PeopleSave: Recommending Effective Drugs Through Web Crowdsourcing

Rahul Majethia, Varun Mishra, Akshit Singhal, Lakshmi Manasa K, Kunchay Sahiti and Vijay Nandwani
Shiv Nadar University, India
Dartmouth College, USA
University of Texas, Arlington

COMSNETS 2016, Bangalore
Summary

1 Introduction
   - About The Paper

2 Motivation
   - Unstructured but Abundant Information
   - PeopleSave: Review Analyses and Drug Recommendation

3 Issues and Challenges

4 Data Extraction

5 DrugElimination

6 Enhanced Recommendations

7 Conclusion
Introduction - About The Paper

- Describes PeopleSave - a drug recommendation and feedback system for doctors on the basis of contextual patient reviews crowd-sourced from the Internet.

- Filters information sources to check for crowdsourcing feasibility and then assess the drug’s effectiveness based on its reported detrimental effect on a patient.

- Recommendations are further refined by analyzing the sentiment behind the opinions of patients who have been administered these drugs in the past.

- Reiterates the need for a feedback system that can possibly go a long way in improving patient experience of a drug.
Motivation - I

- Case Studies and Drug History for Doctors from Drug Portals.
- Why: Availability of true, unbiased data-sets on public drug feedback and review forums.
Motivation - II

- Identified Diabetes as a case study for drug recommendation and therapy monitoring.
- Why: Abundance of Patients in the world. As of 2014, an estimated 387 million people have diabetes worldwide.
Issues and Challenges

- **Issues**
  
  (a) Ensure the reviews used to extract information are not false, self-motivated or unsuitable in any other manner.
  
  (b) Ensure non-credibility is not further aggravated by the presence of anonymity.

- **Solutions:**
  - Analyze the trend of sentiments towards a particular drug across portals.
  - Ensure that the reactions mentioned in patient reviews are sufficiently consistent.
Data Extraction and Analysis

- Reviews and ratings for Type 2 Diabetes drugs scraped from portals like webMD, drugs.com and Askapatient.
- Crowdsourced patient reviews and comments analyzed to obtain contextual sentiment polarity.
- Automatic sentiment analysis achieved by experimenting with SentiStrength and AlchemyAPI.

<table>
<thead>
<tr>
<th>Website</th>
<th>Number of Drugs Reviewed</th>
<th>Number of Reviews Extracted</th>
<th>Features of each Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>webmd.com</td>
<td>7</td>
<td>3476</td>
<td>Ratings, Comment, Age, Gender, Duration</td>
</tr>
<tr>
<td>drugs.com</td>
<td>6</td>
<td>463</td>
<td>Ratings, Comment, Duration</td>
</tr>
<tr>
<td>askapatient.com</td>
<td>6</td>
<td>1326</td>
<td>Ratings, Side Effects, Comments, Gender, Age, Duration, Dosage</td>
</tr>
</tbody>
</table>

TABLE I
Details of Reviews Collected from some of the Portals
## Sentiment Scores and Side Effect Frequencies

### Table III
**Comparison Between Average Sentiment Scores Across Portals**

<table>
<thead>
<tr>
<th>Review Portal</th>
<th>Actos</th>
<th>Byetta</th>
<th>Glucophage</th>
<th>Janumet</th>
<th>Januvia</th>
<th>Victoza</th>
</tr>
</thead>
<tbody>
<tr>
<td>webmd.com</td>
<td>-0.455</td>
<td>-0.391</td>
<td>-0.386</td>
<td>-0.35</td>
<td>-0.396</td>
<td>-0.180</td>
</tr>
<tr>
<td>drugs.com</td>
<td>-0.366</td>
<td>-0.387</td>
<td>-0.527</td>
<td>-0.320</td>
<td>-0.314</td>
<td>-0.394</td>
</tr>
<tr>
<td>askapatient.com</td>
<td>-0.437</td>
<td>-0.303</td>
<td>-0.340</td>
<td>NA</td>
<td>-0.412</td>
<td>-0.338</td>
</tr>
</tbody>
</table>

### Table IV
**Comparison of Normalized Frequencies of Side-Effects of Glucophage Across Portals**

<table>
<thead>
<tr>
<th>Review Portal</th>
<th>Muscle Pain</th>
<th>Gastrointestinal Problems</th>
<th>Weakness or Dizziness</th>
</tr>
</thead>
<tbody>
<tr>
<td>webmd.com</td>
<td>0.1202</td>
<td>0.80</td>
<td>0.796</td>
</tr>
<tr>
<td>drugs.com</td>
<td>0.1003</td>
<td>0.83</td>
<td>0.0697</td>
</tr>
<tr>
<td>askapatient.com</td>
<td>0.0911</td>
<td>0.854</td>
<td>0.0549</td>
</tr>
</tbody>
</table>
Drug Elimination and Recommendation

- Hazard Factor: The hazard factor of each particular side effect of the drug is a general level of adverse effect it may have on any patient as determined by a qualified domain expert.

- Recurrence Factor: It describes the frequency with which a particular side effect has been reported in the data obtained by crowdsourcing. It gives the statistical likelihood of a patient, who is being considered for recommending a drug too, experiencing the same side effect is he/she is administered that drug.

- Personalized Hazard Factor: Characterized by the threat posed by the specific side effect on the individual and is essentially dependent on a patient’s individual case history.
Threat Value

- Without access to a patient’s medical history records and a physician’s cognizance of his allergies or medical aversions, the TV of a distinct side-effect of a particular side-effect $i$, $TV_i$, is defined as:

$$TV_i = RF_i \times HF_i$$

- Given that patient history is available, the TV of a distinct side-effect $i$ for a patient $j$, $TV_{ij}$, is:

$$TV_{ij} = PHF_{ij} \times RF_i \times HF_i$$

- The total threat values is the summation of all the individual threat values.
Average TV across all portals

PeopleSave: Recommending Effective Drugs Through Web Crowdsourcing

Rahul Majethia, Shiv Nadar University
Range of TV of different drugs

- **Victoza**
- **Januvia**
- **Janumet**
- **Glucophage**
- **Byetta**
- **Actos**

**Legend:**
- Actos
- Byetta
- Glucophage
- Janumet
- Januvia
- Victoza

**PeopleSave: Recommending Effective Drugs Through Web Crowdsourcing**

Rahul Majethia, Shiv Nadar University
Enhanced Recommendations

PeopleSave: Recommending Effective Drugs Through Web Crowdsourcing

Rahul Majethia, Shiv Nadar University
The Last Word

(a) Implementing this system can work as a great tool for doctors to make use of a large number of similar past cases on the basis of which to treat a particular patient.

(b) Ensures that each patient is treated in the most case-specific and efficient manner possible, even while making allowance for the fact that each case is unique with respect to the others.
Future Work

- Aims to use continuous sensor streams from clinical machines and smart sensors to gauge the effectiveness rate and period of various similar class drugs across multiple patients.
- Tap into the reservoir of patient data to examine change in glucose levels from HbA1c tests, and other attributes from the patient’s pathological reports, through on-going associations with Hospitals.
Needs of the Hour

- **Non-invasive Smart Sensing**
  - (a) Gluco-wise, low-power radio waves. Kinsa, ADC device uses thermistor.
  - (b) Constant video-analytics and monitoring

- **Ontology and Dr-Eg Collaboration**
  - (a) Help, Data from Doctors - spend time with the creators, entrepreneurs and students!
  - (b) Careers, Internships for Engineers in Medical Institutions.
Questions?

”True intuitive expertise is learned from prolonged experience with good feedback on mistakes.
- Daniel Kahneman