

Interviewleitfaden

Ingrid Tchilibou

10. Februar 2020

Inhaltsverzeichnis

1	Chek-In	3
2	Einleitung	3
3	Evaluation	3
4	Aufgabe	6
5	Interview	6

1 Chek-In

Bevor wir beginnen, lesen und unterschreiben Sie bitte das bereitgestellte Einverständniserklärung.

2 Einleitung

An diesem Tag wollen wir unsere Jupyter Notebook mit verschiedenen Ergebnissen, die wir im Rahmen von Projekt "Konzept Annotation basierend auf WordNet" erzielt haben, auswerten. Das Ziel ist es, den Forschern zu helfen, eine Menge von Ideen zu verstehen, indem sie diese in Kategorien gruppieren. Genauer dann ein Überblick über einer Menge an Ideen zu bekommen. Eine Kategorie ist in unserem Fall ein Wort (synset), das mehrere andere Wörter (synsets) gruppiert, genauer gesagt ein Hypernyme.

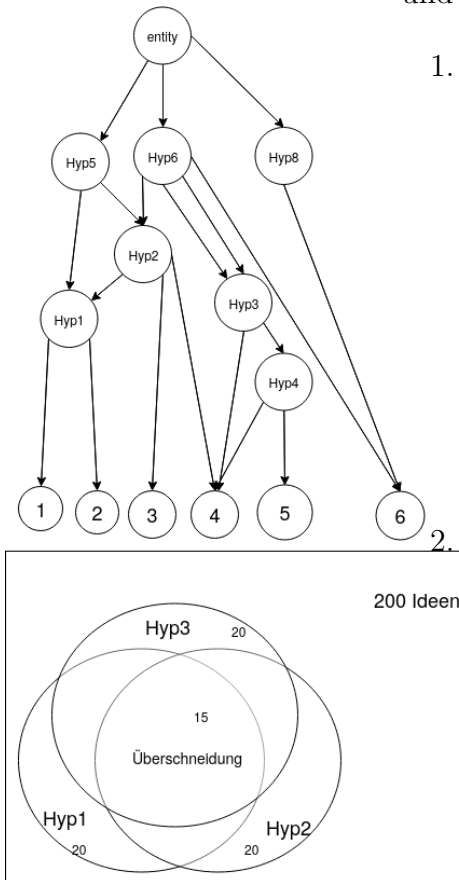
Im Laufe dieses Projekts erhielten wir 200 der 581 Ideen aus dem Bionic Radar Project. Diese Ideen wurden erstmal mit Hilfe des Wordnet-API-Backends von ICV annotiert. Für jede Idee, die wir erhielten, wählten wir das/die Synset(s) aus, das/die die einzelnen Wörter der zu annotierenden Idee am besten beschreibt/beschreiben. Am Ende erhielten wir eine JSON-Datei mit all unseren Entscheidungen. Dann haben wir aus diesem Datensatz die Hypernyme jedes ausgewählten Synset extrahiert und in verschiedenen Wörterbücher gespeichert.

Bevor wir fortfahren, haben Sie Fragen?

3 Evaluation

Wir werden Ihnen zunächst die verschiedenen Punkte erklären, die wir bewerten wollen.

- Zuerst wollen wir mit dieser Evaluation, die Themen(Hypernymen) finden, die die maximale und minimale Ideen beschreibt.



1. **Ansatz 1:** Nehmen wir zum Beispiel sechs Ideen, jede Idee enthält mehrere Hypernymen und als top hypernym ist *entity*. Das Ziel ist es, den Hypernym zu finden, der möglichst viele Ideen enthält. Auf unserer Grafik enthalten Hypernym 6 und Hypernym 5 jeweils 4 Ideen. Dies sind für uns also die besten Themen, die diese 6 Ideen am besten beschreiben. Andererseits enthält Hypernym 8 nur eine Idee (Idee 6), die als unerwartetes Thema betrachtet wird.

2. **Ansatz 2:** Nehmen wir zum Beispiel an, dass wir 3 Hypernymen mit jeweils 20 Ideen (Fenstergröße) haben. Analysieren wir die Schnittmenge dieser 3 Hypernymen, um herauszufinden, wie viele Ideen die Hypernymen 1, 2 und 3 gemeinsam haben.

Frage: Wie groß muss das Fenster sein damit man eine " gute " Liste an Hypernymen bekommt?

Was heißt " gute " ?

Überschneidung > 95%

- Weiterhin haben wir jede Idee auf ein einziges Synset reduzieren, so dass wir die Anzahl der Hypernymen reduzieren und sie besser analysieren können. Dazu haben wir für jede Idee das beste Synset gefunden, das diese Idee am besten beschreibt.

