“We’d like to start out being very involved with you but eventually be drawn away to much more interesting cases on Twitter.”
Hi. I have Graves disease, 6 months ago I increased my thyroxine, and felt better for it. this took my T4 levels to the limit but my T3 is still midway, (apparently acceptable)
Two months ago I was diagnosed as borderline Diabetic 2. I am not convinced of this, as I feel the coincidence between me increasing my thyroxine from 150 to 175 per day to then suddenly becoming diabetic (borderline) to easy. My theory is that by lifting my T4 level as high as it can go (safely) that this as increased my insulin or glucose levels. causing me to be showing as diabetic.
I there anyone out there that has any kno

sean

Judy5

Here is another piece of the puzzle to confuse you - I recently had the genetic test for DI02 and found that I have the heterozygous genetic fault (which means I have a problem with converting T4 to T3) and the genetic counsellor told me there is a link between this fault and Type 2 diabetes, so it is indeed possible for you to have both.

It might be worth considering getting the DI02 test done yourself. Details are on the front page of the Thyroid UK site.
Mining healthcare social media

- Extract
  - symptoms, problems
  - adverse events
  - treatments
  - life-style concepts
  - sentiment
  - psychological indicators
  - quality of life
Mining healthcare social media

When I’m on 60mg prednisolone I can’t sleep and want to eat 24/7. This is my last day on high dosage.

Metabolism and nutrition disorders - Increased appetite

Psychiatric disorders - Insomnia

Mining ADRs and benefits of steroids
Mining healthcare social media

- Can we use Twitter to generate mineable datasets from unsolicited posts regarding risk factors for people with schizophrenia
  - e.g. auditory hallucinatory experiences
  - e.g. sleep-related issues
  - e.g. suicidal thoughts

If hallucinating is thought of as hearing voices that are not actually real, then these painkillers are causing me to hallucinate like mad

So I was convinced I was hearing stuff. It was so funny because the noise was coming from the kitchen but I thought I was hallucinating
Mining healthcare social media

Text classification pipeline

raw (unstructured) text

Text Preprocessing

corrected text

Information Extraction

structured text

Classification

Im hearing a scary voice m, idk if it's in my head or in TV.. craazy

fear expr.

I am hearing a scary voice right now, I don’t know

possible hallucination

audio device

stigmatising lang.

it's in my head or in television..

label: related to hallucinatory experience

POS tagset from Gimpel et al. (2011): O - personal pronoun, V - verb, D - determiner, etc.
Mining healthcare social media

Evaluation

Classification performance of various classification methods on two different sets of features

- **NB**: Proposed features 0.831, Baseline features 0.486
- **SVM**: Proposed features 0.743, Baseline features 0.751
- **AdaBoost**: Proposed features 0.772, Baseline features 0.711

Based on ten experiments of stratified 10-fold cross validation.
Baseline features outperform only with SVM, difference is non-significant (p-value=0.375)
Mining healthcare social media

- **Negative sentiments** significantly associated with posts that indicated the occurrence of auditory hallucinations.

- Posts linked to auditory hallucinations had a **higher proportionate distribution** between the hours of 11pm and 5am.

---

Mining healthcare social media

• Issues:
  – Layman terminology
  – spelling errors
  – subjective...

• “Interpretation” of patient comments
#WhyWeTweetMH

Understanding why people use Twitter to discuss mental health problems

Four main themes:

- **Sense of community**
  “Because I am with friends even when I am unable to go out.”

- **Stigma and awareness**
  “…begin speaking about what’s actually important…”

- **Safe space for expression**
  “…because I’m never dismissed by my Twitter friends as being over sensitive, needing attention or not making enough of an effort.”

- **Coping and empowerment**
  “My Twitter timeline performs as a sort of mood monitor for myself and those who personally know me…”
## Topics in MH-related posts

<table>
<thead>
<tr>
<th>Theme</th>
<th>#Posts</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPD</td>
<td>11,880</td>
<td>Forum to discuss aspects of Borderline Personality Disorder either as a sufferer, someone closely related to a sufferer, or someone interested in this disorder</td>
</tr>
<tr>
<td>bipolar <em>(BipolarSOs, BipolarReddit, bipolar)</em></td>
<td>41,636</td>
<td>Communities to discuss issues surrounding Bipolar Disorder; while bipolar and BipolarReddit focus on sufferers and their support, BipolarSOs invites contributions from people that are in a relationship with someone suffering from Bipolar Disorder</td>
</tr>
<tr>
<td>schizophrenia</td>
<td>4,963</td>
<td>Subreddit to discuss schizophrenia-type disorders and schizophrenia-related issues such as psychosis</td>
</tr>
<tr>
<td>Anxiety</td>
<td>57,523</td>
<td>Forum for anything that is related to an anxiety disorder; does not distinguish between sufferer or someone related to a sufferer</td>
</tr>
<tr>
<td>depression</td>
<td>197,436</td>
<td>A community for helping anyone struggling with depression; posters are not limited to those who have received a diagnosis by their GP/hospital doctor and the emphasis is on supporting others in their struggle with depression</td>
</tr>
<tr>
<td>selfharm <em>(selfharm, StopSelfHarm)</em></td>
<td>17,102</td>
<td>Forums to discuss aspects of people self-harming; while selfharm aims to build a community of sufferers, StopSelfHarm focusses on supporting anyone wanting to stop self-harming even if through a related person</td>
</tr>
</tbody>
</table>

