Visualizing Legal Taxonomy with Class Prevalence and Semantic Role Labeling

Hee Hwang
University of Massachusetts, Amherst
Amherst, MA 01003
hhwang@cs.umass.edu

Abstract

People contact the law in various ways, whether concerning finances, crime, employment, housing, or immigration. Some of these encounters are quickly resolved, while others require sustained engagement with the legal system. We focus on these unmet needs in the online legal community. This paper focuses on quantitative and qualitative analysis of the legal taxonomy model. Our results show that there is an identifiable pattern across legal categories. This study also shows the impact of natural disasters (such as the COVID-19 pandemic) on the legal systems that govern work and employment. To interpret the model, we employ FrameNet([1]) and open-sesame([3]) to acquire the semantic role of given text. The label gives us a possible way to construct a graph that contains rich contextual information, including frame entity(aka role) and frame.

1. Introduction

A significant issue is that people are often unable to identify a problem in their lives as a legal issue. If a person can locate their issue as a legal one, they often do not know what kind of legal issue it may be. Lastly, they do not know how to deal with it. These unmet needs are increasing, and more people are looking for legal help. We are interested in how these needs, and other useful information, might be identified from an online community, specifically, a subreddit.

People post online because they have an issue, whether it be legal or not. Some legal experts can help them and eventually resolve their issues. In current practices, a lawyer or legal advisor may not identify the category of the legal problem that a person is having until they meet or speak with them. My research analyzes and reveals the consumers' needs to connect the people who need help with those who can help them. The consumer will get more support, and the provider will get more insight into the consumer's needs. A large number of unmet legal needs exist and must be resolved. Understanding and predicting these needs will benefit both legal service consumers and providers.

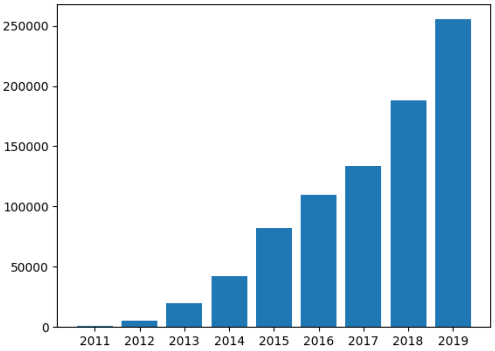


Figure 1. Number of submissions in legal advice subreddit

Figure 1 shows the accumulation of submissions to a legal advice subreddit. What does Reddit's legal advice data tell us about current trends in how people are seeking and identifying forms of legal assistance? What legal needs can we describe and infer through analyzing texts in an online community?

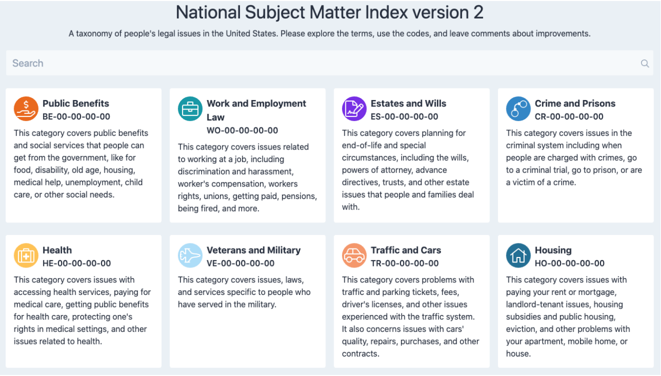
2. Problem Statement

The traditional N-gram model cannot capture context such as long-term dependency because of its limited memory. Recurrent neural network counters the way to process given input sequentially without losing all contexts. However, we still lose a lot of information if we provide sequential input naively. For instance, source code produced by RNN does not do anything useful except showing similar syntactic output. However, source code contains its structures that we can utilize. Moreover, natural language containing events, such as human experience, cannot be encoded with naive word embedding. Inspired by the graph representation of source code, we employ FrameNet that provides rich linguistic traits and convert it to a graph representation of given natural language input. Second, we use this graph representation as an input embedding of the Gated Graph Neural Network to teach the model to predict the correct legal category designed by Stanford Legal

Design Lab and Training Data from Suffolk university's SPOT dataset. We will prove the importance of frame in natural language and the new word embedding created by its Frame Entity and Frame. We will compare our result with widespread attention based pre-trained word embedding BERT. In short, We have two contributions. First, we provide a new word embedding from the local context. Second, we convert natural language into a graph for applying graph neural networks.

