Identifying Patients for Clinical Studies from Electronic Health Records: The TREC Medical Records Track

William Hersh, MD
Professor and Chair
Department of Medical Informatics & Clinical Epidemiology
School of Medicine
Oregon Health & Science University
Email: hersh@ohsu.edu
Web: www.billhersh.info
Blog: informaticsprofessor.blogspot.com

OHSU Research Week – VERY SHORT

You can find a longer version of this talk:
http://oninformatics.com/?p=834
These slides on my web site:
http://www.billhersh.info
There is also more about the TREC Medical Records Track on my blog:
http://informaticsprofessor.blogspot.com

References


Identifying Patients for Clinical Studies from Electronic Health Records: The TREC Medical Records Track

William Hersh, MD
Professor and Chair
Department of Medical Informatics & Clinical Epidemiology
School of Medicine
Oregon Health & Science University
Email: hersh@ohsu.edu
Web: www.billhersh.info
Blog: informaticsprofessor.blogspot.com

Caveat

• I cannot even scratch the surface of this project in 10 minutes
• BUT
• You can find a longer version of this talk
  – http://oninformatics.com/?p=834
• And these slides on my web site
  – http://www.billhersh.info
• As well as materials on my blog
  – http://informaticsprofessor.blogspot.com
Motivations for secondary use of clinical data

- Many “secondary uses” or re-uses of electronic health record (EHR) data, including (Safran, 2007)
  - Personal health records (PHRs)
  - Clinical and translational research – generating hypotheses and facilitating research
  - Health information exchange (HIE)
  - Public health surveillance for emerging threats
  - Healthcare quality measurement and improvement
- Little controlled research on how best to do it
- Opportunities facilitated by growing incentives for “meaningful use” of EHRs in the HITECH Act (Blumenthal, 2011; Blumenthal, 2011)

Information retrieval (IR) evaluation (Hersh, 2009)

- Assessed with *test collections*, which consist of
  - Content – fixed yet realistic collections of content
  - Topics – statements of information need
  - Relevance judgments – by expert humans for which content items should be retrieved for which topics
- Evaluation consists of *runs* using a specific IR approach with output for each topic measured and averaged across topics
- Variety of measures to assess retrieval of “relevant” information, e.g., recall, precision, and aggregations thereof
Challenge evaluations

• A common approach in computer science, not limited to IR
• Develop a common task, data set, evaluation metrics, etc., ideally aiming for real-world size and representation for data, tasks, etc.
• In IR, oldest and largest is Text Retrieval Conference (TREC, trec.nist.gov; Voorhees, 2005) – sponsored by National Institute for Standards and Technology (NIST)
  – Many “tracks” of interest, such as routing/filtering, Web searching, question-answering, etc.
  – Operates on annual cycle of test collection release, experiments, and analysis of results
  – Non-medical, with exception of Genomics Track (Hersh, 2009)

TREC Medical Records Track

• Facilitated by availability of a large-scale, de-identified corpus of medical records from University of Pittsburgh Medical Center (UPMC)
• Task: identify patients for possible inclusion in clinical research studies
• Topic development and relevance assessment carried out by OHSU
• Participation of 29 research groups who could submit up to 8 runs each
  – Total of 127, with 109 automatic and 18 manual
  – Each run scored with bpref averaged over topics
Test collection

VISIT LIST

RECORD-VISIT MAP

Report Extract

Wide variations in number of documents per visit

(Courtesy, Ellen Voorhees, NIST)
Evaluation results for top runs ...

... BUT, wide variation among topics
Easy and hard topics

• Easiest – best median bpref
  – 105: Patients with dementia
  – 132: Patients admitted for surgery of the cervical spine for fusion or disectomy

• Hardest – worst best bpref and worst median bpref
  – 108: Patients treated for vascular claudication surgically
  – 124: Patients who present to the hospital with episodes of acute loss of vision secondary to glaucoma

• Large differences between best and median bpref
  – 125: Patients co-infected with Hepatitis C and HIV
  – 103: Hospitalized patients treated for methicillin-resistant Staphylococcus aureus (MRSA) endocarditis
  – 111: Patients with chronic back pain who receive an intraspinal pain-medicine pump

What approaches did (and did not) work?

• Best results obtained from NLM group (Demner-Fushman, 2011)
  – Top results from manually constructed queries using Essie domain-specific search engine (Ide, 2007) – BPref = 0.658
  – Other automated processes fared less well, e.g., creation of PICO frames, negation, term expansion, etc. – BPref = 0.4822

• Best automated results obtained by filtering by age, race, gender, admission status; terms expanded by UMLS Metathesaurus – BPref = 0.552 (King, 2011)

• Benefits of approaches commonly successful in IR did provided small or inconsistent value for this task

• Nor did manual queries and using ICD-9 codes (Bedrick, 2009)
Conclusions and future directions

• TREC Medical Records Track extended IR challenge evaluation approach to a patient selection triage task
• Initial results show mixed success for different methods – common with a new IR task
• Future work will aim to expand test collection, tasks, and topics – aiming to develop best approaches for variety of tasks