

MINING ELECTRONIC MEDICAL RECORDS

Chris Jermaine
Rice University

Electronic Medical Records

- Increasingly, your medical history is stored electronically
- EMRs are aggregated into “Clinical Data Warehouses”
- Useful to medical researchers
 - Retrospective studies (“is drug A effective?”)
 - Association studied (“is disease B linked to obesity?”)
 - Design of clinical trials

But Not So Easy to Answer Questions

- Might have 100K patient records in a CDW
 - Contain structured data, images, text...
- Want to answer: “Which patients have had breast cancer?”
- How to do this?

But Not So Easy to Answer Questions

- Might have 100K patient records in a CDW
 - Contain structured data, images, and (especially) text...
- Want to answer: “Which patients have had breast cancer?”
- How to do this?
 - Look for breast-cancer-related ICD-9 billing code....
 - ...but this will get you only around 50% of the positive cases
 - Look for cancer-related drugs, do keyword search, etc. etc...
 - ...works, but still have FP rate of 90% at FN rate of 10%
 - In 100K patients, will be ~1K with BC, so ~10K total records to examine...
 - At 30 minutes per record, is around 2 years of human time
 - **How to reduce this?**

Our Idea

- Ask medical researcher to provide a “weak learner”
 - One with very, very few FP
 - Might be a simple billing code check
- Use that learner to identify a small number (~100) positive cases
- Ask medical researcher go through records, label important parts
 - Those parts someone should look at to identify this as a “yes” case

Our Idea

- Ask medical records
 - One with very similar
 - Might be a similar
- Use that learned
- Ask medical records
 - Those parts similar

Human Expert Highlighting

Ms. [redacted] is a [redacted]-year-old female who presents for further evaluation of a right breast mass. The patient first noted the lesion in her right breast on [redacted]. Approximately 3 months prior to that... [long text] ...and a biopsy was recommended. The patient was referred for further management. The patient denies any change in the lesion since she has noticed it and has no history of any pain or nipple discharge and no history of any lesions on the left by the patient report and normal mammogram 3 months prior. PAST MEDICAL HISTORY: Asthma. PAST SURGICAL HISTORY: Tubal ligation, cholecystectomy, and rotator cuff repair. FAMILY HISTORY: She had a grandmother who died of breast cancer in her early 's as well as maternal great aunt who died of breast cancer. No family history of... [long text] On the patient history form, again, she noted a right breast mass, which has not change in size since she has noticed it. The rest of the review of systems is negative per the patient's history... [long text] BREASTS: She has no supraclavicular or infraclavicular or axillary adenopathy. On inspection, she has no erythema or dimpling. No nipple retraction. On palpation of her right breast she a 3 x 4-cm mass located at the 6 o'clock position 6 cm from the nipple. It is firm and consistent with a malignancy. It is moveable, but I believe it is fixed to the muscle. She has no palpable masses in the left breast. IMPRESSION: ...

Similar cases
Important parts

Our Idea

- Ask medical researcher to provide a “weak learner”
 - One with very, very few FP
 - Might be a simple billing code check
- Use that learner to identify a small number (~100) positive cases
- Ask medical researcher go through records, label important parts
 - Those parts someone should look at to identify this as a “yes” case
- Develop a machine learning method that learns how to highlight
 - Examines the “yes” documents
 - Figures out what parts of text are important

Our Idea

- Then for the ~10K cases with ~9K false positives...
 - Automatically highlight to speed up examination time
 - 30 minutes down to 1 minute

Does Our Method Work?

- You be the judge:

Human Expert Highlighting

Ms - is a - year-old female who presents for further evaluation of a right breast mass. The patient first noted the lesion in her right breast on - . Approximately 3 months prior to that... [long text] ..and a biopsy was recommended. The patient was referred for further management. The patient denies any change in the lesion since she has noticed it and has no history of any pain or nipple discharge and no history of any lesions on the left by the patient report and normal mammogram 3 months prior. PAST MEDICAL HISTORY: Asthma. PAST SURGICAL HISTORY: Tubal ligation, cholecystectomy, and rotator cuff repair. FAMILY HISTORY: She had a grandmother who died of breast cancer in her early - 's as well as maternal great aunt who died of breast cancer. No family history of... [long text] On the patient history form , again, she noted a right breast mass, which has not change in size since she has noticed it. The rest of the review of systems is negative per the patient's history... [long text] BREASTS: She has no supraclavicular or infraclavicular or axillary adenopathy. On inspection, she has no erythema or dimpling. No nipple retraction. On palpation of her right breast she a 3 x 4-cm mass located at the 6 o'clock position 6 cm from the nipple. It is firm and consistent with a malignancy. It is moveable, but I believe it is fixed to the muscle. She has no palpable masses in the left breast. MPRESSION: ...

WLR Highlighting

Ms - is a - year-old female who presents for further evaluation of a right breast mass. The patient first noted the lesion in her right breast on - . Approximately 3 months prior to that... [long text] ..and a biopsy was recommended. The patient was referred for further management. The patient denies any change in the lesion since she has noticed it and has no history of any pain or nipple discharge and no history of any lesions on the left by the patient report and normal mammogram 3 months prior. PAST MEDICAL HISTORY: Asthma. PAST SURGICAL HISTORY: Tubal ligation, cholecystectomy, and rotator cuff repair. FAMILY HISTORY: She had a grandmother who died of breast cancer in her early - 's as well as maternal great aunt who died of breast cancer. No family history of... [long text] On the patient history form , again, she noted a right breast mass, which has not change in size since she has noticed it. The rest of the review of systems is negative per the patient's history... [long text] BREASTS: She has no supraclavicular or infraclavicular or axillary adenopathy. On inspection, she has no erythema or dimpling. No nipple retraction. On palpation of her right breast she a 3 x 4-cm mass located at the 6 o'clock position 6 cm from the nipple. It is firm and consistent with a malignancy. It is moveable, but I believe it is fixed to the muscle. She has no palpable masses in the left breast. MPRESSION: ...

That's It!