NLP Application for Life Insurance: Are the Technologies Ready?

Dihui Lai, PhD, ASA
Lead Data Scientist
Sept 2020
Contents

• Overview of Nature Language Process
• Underwriting Process and AI
• Digital Health Data and NLP
• CI Product and Definition Comparison
• Litigation in Claim Operations
• Key Considerations of AI Applications’ Plausibility
Overview of Nature Language Processing

News:

• 2015: Google went one better with FaceNet. On the widely used Labeled Faces in the Wild (LFW) dataset, FaceNet achieved a new record accuracy of 99.63%

• 2018: A tool known as BERT can now beat humans on advanced reading-comprehension tests.

• The NLP market is predicted be almost 14 times larger in 2025 than it was in 2017, increasing from around three billion U.S. dollars in 2017 to over 43 billion in 2025

Reality check - the state of the art NLP algorithms

• Named entity recognition: 93.5%

• Part-of-speech tagging: 97.96%

• The Standing Question Answering (SQuAD): 93.01%
Dissecting the NLP/AI Capability

<table>
<thead>
<tr>
<th>Level</th>
<th>NLP Technology Specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document</td>
<td>Sentiment Analysis; Document Classification;</td>
</tr>
<tr>
<td>Sentence</td>
<td>Relation Extraction; Events Extraction;</td>
</tr>
<tr>
<td>Entities</td>
<td>Named Entity Recognition</td>
</tr>
<tr>
<td>Word</td>
<td>Word2vec; Word count, Similarity Measure</td>
</tr>
</tbody>
</table>

**BMI:** 25-29

Fatigue, **Body mass index 25-29** - overweight, Adult health examination, Family history of malignant melanoma

Diagnoses: LEFT INGROWN TOENAIL [L60.0]


<table>
<thead>
<tr>
<th>Medical Entity I</th>
<th>Medical Entity II</th>
<th>Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other specified effects of reduced temperature, initial encounter</td>
<td>Chilblain lupus erythematosus</td>
<td>0.05</td>
</tr>
<tr>
<td>Discoid lupus erythematosus</td>
<td>Chilblain lupus erythematosus</td>
<td>0.93</td>
</tr>
</tbody>
</table>

**Medical Entity I:**

- Other specified effects of reduced temperature, initial encounter
- Discoid lupus erythematosus
- Chilblain lupus erythematosus

**Medical Entity II:**

- Chilblain lupus erythematosus

**Similarity:**

- 0.05
- 0.93
AI Augmented Underwriting Process

Stage 1 – Data Conversion, Extraction, & Linkage

- AI Function
  - Info Extraction
  - Data Linkage
  - Unstructured → Structured

- ETL
- OCR
- NLP

Aggregate data from multiple sources in multiple formats; convert into structured, linked, & interpretable data

Stage 2 – Risk Assessment

- STP/Rule
- Intelligent Summary
- Accelerated UW
- Quick Decline
- Fully UW
- UW decision

AI Function
- Intelligent Decision
- Recommendation
- Medical Interpret
- Warning/Red Flag

Case triage

Feedback

Additional requirement
### AI Augmented Underwriting Process: Summary

**Underwriting Profile Summary**

20-year-old female with serve medical condition – “Ventricular septal defect”

<table>
<thead>
<tr>
<th>Basic Information</th>
<th>Insured Name: [REDACTED]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Date of Birth: [REDACTED]</td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>U.S.</td>
</tr>
<tr>
<td>Citizenship</td>
<td>U.S.</td>
</tr>
<tr>
<td>Tobacco</td>
<td>No</td>
</tr>
<tr>
<td>Occupation</td>
<td>Student</td>
</tr>
</tbody>
</table>

**Lab Result**

- Height(6ft): 5'11.0
- Weight(lbs): 187.0
- BMI: 26.1 - Overweight
- Blood Pressure: 125/80, 122/80, 122/80 - Normal
- Abnormal Condition: XXXX

