Using Linked Data for Mining Drug-Drug Interactions in Electronic Health Records

Jyotishman Pathak, PhD
Richard C. Kiefer
Christopher G. Chute, MD, DrPH

MEDINFO 2013, Copenhagen, Denmark
Outline

• Discuss study
• Detail workflow
• Highlight results
• Analyze observations
• Take-aways
Drug-Drug Interactions (DDIs)

- It is well-known that adverse drug events are a major health risk, and DDIs are one of the causes of such events.
- While thousands of DDIs have been reported, only a handful is worth paying attention.
- Furthermore, a set of DDIs that suit one medical center or patient care facility might not be entirely appropriate to others.
- Consequently, there is significant research and on-going debate on how DDI information can be leveraged for better care.
Research and applications

High-priority drug–drug interactions for use in electronic health records

Shobha Phansalkar¹,²,³, Amrita A Desai², Douglas Bell⁴,⁵, Eileen Yoshida³, John Doole³, Melissa Czochanski³, Blackford Middleton¹,²,³, David W Bates¹,²,³

Research and applications

Adherence to drug–drug interaction alerts in risk patients: a trial of context-enhanced alerts

Jon D Duke¹,², Xiaochun Li³, Paul Dexter¹,²

Use of an algorithm for identifying hidden drug–drug interactions in adverse event reports

Kyna McCullough Gooden¹, Xianying Pan², Hugh Kawabata³, Jean-Marie Heim⁴
Background

• Use open-source tooling and standardized ontologies for creating virtual RDF graphs from Mayo’s clinical enterprise warehouse.

• Demonstrate federated querying using public data available from the Linked Open Data cloud.

• Selected Warfarin for our study as it is a commonly prescribed anti-coagulant medication with well known DDIs.

• Analyze any trends in DDI prescription practices based on age and gender.
Data Sources: Public and Private

• DrugBank (public) - contains 6711 drug entries where each entry comprising 150 data fields including drug-drug interactions.

• eMERGE (private) - comprises 6758 patients including medication prescription data.

• MCLSS (private) - patient demographics, diagnoses, hospital, laboratory, clinical notes and pathology data obtained from multiple clinical and hospital source systems within Mayo Clinic.

• ICD9CM (public) – contains text descriptions for diagnosis codes
Federated Querying Architecture

The Linked Clinical Data (LCD) Project

©2013 MFMER | slide-7
Creating endpoints – Public data

- **DrugBank**
  - Public endpoint did not have DDI information
  - Database was uploaded into MySQL and exported to RDF format using D2RQ
  - RDF file was imported into Virtuoso

- **ICD9CM**
  - Public endpoint exposed by NCBO BioPortal was used
  - Required API key within the SPARQL “SERVICE” statement
Creating endpoints – Private data

- Used RDF view mappings to interface with the tables in the MySQL database
- Mappings created using R2RML
- SNOMEDCT concept codes replaced column names from the tables to provide a common understanding of what the values represent
  - 396278008 - identification number
  - 422549004 - patient related identification code
  - 432213005 - date of diagnosis
  - 33633005 - medication prescription
  - 8319008 - principle diagnosis
Federated SPARQL Querying Process

Four endpoints were queried to provide our result set.

- **DrugBank**: DrugBank endpoint lists drug-drug interactions with Warfarin.
- **eMERGE**: eMERGE endpoint tells us which patients had been prescribed both drugs on the same visit (date).
- **MCLSS**: Demographics converts patient numbers to internal IDs.
- **BioPortal**: Diagnostics finds diagnosis code clusters based on internal IDs.

ICD9CM graph at the BioPortal endpoint converts diagnosis codes into text descriptions.

