Comparison of automated geocoding methods in a distributed multi-county disease surveillance system

Bryant T Karras MD [a, b], David Bliss [a], Steve Wagner MS [a], Scott Horn [c], Scott Lindquist MD MPH [a, c],

Center for Public Health Informatics

[a] *Department of Health Services, School of Public Health and Community Medicine,*

[b] *Department of Medical Education & Biomedical Informatics, School of Medicine,*

University of Washington, Seattle, WA; [c] *Kitsap County Health District, Bremerton, WA.*

Abstract

This paper describes a distributed multi-county system that remotely processes addresses and identifiers on servers behind hospital firewalls and then transmits geocodes and cryptographically protected identifiers from hospitals in three-counties to a central public health authority. In this study, we compare a new geocoding application's performance (completion rate and speed of execution) to the previous system's. Using both systems, we are able to achieve a 91% success rate using a sample set of 1,978 addresses.

Background

The University of Washington's Center for Public Health Informatics, in collaboration with the Kitsap County Health District and the UW Clinical Informatics Research Group, has developed the Peninsula Syndromic Surveillance Information Collection System (SSIC), a complex second-generation [1,2] distributed database system which collects heterogeneous data from three emergency department / urgent care facilities' computerized electronic admission and discharge diagnosis data. We transform heterogeneous institution-specific data to a standardized XML (eXtensible Markup Language) format, which is then transmitted to and integrated into a central database. Aberration detection algorithms are used to analyze this data so that public health officials can detect higher-than-usual incidences of the clinical syndromes under surveillance.

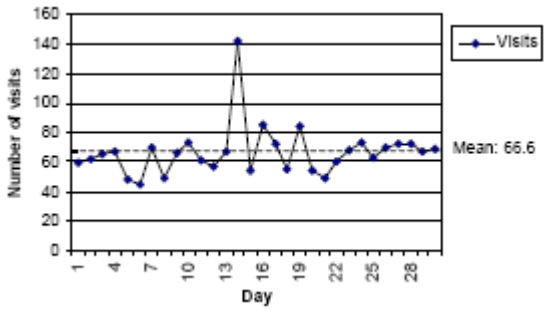
Methods

Patient addresses are coded at the hospital site to the census block group level. This block group is sent to the central server to allow geographic visualization and space-time cluster analysis while preserving patient privacy. Formerly this analysis was performed using a system based on ESRI's ArcGIS product; we have recently changed to a system developed from the ground up in Perl for UW's SSIC system. We compare the completion rate and speed of the two systems using both a typical day (about 70 records) and a typical month (about 2000 records).

Results

The older ArcGIS-based system successfully coded 62 out of 76 records (81.6%) without human intervention. The UW's Perl-based system successfully

Hospital Visits Per Day in 30-Day Period



coded 63 out of 76 records (82.9%). More significantly, the required processing time was 28 seconds for the ArcGIS-based system and 2 seconds for the Perl-based system. On the large data set, completion rates are 87.3% (46 sec) and 83.4% (62 sec) respectively. However, the records successfully coded by each system are significantly non-overlapping; combining the two systems gives a rate of 91.0% with 73.9% coded identically by both systems, 10.8% by only one system, and 6.3% coded differently by both.

Conclusions

Differing approaches to geocoding of syndromic surveillance data can yield similar completion rates at widely differing speeds. The selection of the faster system may depend on the data set size. Combining disparate approaches can give higher completion rates than either system alone, and also generate a confidence measure in the results. "Gold standard" validation remains as future work.

References

[1] Lober WB, Trigg L, Karras BT. Information system architectures for syndromic surveillance. Morbidity and Mortality Weekly Report (MMWR). 2004 Sept 24; 53(Suppl);203-208.

[2] Mandl K, Overhage J, Wagner M, et al. Implementing Syndromic Surveillance: A Practical Guide Informed by the Early Experience. Journal of the American Medical Informatics Association. 2004 Mar-Apr;11(2):141-150.

Further Information: Bryant Karras, bkarras@u.washington.edu, http://cphi.washington.edu.

[3] U.S. Census www.census.gov/geo/www/cob/bg_metadata.html

Further Information: Bryant Karras, bkarras@u.washington.edu, http://cphi.washington.edu