodes in the sensor network share a standard representation of data objects. More specifically, the sensor nodes have a consented set of features that indicate the semantic contents of data objects.

Based on the aforementioned assumptions, we consider the sensor network as an undirected connected graph $G = (N, C, W)$, where $N = \{n_1, n_2, \dots, n_r\}$ is the set of sensor nodes, $C = \{c_1, c_2, \dots, c_s\}$ is the set of wireless connections between neighboring sensor nodes and W is the pair-wise weight (or distance) between the nodes. The problem of content-based retrieval in the sensor network can be viewed as the following: Given a set of data objects $X = \{x_1, x_2, \dots, x_m\}$ disseminated among the set of sensor nodes N , a query data object Q and an integer k , find the minimum subset $N^* = \{n_1^*, n_2^*, \dots, n_s^*\} \subset N$ containing the top k data objects with smallest semantic distances to Q , where the similarity between data objects is evaluated using the Euclidean distance in the semantic space.

Definition 1:

The Euclidean vector norm: Given the set of data objects $X = \{x_1, x_2, \dots, x_m\}$, each object x_i is represented as an n -dimensional semantic vector $\varphi_{xi} = (\omega_{xi}^1, \omega_{xi}^2, \dots, \omega_{xi}^n)$. The Euclidean vector norm of semantic vector φ_{xi} , also called 2-norm, indicates the length of φ_{xi} in the Euclidean geometry:

$$\|\phi_{x_i}\|_2 = \left(\sum_{k=1}^n |\omega_{x_i}^k|^2 \right)^{\frac{1}{2}} = \sqrt{|\omega_{x_i}^1|^2 + |\omega_{x_i}^2|^2 + \dots + |\omega_{x_i}^n|^2} \quad (1)$$

Definition 2:

The scalar product: The scalar product of two semantic vectors ϕ_{x_i} and ϕ_{x_j} , denoted as $\phi_{x_i} \bullet \phi_{x_j}$, is defined as:

$$\begin{aligned} \phi_{x_i} \bullet \phi_{x_j} &= \sum_{k=1}^n (\omega_{x_i}^k * \omega_{x_j}^k) \\ &= \omega_{x_i}^1 * \omega_{x_j}^1 + \omega_{x_i}^2 * \omega_{x_j}^2 + \dots + \omega_{x_i}^n * \omega_{x_j}^n \end{aligned} \quad (2)$$

Definition 3:

The semantic similarity: The semantic similarity between data objects is defined based on the Euclidean distance between their semantic vectors. Formally, the distance between two data objects x_i and x_j is defined as a cosine distance function:

$$\text{dist}(x_i, x_j) = \cos^{-1} \frac{\phi_{x_i} \bullet \phi_{x_j}}{\|\phi_{x_i}\|_2 \|\phi_{x_j}\|_2} \quad (3)$$

Association-based content prediction:

Preliminary concepts: The distribution pattern of the data objects over the sensor nodes can be considered as a many-to-many relationship: Each data object may be distributed among multiple sensor nodes and each sensor node may contain a collection of data objects.

Let $\chi(n_j) = \{x_{j1}, x_{j2}, \dots, x_{jv}\}$ denote the queried data object set of sensor node n_j . It is obvious that $\chi(n_j)$ is a subset of X , i.e. $\chi(n_j) \subseteq X$. Suppose there exists a list of earlier queries $L = \{q_1, q_2, \dots, q_w\}$ where each query q_i is a data object in $\chi(n_j)$. An association rule is a repeatedly happened pattern $R: Q \rightarrow x_{jt}$, where $Q = \{q_{t1}, q_{t2}, \dots, q_{ts}\}$ is a set of queries and $x_{jt} \in \chi(n_j) - Q$ is a data object. Here Q and x_{jt} are called the antecedent and consequence of the association rule, respectively.