Was ist das beste Synset?

Das Synset, das genau einmal in unseren Daten erscheint. Nachdem wir die Synsets, die mehr als einmal in unseren Ideen erscheinen, eliminiert haben. Wir stellten fest, dass es Ideen gab, die völlig verschwunden waren, weil sie nur bereits verwendete Synsets enthielten. Für diejeni-

gen, die wir sie getrennt untersucht haben, haben wir die Anzahl der Wiederholungen jedes Mal erhöht, um die beste Synset zu finden. Für die Ideen von mehrfach vorkommenden Synsets, die nur einmal in unseren Daten vorkommen, haben wir nur das erste Synset genommen.

Frage : Wird es dadurch leichter, die besten Themen zu bekommen?

- Wir wollen unsere Daten auch nutzen, um den Abstraktionsgrad einer Idee zu bewerten.
Aber was ist der Abstraktionsgrad einer Idee?
Nehmen wir 3 Ideen als Beispiel.

Idee 1 hat zwei Synsets: 4 ist der Abstand von Synset 1 zur Entitätswurzel und 5 ist der Abstand von Synset 2 zur Entitätswurzel. Der Abstraktionsgrad der Idee 1 ist gleich $(4+5)/2 = 4,5$

Idee 2 hat drei Synsets: 4 ist der Abstand von Synset 1 zur Entitätswurzel, 5 ist der Abstand von Synset 2 zur Entitätswurzel und 3 ist der Abstand von Synset 3 zur Entitätswurzel. Der Abstraktionsgrad der Idee 2 ist gleich $(4+5+3)/2 = 6$

Idee 3 hat zwei Synsets: 8 die Entfernung von Synset 1 zur Entitätswurzel und 9 die Entfernung von Synset 2 zur Entitätswurzel. Der Abstraktionsgrad der Idee 1 ist gleich $(8+9)/2 = 8,5$

Die abstrakteste Idee ist diejenige mit dem höchstens Abstraktionsgrad und umgekehrt. Bzw. In unserem Beispiel ist Idee 1 die konkreteste und Idee 3 die abstrakteste.

Bevor wir fortfahren, haben Sie Fragen?

Wir haben Ihnen einen Jupyter-Notebook mit verschiedenen Wörterbüchern zur Verfügung gestellt, die Informationen enthalten, die während unserer Analyse der Daten gesammelt wurden. Wir haben in diesem Jupyter-Notebook

jedes Wörterbuch und seine Funktionalität sowie die verschiedenen Funktionen erklärt, die wir Ihnen bei Ihrer Analyse zur Verfügung gestellt haben.

Um fortzufahren, haben Sie 10 Minuten Zeit, das Jupyter-Notebook zu lesen, um zu verstehen, wie diese verschiedenen Punkte umgesetzt wurden und welche verschiedenen Funktionen Ihnen zur Beantwortung der verschiedenen Fragen der Evaluation zur Verfügung stehen.

Bevor wir fortfahren, haben Sie Fragen?

4 Aufgabe

Sie haben nun 20 Minuten Zeit, um die verschiedenen gestellten Fragen zu beantworten.

1. Welche Themenbereiche gibt es in dem Datensatz?
2. Finde unerwartete Ideen oder Outlier (5 Stück)
3. Was sind die abstraktesten Ideen in dem Datensatz? Was sind die konkretesten Ideen in dem Datensatz? (5 Stück)

Bevor wir fortfahren, haben Sie Fragen?

Sie haben nun 10 Minuten Zeit, um ein wenig mit dem Jupyter-Notebook zu spielen, um weitere Informationen zu finden, die wir mit unseren Daten noch erhalten könnten.

Frage: Welche wichtigen Informationen über die Daten können wir mit Hilfe der bereitgestellten Wörterbücher noch erhalten?

5 Interview

Nun werden wir mit einem 30-minütigen Interview fortfahren, um Ihre verschiedenen Eindrücke von dem, was wir gerade getan haben, zu bewerten.

1. Wie konntest du die Aufgaben bearbeiten? Hattest du Probleme bei einer der Aufgaben?

2. Wie schätzt du den Ansatz mit Wordnet ein? Was hat dir gut gefallen, was hat dir nicht gefallen? Hat dir der Ansatz geholfen einen Überblick über die Ideen zu bekommen?
3. Was könnten andere Data-Scientists mit diesem Jupyter notebokk machen?
4. Was könnten die Forscher damit anfangen?
5. Was für Aufgaben hast du mit den Bionic Radar Daten bisher gemacht?