Automated Construction of Formal Normative Models: A Machine Learning Approach for Natural Language Interpretation

Abstract : In both societal and business contexts, people's behavior is often guided by normative texts written in natural language. Examples include laws governing society and organizational policies. The challenge inherent in these normative documents lies in the interpretation process, which often results in multiple, and at times conflicting, understandings. To enhance transparency, formalized constructs such as the Flint Ontology can be employed to make these interpretations explicit.

However, the time-intensive nature of creating normative models and the scarcity of experts capable of interpreting these documents pose significant challenges, making the task difficult to scale.

To address this hurdle, our project aims to automate the extraction of formalized constructs by leveraging the capabilities of transformer-based machine learning. Our system utilizes these models to construct formalized representations, incorporating expert's feedback to iteratively retrain the models and enhance its adaptability to various types of normative text.

While primarily trained on game rule sentences, our models exhibit promising accuracy in generating formalized constructs based on game rule sentences and the retraining mechanism demonstrates adaptability to other types of normative texts. However, to enhance generalization, there is a need for more diverse data.

Our Team (Group D2)
- Efstathios Trigkas
- Dimitrios Tsiamouras
- Atanas Yonkov
- Lucas Belderink
- Tawfeek Alkanbar

Acknoledgements
We would like to thank Tom van Engers, Giovanni Sileno (UvA) and Robert van Doesburg (TNO) in guiding the project.

University of Amsterdam
TNO innovation for life

PROBLEM STATEMENT

To what extent can state-of-the-art machine learning tools be leveraged to automate the accurate interpretation of normative text into a formal structure, and how does this automation contribute to scalability in the extraction of formalized constructs?

Given the difficulty in finding a validated dataset suitable for training, to what extent can we rely on the feedback loop of retraining to improve the generalization of models?

OUR SOLUTION

To address this problem, we developed a system that assists users in automatically constructing normative models in Flint ontology from natural language while providing a feedback loop of retraining using the expert's input to enhance the accuracy of underlying predictive models over time.

A typical use case of the system involves an expert inputting normative text in natural language. The system uses pre trained models to predict the formalized constructs, initiating a dialogue between the user and the system to refine and finalize the produced flint frames. The accurate representations are then stored for future training iterations, contributing to continuous improvement in the system's performance.

LIMITATIONS

- **Sentences are treated individually:** context is lost if information is implied or present in another part of the text. Experts need to manually fill in these parts of the frames.
- **Token classification contingent on correct sentence classification:** because of the model's architecture, incorrect sentence classification has cascading effect on both frame and token prediction accuracy.

METHODOLOGY

Data Collection & Annotation
- Collected game rule sentences from official game rule websites.
- Manually annotated 200 sentences, classifying each as Act, Fact, or Duty.
- Labeled each word in the sentence based on its corresponding slot.
- Evaluated annotations with normative experts to ensure consistency.

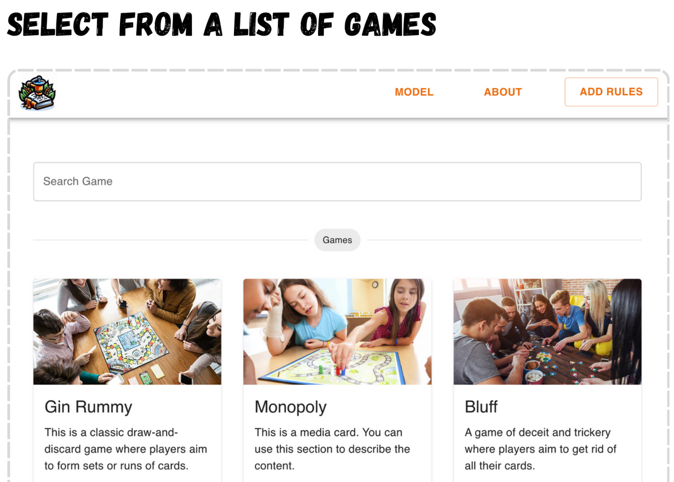
Model Training
- Utilized BERT models for interpretation.
- Trained a BERT model for sentence-level categorization (ACT, FACT, DUTY).
- Trained separate BERT models for word-level semantic role assignment in Act and Duty categories.
- Conducted hyperparameter optimization for enhanced model performance.
- Involved stakeholders and domain experts throughout the process for validation and refinement.

Application
We adeveloped a web interface for experts to interact with the models. In the backend of the system, the models make predictions for user input while also allowing for further retraining.

