An Analysis of RxNORM Clinical and Branded Drug Similarity

Maria Yang
Lister Hill National Center for Biomedical Communications
Outline

- What is RxNorm?
- Key Components of RxNorm Drugs
- What makes two drugs similar
- Gathering Data
- Transforming Raw Data to Usable Data
- Developing a Similarity Function
- Results
- So What?
What is RxNorm?

- RxNorm is a database containing information on all medications available on the US Market
- We use it to extract data on all clinical and branded drugs to conduct comparisons
  - Semantic Clinical Drug (SCD)
    - diphenhydramine hydrochloride 12.5 mg chewable tablet
  - Semantic Branded Drug (SBD)
    - Benadryl 12.5 mg chewable tablet
Key Components of RxNorm Drugs

- Ingredient
  - Some drugs can contain multiple ingredients
- Dose Form
- Strength
- Quantity
- Quality
  - Very new factor and is available for a few select drugs
- Brand Name vs. Generic
What makes two drugs similar?

- An Ingredient Match
- Topical vs. Systemic
- Close Strengths
- Whether or not the drug contains multiple ingredients
- Close quantity values
- The same quality
Gathering Raw Data/Implementation

- Wrote Java code using the RxNorm RESTful API to extract SCDs and SBDs, access attributes and properties of each drug
- Isolate each component
  - RxCUI #
  - Multiple ingredient
  - Strength (keeping units)
  - Dose form RxCUI/Dose Form
  - Quantity
  - Quality
  - Brand Name RxCUI/Brand Name
- Write to text file delimited by tabs
- Export to Excel to have filtering functionality
Transforming Raw Data to Usable Data

- Ingredient - no transformation, exact match
- Dose form - binarization and Boolean evaluation
  - The dose form could either be topical or systemic
- Strength - quantitative measure
  - Ensure that units are the same
- Multiple ingredient - binarization and Boolean evaluation
  - Creation of Boolean value that says whether there are multiple ingredients or not
- Quantity - quantitative measure
- Quality - no transformation, exact match
Calculating the Components based on Exact Match

- If the Strings are exact matches, we give it the full weight
  - $ingredientScore = 1$
  - $qualityScore = 1$
- Otherwise, it doesn’t get any of the weight
  - $ingredientScore = 0$
  - $qualityScore = 0$
Calculating the Components Based on Binarization and Boolean Evaluation

- To get binary dose form types, a hashmap was created with each Dose Form RxCUI associated with whether it was systemic or topical.
- If the binary dose form types are the same, we give it the full weight:
  - \( doseFormScore = 1 \)
- Otherwise, it doesn’t get any of the weight:
  - \( doseFormScore = 0 \)
Calculating the Components Based on Binarization and Boolean Evaluation

- The multiple ingredient Boolean is determined on if more than one ingredient were present or not. The exact number didn’t matter.
- If the drugs were either both multiple ingredients or both single ingredients, we give it the full weight
  - $\text{multipleIngredientScore} = 1$
- Otherwise, it doesn’t get any
  - $\text{multipleIngredientScore} = 0$
Calculating the Components Based on Binarization and Boolean Evaluation

- The binarization is whether the drug is branded or generic
- If the drugs were either both branded or both generic, we give it the full weight
  - $brandScore = 1$
- Otherwise, it doesn’t get any
  - $brandScore = 0$
Calculating the Components Quantitatively (Strength, Quantity)

- We want a quantitative measure instead of an all-or-nothing method.
- Compare the units on each drug’s strength to determine if there’s a match. If not, we set the strength factor to 0.
  - \( strengthScore = 0 \)
- Otherwise, we want a fraction to measure how close two numbers are. If they’re the same we want the function to return 1. The closer the numbers are, the closer to 1 the factor should be.

\[
strengthScore = \frac{MIN(s_1, s_2)}{MAX(s_1, s_2)}
\]

\[
quantityScore = \frac{MIN(q_1, q_2)}{MAX(q_1, q_2)}
\]
Developing a Similarity Function

- Determining weights for each component
  - Different weights for various use cases
  - Built separate weight file for user input
  - The weights that we went with:
    - Ingredient (50%)
    - Binary Dose Form (15%)
    - Strength (10%)
    - Multiple ingredient boolean (10%)
    - Quantity (5%)
    - Quality (5%)
    - Brand (5%)
Getting the Total Similarity Score

- Add up all the component scores that were multiplied by their weights to get the final similarity score

\[ score = 100 \times (\text{ingredientScore} \times \text{ingredientWeight} + \text{doseFormScore} \times \text{doseFormWeight} + \text{strengthScore} \times \text{strengthWeight} + \text{multipleIngredientScore} \times \text{multipleIngredientWeight} + \text{quantityScore} \times \text{quantityWeight} + \text{qualityScore} \times \text{qualityWeight} + \text{brandScore} \times \text{brandWeight}) \]
Dealing with Multiple Ingredients

- Decided to run the similarity test for each individual ingredient and take the highest match from all the comparisons.