3. Data

3.1. NSMI v2



National Subject Matter Index version 2

A taxonomy of people's legal issues in the United States. Please explore the terms, use the codes, and leave comments about improvements.

Public Benefits
BE-00-00-00-00
This category covers public benefits and social services that people can get from the government, like for food, disability, old age, housing, medical help, unemployment, child care, or other social needs.

Work and Employment Law
WO-00-00-00-00
This category covers issues related to working at a job, including discrimination and harassment, worker's compensation, workers rights, unions, getting paid, pensions, being fired, and more.

Estates and Wills
ES-00-00-00-00
This category covers planning for end-of-life and special circumstances, including the wills, powers of attorney, advance directives, trusts, and other estate issues that people and families deal with.

Crime and Prisons
CR-00-00-00-00
This category covers issues in the criminal system including when people are charged with crimes, go to a criminal trial, go to prison, or are a victim of a crime.

Health
HE-00-00-00-00
This category covers issues with accessing health services, paying for medical care, getting public benefits for health care, protecting one's rights in medical settings, and other issues related to health.

Veterans and Military
VE-00-00-00-00
This category covers issues, laws, and services specific to people who have served in the military.

Traffic and Cars
TR-00-00-00-00
This category covers problems with traffic and parking tickets, fees, driver's licenses, and other issues experienced with the traffic system. It also concerns issues with cars' quality, repairs, purchases, and other contracts.

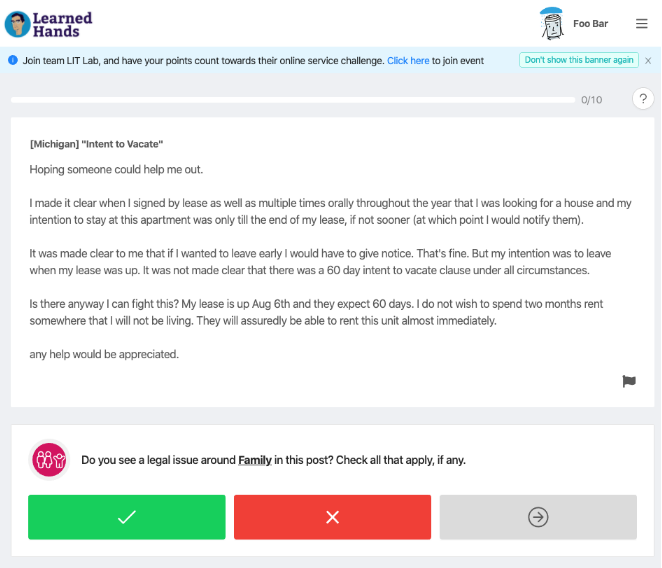
Housing
HO-00-00-00-00
This category covers issues with paying your rent or mortgage, landlord-tenant issues, housing subsidies and public housing, eviction, and other problems with your apartment, mobile home, or house.

Figure 2. Legal Taxonomy

The National Subject Matter Index(NSMI) v2[1] is a legal taxonomy classified by legal experts. The object of this classification is to provide an understandable and general way to categorize legal terms. Currently, we have 20 categories, including public benefit, work and employment law, health, and housing. Each group has its subclasses resulting in more than 300 subclasses.

3.2. Learned Hands

Researchers from Stanford and Suffolk university created a game[2] called Learned Hands [2]. The game presents a text from the legal community, lawyers, law students, and others who label the issue into a legal category when they agree about categorizing them. The legal class is called NSMI v2(taxonomy.legal), which contains twenty types: work, housing, crime, court, traffic, and accidents. These categories have subterms which determine the actual legal cases defined by law experts. From this game, the extracted data contains 2777 labeled Reddit submissions collected from 7/8/2017 to 1/20/2018[3]. We use this as training data and conduct an exploratory analysis. [4]



Join team LIT Lab, and have your points count towards their online service challenge. Click here to join event

[Michigan] "Intent to Vacate"

Hoping someone could help me out.

I made it clear when I signed by lease as well as multiple times orally throughout the year that I was looking for a house and my intention to stay at this apartment was only till the end of my lease, if not sooner (at which point I would notify them).

It was made clear to me that if I wanted to leave early I would have to give notice. That's fine. But my intention was to leave when my lease was up. It was not made clear that there was a 60 day intent to vacate clause under all circumstances.