Definition 4:

The dataset support: The support of a data query set Q , denoted as $\psi(Q)$, can be defined as the percentage in the list L of query sets that include Q as their sub sets. The dataset support can be formalized as:

$$\psi(Q) = \frac{\sum_{|X|=|Q|} |\text{truesubset}(X, Q)|}{\sum_{|X|=0}^{|\mathcal{L}|} |\text{subset}(X)|} \text{ for } \forall X \subseteq L \quad (4)$$

where, $\text{subset}(X)$ returns all possible subsets of X , while $\text{truesubset}(X, Q)$ only returns the subsets of L that are also supersets of Q .

Definition 5:

The association support: The support of an association rule $R: Q \rightarrow x_{jt}$ is considered as the percentage of query sets that include both Q and x_{jt} , or formalized as:

$$\Psi(R \subseteq) = \frac{\sum_{|X|=|Q|} |\text{truesubset}(X, \{x_{jt}\} \cup Q)|}{\sum_{|X|=0}^{|\mathcal{L}|} |\text{subset}(X)|} \text{ for } \forall XL \quad (5)$$

Definition 6:

The association confidence: The confidence of an association rule $R: Q \rightarrow x_{jt}$, denoted as $\zeta(R)$, can be considered as the ratio of the association support in comparison with the dataset support:

$$\zeta(R) = \frac{\Psi(R)}{\psi(Q)} = \frac{\Psi(\{x_{jt}\} \cup Q)}{\psi(Q)} \quad (6)$$

Definition 7:

The association rule set: Given a query history list L and a confidence threshold ζ_{th} , the association rule set $\text{Rule}(L, \zeta_{th})$ is the set of association rules having confidence no smaller than ζ_{th} :

$$\text{Rule}(L, \zeta_{th}) = \{ R: Q \rightarrow x_{jt} \mid \zeta(R) \geq \zeta_{th} \} \quad (7)$$

Definition 8:

The validated node content: Given a sensor node n_i and a query history L , the data objects in the query results that are resolved from node n_i are considered as the validated node content of n_i :

$$v(n_i) = \{x_j \mid x_j \in Q, \text{ for } \forall Q \in L\} \quad (8)$$

Definition 9:

The estimated node content: Given a sensor node n_i and a rule set $\text{Rule}(L, \zeta_{th})$, the estimated content of n_i is the set of data objects that are the consequences of the association rules after applying the validated node content of n_i as the antecedents:

$$\xi(n_i) = \{x_j \mid R: Q \rightarrow x_j, \text{ for } \forall Q \in v(n_i) \wedge R \in \text{Rule}(L, \zeta_{th})\} \quad (9)$$

Definition 10:

The estimated node interest: Given a sensor node n_i and a rule set $\text{Rule}(L, \zeta_{th})$, the estimated interest of n_i is the set of data objects that are the consequences of the association rules after applying the node interest of n_i as the antecedents:

$$\lambda(n_i) = \{x_j \mid R: Q \rightarrow x_j, \text{ for } \tau \forall Q \in \xi(n_i) \wedge R \in \text{Rule}(L, \zeta_{th})\} \quad (10)$$

Definition 11:

The interest-content overlap: Given a sensor node n_i and the query history L , the node set that contains data contents related with the interest of node n_i is considered as the overlap-interest node set:

$$\theta(n_i) = \{ n_k \mid \xi(n_k) \cap \lambda(n_i) \neq \emptyset \} \quad (11)$$

Prediction algorithm: The fundamental idea is to first allow the sensor nodes to record the queries and their results passing by it as the query history and then gradually increase the size of association set based on the appearance ratio of data objects in the queries (i.e., antecedents) and results (i.e., consequences). Table 1 shows the notations used in the association rule set generation algorithm. Algorithm 1 shows the details of the rule generation process.