<table>
<thead>
<tr>
<th>ddi drug</th>
<th>diagnosis</th>
<th>frequency</th>
<th>description</th>
</tr>
</thead>
</table>

One SPARQL query results in a list of drugs which interact with Warfarin and determines the frequency of diagnoses for each DDI combination prescribed during the clinic visit.
The Linked Clinical Data (LCD) Project

Federated SPARQL Query

```
PREFIX drug: <http://bmidev4:8890/drugbank#>
PREFIX emerge: <http://hsrdev02:8890/emerse/>
PREFIX demo: <http://edison:8890/demo/>
PREFIX diag: <http://edison:8890/diag/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX snomed: <http://purl.bioontology.org/ontology/SNOMEDCT#>

SELECT ?ddiDrug1 ?diag ?label (COUNT(?diag) AS ?diagCount) 
WHERE { 
  SERVICE <http://bmidev4:8890/sparql> { 
    ?s1 drug:name "Warfarin" .
  } 
  SELECT distinct ?drug1 ?patientId2 
  WHERE { 
    SERVICE <http://hsrdev02:8890/sparql> { 
      ?s2 snomed:33633005 "Warfarin" .
      ?s2 snomed:432213005 ?date2 .
      ?s3 snomed:422549004 ?patientId3 .
      ?s3 snomed:432213005 ?date3 .
      filter(?patientId2=?patientId3)
      filter(?date2=?date3)
    } 
  } 
  SERVICE <http://edison.mayo.edu:8890/sparql> { 
    ?s5 snomed:396278008 ?internalId5 .
    ?s5 snomed:8319008 ?diag 
    filter(?internalId4=?internalId5) 
  } 
  SERVICE <http://sparql.bioontology.org/sparql?apikey=24e0350e-54e0- 
  GRAPH <http://bioportal.bioontology.org/ontologies/ICD9CM> { 
  } 
} 
GROUP BY ?drug1 ?diag 
ORDER BY ?drug1 DESC(?diagCount)
```

DDI’s

Patients

Diagnoses

Translated
Results – Gender distribution

The chart shows the gender distribution of patients for various drugs interacting with Warfarin. The x-axis represents different drugs, while the y-axis represents the total number of patients. The chart uses two bars for each drug: red for females and blue for males. The gender distribution varies significantly across the different drugs.
Observations - Gender

• Males were almost 3 times more likely to be prescribed Allopurinol
  • Commonly prescribed for hyperuricemia
  • Studies have associated it to increased risk for Stevens-Johnson Syndrome (SJS)

• Males were almost 4 times more likely to be prescribed Amiodarone
  • Commonly prescribed for arrhythmias
  • Studies have associated it to increased risk for cancer in males

• Females were 10 times more likely to be prescribed Fluconazole
  • Commonly prescribed for fungal infections
  • Studies have associated it to increased risk for rare birth defects for infants
Results – Age distribution

The graph shows the age distribution of patients taking various drugs that interact with Warfarin. The x-axis represents different drugs, and the y-axis shows the total number of patients. The age groups are color-coded as follows:

- Age: >=71 - Purple
- Age: 51–70 - Green
- Age: 31–50 - Red
- Age: 18–30 - Blue

The graph indicates that the majority of patients taking warfarin are aged 51–70 years old.
Observations - Age

- Prednisone was the most commonly prescribed medication among patients in the 18-30 year old grouping.
- Dicloxacillin was given to age groups 18-30 and 71+, but not to 31-50 or 51-70 year olds.
- Older patients were more likely to take Clopidogrel – as usually prescribed to patients receiving stents.
- Patients in the age group of 71 years or above were only given Phytonadione (Vitamin K).
## Results – Sample of top 5 diagnoses

