

A Case Manager Tool for Anomaly Investigation in BioSurveillance

Colin Goodall, Ph.D.¹, Guy Jacobson, Ph.D.², Arnold Lent, Ph.D.¹, Sam Parker, Ph.D.², Simon Tse, Ph.D.², John Allegra, M.D., Ph.D.³, Dennis Cochrane, M.D.³, Eileen Koski, M.Phil.⁴, Patrick Tellez, M.D., M.P.H.⁵

¹*AT&T Labs – Research, Middletown, NJ*, ²*AT&T Labs – Research, Florham Park, NJ*, ³*Emergency Medical Associates Research Foundation, Livingston, NJ*, ⁴*Quest Diagnostics Incorporated, Lyndhurst, NJ*, ⁵*MedPlus, A Quest Diagnostics Company, Mason, OH*

OBJECTIVE

Effective anomaly detection depends on the timely, asynchronous generation of anomalies from multiple data streams using multiple algorithms. Our objective is to describe the use of a case manager tool for combining anomalies into cases, and for collaborative investigation and disposition of cases, including data visualization.

BACKGROUND

Current research at AT&T includes the analysis, understanding, early event detection, and visualization of large and massive streams of data, such as provisioning, call detail, and IP traffic data. The AT&T technologies for early event detection [EED] in health data have their genesis in a background and institutional experience in large scale computation and spatial and temporal data mining. Previously, we have shown applications to early event detection in large data bases of health data, working with a national provider of laboratory tests, and, in a related project, to hospital emergency department [ED] data, including text-based chief complaint data.

The early event detection tools include modules for: (1) data extraction, transformation, and loading [ETL], (2) a database [DB] to support very large scale data storage and access, (3) multiple anomaly detection methods, including text mining, for EED, (4) a flexible platform for investigation and management of anomalies, and (5) tools for spatial and temporal data visualization.

In this report, we focus on (4), the generation, management, and disposition of cases. A case here is a logical combination of multiple anomalies. We describe technology, methods, and practice of case management.

METHODS

Emergency department data from Emergency Medical Associates of New Jersey include some 4.2 million records for ED visits to 24 hospitals in the New Jersey and New York, for the years 1996-2005. The record for each visit includes hospital, patient geography (city/state/ZIP5), age, sex, ICD9 diagnoses, chief complaint, and other fields. Data extraction was from an Oracle® DB at EMA to a Daytona® DB at AT&T. For this study we also used a subset of several million electronic laboratory records [ELR] from the Quest Diagnostics Incorporated Data Warehouse from the New

York City metropolitan area. The records spanned a 15 month period and are from two syndromic groups, respiratory and gastro-intestinal.

A suite of anomaly detection algorithms were applied to both sources of data. The resulting anomalies were combined into cases using the logic of the Yoix® case manager tool. A case is defined based on (i) medical condition, including but not limited to syndrome, chief complaint, ICD9 codes, laboratory test types and results, (ii) demographic variables such as age and sex, (iii) geographical region such as ZIP5s, and (iv) time interval.

The case manager is a configurable Java application using established Yoix technology that, elsewhere at AT&T and in various configurations, is used by thousands of associates at tens of work centers. The case manager tool for health data comprises two linked configurations, one for managing cases and the other for drilling down and visualization of detail data. The case manager supports consequence management with functions for case investigation, case annotation, and case disposition – dismiss, defer, alert. Data visualization comprises time series and geographic views, capabilities for defining subsets dynamically and recursively, and structural views of data, notably of ICD9 codes, of word associations in text data, and of syndromic groupings.

RESULTS

The visualization tools allowed us quickly and painlessly to investigate in real time each anomaly, such as the meningitis scare, the “Clinton” chest pain effect, ED visits related to possible anthrax exposure around October 2001, and ZIP3-level numbers of viral test requisitions. We explore each anomaly by hospital, by patient ZIP5, age, gender and ICD9 codes, singly and in combination, with a few clicks of the computer mouse.

CONCLUSIONS

The case manager provides a disciplined and flexible central element in the EED process. This tool can be even more effective for discriminating and categorizing across various levels of alerts when connected to an alert management system. The use of the case manager for ED visits and ELR supports our operational experience of data mining at scale and with predictable performance, security, and availability.