Algorithm 1: Constructing association rule set:

Input: query history L and confidence threshold ζ_{th}
 Output: association rule set $Rule(L, \zeta_{th})$

- 1- $L \leftarrow \emptyset$
- 2- $Rule(L, \zeta_{th}) \leftarrow \emptyset$
- 3- for $i \leftarrow 1$ to r
- 4- if $(v(n_i) \not\subset L) L \leftarrow L \cup v(n_i)$
- 5- for $i \leftarrow 1$ to $|L|$
- 6- $Q \leftarrow qry(L)$
- 7- $cons \leftarrow sup(rslt(Q) \cup Q)$
- 8- $ante \leftarrow sup(Q)$
- 9- if $(cons/ante > \zeta_{th})$
- 10- $Rule(L, \zeta_{th}) \leftarrow Rule(L, \zeta_{th}) \cup \{ R: Q \rightarrow rslt(Q) \}$
- 11- return $Rule(L, \zeta_{th})$

Table 2 shows the notations used in interest node generation. Algorithm 2 for computing interest node set is listed as follows.

Algorithm 2: Generation interest node set:

Input: sensor node n_i
 association rule set
 Output: Interest node set of n_i

- 1- $inst(n_i) \leftarrow \emptyset$
- 2- while $(Hq(n_i) \neq \emptyset)$
- 3- $Q \leftarrow \mathcal{E}(n_i)$
- 4- $Hq(n_i) \leftarrow Hq(n_i) - \{Q\}$
- 5- for $i \leftarrow 1$ to $|node(Q)|$
- 6- select a node n_k from $node(Q)$
- 7- if $(est(n_k) \cap \zeta_{ist}(n_i) \neq \emptyset)$
- 8- $inst(n_i) \leftarrow inst(n_i) \cup \{ n_k \}$
- 9- return $inst(n_i)$

Table 1: Notations for rule generation

Items	Notations
r	The number of nodes in the sensor network
$qry(L)$	The function of selecting a query from L
$rslt(Q)$	The function of obtaining the result of query Q
$sup(X)$	The function of computing the support of data set X

The proposed algorithms generate the interest node set for a node based on its query history and the association rules. For the simplicity of computation, the algorithm employs the estimated interest of the input node to filter out other nodes that do not contain overlapping data contents. Therefore, a node n_i only shares its interest with a small collection of nodes. When a new query is issued, n_i will first forward it to the interest-related node set, with the purpose of resolving the query with a small portion of the network. If the query cannot be resolved in this interest node set, then use flooding to find the result. Due to the semantic locality of queries, most queries will be processed in the small set of nodes.

In the aforementioned algorithms we address finding the content distribution knowledge of the sensor network to facilitate similarity-based data retrieval. In comparison with the conventional centralized or flooding methodologies, the proposed content prediction approach has the following properties:

- The fundamental idea is based on the association rules generated from the analysis of the coincidence pattern of data objects in the earlier queries and their results. Because in most cases the data content of a node is semantically related with its interest of queries, the sensor network is therefore partitioned into small groups of common interest nodes and most of the queries can be resolved within these groups. Consequently, blind search approach based on flooding can be replaced by the heuristic-based unicasting or multicasting schemes, which drastically reduces the system cost of storage space, network bandwidth and computation power
- The proposed content prediction approach does not rely on any fixed or overlaid infrastructure, which not only adheres to the infrastructure-free nature of sensor networks, but also greatly reduces the system overhead of updating the indexing or mapping information in the content/location servers, improving the robustness and scalability

Bayesian-based content prediction:

Methodology statement: Suppose we are given a collection of data objects $X = \{x_1, x_2, \dots, x_m\}$ that are disseminated among the sensor nodes. Each data object is assigned a probability of being accessed. After a period of time, some data objects are requested in the query history, which changes the access probabilities.

Table 2: Notations for interest node set generation

Items	Notations
inst(n_i)	The interest node set of n_i
Hq(n_i)	The history of queries issued by node n_i
$\xi(n_i)$	The function of obtaining a query issued by n_i
node(Q)	The function of finding the nodes resolving Q
est(n_k)	The function of estimating the content of node n_k
$\xi_{\text{ist}}(n_i)$	The estimated the interest of node n_i

Table 3: Notations for Bayesian content prediction

Items	Notations
$P(x_i \in n_i)$	The probability of data object x_i existing in n_i
$P(X n_i, t_k)$	The mapping of all X's data objects in node n_i
query(t_k)	The function of taking a query from iteration t_k
obj(Q)	The function of taking an object from query Q
satisfy(Q, n_i)	Whether the query Q is satisfied at node n_i

