**Why mine healthcare text?**

**Anatomy of what patients currently use:**
- webmd (e.g. drugs.com, yahoo health, etc.)

**Q/A**
- original content
- fresh
- simple to understand
- good coverage in the head

**Topic Pages**
- original content
- moderately authoritative
- simple to understand
- good coverage in the head

**News + Media**
- original content
- typically aligned with google searches
- i.e. treatment for X, symptoms for Y

**Experts**
- original content
- provides answers

**Discussions**
- original content
- provides answers

**Original content=important**
- providing answers=important
- head=developed
- fresh=important
- authority=important

**Types of text mining**
- Document Classification
- Information Extraction
- Text Summarization
- Keyphrase Extraction

**Automated content generation**
- cost effective
- structured content
- statistically based
- scales to millions of patients
- scales to long tail treatments + conditions
- authoritative / citation driven
- fresh

**Alliance health networks**

HCP/advanced patients
- medify.com – 10,000% user growth past year

Patients
- diabeticconnect.com - #1 online diabetes site @ ~1.4M uniques
- *connect.com – content sites + guided social networks

Health care industry
- patient surveying, matchmaking and analysis

**Deep Dhillon**

Chief data scientist | ddhillon@alliancehealth.com | twitter.com/zang0

**Mining unstructured healthcare data**
medify demo

how do we mine text?

annotate in the product:

- Easier for people to give explicit GT feedback
- Channels high visibility error angst productively
- Override high visibility system mistakes with human data

gt demo:

- http://www.diabeticconnect.com/discussions/4790
- https://www.medify.com/internal/annotate/abstract?abstractId=8181575
Detailed Signals Preliminary Mining From Discussion Threads

Distribution of Signals Mined From Diabetic Connect Discussion Threads

knowledge base

knowledge base

message

sentence split

shallow tagging

dependency tree parse

triple extraction

classification

concluder

confidences

feature engineering

- entities, i.e. normalize synonyms, id new entity types, like social relations
- entity types, i.e. metformin > treatment_medication
- phrase driven cues, i.e. [have] [you] [considered] > suggestion_indicator
anaphora resolution

• relation structure, i.e. [person] takes it
  it refers to treatment (i.e. not condition/symptom), and specifically: medication but not device
• statistically driven, manually curated cues
  i.e. drug > treatment/medication
• filter
  — non matching antecedent candidates
  — singular/plural agreement
• score candidates:
  — antecedent occurrence frequency
  — distance (#sents) from antecedent to anaphor
  — co-occurrence of anaphor w/ antecedent

technology

• Lang: Java + Python + Ruby
• DB: Solr 4, Mongo DB, S3
• Work: Map Reduce
• Dependencies: Malt Parser, Stanford Parser
• Misc: Tomcat, Spring, Mallet, Reverb, Minor Third
• Tagging: Peregrine + home grown

Data Pipeline

request transaction flow

questions?
Abstract Summary

Prognostic Factors: The general consensus and health characteristics that have a bearing on the course of a disease and its eventual outcome. 

BREAST CANCER

Symptoms and Side Effects

Most Commonly Used Treatments

Terminated Treatments

Replacement Treatments
Advair: Experts vs. Patients

- "medicalese" vs. patients' words
- more granularity
- a story-like perspective w/ words of inspiration

My Pulmonologist today said that he just came from the hospital bedside of a patient with COPD symptoms who is in oxygen and not doing well. I hadn't been to a Pulmonologist in a while. It's good to see the patients I care for doing well.

As the man, my Pulmonologist is so busy with his own patients that he doesn't get much opportunity to see patients more broadly. He said that most of his patients are doing well. He had a patient this week for example who had managed to improve their symptoms. They're doing better.

He also said that people who take their medication regularly and do what they need to do to manage their disease tend to do better than those who don't.

By the way, my Pulmonologist still gets upset when he sees people who are not doing what they need to do to manage their disease. He said that people who are doing well are the exception rather than the rule.

My response to this: Why are people not doing what they need to do to manage their disease? They need to take control of their own health or no one else will.