

A Practical Approach for Writer-Dependent Symbol Recognition Using a Writer-Independent Symbol Recognizer

- Why pair-wise symbol recognition in AdaBoost?
 - If you are only comparing two characters, it should be easy to recognize either character.
 - Extend this idea to have a recognizer for each character pairing.
 - Get the best answer from all the recognizers.
 - You will need many classifiers, but using a recognizer (in this case the Microsoft recognizer) helps to reduce the number of classifiers needed.
- The Microsoft recognizer only helps with characters it knows.
 - You could use any recognizer that supplies an n-best list.
- No rejection criteria.
- Professor LaViola wanted to make a system that would learn as you went along
- Is it easier to train a system for natural language or mathematical expressions?
 - Natural language is easier because you can use information about previous characters to help determine the next character.
 - Ex. Trigrams – “th” is often followed by an “e.”
- Improvements
 - Multi-stage classifier.
- There is a paper coming out which talks about the mechanisms of Microsoft’s recognizer.
 - Using the n-best list helps reduce the time for recognition, as you only have to look at 5 or 6 symbols.
 - There are Microsoft recognizers for many different languages.
- PCA – Principle Components Analysis
 - Computes eigenvalues/eigenvectors for matrix of samples’ vectors.
- Refereed journal
- AdaBoost algorithm is a bit different than normally used in AdaBoost.
 - For loop uses $T \cdot J$
- LaViola chose the features to be his weak-learners in AdaBoost.
- No distributional assumptions for AdaBoost.
- AdaBoost
 - # of rounds
 - Cap alpha?
 - Lots of trial and error to get correct parameters for machine learning algorithm.

October 3, 2007

Ink Features for Diagram Recognition

- Something is wrong in the paper – It says that Rubine used HMMs (Hidden Markov Models) in his paper, but he used linear classifiers.
- Recognition is not just recognition of a character or drawings.
 - Need to know what section has characters and what section has drawings.
- Appendix is very useful – lots of features.
- Didn't do much.

Template-based online character recognition

- Anil K. Jain is a famous recognition researcher.
- Not feature-based.
- String matching – like curve matching.
- Dynamic program
- Combines different recognition methods.
 - At what point do you stop seeing results when you combine recognizers?
- Seems like you can swap in different classifiers.
- Recognition rate not very high – completely independent recognizer.
 - Ran on many characters.
 - Large study
 - Independent recognition is harder to do.