Assume that for each data object x_i there exists some underlying probability distribution $P(\cdot|x_i)$, referred as the fundamental relevance probability. The initial probability model can be constructed using many existing analysis systems, e.g., Latent Semantic Analysis (LSA)^[5,10]. The data objects will then be retrieved with a collection of queries, which can be considered as a series of iterations-each iteration containing the query results and their relevance to the query objects. For a specific iteration t_k , when a new query Q comes, the data retrieval system can use a deterministic strategy to select the objects with the highest Bayesian probabilities $P(x = Q|t_1 \dots t_k)$.

Definition 12:

The Bayes' rule: Based on above descriptions of data retrieval, the Bayes' rule can be defined as:

$$P(x = Q | t_1 \dots t_k) = \frac{P(t_1 \dots t_k | x = Q)P(x = Q)}{P(t_1 \dots t_k)}$$

$$= \frac{\sum_{j=1}^m \sum_{w=0}^{k-1} P(t_w | t_{w+1})P(t_k | x = Q)}{\sum_{i=1}^m \prod_{v=0}^{k-1} P(t_v | t_{v+1})P(t_k | x = x_i)} \quad (12)$$

where, $P(t_0) = 1$.

Content prediction: The Bayesian-based content prediction is a process of estimating the probability of the user's next query being satisfied by the data content of a sensor node according to the observation of its resolution of earlier queries. For a node n_i , the query history is divided into a series of iterations t_1, \dots, t_k , each iteration containing a collection of queries and the results (i.e. whether the requested data objects are found at n_i). Table 3 shows the notations used for Bayesian content prediction. The detailed process is described in algorithm 3.

Algorithm 3: Bayesian content prediction:

Input: Query iterations t_k
 data content mapping $P(X|n_i, t_{k-1})$ before t_k
 Output: data content mapping $P(X|n_i, t_k)$ after t_k

- 1- for $i \leftarrow 1$ to $|t_k|$
- 2- $Q \leftarrow \text{query}(t_k)$
- 3- for $j \leftarrow 1$ to $|Q|$
- 4- $x \leftarrow \text{obj}(Q)$
- 5- if (satisfy(Q, n_i))
- 6- $P(X|n_i, t_k) \leftarrow P(X|n_i, t_{k-1})P(x_j \in n_i)$
- 7- else
- 8- $P(X|n_i, t_k) \leftarrow P(X|n_i, t_{k-1})(1 - P(x_j \in n_i))$
- 9- return $P(X|n_i, t_k)$

With the help of Bayesian-based content prediction, the retrieval of data objects becomes a clearly aimed process, navigated by the probabilistic distribution of data objects among the nodes:

- When a query (i.e., data object x_q) is issued to node n_i , first check whether it can be resolved locally
- If x_q is not locally resolved, compare the probabilities $P(x_q \in n_1), \dots, P(x_q \in n_r)$, find the node with highest probability
- Update the data content mapping $P(X|n_i, x_q)$ using the content prediction algorithm

Based on this strategy, the queries may be resolved with reduced search cost in terms of network traffic and computation complexity, because the query messages are only forwarded to the nodes with highest Bayesian probabilities. This approach is also self-stabling due to its capability of updating the probabilistic information after the resolution of queries. This maintenance process may cause system overhead; however, this overhead can be amortized on each query can be alleviated.

RESULTS

Performance analysis: We compared the performance of the two proposed schemes using a simulator implemented in ns-2 environment. The simulated sensor network consists of up to 200 sensor nodes joining/leaving the network according to a Poisson process during a period of 8000 sec. The data object set comprises up to 12,000 data points whose semantic vector values are assigned by a random number generator abiding by normal distribution in the interval [0,1). Each node randomly selects waypoints within a 2000*1000 m flat area. The node density can be adjusted by changing the number of nodes in the area. The query generation pattern follows the exponential

distribution, which is similar to the previous study^[8]. To provide the storage for recording earlier queries, we also assigned each node with a 1MB fixed-size cache.