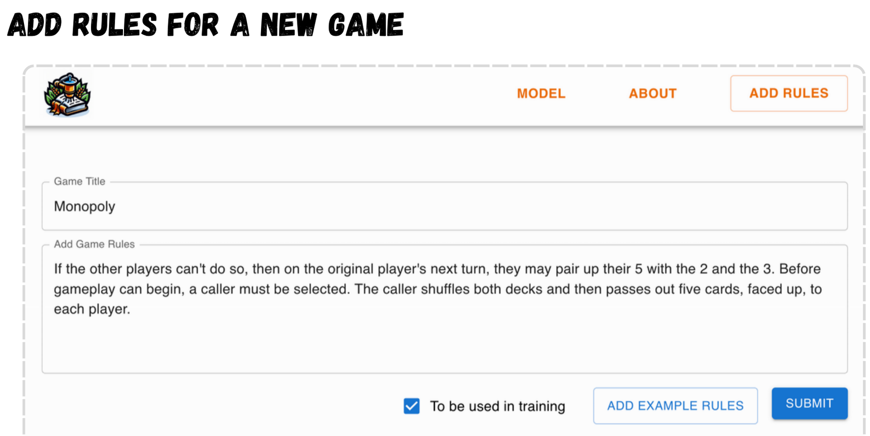
Validation
- **Ground Truth Validation:** Results validated by domain experts, ensuring alignment with expert interpretations.
- **Summative Usability Testing:** Obtained stakeholders' insights, refining functionality to meet their needs.

THE SYSTEM

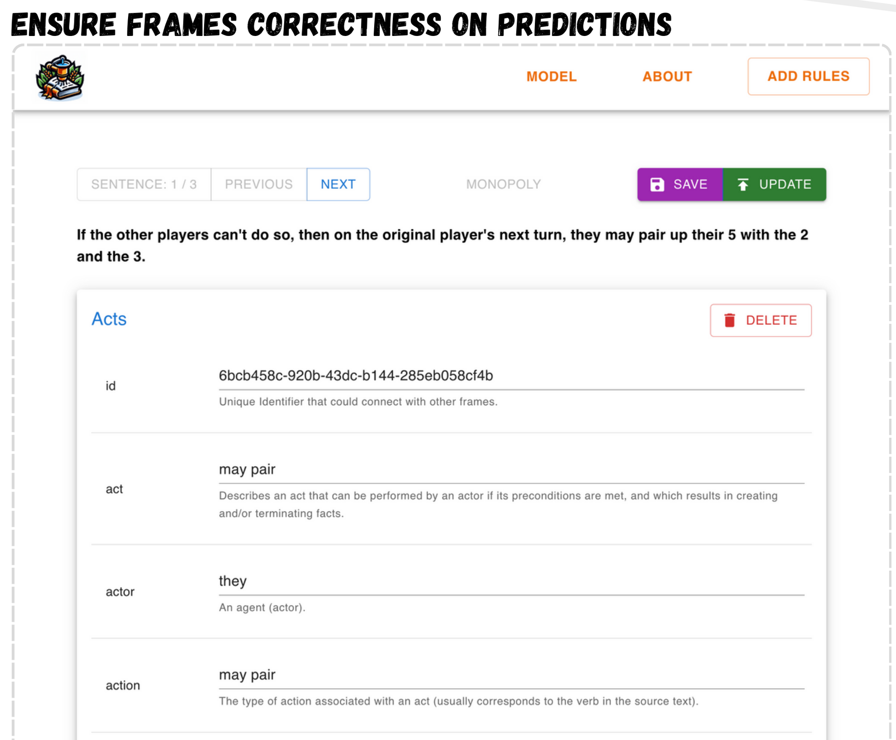
SELECT FROM A LIST OF GAMES



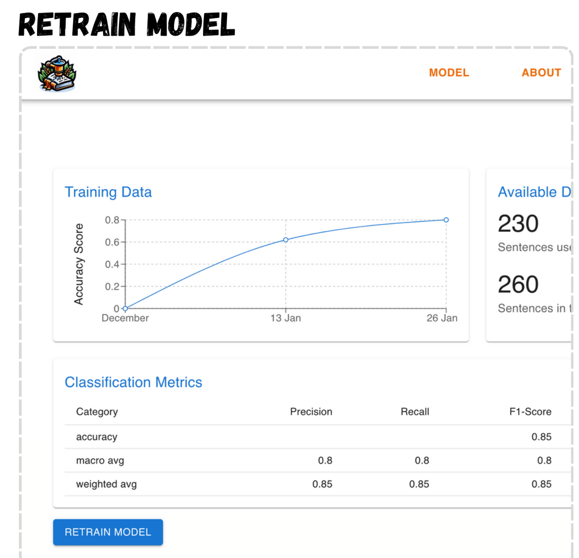
ADD RULES FOR A NEW GAME



ENSURE FRAMES CORRECTNESS ON PREDICTIONS



/model
RETRAIN MODEL



RESULTS

- **F1-scores for sentences in training domain:** 93% for Fact, 80% for Act and 67% for Duty frames.
- **F1-scores for sentences in new domains:** 82% for Fact, 77% for Act and 52% for Duty frames.
- **Act token classification:** 82% for Preconditions.
- **Duty token classification:** 87% for Enforcing Fact and 85% for Creating Act.

CONCLUSION

- **Promising Generalization:** Although initially trained on game rules, our platform shows potential when applied to diverse domains, suggesting a robust foundation.
- **Retraining for Diversity:** Enhancing the model's understanding across various contexts requires retraining with a more diverse set of examples.
- **Consider Architecture Change:** Exploring a shift in the model architecture to treat entire paragraphs instead of individual sentences may allow for a more holistic understanding of contextual information.