**Characterisation of mental health conditions in social media using Informed Deep Learning**

George Gkotsis*, Anika Oellrich, Sumithra Velupillai*, Maria Liakata, Tim J. P. Hubbard, Richard J. B. Dobson*, & Rina Dutta*
<table>
<thead>
<tr>
<th>Topics in MH-related posts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SuicideWatch</strong></td>
</tr>
<tr>
<td><strong>addiction</strong></td>
</tr>
<tr>
<td><strong>cripplingalcoholism</strong></td>
</tr>
<tr>
<td><strong>Opiates (OpiatesRecovery, opiates)</strong></td>
</tr>
<tr>
<td><strong>autism</strong></td>
</tr>
<tr>
<td><strong>Non-mental health</strong></td>
</tr>
</tbody>
</table>
Text mining patient feedback

• Patient feedback on services
  – treatment effectiveness
  – side effects
  – safety concerns
  – healthcare environment
  – communication and involvement
  – coordination of care

• Use sentiment analysis and topic modelling to identify topics and associated experience

This project is funded by the National Institute for Health Research (NIHR) HS&DR programme, project **14/156/16**. The views and opinions expressed are those of the authors and do not necessarily reflect those of the NIHR, the NHS or the Department of Health.
Text mining patient feedback

This project is funded by the National Institute for Health Research (NIHR) HS&DR programme, project 14/156/16. The views and opinions expressed are those of the authors and do not necessarily reflect those of the NIHR, the NHS or the Department of Health.
State of the art

i2b2 challenges in information extraction from clinical narratives

- De-identification (2006, 2014)
- Smoking Status (2006)
- Obesity and disease status (2008)
- Medications (2009)
- Concepts, assertions, relations (2010)
- Coreference resolution (2011)
- Temporal relations (2012)
- Risk factors (2014)
- Symptom severity in psychiatric notes (2016)
State of the art

Progression of CAD Risk Factors in Diabetic Patients Results

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Mean</th>
<th>Median</th>
<th>Max</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micro Precision</td>
<td>.455</td>
<td>.808</td>
<td>.852</td>
<td>.913</td>
<td>.119</td>
</tr>
<tr>
<td>Micro Recall</td>
<td>.203</td>
<td>.835</td>
<td>.908</td>
<td>.969</td>
<td>.175</td>
</tr>
<tr>
<td><strong>Micro F1</strong></td>
<td>.305</td>
<td>.815</td>
<td>.872</td>
<td>.928</td>
<td>.145</td>
</tr>
<tr>
<td>Macro Precision</td>
<td>.455</td>
<td>.800</td>
<td>.849</td>
<td>.914</td>
<td>.121</td>
</tr>
<tr>
<td>Macro Recall</td>
<td>.258</td>
<td>.834</td>
<td>.904</td>
<td>.968</td>
<td>.162</td>
</tr>
<tr>
<td>Macro F1</td>
<td>.365</td>
<td>.812</td>
<td>.870</td>
<td>.928</td>
<td>.137</td>
</tr>
</tbody>
</table>

i2b2 2014 challenge
# State of the art

<table>
<thead>
<tr>
<th>F measure</th>
<th>Precision</th>
<th>Recall</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.69</td>
<td>0.71</td>
<td>0.67</td>
<td>Rule based pair selection+CRF+SVM</td>
</tr>
<tr>
<td>0.69</td>
<td>0.75</td>
<td>0.64</td>
<td>MaxEnt+SVM+rule based</td>
</tr>
<tr>
<td>0.68</td>
<td>0.66</td>
<td>0.71</td>
<td>SVM</td>
</tr>
<tr>
<td>0.63</td>
<td>0.76</td>
<td>0.54</td>
<td>SVM+rule-based</td>
</tr>
<tr>
<td>0.61</td>
<td>0.54</td>
<td>0.72</td>
<td>CRF</td>
</tr>
<tr>
<td>0.59</td>
<td>0.65</td>
<td>0.54</td>
<td>MaxEnt/Bayes</td>
</tr>
<tr>
<td>0.56</td>
<td>0.57</td>
<td>0.56</td>
<td>Rule based+MaxEnt</td>
</tr>
<tr>
<td>0.56</td>
<td>0.48</td>
<td>0.66</td>
<td>SVM</td>
</tr>
<tr>
<td>0.55</td>
<td>0.51</td>
<td>0.59</td>
<td>SVM</td>
</tr>
<tr>
<td>0.43</td>
<td>0.34</td>
<td>0.59</td>
<td>MaxEnt</td>
</tr>
</tbody>
</table>

i2b2 2012 – extraction of temporal links  
Sun et al. J Am Med Inform Assoc 2013;
Mining clinical narratives

