FDA BI Research Weekly Report

Yage Wang, Zheng Gong 20170823

1. **Integrated BI text mining method report**

**Step 1: Extract data from data source**

The current dataset comes from MAUDE database and all records whose product code equals to “FWM” or “FTR” are selected out from table “Foidev” as the initial dataset.

Then we go through table “Foitext” and “Mdrfoi” to select all records who matches the initial dataset on [MDR\_REPORT\_KEY] column, and add these records to our dataset.

We now have 3 tables in our dataset, each having around 27,000 records using [MDR\_REPORT\_KEY] column as the primary identifier.

Currently, we focus on following columns in each table and we will use [Column Name] to stand for the column in following content.

|  |  |
| --- | --- |
| Table | Column |
| Mdrfoi | - |
| Foidev | MANUFACTURER\_D\_NAME, BRAND\_NAME, GENERIC\_NAME |
| Foitext | FOI\_TEXT |

Note: all columns above is text type. Our method will mainly focus on natural language processing techniques.

**Step 2: Data preprocessing and cleaning**

1. Convert all text to lower/upper case

All contents in MAUDE database are already upper case, so we skip this sub-step.

1. Remove stop words

Sometimes, some extremely common words which would appear to be of little value in helping select documents matching a user need are excluded from the vocabulary entirely. These words are called stop words.

1. Remove punctuations