One of important metrics for data retrieval system is the accuracy. The performance of a search scheme can be evaluated by its accuracy when accessing a restricted portion of nodes. In the simulation we restricted the search schemes to visit less than 10% of the nodes. Figure 1 compares the accuracy of query results returned by Bayesian, association and flooding schemes as the number of queries increases.

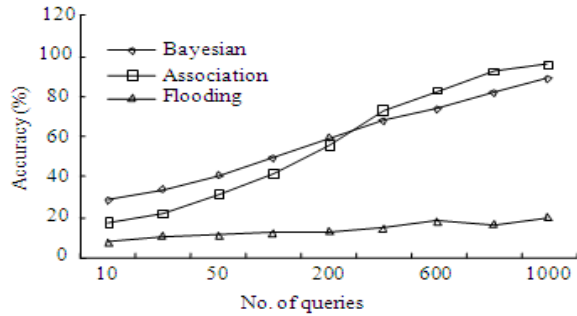


Fig. 1: The impact of query sample set size

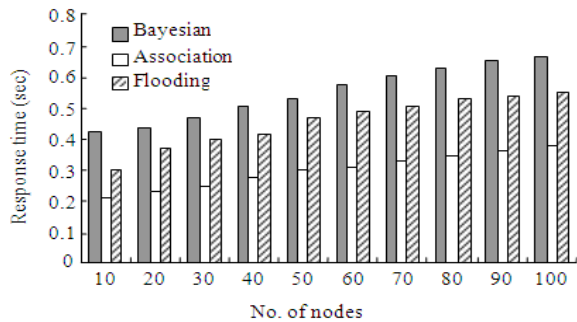


Fig. 2: The impact of node density on response time

Generally speaking, more query history improves the accuracy of all three schemes. Flooding schemes improves its accuracy due to larger number of earlier queries being cached, which increases the possibility of local resolution. The association-based scheme performs worse than Bayesian at the beginning due to the poor stability of association rules generated with small sample set of queries. As the query history becomes longer, the association rules precisely describe the relationships among data objects and therefore help improving the retrieval accuracy.

In another simulation run, we measured the impact of node density on query response time. Figure 2 shows the result. As can be seen from Fig. 2, the association-

based scheme achieves shortest response time. This is due to the fact that data objects can be prefetched with the help of association rules, reducing the average search delay. Note that the flooding scheme outperforms the Bayesian scheme on response time due to its capability of finding the shortest path to data sources using query broadcasting. However, this shorter response time of is obtained at the price of large message complexity. In Fig. 3 we compare the average number of messages required to resolve a query. As one can conclude, both Bayesian scheme and association scheme require much less message complexity than flooding, which drastically reduces the consumption of system resources.

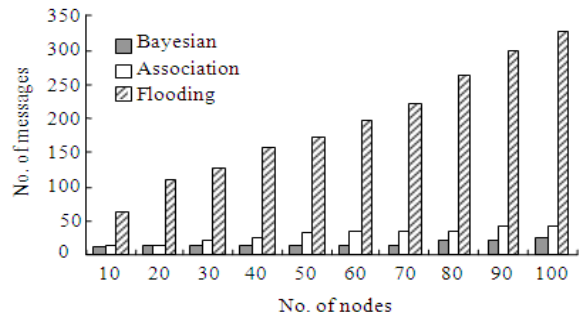


Fig. 3: The message complexity for solving queries

DISCUSSION

The proposed scheme makes use of data content distribution in sensor networks to resolve video queries without incurring flooding in the network. Through theoretical and experimental study, we found that the proposed scheme has the following features:

- Association and Bayesian probability based content estimation
- Constraint-based representation method showing the semantic similarity between query objects
- Non-flooding query processing
- Adaptive accuracy and search cost performance

CONCLUSION

Content distribution knowledge can be used to improve the system performance of content-based data retrieval in sensor networks. In this study we proposed Association-based Content Prediction (ACP) scheme and Bayesian-based Content Prediction scheme to facilitate the efficient light-weight data retrieval. Association rules and Bayesian probabilities are generated to indicate the content distribution in the sensor network. The simulation result shows that these

schemes drastically improve system performance in terms of accuracy and search cost.