• **Challenges**
  – Highly condensed text, often without
    • proper spelling
    • proper sentences
    • specific discourse
  – Terminological variability and ambiguity
    • orthographic, acronyms, local conventions
    • mapping to standardised terminologies
Mapping to terminologies

I. Spasic
Automated coding

- E.g. using SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms)
  - must be adopted by all GPs before 1 April 2018.
  - Secondary Care, Acute Care, Mental Health, Community systems, Dentistry and other systems used in direct patient care by 1 April 2020.

- Can we automatically code

  "Chronic renal impairment (eGFR 44)"

  Chronic kidney disease stage 2 (disorder)
  SCTID: 431856006
Mining MRI reports

- Convert a radiologist's interpretation of the image into a structured form
- This information is then used by a clinician to support decision making on appropriate treatment

**HISTORY** Injury two weeks ago, ACL and lateral meniscal tear

**MRI LT KNEE** There has been a complete ACL tear in its mid portion. The medial meniscus is intact. There is a radial tear in the lateral meniscus. The PCL is intact. There is bone marrow oedema in the lateral femoral condyle consistent with trauma from a rotational injury. In addition there is a fragment following cartilage signal intensity lying just medial to the PCL insertion possibly representing a cartilage fragment from the lateral femoral condylar notch. There is a large joint effusion. The lateral ligamentous complex is intact. There is oedema surrounding the MCL consistent with a sprain but the ligament is intact. The posterolateral corner is intact. The patella cartilage is unremarkable.

**CONCLUSION** Complete ACL tear, radial tear in the lateral meniscus, MCL sprain, depression of the lateral femoral condylar notch with bone marrow oedema and a small cartilaginous fragment at the medial aspect of the PCL insertion.
TRAK ontology

Taxonomy for RehAbilitation of Knee conditions

I. Spasic
MRI LT KNEE - This confirms a tear of the posterior portion of the lateral meniscus which is intact.
The medial meniscus, cruciate and collateral ligaments are intact.
Results ~~~~~~~~ MRI RT KNEE - Both menisci are intact.
There is slight increased signal and widening of the ACL.
There are however intact fibers of the ACL and this probably represents the PCL is intact.
There is bone bruising in the lateral femoral condyle probably in relation to the lateral ligamentous complex is intact.
The collateral ligaments are intact.
There is however mild subcutaneous edema adjacent to the illo-tibial Extensor tendons are intact.
The articular cartilage is grossly intact.
CONCLUSION - ACL sprain and MCL partial tearing.
Sprain of the postero-lateral corner.

MRI LT KNEE - There has been a complete ACL tear.
There is no associated bone bruising demonstrated.
The PCL is intact.
The lateral meniscus as well as the popliteus tendon are unremarkable.
There is some increased linear signal in a vertical fashion in relation to the extensor tendons which could represent a very peripheral vertical tear. Please correlate clinical findings.
The extensor tendons are intact.
The articular cartilage is within normal limits.
The collateral ligaments are intact.
CONCLUSION - Complete ACL tear, possible vertical tear in the periphery of the posterior horn of the medial meniscus.

HISTORY ~~~~~~~~
medial meniscal tear MRI LT KNEE - There has been an ACL repair.
The graft is intact.
There is degeneration in the medial compartment with a very small residual medial meniscus.
This is probably due to meniscal surgery but there also has been re-tearing at the posterior horn.
There are marked cartilage irregularities with a local defect at the medial femoral condyle.
The lateral and patello-femoral compartmental cartilage is intact.
Normal lateral meniscus and PCL.
The extensor tendons and collateral ligaments are unremarkable.

I. Spasic
Challenges

• Negation identification
  – there is currently no evidence of a significant meniscal cyst
  – The low signal of the anteromedial bundle seen in a normal ACL is completely absent

• Suspected diagnosis
  – Likely primary Raynaud’s in hands and feet

• Family history
  – Her mother's brother was diabetic.

• Patient generated data
  – Once I start moving around or exercise the joint stiffness easies.
  – Constant pain weather sitting or standing.
# Challenges

- **Coordination** medial and lateral meniscus
- **Coreference** the medial meniscus ... the meniscus

- **Temporal information extraction**
  - recent scan, doesn’t feel well recently – is it the same ‘recent’
  - check the serum levels in 3 months – 90 days?
  - take 1 tablet with every meal – how many times?

