Resource Allocation in Wireless Networks

Dimitrios Katsaros
Aristotle University, Greece

Gökhan Yavas
Bilkent University, Turkey

Alexandros Nanopoulos
Aristotle University, Greece

Murat Karakaya
Bilkent University, Turkey

Özgür Ulusoy
Bilkent University, USA

Yannis Manolopoulos
Aristotle University, Greece

INTRODUCTION

During the past years, we have witnessed an explosive growth in our capabilities to both generate and collect data. Advances in scientific data collection, the computerization of many businesses, and the recording (logging) of client's access to networked resources have generated a vast amount of data. Various data mining techniques have been proposed and widely employed to discover valid, novel and potentially useful patterns in these data.

Traditionally, the two primary goals of data mining tend to be *description* and *prediction*, although description is considered to be more important in practice. Recently though, it was realized that the prediction capabilities of the models constructed by the data mining process can be effectively used to address many problems related to the allocation of resources in networks. For instance, such models have been used to drive prefetching decisions in the World Wide Web (Nanopoulos, Katsaros, & Manolopoulos, 2003) or to schedule data broadcasts in wireless mobile networks (Saygin & Ulusoy, 2002). The intrinsic attribute of these environments is that the network records the characteristics, for example, movements, data preferences of its clients. Thus, it is possible to infer future client behaviors by mining the historical information, which has been recorded by the network.

The present article will highlight the data mining techniques that have been developed to achieve efficient allocation of resources, for example bandwidth, to wireless mobile networks or the data mining methods that have been used in order to reduce the latency associated with the access of data by wireless mobile clients.

BACKGROUND

We consider a typical wireless Personal Communications Systems (PCS) (see) with architecture similar to those used in *EIA/TIA IS*-41 and *GSM* standards. The PCS serves a geographical area, called *coverage area*, where mobile users (MU) can freely roam. The coverage area served by the PCS is partitioned into a number of non-overlapping regions, called *cells*. At the heart of the PCS lies a fixed backbone (wireline) network. A number of fixed hosts are connected to this network. Each cell is usually served by one *base station* (BS), which is connected to the fixed network and it is equipped with wireless transmission and receiving capability. We assume that each base station serves exactly one cell. MUs use radio channels to communicate with BSs and gain access to the fixed or wireless network. The BS is responsible for converting the network signaling traffic and data traffic to the radio interface for communication with the MU and also for transmitting paging messages to the MU. Finally, a *cell site switch* (CSS) will govern one or more base stations. CSS will provide access to the serving mobile network, will manage the radio resources and provide mobility management control functions (for example, location update).

The coverage area consists of a number of *location areas* (LA). Each location area consists of one or more cells. The MU can freely roam inside a location area

Figure 1. Architecture of a wireless PCS
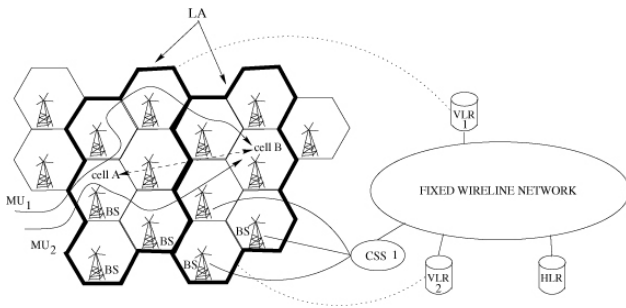
without notifying the system about its position. Whenever it moves to a new location area it must update its position, reporting the location area it entered. This procedure is called *location update*. This is done as follows: each mobile user is assigned to one database, the *home location register* (HLR) (one for each PCS network), which maintains the *profile* information regarding the user, such as authentication, access rights, billing, position, and etcetera. Each location area is associated to one *visitor location register* (VLR), which stores the profile of the MUs currently residing in its respective location area. We assume that each VLR is associated to one location area and vice-versa. The search for mobile clients is performed by broadcasting *paging messages* to the cells, where the clients might have moved, until the client is located or the whole coverage area is searched.

The identity of the cell is continuously being broadcast by the cell's BS, thus the terminal is aware of the cell it resides. If each mobile terminal records the sequence of cells it visits and communicates them back to the network every time it reports a location area crossing, then the network can have accurate information for the *mobile user trajectories* inside the coverage region.

The concept of resource allocation in wireless mobile environments covers aspects of both network and data management issues. With respect to network management, the issue of *dynamic bandwidth allocation* is of particular importance. Instead of granting a fixed frequency spectrum to each cell, irrespective of the number and needs of the clients residing therein, the allocated spectrum varies according to the clients' demands. This necessitates prediction of both future clients' movements and future data needs. The issue of client movement prediction is also related to the order according to which the paging messages should be broadcast to the cells, so as to guarantee minimum energy consumption and, at the same time, fast client location determination.

