AlpacaTag: An Active Learning-based Crowd Annotation Framework for Sequence Tagging

Bill Yuchen Lin*,1, Dong-Ho Lee*,1, Frank F. Xu2, Ouyu Lan1, and Xiang Ren1

1University of Southern California  2Carnegie Mellon University

http://inklab.usc.edu/AlpacaTag

Easy-to-use Intelligent Recommendations

Overall Workflow of AlpacaTag

Learning Backend Model with Consolidation

Key Features

- **Active Intelligent Recommendation**
  Dynamically suggesting annotations and actively sampling the most informative unlabeled instances.

- **Annotation Consolidation**
  Enhancing inter-annotator agreement by consolidating multiple personal backend models.

- **Real-time model deployment**
  Users can deploy backend models in downstream system via APIs while annotators are tagging data.

### Framework Implementations

- **UI-design**: doccano
- **Front-end**: Vue.js
- **Back-end**: Django
- **Database**: SQLite
- **Requests**: OMQ
- **Model Server**: Worker
- **Annotation Server**: Worker
- **Ventilator**: Worker

### Extensible Configurations

- **Embedding**: Glove, Word2Vec, Fasttext, ELMo, BERT
- **Recommendation**: Noun Chunk, Online Learning, ELMo, Dictionary Match
- **Active Learning**: Random, MNLP, MNLP + MC, MNLP + BB

- Users can optionally enable each source in the Settings.
- Final recommendations are merged from three options with the priority:
  “Noun Chunk < Model Inference < Dictionary Match”