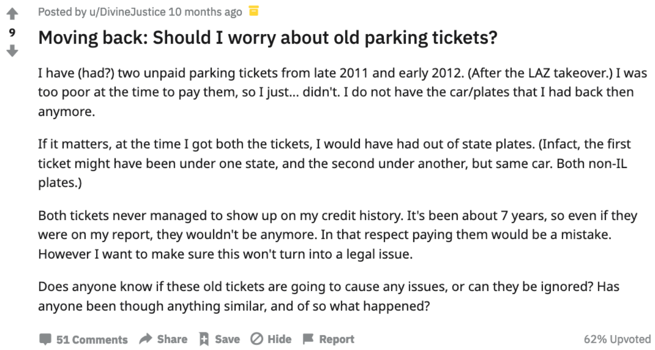
Is there anyway I can fight this? My lease is up Aug 6th and they expect 60 days. I do not wish to spend two months rent somewhere that I will not be living. They will assuredly be able to rent this unit almost immediately.

any help would be appreciated.

Do you see a legal issue around Family in this post? Check all that apply, if any.

Figure 3. Number of submissions in legal advice subreddit

3.3. Reddit Legal Advice



Posted by u/DivineJustice 10 months ago

Moving back: Should I worry about old parking tickets?

I have (had?) two unpaid parking tickets from late 2011 and early 2012. (After the LAZ takeover.) I was too poor at the time to pay them, so I just... didn't. I do not have the car/plates that I had back then anymore.

If it matters, at the time I got both the tickets, I would have had out of state plates. (Infact, the first ticket might have been under one state, and the second under another, but same car. Both non-IL plates.)

Both tickets never managed to show up on my credit history. It's been about 7 years, so even if they were on my report, they wouldn't be anymore. In that respect paying them would be a mistake. However I want to make sure this won't turn into a legal issue.

Does anyone know if these old tickets are going to cause any issues, or can they be ignored? Has anyone been though anything similar, and of so what happened?

Figure 4. A submission on Reddit /r/legaladvice

Legal advice subreddit[5] is an online community where people can post their issues, which may or may not be legal issues. People submit their issues, location, and other users who have had similar experiences offer their advice. They are not necessarily legal experts, and we do not know if their answers are valid or not. The first submission was on April 20, 2010, and currently, there are more than 900,000 submissions. To retrieve these, we construct a downloader[6] using Pushshift Reddit API.[7]. The API allows a user to scrap

[1] taxonomy.legal

[2] https://learnedhands.law.stanford.edu/

[3] https://spot.suffolklitlab.org/

[4] https://github.com/heeh/legal_advice/blob/master/EDA.ipynb

[5] https://www.reddit.com/r/legaladvice/

[6] https://github.com/heeh/subreddit_downloader/

[7] https://github.com/pushshift/api

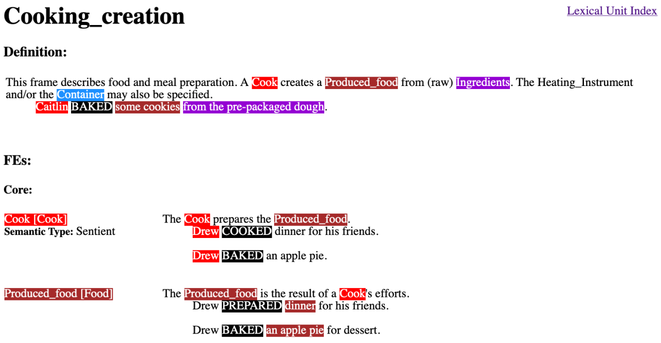
1000 submissions per request and 200 requests per minite. For example, the following API call retrieves 1000 submissions with specified fields.

4. Technical Approach

4.1. Legal Category Classification

Given online text data, we predict the legal category. The model is a simple TF-IDF GloVe word embedding model with logistic regression. We used SPOT data for training and applied the trained model to Reddit legal advice submissions.

4.2. FrameNet



Cooking_creation

Lexical Unit Index

Definition:

This frame describes food and meal preparation. A Cook creates a Produced_food from (raw) Ingredients. The Heating_Instrument and/or the Container may also be specified.

Caitlin BAKED some cookies from the pre-packaged dough.

FEs:

Core:

Cook [Cook]
Semantic Type: Sentient

The Cook prepares the Produced_food.

Drew COOKED dinner for his friends.

Drew BAKED an apple pie.

Produced_food [Food]

The Produced_food is the result of a Cook's efforts.

Drew PREPARED dinner for his friends.

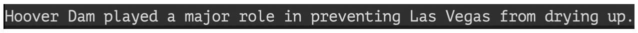
Drew BAKED an apple pie for dessert.

