Project Proposal: Active Learning for Bootstrap Learning Quality Control

Andrew Carlson, Sue Ann Hong, Kevin Killourhy, Sophie Wang

The goal of our project is to build a system for inferring the quality of facts and patterns (rules) gathered by a bootstrap learning system, with the aid of a human labeler in an active learning setting. There are two main parts to our project: to construct a general (probably probabilistic) framework for confidence scoring and to utilize active learning to enhance confidence scoring. Our system will not learn any specific facts or relations. It is most appropriate for our system to be placed inside a general (or specific) bootstrap learning system to aid in quality control of both facts and patterns the system gathers. Put differently, our system is to help a bootstrap learning system from perishing or diverging far away from the truth (the word truth is used loosely as a concept – both facts and patterns – the corpus and the labeler believe to be true) due to patterns and facts gathered by the system that are inappropriate for the given task. We plan to collaborate with the general bootstrap system group to realize this picture.

The input to our system will be a set of extracted facts and a set of patterns that extract, along with how many times each fact was extracted by each rule. (Note active learning also requires some string representation of the patterns, extracted facts, and the relation or predicate the system is extracting instances of.) This reflects a limited form of bootstrapping termed Meta-bootstrapping by [Ghani & Jones]. If time permits, we may be able to extend it to other bootstrap learning settings. Based on the inputs and the feedback from user, the system will output confidence scores for each fact and pattern. The exact mechanism of confidence score inference is yet to be determined, but we are currently considering two main possibilities.

The first model is an extension of the Urns model by [Downey et al.]. Know-It-All, the bootstrap learning (really?) system which uses the Urns model for information extraction, does not extract new patterns but only facts. Hence the Urns model of estimating the probability of correctness of facts based on pre-defined rules (or "urns") is adequate. However, the bootstrap learning scheme we consider not only learns new facts but also new patterns with which we can discover more facts, and thereby violates crucial assumptions in the Urns model. The foremost of these assumptions is that the probability of any particular correct example being drawn is greater than the probability of any particular error example from a Urn, at least given uniform distribution. The bootstrapping mechanism may easily discover new patterns, or Urns, that does not satisfy this assumption. We can employ active learning here to determine which of the learned patterns are appropriate ("good") to be used to judge extracted facts under the Urns model. We name this model Active LURNS (Active Learning URNS).

While Active LURNS provides us with a solid foundation that is the Urns model we can work with, it may be more beneficial to use a more general framework. We are considering a simple graph of patterns and facts as nodes, where an edge e:v-w implies that v extracted w (or vice versa). We have not worked out the details, but we envision the nodes and edges having weights, or confidence scores, where each confidence score is determined based on the neighboring nodes and edges. For example, a label from the user on a node would "propagate" confidence throughout the graph with some decaying function with respect to the distance from the labeled node.

The premise of active learning that by using an intelligent scheme one can decrease the required number of labeled examples. Hence the main question for active learning is for which examples to we want the labels. The exact scheme should depend on the confidence model to be able to take full advantage of active learning. We will start with the schemes suggested by [Jones et al.] and [Muslea 2002], some of which are generally applicable (e.g. uniform random selection and high frequency selection) and some of which will require modification to fit our model (e.g. feature set disagreement scheme by [Jones et al.] only works within the co-testing setting). Other questions such as for how many examples we must ask labels will be answered based on the specific example selection scheme.

Note that we purposely do not define what an example is at this point. It may be an extracted fact, an extraction pattern, or even a black box system that presents its opinions on the correctness of facts or patterns. One could even imagine the user (labeler) as an example, whose confidence score may be inferred by other examples. As a starting point we will consider only extracted facts as examples for simplicity, which the user giving only yes, no, or do-not-know answers. We will then take user's answers to be the ground truth.

Bibliography

[Ghani & Jones] Rayid Ghani and Rosie Jones. A Comparison Of Efficacy And Assumptions Of Bootstrapping Algorithms For Training Information Extraction Systems. pdf

[Jones et al.] Jones, R.; Ghani, R.; Mitchell, T.; and Riloff, R. Active Learning for Information Extraction with Multiple View Feature Sets. 2003. pdf

[Muslea 2002] I. Muslea, S. Minton, and C. Knoblock. Active + Semi-Supervised Learning = Robust Multi-View Learning. ICML 2002. pdf

I. Muslea, S. Minton, and C. Knoblock. Selective sampling with redundant views. AAAI 2000. CiteSeer

Raghavan, H.; Madani, O.; and Jones, R. InterActive Feature Selection. IJCAI 2005. pdf

[Downey et al.] Downey, Etzioni, and Soderland. A Probabilistic Model of Redundancy in Information Extraction. IJCAI 2005 pdf.

About Synergy Scenarios

Our system should be able to aid scenarios in general given that there are clear patterns or extraction mechanisms that can be presented to a user. For example, given a bootstrap system that finds a biology faculty member's publications, our system can infer a confidence value for an extracted pair (professor X, publication Y) based on the rules that extracted it and (sometimes) what the user thinks about the pair, thereby aiding the bootstrap system to make better decisions for choosing which patterns and facts to bootstrap with.