Data Mining in Medical Record Access Logs
Phillip V. Asaro, M.D., Integrated Technology Services, Health Sciences Center James E. Ries, M.S., Department of Health Management and Informatics and Department of Computer Engineering and Computer Science University of Missouri, Columbia, Missouri

Background. The necessary logging of all transactions of a computerized medical record system creates voluminous access logs. One use for access logs is the auditing mechanism that is described by the proposed HIPAA security regulations1. Because of the typically large numbers of access records, it is crucial to focus the attention of auditing efforts. Data mining has been studied and used extensively in security applications, particularly in attempts to identify network intrusion2. Applying the techniques at the application level in healthcare is more difficult due to the complexity of healthcare activities. Nonetheless, there are good reasons to make every reasonable effort to protect the confidentiality of private healthcare records including patient confidence, regulatory mandates, and cost-avoidance from a legal liability standpoint, not to mention a clear ethical obligation to do so3.

Methods. Our attempt at using a data mining approach to characterize user behavior in the computerized medical record begins by viewing the access logs using multiple dimensions or concept levels4. The data can be viewed at the transaction level, the session level, the user level (meaning user activity over a period of time), or the patient level (meaning activity within a particular patient’s chart over a period of time). For the purpose of this study we focused on the session level and user level.

The subject of our study was access data, over the period October 1999 to September 2000, from the System for Text Archiving and Retrieval (STAR) at University of Missouri Health Sciences Center. A series of SQL scripts were constructed which used the access-log table along with other tables in STAR to create a transaction level data table, a session level data table, a user level data table. The user level table was then joined to the session level table and we then used the SGI product, MineSet (Windows version 3.1) to analyze this single-table database. We used the MineSet clustering tool. We initially clustered on user level attributes. We found that a few of the clusters contained users with much higher numbers of document views than the remainder of the database. Focusing on just these clusters, we then clustered on session level attributes.

Results. Among the observations made of the resulting clusters was that one of the clusters had an unusually high (13.7) ratio of the number of distinct patient charts opened during a session to the number of distinct types of documents viewed during the session. Considering the other characteristics of these sessions, at least some of these sessions were atypical for clinically related activities. For example, in one 34-minute session, the user accessed the charts of 45 different patients, viewed 29 documents of only two different document types averaging 18 days old, and printed 26 documents.

Conclusion. Our results anecdotally suggest that data mining techniques may be useful in revealing inappropriate access patterns in medical record access logs. Further work will be needed to develop practical methodologies for applying these techniques.

Acknowledgements. This work was partially supported by the National Library of Medicine training grant LM-0708-09. The development of STAR was supported by IAIMS grant LM05415-7 and Telemedicine grant LM-6-3538.

References