Figure 5. FrameNet

FrameNet[8] contains more than 1,200 frames, 10,000 Frame Entities, and 200,000 manually annotated sentences that defines the roles and frame of a sentence. Each word has an associated frame entity and eventually form a frame. We want to create a graph for each sentence. Nodes are frame entity core, and the edge is the frame that shows the relation between two nodes in a sentence. We combine lexical word embedding and the frame entity core to produce the distributed representation of nodes. We will take into account the given frame's supertypes to avoid new frames' problems.

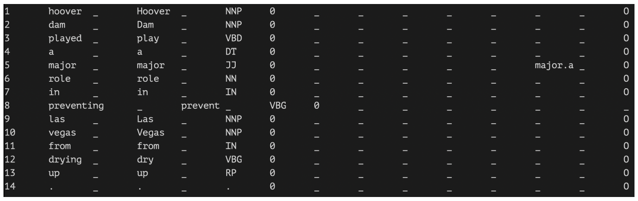
4.3. Open Sesame: Frame-Semantic Parser

open-sesame([3]) is a frame-semantic parser that detects FrameNet frames and frame elements from a sentence. The process is similar to a semantic role labeling task, and This is a useful tool for preprocessing data. This parser internally uses dynet for backpropagation, but unfortunately, it is outdated and poor support on GPU. Worse yet, its pretrained model has a dimensionality issue that I had to train from scratch. I spent a ridiculous amount of time resolving compatibility and finally able to run training. I will use the output structure and create a graph that represents the text.



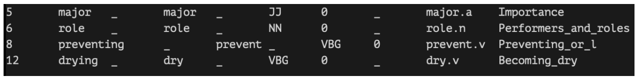
```
Hoover Dam played a major role in preventing Las Vegas from drying up.
```

Figure 6. sentencs.txt (input)



```
1       hoover  _       Hoover  _       NNP     0       _       _       _       _       _       _       _       0
2       dam     _       Dam     _       NNP     0       _       _       _       _       _       _       _       0
3       played  _       play    _       VBD     0       _       _       _       _       _       _       _       0
4       a       _       a       _       DT      0       _       _       _       _       _       _       _       0
5       major   _       major   _       JJ      0       _       _       _       _       _       major.a _       0
6       role    _       role    _       NN      0       _       _       _       _       _       _       _       0
7       in      _       in      _       IN      0       _       _       _       _       _       _       _       0
8       preventing      _       prevent _       VBG     0       _       _       _       _       _       _
9       las     _       Las     _       NNP     0       _       _       _       _       _       _       _       0
10      vegas   _       Vegas   _       NNP     0       _       _       _       _       _       _       _       0
11      from    _       from    _       IN      0       _       _       _       _       _       _       _       0
12      drying  _       dry     _       VBG     0       _       _       _       _       _       _       _       0
13      up      _       up      _       RP      0       _       _       _       _       _       _       _       0
14      .       _       .       _       .       0       _       _       _       _       _       _       _       0
```

Figure 7. Target Prediction



```
5       major   _       major   _       JJ      0       _       major.a    Importance
6       role    _       role    _       NN      0       _       role.n     Performers_and_roles
8       preventing      _       prevent _       VBG     0       prevent.v  Preventing_or_l
12      drying  _       dry     _       VBG     0       _       dry.v      Becoming_dry
```

Figure 8. Frame Prediction

5. Experiment

5.1. Model: Binary Classification of Legal Taxonomy

The following result comes from the previous semester's research using raw text input with TF-IDF and GloVe word embedding on 2,777 Reddit submissions annotated with the corresponding legal category. We conducted a binary classification task and acquired the accuracy, precision, recall, and F1 score. Keep in mind that the baseline does not consider syntactic and semantic information, such as the order of words or semantic roles' relationship. Accuracy is pretty high because this is a binary classification of each legal category. A low F1 score implies a massive number of false-positives.

Currently, there are more than a hundred classes, and eventually, We end up with sixteen legal categories that give a consistent result. Given sixteen legal categories, we build seperate classifiers for each legal category. Each classifier predicts whether given documents are related to a certain legal category. We test TF-IDF and GloVe vector representation and decide the TF-IDF that uses the logistic regression with cross entropy loss and L1 regularization.

$$\begin{gathered} score(\lambda) = loss(\mathbf{x}^{de}, \mathbf{y}^{de}, \hat{\theta}) \\ \hat{\theta} =_{\theta} \log P_{\theta}(\mathbf{y}^{tr} |\, \mathbf{x}^{tr}) - \lambda|\theta| \end{gathered}$$

The following is the performance of the logistic regression classifier with different word embedding. [9]

