EXTENDING TRANSFORMATION-BASED LEARNING TO COMPUTE DIALOGUE ACTS

We have developed an implementation that addresses a natural language understanding task called *Dialogue Act Tagging* with a machine learning method called *Transformation-Based Learning.* We are focusing on the important goal of interpreting *dialogue acts*, where a dialogue act is a concise abstraction of the speaker's *intention.* This sample dialogue shows how each utterance may be labeled with a dialogue act, such as GREET, SUGGEST, or ACCEPT.

To achieve the Dialogue Act Tagging goal we are using Brill's *Transformation-Based Learning* machine learning method. Given a tagged training corpus, Transformation-Based Learning produces a learned model that consists of a sequence of *rules.* Here are the first seven rules in a sequence produced by our implementation. These rules can be used to label a dialogue by applying each rule, in order, to every utterance in the dialogue.

Three significant problems arose in the course of this research. The first issue dealt with *efficiency.* In training, Transformation-Based Learning must generate an *intractably* large number of rules. To solve this problem, we developed a *Monte Carlo* method that randomly samples only R of the rules that can be used to fix a given tag, where R is a predefined constant. As a result, the complexity of the method *drops significantly.* In addition, our version is still highly likely to find an *accurate* sequence of rules. We believe that this is because we are still searching through a significant percentage of the space of possible rules, there are several rules that are about equally effective, and our method is *more* likely to find the rules that are *more* effective. Our experimental results show that the Monte Carlo version of Transformation-Based Learning dramatically decreases the training time, without compromising accuracy on held-out data.

A significant limitation of Transformation-Based Learning is that it fails to offer any measure of *confidence* in the labels that it produces. To remedy this, we have adapted Dagan and Engelson's Committee-Based Sampling approach and Freund and Shapire's Boosting technique, which were developed for entirely different purposes. The basic idea is that, rather than generating just one model of the training data, we run the training algorithm *multiple* times, to produce a few different models, which we call *committee members.* To generate *diverse* and *reasonable* committee members, we follow the Boosting approach. The first committee member is trained by running Transformation-Based Learning in the normal way, and then the other committee members look at which instances were tagged incorrectly by the preceding committee members, and they pay greater attention to those instances during training. After the committee members are trained, given untagged data, each committee member independently determines the tags, and then their answers are compared. The confidence of a tag is based on the degree of *agreement* between the different committee members on that tag. Our experimental results show that the committee method produces useful confidence measures and increases the overall tagging accuracy of the system.

Our third contribution is to develop an automatic approach for selecting *dialogue act cues*, which are the most relevant words and phrases for computing dialogue acts in a particular domain. We are currently investigating several different metrics for evaluating the *predictive power* of a given word or phrase, by considering the *distribution of dialogue acts* that co-occur with that word or phrase in a training corpus. We ran some experiments that show that the traditional sets of *cue phrases* that are found in the literature are not sufficient for our purposes, while the exhaustive set of *word sequences* in a training corpus overwhelms the system, limiting accuracy. However, our initial experiments have already demonstrated that it is possible to automatically select a *relatively small* set of dialogue act cues that achieves a significantly *higher accuracy score* in tagging held-out data with dialogue acts.

So, throughout this research, our motivating goal was to develop a system that could effectively label utterances with dialogue acts. Previously the best reported score of any method on the Dialogue Act Tagging task was achieved by Reithinger and Klesen, whose system could label *74.7%* of the utterances with the correct dialogue acts. And our system's results *on the same corpus* are comparable.

So, to summarize our main contributions, we presented a *Monte Carlo* version of Transformation-Based Learning to make the method *tractable* for our task, which decreases the labor demands on the human developer and increases the range of tasks where Transformation-Based Learning can apply effectively. We adapted a *committee method* to compute useful *confidence measures* and increase the overall tagging accuracy of our system. And, we're developing a strategy for selecting *dialogue act cues* automatically, from a training corpus. This approach collects words and phrases that are tuned to a particular task and domain, and it improves the system's performance.