Legal Semantic Modeling

Understanding the Structure of Sentences in BVA PTSD Claims Cases

Harout Boujakjian, Clara Buss, Anh Hoang, Elizabeth Kenis, Tim Rogers

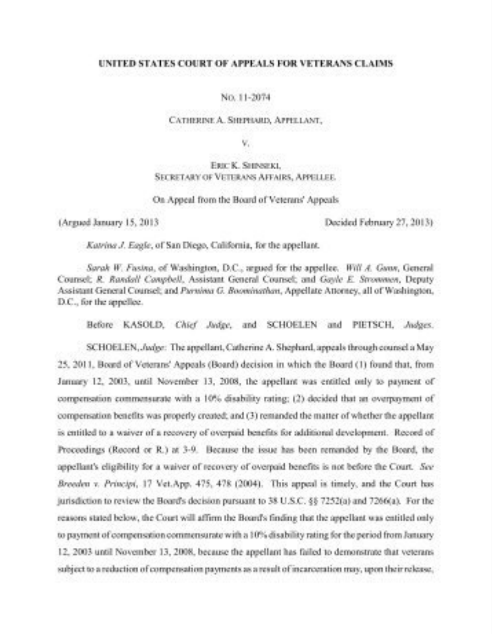
June 20, 2020

George Mason University

Background

- Veterans submit tons of appeals each year to BVA (Board of Veterans' Appeals) regarding PTSD diagnosis.
- BVA must sift through all of these cases.
- Automating argument mining can be very helpful!
- Do legal documents have any sort of structure?

BVA sample raw case



UNITED STATES COURT OF APPEALS FOR VETERANS CLAIMS

No. 11-2074

CATHERINE A. SHEPHARD, APPELLANT,

V.

ERIC K. SHINSEKI,
SECRETARY OF VETERANS AFFAIRS, APPELLEE.

On Appeal from the Board of Veterans' Appeals

(Argued January 15, 2013 Decided February 27, 2013)

Katrina J. Eagle, of San Diego, California, for the appellant.

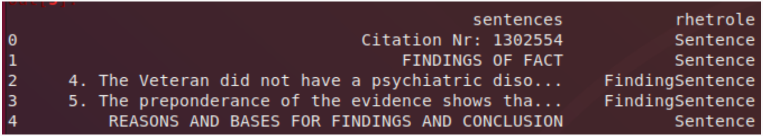
Sarah W. Fusina, of Washington, D.C., argued for the appellee. *Will A. Gunn*, General Counsel; *R. Randall Campbell*, Assistant General Counsel; and *Gayle E. Strommen*, Deputy Assistant General Counsel; and *Purnima G. Boominathan*, Appellate Attorney, all of Washington, D.C., for the appellee.

Before KASOLD, *Chief Judge*, and SCHOELEN and PIETSCH, *Judges*.

SCHOELEN, *Judge*: The appellant, Catherine A. Shephard, appeals through counsel a May 25, 2011, Board of Veterans' Appeals (Board) decision in which the Board (1) found that, from January 12, 2003, until November 13, 2008, the appellant was entitled only to payment of compensation commensurate with a 10% disability rating; (2) decided that an overpayment of compensation benefits was properly created; and (3) remanded the matter of whether the appellant is entitled to a waiver of a recovery of overpaid benefits for additional development. Record of Proceedings (Record or R.) at 3-9. Because the issue has been remanded by the Board, the appellant's eligibility for a waiver of recovery of overpaid benefits is not before the Court. *See Breeden v. Principi*, 17 Vet.App. 475, 478 (2004). This appeal is timely, and the Court has jurisdiction to review the Board's decision pursuant to 38 U.S.C. §§ 7252(a) and 7266(a). For the reasons stated below, the Court will affirm the Board's finding that the appellant was entitled only to payment of compensation commensurate with a 10% disability rating for the period from January 12, 2003 until November 13, 2008, because the appellant has failed to demonstrate that veterans subject to a reduction of compensation payments as a result of incarceration may, upon their release,

Data set

- Data set contains 20 cases (2,526 sentences in total)[1]
- Before preprocessing, the data set only has two columns (sentence, rhetorical role)
- Each sentence has labeled rhetorical role by LLT lab
- Not your typical "Big Data" NLP problem! Falls under multiclass classification.



```
                                          sentences           rhetrole
0                             Citation Nr: 1302554           Sentence
1                                 FINDINGS OF FACT           Sentence
2     4. The Veteran did not have a psychiatric diso...    FindingSentence
3     5. The preponderance of the evidence shows tha...    FindingSentence
4        REASONS AND BASES FOR FINDINGS AND CONCLUSION           Sentence
```

[1] Full data set: https://github.com/vernrwalker/VetClaims-JSON

Rhetorical Roles

Examples

- **Finding of Fact**: The Veteran is not service-connected for any disability.
- **Reasoning**: He is a lay person, as there is no indication that he possesses medical knowledge, training, or experience.
- **Evidence**: Diagnoses of schizoaffective disorder were made in VA treatment records and at the February 2008 VA mental disorders examination.
- **Legal Rule**: A current disability exists when there is a disability when a claim for it is filed or at any time during the pendency of such claim.
- **Citation**: McClain v. Nicholson, 21 Vet. App. 319 (2007).

Problem

- Ability to predict rhetorical role for sentences
- This is important because most legal documents follow the same structure.

Preprocessing

- Remove stop words and convert all words to lower case
- In order to create predictors, the sentences must be tokenized into individual words.
- The tokens will then be transformed using Bag of Words or TD-IDF
- This will create the dataframe (or dataframe) of numerical values to run the models on

Modeling and Metrics

- Potential models: regularized logistic regression, neural network, random forest
- Confusion matrix and classification report (precision, recall, F1-score)