[8] https://framenet.icsi.berkeley.edu/fndrupal/

[9] https://github.com/heeh/legal_advice/blob/master/comparison.ipynb

Classifier	Acc.	Prec.	Rec.	F1
TF-IDF L1	0.97	0.52	0.41	0.46
TF-IDF L2	0.97	0.55	0.22	0.28
GloVe(50) L1	0.93	0.25	0.54	0.32
GloVe(50) L2	0.92	0.24	0.56	0.31
GloVe(300)L1	0.96	0.37	0.52	0.42
GloVe(300)L2	0.97	0.40	0.51	0.44

First, we classify each submission into the sixteen legal categories. We used a simple TF-IDF logistic regression model with a grid search of the power of two. We conduct ten-fold validation on the Learned Hands data and apply this model to predict the legal category of each submission.
For example, given the previous parking ticket submission on Figure 4, the classifier produces following results. In short, this issue is related to traffic, court, and money.

Legal Class	Probability(%)
TR-00-00-00-00	95.61
CO-00-00-00-00	64.52
MO-00-00-00-00	37.11
BU-00-00-00-00	4.86

5.2. Quantitative Analysis

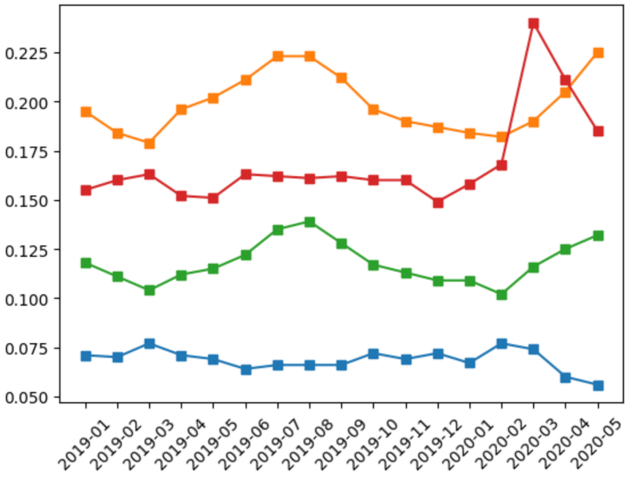


Figure 9. Legal Taxonomy Prevalence on Reddit Data

Using the previous classification model, we measure the legal category prevalence on Reddit legal advice from 2019 to 2020. The graph above shows the legal category prevalence changes during 2019-2020. The red, orange, green, and blue line corresponds to work and employment law, housing, renting and leasing, and health, respectively. We see a sharp increase in work and employment law demand in March.

5.3. Qualitative Analysis: Word Cloud

The quantitative analysis shows the prevalence of category on Reddit online community. However, it is still not enough to understand each legal taxonomy intuitively. To tackle this problem, one approach is to use word cloud.

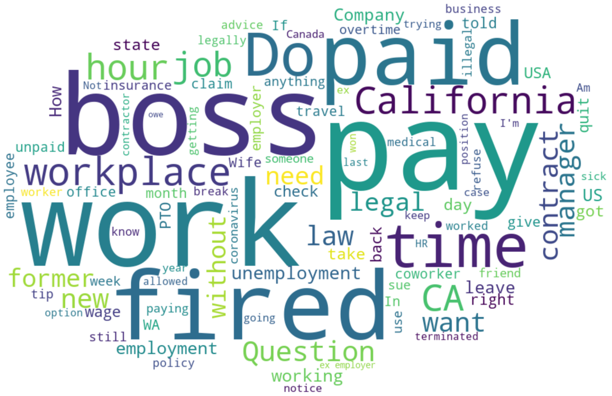


Figure 10. Work and Employment(2020 Week 10)

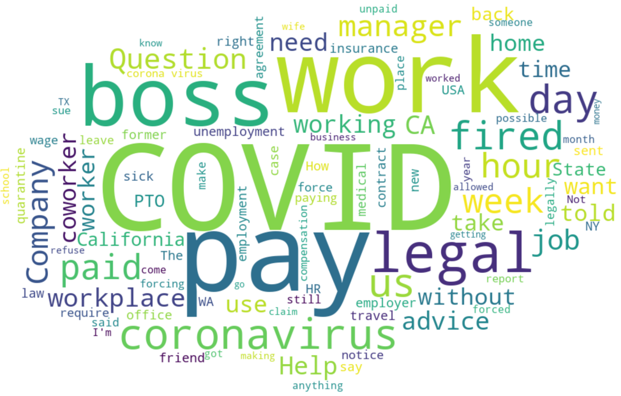


Figure 11. Work and Employment(2020 Week 11)

5.4. Qualitative Analysis: Semantic Role Labeling

We want to convert a text to a graph that represents the relationship between their roles. open-sesame[10] takes a sentence and does the semantic role labelling using FrameNet. This model internally uses bidirectional LSTM to predict the role. I was able to get it working and train the three different models, target words, frames, and arguments(Frame Elements). We ran the training and prediction task on a google cloud platform, and it took a week to process full SPOT data, 2777 submissions. The result is as follows.