<table>
<thead>
<tr>
<th>RxCU</th>
<th>MultipleIngredient</th>
<th>Ingredient</th>
<th>Strength</th>
<th>DoseForm</th>
<th>DoseFormBlind</th>
<th>Quantity</th>
<th>Quality</th>
<th>BrandName</th>
</tr>
</thead>
<tbody>
<tr>
<td>148531</td>
<td>TRUE</td>
<td>82003</td>
<td>Docusate</td>
<td>55 MG</td>
<td>Oral Capsule</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>148531</td>
<td>TRUE</td>
<td>47628</td>
<td>Tricalcium phosphate</td>
<td>160 MG</td>
<td>Oral Capsule</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>148531</td>
<td>TRUE</td>
<td>1151</td>
<td>Ascorbic Acid</td>
<td>25 MG</td>
<td>Oral Capsule</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>148531</td>
<td>TRUE</td>
<td>237099</td>
<td>alpha Tocopherol</td>
<td>30 UNT</td>
<td>Oral Capsule</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>148531</td>
<td>TRUE</td>
<td>684879</td>
<td>pyridoxine</td>
<td>25 MG</td>
<td>Oral Capsule</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>148531</td>
<td>TRUE</td>
<td>24941</td>
<td>Ferrous fumarate</td>
<td>30 MG</td>
<td>Oral Capsule</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>148531</td>
<td>TRUE</td>
<td>1006469</td>
<td>doconexnet</td>
<td>265 MG</td>
<td>Oral Capsule</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>148531</td>
<td>TRUE</td>
<td>4511</td>
<td>Folic Acid</td>
<td>1.2 MG</td>
<td>Oral Capsule</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>1151</td>
<td>Ascorbic Acid</td>
<td>50 MG</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>2418</td>
<td>Cholecalciferol</td>
<td>400 UNT</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>11246</td>
<td>Vitamin B 12</td>
<td>0.002 MG</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>4511</td>
<td>Folic Acid</td>
<td>0.8 MG</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>11246</td>
<td>Vitamin A</td>
<td>4000 UNT</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>7405</td>
<td>niacinamide</td>
<td>10 MG</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>1925</td>
<td>Calcium Sulfate</td>
<td>250 MG</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>21579</td>
<td>Copper Sulfate</td>
<td>0.15 MG</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>24941</td>
<td>Ferrous fumarate</td>
<td>27 MG</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>285155</td>
<td>Kelp preparation</td>
<td>0.01 MG</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>6582</td>
<td>Magnesium Oxide</td>
<td>0.15 MG</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>6585</td>
<td>Magnesium Sulfate</td>
<td>0.05 MG</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>34323</td>
<td>potassium sulfate</td>
<td>0.34 MG</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>684879</td>
<td>pyridoxine</td>
<td>1 MG</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>9346</td>
<td>Riboflavin</td>
<td>2 MG</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>802792</td>
<td>TRUE</td>
<td>39554</td>
<td>Zinc Sulfate</td>
<td>0.85 MG</td>
<td>Oral Tablet</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Results

Examples:

<table>
<thead>
<tr>
<th>Drug 1</th>
<th>Drug2</th>
<th>Similarity Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05 ML aflibercept 40 MG/ML Injection</td>
<td>0.05 ML ranibizumab 10 MG/ML Injection</td>
<td>42.5%</td>
</tr>
<tr>
<td>0.2 ML Bemiparin sodium 12500 UNT/ML Prefilled Syringe</td>
<td>0.2 ML Bemiparin sodium 17500 UNT/ML Prefilled Syringe</td>
<td>90%</td>
</tr>
<tr>
<td>0.4 ML lufenuron 100 MG/ML Prefilled Syringe</td>
<td>0.4 Methotrexate 18.8 MG/ML Auto-Injector</td>
<td>40%</td>
</tr>
</tbody>
</table>
So What?

- We were able to develop a proof-of-concept for getting similarity scores for drugs
Limitations

- We used arbitrary weights for each component
- Similarity function is application-dependent
  - Each use case may need different sets of weights
  - In the Java code, the weights are in a property file, not in the code
Acknowledgements

- Dr. Olivier Bodenreider
- Ms. Keyla Cooper
- ORISE