Annotation Guidlines for Explainability Annotations for Legal Judgment Prediction in Switzerland

Nina Baumgartner

1 Annotation Goal

Recently Niklaus, Chalkidis, and Stürmer (2021) presented a dataset for legal Judgment prediction including 85K Swiss Federal Supreme Court decisions. Using Hierarchical BERT they achieved a Macro-F1 Score of approximately 68-70%. Nevertheless the inner workings of such models are still mostly unknown and the results thus not interpretable. For this reason this annotation task focuses on the explainability of these predictions. With your annotation you will give your insight as a legal expert and tag parts of the facts that support or oppose the judgment. You will be using the two labels *Supports judgment* and *Opposes verdict* to tag parts of the facts of a court decision. For your annotation we have constructed three datasets (one for each language) from the Swiss Court Ruling Corpus. Each containing six cases over six years. With each year having two cases per legal area: One with a verdict approved and one with a the verdict dismissed.

2 Annotation specification

The following annotation specification and guidelines will help you identify the important elements (terms and sections) of the facts of the court decisions in the dataset and help you to label them correctly. Note that Natural Language Annotation is an Iterative task and this Guidelines are just a first draft.

2.1 Model

Firstly let us introduce an abstraction of the task ahead. In the following section we will define a model as the triple M = ¡T,R,I¿ with T being vocabulary of terms, R the relations between these terms and I being their interpretation (Pustejovsky and Stubbs (2012)).

For the legal judgment predication we have a binary classification task with the categories (dissmissal, approval) (Niklaus et al. (2021)). These label are assigned by a court on the basis of the facts and the task for the Models where to predict them correctly using the facts.

```
T = {Judgment, approval, dissmissal}
R = {Judgment::= approval|dissmissal}
I = {approval="request is deemed valid", dissmissal ="request is denied"}
```

For our annotation task we use an adaption of the model above. For our dataset we have chosen 36 cases where the model decided correctly. We now want to label which word or phrase of the facts are important for the judgment using the labels *(Supports judgment,Opposes verdict)* and therefore indicate what they denote in the context of the judgment.

```
T = {Facts_Section, Supports judgment,Opposes verdict}
R = {Facts_Section::= Supports judgment|Opposes verdict}
I = {Supports judgment="this section/term supports the judgment",
      Opposes verdict="this section/term opposes verdict"}
```

2.2 Guidelines

As mentioned above you will only annotate the facts of the court decision, but you will have access to the full document (via a link). The judgment is indicated clearly on the interface, as are the considerations and the ruling. Even if you only annotate the fact you can and should use these other resources as an indicator on which part of the facts are of greatest importance.

2.2.1 Semantic Roles

Pustejovsky and Stubbs described different semantic roles associated with different verbs:

Agent

The event participant that is doing or causing the event to occur

Theme/figure

The event participant who undergoes a change in position or state

Experiencer

The event participant who experiences or perceives something

Source

The location or place from which the motion begins; the person from whom the theme is given

Goal

The location or place to which the motion is directed or terminates

Recipient

The person who comes into possession of the theme

Patient

The event participant who is affected by the event

Instrument

The event participant used by the agent to do or cause the event

Location/ground The location or place associated with the event itself (Pustejovsky and Stubbs (2012))

For our task we adapt this thinking to better explain which part of the text should be annotated. Please note that due to my own limited legal education. I can only describe the thinking process using knowledge from Articles covered in the StGB Allgemeiner Teil and in the Besondere Bestimmungen I. So the following question that could be asked are inspired by this knowlege and should be adapted to the cases of other legal area.

So when we now apply the semantic role thinking to our annotation task we can probably quite easily identify the verb indicating the prediction (see example below) and thus a starting point for our annotation is to label it. Associated with the verb may be an object. We can ask ourselves is this object relevant to the judgment? Does this object make the offense possible at all? Is this object an indicator of privileges or qualifications relevant for the verdict?

If we now ask about the parties involved, for example, plaintiff, court, victim, etc. Which of these parties or persons is particularly relevant for the judgment? Is it a crime that requires a particular subject or object? Which courts are involved and is one instance especially relevant to the judgment. is the plaintiff relevant to the outcome or could his person be replaced by any person?

Looking at the localities now are mentioned localities relevant to judgment, inland foreign countries etc.?

All these questions can be answered with a certain verb, noun, name etc. and should be marked with the label supports judgment if they are relevant and supportive for the judgment. If the context requires it, prepositions and adverbs should also be marked. If there is a contradictory phrase in the facts, which may be ignored in the verdict or dismissed as irrelevant, it should be marked with the label opposes verdict. Everything that is not marked has a neutral value in the evaluation. As mention above the emphasize of your annotation should be on the explainability. There is no right or wrong annotation. Your labeling and expertise provides more context to results.

2.3 A simple Example

We now look at a very simple example of a basic annotation. The following facts and ruling from are from a case from 2007 (id in scrc 433714) with the judgment being "dismissal".

2.3.1 Facts

Sachverhalt: Mit Urteil vom 30. Oktober 2006 wies das Eidgenössische Versicherungsgericht die von C._ gegen den Entscheid des Sozialversicherungsgerichts des Kantons Zürich vom 27. April 2006 (betreffend Invalidenrente) erhobene Verwaltungsgerichtsbeschwerde ab. C._ ersucht unter Hinweis auf bisher "unbekannte neue Fakten" um Revision des letztinstanzlichen Urteils vom 30. Oktober 2006 (Eingabe vom 25. April 2007).

2.3.2 Ruling

Demnach erkennt das Bundesgericht: 1. Das Revisionsgesuch wird abgewiesen. 2. Die Gerichtskosten von Fr. 500.- werden dem Gesuchsteller auferlegt. 3. Dieses Urteil wird den Parteien, dem Sozialversicherungsgericht des Kantons Zürich und dem Bundesamt für Sozialversicherungen zugestellt.

In this case, the verb that refers to the rejection, as well as who rejects and what was rejected, must certainly be annotated with the label supports judgement.

[...] wies $_{\text{supports judgment}}$ [...] Eidgenössische Versicherungsgericht $_{\text{supports judgment}}$ [...] erhobene Verwaltungsgerichtsbeschwerde ab $_{\text{supports judgment}}$ [...]

There could be more possibilies of annotating the above facts this is only an illustration of the very basic elements that should always be labeled.

References

Niklaus, J., Chalkidis, I., & Stürmer, M. (2021). *Swiss-judgment-prediction: A multilingual legal judgment prediction benchmark.* arXiv. Retrieved from https://arxiv.org/abs/2110.00806 DOI: 10.48550/ARXIV.2110.00806

Pustejovsky, J., & Stubbs, A. (2012). *Natural language annotation for machine learning* (Nos. Bd. 9,S. 878). O'Reilly Media, Incorporated.