With respect to the data management issues, a prominent problem is that of reducing the average latency experienced by the clients while retrieving data from the underline infrastructure network or the wireless network. This problem is tightly related to caching data at various places of the network. The caches can be air-located, that is, data broadcasting (Karakaya & Ulusoy, 2001), or relocated to specific base stations. Thus, we need to identify data that will be requested together by the same client or group of clients, so as to broadcast them during short time intervals. In addition, we need to deduce future client movements to cells so as to "*push-cache*" data to the base stations serving these cells (Hadjiefthymiades & Merakos, 2003; Kyriakakos et al., 2003).

DATA MINING AT THE SERVICE OF THE NETWORK

We will present the most important research areas where data mining methodologies are employed to improve the efficiency of wireless mobile networks. The first area is the mobile user location prediction and aims at deducing future client movements. Location prediction is important for both bandwidth allocation and data placement. The second area is the data broadcast schedule creation and aims at recognizing groups of data items that are likely to be requested together or during small time intervals, so as to place them "closely" in the broadcast program.

LOCATION PREDICTION

The issue of predicting future mobile client positions has received considerable attention (e.g., Aljadhai & Znati, 2001; Liang & Haas, 2003) in the wireless mobile networks research community. The focus of these efforts is the determination of the position of a mobile, given some information about its velocity and direction. Though most (if not all) of these works make unrealistic assumptions about the distribution of the velocity and direction of the mobile terminals. Only recently, data mining techniques have been employed in order to predict future trajectories of the mobiles.

Data mining techniques capitalize on the simple observation that *the movement of people consists of random movements and regular movements* and the majority of mobile users has some regular daily (hourly, weekly,…) movement patterns and follow these patterns more or less every day.

Several efforts targeting at location prediction exploited this regularity. The purpose of all these efforts is to discover movement regularities and code them into some form of "knowledge," say, sequences of cell visits. Thus, for a considered mobile user the system tries to match its current trajectory with some of the already

discovered sequences and provide appropriate predictions.

Yavas et al. (2004) proposed a method to derive such sequences in the form of association rules, which describe the frequently followed paths by mobile users. This method is a level-wise algorithm, like the *Apriori* (Agrawal & Srikant, 1994), but takes into consideration the cellular structure of the PCS system. In each iteration, instead of generating all possible candidate paths by combining the frequent paths discovered in the previous iteration, it generates only the candidates that comprise legal paths over the coverage area. Thus, it achieves a significant reduction in the processing time. Similar reasoning was followed by Lee & Wang (2003) and Peng & Chen (2003), though the latter method is based on the application of the sequential patterns paradigm (Srikant & Agrawal, 1996).

Unlike the aforementioned works, which treated the location prediction problem as an association rule generation problem, the works by Katsaros et al.,(2003) and by Wu et al. (2001) investigated solutions for it based on the clustering paradigm. They treated the trajectories of the mobile users as points in a *metric space* with an associated distance function. The first work treated the trajectories as sequences of symbols (each cell corresponds to a symbol), utilized as distance function the string-edit distance and applied hierarchical agglomerative clustering in order to form clusters of trajectories. Each cluster is represented by one or more *cluster representatives* and each representative is a sequence of cells. Similar methodology was followed by Wu et al. (2001), but they used the standard Euclidean space and Euclidean distance function, that is, the L_2 norm.

SCHEDULING BROADCASTS

For the purpose of discovering data dependencies and subsequently scheduling the broadcast of these items closely in time, Saygin & Ulusoy (2002) proposed the use of association rule mining to the log files of the base servers, which record the data requests. Having discovered these dependencies, then a *correlation graph* is constructed, which depicts the correlated data requests. Applying a topological sorting over this graph, the authors derive the broadcast schedule. The main characteristic of this schedule is that items, which are frequently requested together by the clients, are broadcasted either consecutive or with very small distance in time. In this way, the average latency of the clients' requests is significantly reduced.

Aiming to reduce the mobile client latency associated with data retrieval, the work of Song & Cao (2004) designed a prefetching scheme for mobile clients. Since prefetching also consumes system resources such as bandwidth and power, they considered the system overhead when designing the prefetching scheme and proposed the cache-miss-initiated prefetch (CMIP) scheme to address this issue. The CMIP scheme relies on two prefetch sets: the always-prefetch set and the miss-prefetch set. The always-prefetch set consists of data that should always be prefetched if possible. The miss-prefetch set consists of data that are closely related to the cache-missed data item. When a cache miss happens, instead of sending an uplink request to ask for the cache-missed data item only, the client also requests for the data items, which are within the miss-prefetch set. This reduces not only future cache misses but also the number of uplink requests. Song & Cao proposed novel algorithms to mine the association rules and used them to construct the two prefetch sets.

FUTURE TRENDS

The application of data mining techniques to the improvement of wireless networks performance proved to be an effective tool; though, the proposed techniques to date are straightforward. More sophisticated methods are needed to support, for example, data allocation schemes that utilize the knowledge of user moving patterns for proper allocation of shared data in a mobile computing system. In addition, the knowledge discovered from telecommunication alarm data can be used in finding problems in networks and possibly in predicting severe faults or detecting intrusion attempts. For these application areas new mining procedures are needed.

