Causality Driven Data Integration for Adverse Drug Reaction Discovery

Chen Wang and Sarvnaz Karimi

3 April 2014
Adverse Drug Reaction (ADR)

• ADR is costly
  • 400,000 visits to GPs and 190,000 hospital admissions each year are due to adverse drug events in Australia
    – National Safety and Quality Health Service Standards (Sept. 2012)

• ADR detection is complicated
ADR Detection: Passive vs. Active

• **Passive** ADR detection
  • Rely on manual reports submitted from health professionals, pharmaceutical companies or consumers → Spontaneous Report Systems (SRS)
  • Drawback: delay, under-reporting

• **Active** ADR detection
  • Automatically generate reports based on monitoring ADR related data from a variety of sources:
    – Patient health records, administrative and insurance claim databases, medical literature, social media…
  • Challenge: difficulty in generating high quality signals

• A high quality signal reveals the causal relationship between a drug and ADR caused by the normal use of the drug at the normal dose
Causality

• Probabilistic causality
  • \(P(\text{Effect} \mid \text{Cause}) > P(\text{Effect})\)

• Bradford-Hill criteria
  • **Strength**: how strong is the association between cause and effect;
  • **Consistency**: the relationship is presented in multiple places and the results are replicable;
  • **Specificity**: whether the cause leads to a particular effect or group of effects;
  • **Temporality**: the cause precedes the effects;
  • **Biological gradient**: does the level of the effect increase with an increase in the level of the cause?
  • **Plausibility**: is there some mechanism that could potentially connect cause and effect?
  • **Coherence, experiment, analogy.**
Our Work

1. Causality rule definition
2. Causality rule based information extraction
3. Causality inference
Step 1: Define causal rules

• Naranjo probability scale based rules
  • A well-designed questionnaire that reduces ADR judgment variability
  • The method assigns a weighted score to each component that must be considered in establishing causal association between drug D and reaction R

• Causal association rules
  • Discontinued D, R still existed;
  • Discontinued D, R improved;
  • Readministered D, R reappeared;
  • Increased the dose of D, R became more severe;
  • Decreased the dose of D, R became less severe;
  • Factors F1, F2 and F3 cause R.
Workflow

- Two types of data
  - Spontaneous Report System maintained by FDA, WHO and TGA
  - Medical Literature (PubMed)
Step 2: Information Extraction

- Extracting causal information from paper abstracts in PubMed according to these rules

1. Generate a set of **drug names** that refer to the given drug;
2. Generate a set of synonyms of the given **adverse drug reaction**;
3. Define a set of **trend terms** indicating the change of reactions:
   - {"stop", "improv", "increas", "decreas", "restor", "reduc"}
4. Define a set of terms indicating **temporal relationship**:
   - {"before", "after", "when", "during", "simultaneous", "begin", "end"}
5. Define a set of **causal terms**:
   - {"induc", "result", "associat", "indicat"}
6. Pre-process paper abstracts returned from PubMed: **sentence detection**, **stemming**;
7. Detect co-occurrence of drug name and ADR pairs with causal terms, temporal terms and trend terms.
Step 2: Information Extraction

- (warfarin, blue toe)
  - PMID 11885774
    “Skin discoloration and renal failure were improved after stopping warfarin potassium administration.”
  - PMID 19614996
    “The skin lesions improved after the warfarin dose was reduced.”

- (diclofenac, hepatitis)
  - PMID 10414481
    “Diclofenac was immediately stopped, leading to a complete restoration of liver functions over the course of the next few months.”
  - PMID 22234756
    “Certain drugs including oxyphenisatin, methyldopa, nitrofurantoin, diclofenac, interferon, infliximab, pemoline, minocycline, atorvastatin, and rosvastatin can induce hepatocellular injury that mimics autoimmune hepatitis.”
  - PMID 11322018
    “Naproxen and diclofenac were associated with a higher frequency of liver injuries, respectively 15.7 per cent and 11.5 per cent.”
Step 2: Information Extraction

• Example of inaccurate results (diclofenac, hepatitis)
  • PMID 10525281

  “Preincubation of human liver microsomes with dihydralazine in the presence of NADPH resulted in decreases in phenacetin O-deethylase activity (an indicator of P450 1A2 activity) and testosterone 6beta-hydroxylase activity (P450 3A4), but not in diclofenac 4'-hydroxylase activity (P450 2C9), an indication of inactivation of P450s 1A2 and 3A4 during the dihydralazine metabolism.”
### Step 3: Causality Inference

<table>
<thead>
<tr>
<th>Report</th>
<th>D discontinued</th>
<th>D readmin’ed</th>
<th>Dose change</th>
<th>Other factors</th>
<th>R change</th>
<th>D causes R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>None</td>
<td>Improved</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Unknown</td>
<td>No improvement</td>
<td>Unknown</td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>F2</td>
<td>No improvement</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>No</td>
<td>No</td>
<td>Decreased</td>
<td>F3</td>
<td>Improved</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>No</td>
<td>Decreased</td>
<td>Unknown</td>
<td>No improvement</td>
<td>Unknown</td>
</tr>
</tbody>
</table>
Step 3: Causality Inference
Conclusion

• Causality discovery is essential to detect potential adverse drug reactions.
  • Currently it is mainly a manual process.
• There is plenty room to partially automate the process.
  – Extracting information based on well-designed rules;
  – Assisting decision making based on a variety of information extracted.
CSIRO Drug Side-Effect Discovery Project

ADR Services
- Data mining
- Causality inference

Information extraction

Scalable data platform

FDA FAERS
WHO VigiBase
TGA SRS

Data acquisition / Data quality control (duplicate detection) / Data integration

Drug Forum
Social Media
Medical literature

Health records
Admin DB
Insurance claim DB

Scalable computing platform
CSIRO Drug Side-Effect Discovery Project

• Project Scope
  • Information extraction and monitoring for specific drugs
  • Data quality improvement for ADR systems
    – Duplicate detection for adverse drug event reports
  • Data integration and causality inference for ADR assessment

• Project Team
  • Dr Sarvnaz Karimi, Research Scientist and Project Leader
  • Dr Chen Wang, Senior Research Scientist
  • Dr Cecile Paris, Principal Research Scientist
  • Dr Alejandro Metke, Postdoctoral Researcher
  • Dr Raj Gaire, Research Engineer

• We are seeking collaboration opportunities.