

Natural language processing of Echocardiography reports

Mangaraju Chakka

A. Study purpose and rationale

Clinical information is largely locked up in narrative reports like discharge summaries, admission history and physical examination reports, reports of diagnostic tests. Though some of these reports are available in electronic form on large hospital systems, this information can't be manipulated easily. Extracting this information by automation has great clinical significance.

Natural language processor is an effective solution. The processor converts narrative reports to a coded description that is appropriate to automated systems. Encoded data can be used in wide range of applications like Decision support, Quality assurance, Retrospective and prospective studies, Outcomes assessment, Resource management, Patient management, Summarization of patient data, Generating problem lists, Automated coding for information retrieval purposes.

MedLEE - a medical language extraction and encoding system is part of our web based clinical information system. The system has been trained and validated for chest x-ray reports and mammography reports. It is trained for processing discharge summaries.

In this study, we would train MedLEE for echocardiography reports and validate the system,

B. Study design and statistical analysis

Echocardiography has domain specific vocabulary. The initial step is to identify the vocabulary and include it in the Lexicon. The training process is an iterative one using echocardiography reports prior to 1996.

Validation is crucial prior to using the system for various potential applications. MedLEE had been validated for CXR reports and Mammography reports. It was shown to be very close to physicians in deciding if a chest x-ray report has a particular clinical finding. The sensitivity was 98.3(97.4 to 99.0) and specificity was 80.6(72.8 to 87.2).

We plan to validate MedLEE based on six cardiac clinical conditions. They are Left ventricular wall motion abnormality, Ejection fraction, Mitral valve regurgitation, Aortic stenosis, Regional wall motion abnormality, Left atrial enlargement. Hundred echocardiography reports done after 1996 will be randomly chosen. Logic modules will be written to identify these clinical conditions in the hundred reports. The result would be that each condition is present or not.

Two cardiologist who are not involved in the process of training the system would be asked to review and mark if each of the conditions are present. If there is a disagreement, a third cardiologist would be asked to break the tie. The results of the system would be compared against that of cardiologists to calculate the sensitivity and specificity.

C. Study procedures

None

D. Study drugs

None

E. Medical devices

None

F. Study questionnaires

None

G. Study subjects

Echocardiography reports will be used for the study

H. Recruitment of subjects

Random selection of the echocardiography reports.

I. Confidentiality of study data

Names of the patients will be masked while handing the reports to the cardiologists.

J. Potential conflict of interest

None

K. Location of the study

New York Presbyterian hospital

L. Potential risks

None

M. Potential benefits

Application in improving patient care

N. Alternative therapies

Not applicable.

O. Compensation to subjects

Not applicable.

P. Costs to subjects

None.

Q. Minors as research subjects

None

R. Radiation or radioactive substances

None

S. Bibliography

- a. Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. *Ann Intern Med* 1995;122:681-8
- b. Hripcsak G, Kuperman GJ, Friedman C. Extracting findings from narrative reports: software transferability and sources of physician disagreement. *Method Inform Med*. 1993;37:1-7
- c. Friedman C, Hripcsak G, DuMouchel W, Johnson SB, Clayton PD. Natural language processing in an operational clinical information system. *Natural Language Engineering*. 1995;1:83-108
- d. Knirsch CA, Jain NL, Pablos-Mendez A, Friedman C and Hripcsak G. Respiratory isolation of tuberculosis patients using clinical guidelines and an automated clinical decision support system. *Infection Control and Hospital Epidemiology*. 1998;19(2):94-100