**Rx Score:** 4

- Selected Prescription Drugs:
  - Dapagliflozin Propanediol Tab 5 MG (Base Equivalent) – High
  - Glimepiride Tab 4 MG – High
  - Ondansetron HCI Tab 4 MG – Medium
  - Fluconazole Tab 200 MG – Medium
  - Levothyroxine Sodium Tab 100 MG – Low
  - Amoxicillin & K Clavulanate Tab 875 – 125 MG – Low
  - Simethicone Chew Tab 80 MG – Low

**MVR Record:** 7

- Violation Code: 52110 - serious Offences
- Violation Date: 08/29/2014
- Points: 2

**APS Information:** 8

- Highest Severity Score: 7
- ICD Code with Highest Severity Score:
  - G21.0 - Ventricular septal defect
  - 745.4 - Ventricular septal defect

- Extract demographic information from multiple sources (PDF, xml etc.)
- Provide an overview of the applicants' underwriting profile
- Assess the risks for multiple cohorts: lab, prescription drug, APS, driving records etc.
- Provide severity score if possible
- Linkages to the source information e.g. BMI
AI Augmented Underwriting Process: Medical History

Data Visualization

Medical History

- Extract medical information with a time stamp
- Provide medical history at a quick glance
- Color coded severity medical condition
- Allow quick linkage to the source data
AI Augmented Underwriting Process: Navigation and Decision Making

- Bookmark generated by document classification algorithm
- Allows easy navigation across hundreds of application pages.

Decision Making

- Machine learning based underwriting decision model
- Improving overtime with accumulated data
Digital Health Data (DHD) and NLP

- DHD contains both structured/unstructured data
- Unstructured data contains comprehensive information about individual health
- Read through all unstructured data is time consuming
- Unstructured data could contain information that is not covered in the structured data
- In one study, we use NLP and are able to identify ~15% of the cases that contain more severe medical conditions from unstructured text.
Product Definition Comparison

When it comes to CI product, companies share a lot of common features however, every company could have varied definitions on detailed exclusions, time span, surgery procedures etc. The review process of comparing CI definitions could be challenging and tedious. Could NLP offer some help?

Surgery to Aorta - Definition 1

The actual undergoing of laparotomy or thoracotomy to repair or correct an aneurysm, narrowing, obstruction or dissection of the aorta through surgical opening of the chest or abdomen.

For the purpose of this definition, aorta shall mean the thoracic and abdominal aorta but not its branches. Surgery performed using only minimally invasive or intra arterial techniques are excluded. Traumatic injury of the aorta is excluded. The surgery must be confirmed as necessary by an Cardiothoracic surgeon.

Surgery to Aorta - Definition 2

Surgery to Aorta is defined as the actual undergoing of a laparotomy or thoracotomy to repair or correct an aneurysm, narrowing, obstruction or dissection of the aorta. Aorta means the thoracic and abdominal aorta but not its branches.

For the above definition, the following are not covered: Intra arterial surgeries such as percutaneous endovascular aneurysm repair; Any peripheral vascular surgery including but not limited to aortofemoral bypass surgery.

Similarity: 0.167
Claim Notes Classification

Sample Claim Notes: 2012.5.6投保822，2016.12.16投保1183、1185，本次于2017.10.27因脑白质病变申请理赔，本次疾病不符合条款约定的重疾，无对应重疾责任，已协谈，本次拒付处理。

Claim Note -> Benefit Classifier

- ESCI
- LSCI
- DEA

Use machine learning based model to categorize the claim notes into appropriate type.
Claim Operations: Litigation Detections

• The claim operation team needs to make sure they pay sufficient attentions to the claim cases that contain litigation/law suits.
• It is not very efficient to review each page of a claim document manually to identify litigation.
• We use document classification to automatically detect pages that could contain litigation files.
• This piece of AI serves as an augmented method to increase the efficiency of the manual review process. With the potential of automate the manual process fully.
The Importance of Data Quantity & Quality

Data Quality

• Scanned Document
• General Language
• Internet Language
• Professional Language
• Handwriting ×

Data Quantity
Learnings on Building Useful AI Application

- Is the technology ready? Yes and no.
- Accuracy v.s. error tolerance (e.g. is 90% recall rate enough?)
- Human input is critical in the equation: from both user and development perspectives.
- Quality data with labeling
- AI system is going to improve overtime, significantly, e.g. face recognition, Google
- IT system compatibility
Questions?