- **Approximate expressions**
  - pea-sized nodule in the neck – how big is it?
State of the art – systems

- Open Health NLP - [http://www.ohnlp.org/](http://www.ohnlp.org/)
- GATE infrastructure
- NLTK
- Text mining with R
- A number of commercial products
State of the art – cTAKES

• cTAKES - http://ctakes.apache.org/

• Demo at

  http://chipweb2.chip.org/cTakes_webservice_trunk/index.html
State of the art

- A number of libraries for basic text processing, frequencies, word clouds, finding associations
State of the art

- **Word embeddings** – language models where words are represented as vector of numbers, which are learnt from a large corpus

If we use word context vectors...

In the rest of this talk, we refer to this notion of relatedness as *Typical* (by-type) similarity.
State of the art

- IBM Watson Development Cloud (ex AlchemyAPI)

200 medical textbooks and 300 medical journals.

http://drevidence.com/

Support to clinical decision support:

- Parse doctor’s query
- Parse EHR
- Parse guidelines, other data to test hypothesis
- Suggest individualised treatment options
Use case scenarios
State of the art

• Some off-the-shelf tools are useful for basic NLP tasks

• But, “there is no such thing as "THE TOOL" for text mining”
  – There will be always the need for tailoring
  – Some tools are more suitable for a given task, but less for the other

• Successful “text mining is 20% engineering, 40% algorithms and 40% science/statistics”
  – there is always going to be new challenges in every new problem you work on (even if it is similar to the previous one)

http://text-analytics101.rxnlp.com/
Group reflection: Needs

• What kind of text mining application would you like to have?
• What would be the opportunities and challenges?
Summary

• Loads of information is in healthcare free text
  – Clinical narrative, social media
  – Guidelines (e.g. NICE)

• Clinical language(s)
  – condensed text, overloaded with terminology
  – spelling errors, abbreviations (local?)
  – implicit information/assumptions

• Healthcare text mining
  – identify (key) entities and relations of interest
  – place the results in context
Summary

• What can we do with this data?
  – Support personalised medicine
    • E.g. tailor the therapy for an individual based on social and medical history, environment, allergies, genotype, etc.
  – Improve our understanding of the diseases
    • Identify patterns in genotypes and phenotypes
  – For audit, monitoring and surveillance
    • Addressing some legal obligations

• Text analytics will be an essential part of Learning Health Systems

→ Improve both clinical practice and science
Mental health
Research

Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project

Richard G Jackson, Rashmi Patel, Nishamali Jayatilleke, Anna Kolliakou, Michael Ball, Genevieve Gorrell, Angus Roberts, Richard J Dobson, Robert Stewart

Negative symptoms in schizophrenia: a study in a large clinical sample of patients using a novel automated method


Analysis of diagnoses extracted from electronic health records in a large mental health case register

Yevgeniya Kovalchuk, Robert Stewart, Matthew Broadbent, Tim J. P. Hubbard, Richard J. B. Dobson

European Child & Adolescent Psychiatry

Clinical predictors of antipsychotic use in children and adolescents with autism spectrum disorders: a historical open cohort study using electronic health records
Healtex

• UK healthcare text analytics research network
  – AIM: unlock the evidence contained in healthcare text

• Multi-disciplinary community
  – data/text analysts
  – clinicians, epidemiologists
  – semantic technologies
  – legal and data protection
  – NHS and industry
Healtex

- Partners and members

[Logos of various organizations]
Healtex

- Workshops
- Datathons
- Hackathons
- Working groups

- Call for pilot projects and feasibility studies (May 2017 and 2018)
Future workshops

<table>
<thead>
<tr>
<th>Workshop</th>
<th>Date (TBC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing patient generated data – opportunities, barriers and challenges</td>
<td>October 2017</td>
</tr>
<tr>
<td>What’s in a clinical narrative – how clinicians compose a free text narrative and why</td>
<td>November 2017</td>
</tr>
<tr>
<td>Preserving privacy and facilitating sharing of healthcare free text – is there a best practice for accessing clinical text?</td>
<td>February 2018</td>
</tr>
<tr>
<td>Patients’ view on using healthcare narrative for research – are patients concerned with sharing the narrative?</td>
<td>April 2018</td>
</tr>
<tr>
<td>Challenges of knowledge- and data-intensive text analytics etc. – what are the open problems in processing healthcare text analytics?</td>
<td>September 2018</td>
</tr>
<tr>
<td>Integration of clinical text into actionable healthcare analytics – how to make sense of free text?</td>
<td>January 2019</td>
</tr>
</tbody>
</table>

- First UK conference on healthcare text analytics – early 2018
- Workshop at Informatics for Health conference: “Extracting evidence from clinical free text: opportunities and challenges”
  Tuesday 9:30 (Exchange 2)
Join the network