5.4.1 Input data from SPOT dataset

The text below is taken from the Reddit Legal Advice forum. We use this text to label semantic roles using

[10] https://github.com/swabhs/open-sesame

FrameNet[1] and open-sesame[3]

”Moving back: Should I worry about old parking tickets? I have (had?) two unpaid parking tickets from late 2011 and early 2012. (After the LAZ takeover.) I was too poor at the time to pay them, so I just... didn’t. I do not have the car/plates that I had back then anymore. If it matters, when I got both the tickets, I would have had out of state plates. (In fact, the first ticket might have been under one state, and the second under another, but same car. Both non-IL plates.) Both tickets never managed to show up on my credit history. It’s been about seven years, so even if they were on my report, they wouldn’t be anymore. In that respect, paying them would be a mistake. However, I want to make sure this won’t turn into a legal issue. Does anyone know if these old tickets will cause any issues, or can they be ignored? Has anyone been through anything similar, and if so, what happened?”

5.4.2 Target Prediction

worry.v two.num late.a early.a too.adv back.v get.v state.n plate.n might.v state.n second.a car.n never.adv manage.v show.v up.prep respect.n pay.v want.v make.v sure.a turn.v know.v go.v cause.v can.v happen.v

Using open-sesame, we get frames in this text. Firstly, the program predicts the ”target” which invokes frames. After, it indicates the actual frame from the target. Vehicle, Commerce-pay, Cause-change, Point-of-dispute can be seen as an important relationship that determines this text’s type. This text is a traffic-related legal issue.

5.4.3 Frame Prediction

Emotion-active, Age, Placing, Cardinal-numbers, Placing Temporal-subregion, Temporal-subregion, Time-vector, Sufficiency Commerce-pay, Taking-sides, Temporal-collocation Transition-to-state, Leadership, Armor, Ordinal-numbers Likelihood, Leadership, Ordinal-numbers, Identicality, Vehicle Armor, Frequency, Successful-action, Evidence, Locative-relation History, Calendric-unit, Statement, Judgment, Commerce-pay Desiring, Manufacturing, Certainty, Cause-change, Legality Point-of-dispute, Awareness, Age, Motion, Causation Point-of-dispute, Capability, Similarity, Event

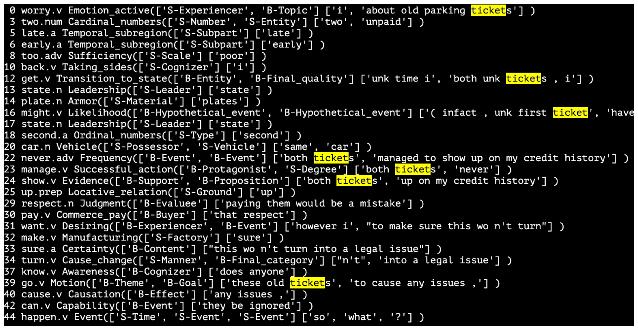
These frames inform us about the frame invoked by the target words. The frames may contain several Frame Elements associated with this frame, such as agents, tools, or entities.

5.4.4 Frame Element Prediction

I-Evaluee, B-Theme, S-Manner, S-Number, I-Cognizer, I-Protagonist, B-Goal, I-Buyer, B-Experiencer, S-Ground, S-Factory, S-Vehicle, B-Protagonist, S-Subpart, I-Proposition, B-Final-quality, S-Leader, S-Type, S-Entity, I-Theme, B-Cognizer, S-Degree, I-Final-quality, S-Cognizer, S-Event, I-Final-category, S-Experiencer, B-Evaluee, I-Content, B-Event, B-Proposition, B-Buyer, B-Hypothetical-event, S-Scale, I-Topic, B-Effect, I-Support, B-Final-category, S-Possessor, B-Content, B-Support, I-Experiencer, B-Entity, S-Time, I-Effect, I-Hypothetical-event, S-Material, I-Event, I-Entity, B-Topic, I-Goal

The frame elements are represented with CoNLL-2009 format. B stands for the beginning of the word, and ’I’ means intermediate words. These entities work with the frame from the previous section.