CONCLUSION

Traditionally, the data mining process has been used to develop models, which describe the data. Recently though, it was realized that *these models can be effectively used to predict characteristics of the data*. This observation has led to a number of data mining methods used to improve the performance of wireless mobile networks. The aim of the present article is to present the fields where these methods can be applied and also to provide an overview of the particular data mining techniques, which have been developed into these fields.

REFERENCES

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *Pro-*

ceedings of the International Conference on Very Large Data Bases (VLDB'94) (pp. 487-499).

Aljadhai, A., & Znati, T. (2001). Predictive mobility support for QoS provisioning in mobile wireless environments. *IEEE Journal on Selected Areas in Communications*, *19*(10), 1915-1930.

Hadjiefthymiades, S., & Merakos, L. (2003). Proxies + path prediction: Improving Web service provisioning in wireless mobile communication. *ACM/Kluwer Mobile Networks and Applications*, *8*(4), 389-399.

Karakaya, M., & Ulusoy, O. (2001). Evaluation of a broadcast scheduling algorithm. In *Proceedings of the Conference on Advances in Databases and Information Systems (ADBIS'01)* (pp. 182-195).

Katsaros et al. (2003). Clustering mobile trajectories for resource allocation in mobile environments. In *Proceedings of the 5th Intelligent Data Analysis Symposium (IDA)*. Lecture Notes in Computer Science (2810) (pp. 319-329).

Kyriakakos, M., Frangiadakis, N., Merakos, L., & Hadjiefthymiades, S. (2003). Enhanced path prediction for network resource management in wireless LANSs. *IEEE Wireless Communications, 10*(6), 62-69.

Lee, A.J.T., & Wang, Y.-T. (2003). Efficient data mining for calling path patterns in GSM networks. *Information Systems*, *28*, 929-948.

Liang, B., & Haas, Z. (2003). Predictive distance-based mobility management for multidimensional PCS networks. *IEEE/ACM Transactions on Networking*, *11*(5), 718-732.

Nanopoulos, A., Katsaros, D., & Manolopoulos, Y. (2003). A data mining algorithm for generalized Web prefetching. *IEEE Transactions on Knowledge and Data Engineering, 15*(5), 1155-1169.

Peng, W.-C., & Chen, M.S. (2003). Developing data allocation schemes by incremental mining of user moving patterns in a mobile computing system. *IEEE Transactions on Knowledge and Data Engineering*, *15*(1), 70-85.

Saygin Y., & Ulusoy, Ö. (2002). Exploiting data mining techniques for broadcasting data in mobile computing environments. *IEEE Transactions on Knowledge and Data Engineering*, *14*(6), 1387-1399.

Srikant, R., & Agrawal, R. (1996). Mining sequential patterns: Generalizations and performance improvements. In *Proceedings of the International Conference on Extending Database Technology (EDBT'96)* (pp. 3-17).

Song, H., & Cao, G. (2004). Cache-miss-initiated prefetch in mobile environments. In *Proceedings of the International Conference on Mobile Data Management (MDM'04)*.

Wu, H.-K., Jin, M.-H., Horng, J.-T., & Ke, C.-Y. (2001). Personal paging area design based on mobile's moving behaviors. In *Proceedings of the IEEE Conference on Computer and Communications (IEEE INFOCOM'01)* (pp. 21-30).

Yavas, G., Katsaros, D., Ulusoy, O., & Manolopoulos, Y. (2004). *A data mining approach for location prediction in mobile environments*. Submitted for publication.

KEY TERMS

Calling Path: A calling path $\langle c_1, c_2, \ldots, c_n \rangle$ $n \geq 2$, is a sequence o visited cells during a mobile phone call, where $c_1, c_2, \ldots, c_n$ are cell IDs.

Handoff Or **Handover:** It is the process of changing some of the parameters of a channel (frequency, time slot, spreading code, or a combination of them) associated with the current connection in progress. Handoffs are initiated by a client's movement, by crossing a cell boundary, or by a deteriorated quality of signal received on a currently employed channel.

Hierarchical Agglomerative Clustering (HAC): A family of clustering algorithms, which start with each individual item in its own cluster and iteratively merge clusters until all items belong in one cluster.

MANET: A Mobile Adhoc Network (MANET) is a local network with wireless or temporary plug-in connection, in which mobile or portable devices are part of the network only while they are in close proximity.

Prefetching: It is the technique of deducing future client requests for objects based on the current request, and bringing those objects into the cache in the background before an explicit request is mode for them.

Push-Caching: The technique of pushing data closer to consumers by making an informed guess as to what the clients may access in the near future. The concept of push-caching is closely related to prefetching, but prefetches are always initiated in response to an on-demand request.

String Edit Distance: The edit distance between two strings is defined as the minimum number of edit operations – insertions, deletions, and substitutions – needed to transform the first string into the second (matches are not counted).