FDA BI Research Weekly Report

Yage Wang, Zheng Gong 20170816

1. **What is the text mining method?**

Method: Turn keywords to vectors

Data source: foidev table and foitext table for all BI records

Keywords list: (We add some new words to the list which are marked with shading)

manufacture = ['MENTOR', 'ALLERGAN', 'IDEAL', 'SIENTRA', 'MCGHAN', 'INAMED', 'SILIMED', 'NAGOR']

fill\_type = ['SALINE', 'SILICONE', 'GEL', 'COHESIVE']

surface\_type = ['SMOOTH', 'TEXTURED', 'BIOCELL']

implantation\_indication = ['AUGMENTATION', 'RECONSTRUCTION', 'COSMETIC']

ALCL = ['ALCL', 'ANAPLASTIC LARGE CELL LYMPHOMA', 'LYMPHOMA', 'T-CELL LYMPHOMA', 'B-CELL LYMPHOMA']

side = ['LEFT', 'RIGHT']

symptom = ['breast pain', 'breast swelling', 'breast cyst', 'breast calcification', 'capsular contracture', 'lymph node enlargement', 'firmness of breast', 'hematoma', 'mass', 'lump', 'rupture', 'deflated’, ‘infection', 'abscess', 'leukopenia', 'nodules', 'skin discoloration', 'skin lesion', 'seroma', 'effusion', 'fluid']

Fuzzy Match Rate: 95%

We go through all records to find the word that matches keywords list, and label each record by these keywords.

E.g. Original records

|  |  |
| --- | --- |
| MDR\_REPORT\_KEY | FOI\_TEXT |
| 24568502 | WE HAVE A CONFIRMED CASE OF BI-ALCL IN A PT THAT RECEIVED BREAST RECONSTRUCTION AFTER SUBCUTANEOUS MASTECTOMY 2011 VIA SILICONE GEL IMPLANTS. IN THE CYTOLOGY OF PERIPROSTHETIC FLUID IN THE IMPLANT CAPSULE WE COULD IDENTIFY CD 30. |

After word-to-vector transformation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| MDR\_REPORT\_KEY | Fill\_type | | implantation\_indication | symptom | ALCL |
| SILICONE | GEL | RECONSTRUCTION | FLUID | ALCL |
| 24568502 | 1 | 1 | 1 | 1 | 1 |

For other keywords we don’t find, mark it as 0 in the record.

Pros:

1. Computational efficiency
2. Text content is still visible after transformation
3. Easy for developing models (statistical, regression, deep learning)

Cons:

1. Difficult to analyze complicated expressions for negation and comparison

e.g. The case is suspected ALCL but it’s confirmed not after further test.

e.g. The left implant device was removed but the other remained.

1. **What columns to look at?**

|  |  |  |
| --- | --- | --- |
| Keywords | Foidev | Foitext |
| Manufacture | MANUFACTURER\_D\_NAME | - |
| fill\_type | BRAND\_NAME, GENERIC\_NAME | FOI\_TEXT |
| surface\_type | BRAND\_NAME, GENERIC\_NAME | FOI\_TEXT |
| ALCL | - | FOI\_TEXT |
| side | - | FOI\_TEXT |
| symptom | - | FOI\_TEXT |

1. **Word frequency analysis?**

See attachment [1\_Keywords\_Word\_Frequency.xlsx]