<table>
<thead>
<tr>
<th>Drug Interacting with Warfarin</th>
<th>ICD-9-CM Diagnosis codes</th>
<th>Top 5 observed ICD-9-CM diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetaminophen</td>
<td>401.9</td>
<td>Unspecified essential hypertension</td>
</tr>
<tr>
<td></td>
<td>V58.61</td>
<td>Encounter for long-term (current) use of anticoagulants</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>Diabetes mellitus without mention of complication</td>
</tr>
<tr>
<td></td>
<td>272.4</td>
<td>Other and unspecified hyperlipidemia</td>
</tr>
<tr>
<td>Clopidogrel</td>
<td>414</td>
<td>Coronary atherosclerosis of nonautologous biological bypass graft</td>
</tr>
<tr>
<td></td>
<td>401.9</td>
<td>Unspecified essential hypertension</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>Diabetes mellitus without mention of complication</td>
</tr>
<tr>
<td></td>
<td>272.4</td>
<td>Other and unspecified hyperlipidemia</td>
</tr>
<tr>
<td></td>
<td>414.01</td>
<td>Coronary atherosclerosis of native coronary artery</td>
</tr>
<tr>
<td>Doxycycline</td>
<td>401.9</td>
<td>Unspecified essential hypertension</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>Diabetes mellitus without mention of complication</td>
</tr>
<tr>
<td></td>
<td>414</td>
<td>Coronary atherosclerosis of nonautologous biological bypass graft</td>
</tr>
<tr>
<td></td>
<td>272.4</td>
<td>Other and unspecified hyperlipidemia</td>
</tr>
<tr>
<td></td>
<td>714</td>
<td>Felty's syndrome</td>
</tr>
<tr>
<td>Gemfibrozil</td>
<td>250</td>
<td>Diabetes mellitus without mention of complication</td>
</tr>
<tr>
<td></td>
<td>V58.61</td>
<td>Encounter for long-term (current) use of anticoagulants</td>
</tr>
<tr>
<td></td>
<td>V42.0</td>
<td>Kidney replaced by transplant</td>
</tr>
<tr>
<td></td>
<td>401.9</td>
<td>Unspecified essential hypertension</td>
</tr>
<tr>
<td></td>
<td>272.4</td>
<td>Other and unspecified hyperlipidemia</td>
</tr>
</tbody>
</table>
Observations – Diagnosis frequency

• These findings further illustrate a higher prevalence of cardiovascular diseases and its risk factors in this cohort
  • A large number of patients diagnosed with Essential Hypertension, followed by Diabetes Mellitus and Hyperlipidemia
  • This would correlate with the fact most of the patients were prescribed cardiovascular drugs including statins and antiplatelets

• With this particular cohort having a mean age of 60.3 years, we also observe higher rates of prescriptions for antibiotics and antivirals
Acknowledging “clinical” reality

• DDIs relevant to Warfarin are often known to the clinical care providers, and are in fact prescribed in combinations, although with appropriate dosing considerations and daily blood tests.
  • 37 out of 38 patients prescribed Clopidogrel and Warfarin were 50 years old or above. This would make “clinical sense” since individuals in this age group frequently receive drug-eluting stents
  • Concomitant prescription for Ciprofloxacin was observed in 51 out of 55 patients that were above 50 years since it is an important antibiotic used widely to treat several respiratory, urinary tract, gastrointestinal and abdominal infections
So...what does this use case highlight?

- Web-scale queries “work”
- Ability to integrate “private” and “public” linked data
- Ability to “re-use” standardized ontologies and identifiers
- Querying and visualization tools are getting “better”
But…significant challenges still remain

• Performance and security issues
  • Still substantial penalties compared to relational
• Stale and unstable LOD endpoints
  • Quality of existing public SPARQL endpoints in the LOD cloud are a concern for “production” system
• Authoring SPARQL queries is still non-trivial
  • Graphical query builders are meant for expert users
• Schema mapping not always straightforward
  • Need for complex data structures; Clinical Archetypes
• SPARQL and handling of clinical narratives/text
Future

• Our immediate goal is to expand the DDI pair list that are of clinical significance and consider both drug prescription and administration data.

• Explore visual SPARQL editing tools, such as SPARQLMotion so clinicians may use the technology without having to learn SPARQL

• Use the Linked Data API for creating our service layer to provide application developers a friendlier access to the data, for example, using JSON
Thank You!

http://informatics.mayo.edu/LCD

Acknowledgment. This research is supported in part by the Mayo Clinic Early Career Development Award (FP00058504) and the eMERGE consortia (U01-HG-006379). The authors would also like to thank Jamie Kiefer for her inputs and feedback on the manuscript.
Discussion

• Our use case for federated querying of DDI information from DrugBank and the NCBO BioPortal SPARQL endpoints demonstrated the applicability of such a system and the benefits of interlinking multiple, heterogeneous Web data sources that are publicly available, with private (and institution-specific) patient information.

• We hypothesize that further development of such a system can immensely facilitate, and potentially accelerate scientific findings in clinical and translational research.