REFERENCES

1. Auffret, G., J. Foote, C. Li, B. Shahraray, T.S. Mood and H. Zhang, 1999. Multimedia access and retrieval: the state of the art and future directions. Proceedings of the ACM international Conference on Multimedia, Oct. 3-Nov. 5, ACM Press, Orlando, Florida, United States, pp: 443-445. <http://portal.acm.org/citation.cfm?doi=319463.319684>
2. Yang, B. and A.R. Hurson, 2004. An extendible semantic-based content representation method for distributed image databases. Proceedings of the IEEE Symposium on Multimedia Software Engineering, Dec. 13-15, IEEE Xplore Press, USA., pp: 222-226. DOI: 10.1109/MMSE.2004.26
3. Woodrow, E. and W. Heinzelman, 2002. SPIN-IT: A data centric routing protocol for image retrieval in wireless networks. Proceedings of the IEEE International Conference on Image Processing, June 24-28, IEEE Xplore Press, USA., pp: 913-916. DOI: 10.1109/ICIP.2002.1039121
4. Li, D., N. Dimitrova, M. Li and I.K. Sethi, 2003. Multimedia content processing through cross-modal association. Proceedings of the ACM International Conference on Multimedia, Nov. 20-8, ACM Press, Berkeley, CA., USA., pp: 604-611. <http://portal.acm.org/citation.cfm?id=957143>
5. Tang, C., Z. Xu and S. Dwarkadas, 2003. Peer-to-peer information retrieval using self-organizing semantic overlay networks. Proceedings of the ACM International Conference on Applications, Technologies, Architectures and Protocols for Computer Communications, Aug. 25-29, ACM Press, Karlsruhe, Germany, pp: 175-186. <http://portal.acm.org/citation.cfm?id=863955.863976>
6. Pei, G., M. Gerla and X. Hong, 2000. LANMAR: Landmark routing for large scale ad hoc networks with group mobility. Proceedings of the 1st International Symposium on Mobile Ad Hoc Networking and Computing, (MAHNC'00), Boston, MA., USA., pp: 11-18. DOI: 10.1109/MOBHOC.2000.869208
7. Stoica, I., R. Morris, D. Karger, M. F. Kaashoek and H. Balakrishnan, 2001. Chord: A scalable peer-to-peer lookup service for internet applications. Proceedings of the Conference on Applications, Technologies, Architectures and Protocols for Computer Communications, (APAPCC'01), ACM Press, San Diego, California, United States, pp: 149-160. <http://portal.acm.org/citation.cfm?id=383071>
8. Yin, L. and G. Cao, 2006. Supporting cooperative caching in ad hoc networks. IEEE. Trans. Mobile Comput., 5: 77-89. DOI: 10.1109/TMC.2006.15
9. Gionis, A., D. Gunopulos and N. Koudas, 2001. Efficient and tunable similar set retrieval. Proceedings of the ACM SIGMOD International Conference on Management of Data, May 21-24, ACM Press, Santa Barbara, California, United States, pp: 247-258. <http://portal.acm.org/citation.cfm?id=375663.375689>
10. Letsche, T.A. and M.W. Berry, 1997. Large-scale information retrieval with latent semantic indexing. Inform. Sci., 100: 105-137. <http://portal.acm.org/citation.cfm?id=254811>
11. He, X., W. Ma and H. Zhang, 2004. Learning an image manifold for retrieval. Proceedings of the ACM International Conference on Multimedia, Oct. 10-16, ACM Press, New York, USA., pp: 17-23. <http://portal.acm.org/citation.cfm?id=1027532>
12. Broch, J., D. Maltz, D. Johnson, Y. Hu and J. Jetcheva, 1998. A performance comparison of multi-hop ad hoc wireless network routing protocols. Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking, Oct. 25-30, ACM Press, Dallas, Texas, United States, pp: 85-97. <http://portal.acm.org/citation.cfm?id=288256>