5.5. Graph structure

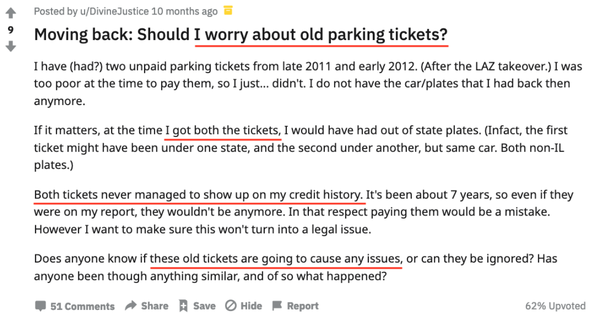


```
 0 worry.v Emotion_active(['S-Experiencer', 'B-Topic'] ['i', 'about old parking tickets'] )
 3 two.num Cardinal_numbers(['S-Number', 'S-Entity'] ['two', 'unpaid'] )
 5 late.a Temporal_subregion(['S-Subpart'] ['late'] )
 6 early.a Temporal_subregion(['S-Subpart'] ['early'] )
 8 too.adv Sufficiency(['S-Scale'] ['poor'] )
10 back.v Taking_sides(['S-Cognizer'] ['i'] )
12 get.v Transition_to_state(['B-Entity', 'B-Final_quality'] ['unk time i', 'both unk tickets , i'] )
13 state.n Leadership(['S-Leader'] ['state'] )
14 plate.n Armor(['S-Material'] ['plates'] )
16 might.v Likelihood(['B-Hypothetical_event', 'B-Hypothetical_event'] ['( infact , unk first ticket', 'have
17 state.n Leadership(['S-Leader'] ['state'] )
18 second.a Ordinal_numbers(['S-Type'] ['second'] )
20 car.n Vehicle(['S-Possessor', 'S-Vehicle'] ['same', 'car'] )
22 never.adv Frequency(['B-Event', 'B-Event'] ['both tickets', 'managed to show up on my credit history'] )
23 manage.v Successful_action(['B-Protagonist', 'S-Degree'] ['both tickets', 'never'] )
24 show.v Evidence(['B-Support', 'B-Proposition'] ['both tickets', 'up on my credit history'] )
25 up.prep Locative_relation(['S-Ground'] ['up'] )
29 respect.n Judgment(['B-Evaluee'] ['paying them would be a mistake'] )
30 pay.v Commerce_pay(['B-Buyer'] ['that respect'] )
31 want.v Desiring(['B-Experiencer', 'B-Event'] ['however i', "to make sure this wo n't turn"] )
32 make.v Manufacturing(['S-Factory'] ['sure'] )
33 sure.a Certainty(['B-Content'] ["this wo n't turn into a legal issue"] )
34 turn.v Cause_change(['S-Manner', 'B-Final_category'] ["n't", 'into a legal issue'] )
37 know.v Awareness(['B-Cognizer'] ['does anyone'] )
39 go.v Motion(['B-Theme', 'B-Goal'] ['these old tickets', 'to cause any issues ,'] )
40 cause.v Causation(['B-Effect'] ['any issues ,'] )
42 can.v Capability(['B-Event'] ['they be ignored'] )
44 happen.v Event(['S-Time', 'S-Event', 'S-Event'] ['so', 'what', '?'] )
```

Figure 12. Graph Representation of the parking ticket text

Now that we have targets, frames, and FEs(frame entities), we can construct a graph data structure. For example, the first word is a target, the invoker of the frame. After that, we have a frame followed by its frame elements enclosed with square brackets. The frame becomes the edge of the graph, and the frame elements are nodes that use the edge.

5.6. Converting semantic labeling to a graph



Posted by u/DivineJustice 10 months ago

Moving back: Should I worry about old parking tickets?

I have (had?) two unpaid parking tickets from late 2011 and early 2012. (After the LAZ takeover.) I was too poor at the time to pay them, so I just... didn't. I do not have the car/plates that I had back then anymore.

If it matters, at the time I got both the tickets, I would have had out of state plates. (Infact, the first ticket might have been under one state, and the second under another, but same car. Both non-IL plates.)

Both tickets never managed to show up on my credit history. It's been about 7 years, so even if they were on my report, they wouldn't be anymore. In that respect paying them would be a mistake. However I want to make sure this won't turn into a legal issue.

Does anyone know if these old tickets are going to cause any issues, or can they be ignored? Has anyone been though anything similar, and of so what happened?

Figure 13. Frame Graph Example(Ticket Entity)

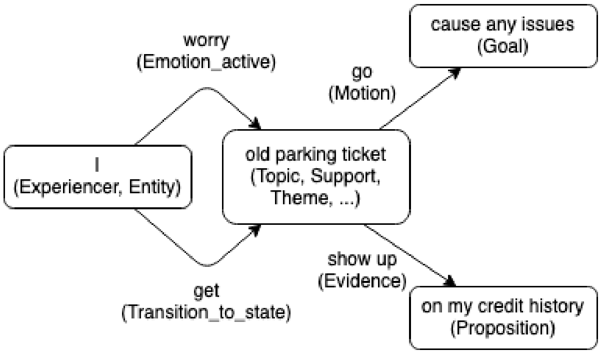


Figure 14. Frame Graph Example(Ticket Entity)

Open-sesame successfully tagged semantic roles. Now, we use these to construct a graph. Let us define a graph $G = (N, E)$ where N is node and E is an edge. In this paper, we take Frame Elements as nodes and Frames as edges. For example, a lexical unit 'worry' triggers a frame 'Emotion-active' that takes frame arguments Experiencer(I) and 'Topic'(old parking ticket). There are several problems here. First, the model did not resolve duplicate frame elements. Hense, our graph structure does not encode unique nodes. In natural language processing, this problem is called coreference resolution and require different models. Second, several frames do not take any frame elements at all. To explain this in a mathematical term, the edge of the graph does not have nodes. Having an edge without corresponding nodes is nonsense, and it is not clear how to treat these node-less edges.

Third, some lexical units trigger the wrong frame. In the textbox below, we can see the incorrect frame invocation by state(Leadership),

Fourth, each frame does not contain enough information to convey the relationship between entities in our graph. I believe that this happens because of the complexity of natural language. For instance, every line counts informal language. It has input, output, and operator, which can be specified. On the other hand, the natural language may lack meaning.

Finally, each frame evokes different Frame Elements, and they may refer to the same entity. Multiple instances of the same entity confuse the model, which is another big challenge of text graph-conversion.

6. Conclusion

There are two main takeaways while doing this project. First, it is vital to set up a clear goal at the beginning. I was planning to do the problem related to computer graphics and vision at the beginning of the semester. I changed it to network visualization on text data without thinking much about how hard it will be.

I was planning to convert the .conll file JSON format for the Gated Graph Neural Network and acquire each word's semantic embedding. We do not consider words that do not evoke frames. Converting a given legal text into the graph structure was the most critical part of the project. However, I realized that the graph structure is not powerful enough to be used as an input to the model. Thus, the FrameNet and Gated Neural Network output is not available because of the failure to create a graph data structure from the semantically labeled data.

I tried semantic role labeling and gated graph neural network. I successfully processed the semantic role labeling but could not enhance the previous classification model using a gated graph neural network. This experience taught me problem-solving skills as well as how to deal with frustration.

I learned several traits of natural language. Dealing with unstructured data is challenging, even if we use semantic information of the text. Unlike other data, such as image, video, or relational data, natural language needs to be preprocessed and modified extensively depending on the task that I am interested in.

7. Future Works

A neural network is a massive structure that computes numerous vectors inside. A large portion of what they are doing is unknown. In computer vision, there is a concept called network visualization that shows what feature learns during the process. It will be exciting to find a way to visualize features for text and graph data structure.

Lastly, I still believe that there must be a better way to encode natural language. For example, most natural language tasks, including the latest neural net, rely on the word and sequence. I believe that contextual information can be encoded as a graph. I want to work more on processing natural language so that we can get a better result.

8. For Instructors

The current semester's object was to acquire better intuition on what's going on inside the model. For this, we focused on qualitative analysis using semantic role labeling based on FrameNet and open-sesame, a bidirectional LSTM model. I want to clarify that most quantitative analysis in this paper comes from the previous semester's legal advice research.

References

[1] C. F. Baker, C. J. Fillmore, and J. B. Lowe. The Berkeley FrameNet project. In *COLING-ACL '98: Proceedings of the Conference*, pages 86–90, Montreal, Canada, 1998.

[2] M. Hagan and D. Colarusso. You say potato, we say legal issue: Adapting a digital epidemiology approach to access to justice. unpublished, unpublished.

[3] S. Swayamdipta, S. Thomson, C. Dyer, and N. A. Smith. Frame-Semantic Parsing with Softmax-Margin Segmental RNNs and a Syntactic Scaffold. *arXiv e-prints*, page arXiv:1706.09528, June 2017.