MATH189: Paper Summary

The paper is titled 'The Infinite Markov Model' and the link is https://papers.nips.cc/paper/3281-the-infinite-markov-model.pdf.

This paper presents a nonparametric method of estimating variable order Markov processes up to a theoretically infinite order. Markov models grow exponentially with respect to their order and therefore it is often necessary to have a low order. In Markov models using words, the order is usually capped at 5. However, it may very likely be the case that these word dependencies extend for longer than 5 words.

Methods have been proposed that involve a variable order. This method has several limitations such as needing to prune a very large tree if deeper connections are to be analyzed. Also, the parameter thresholds must often be specific and the optimal values are generally empirical and have no theoretical backing.

With the introduction of hierarchical Poisson-Dirichlet processes, Markov models can handle sparser distributions. We then consider Chinese restaurant processes with an infinite depth suffix tree. We assume that each node i has a hidden probability q_i of stopping at node i when following a path from the root of the tree to add a customer. This means $(1 - q_i)$ is the penetration probability when descending an infinite depth suffix tree from its root. Since this probability is hidden, we must find a way to estimate it. This is done with a Gibbs sampler, a Markov chain Monte Carlo algorithm for obtaining a sequence of observations when direct sampling is difficult. This can then give us an estimation of the Markov order from which each word was generated.

This allows us to extract 'stochastic phrases', which are words that tend to go together in text, and they can be aggregated based on topic. Some of the phrases extracted from a sampling of a dataset of NIPS papers include 'primary visual cortex' and 'american institute of physics'.

Character-based Markov models were also investigated. This has applications in language processing and unknown word recognition. With the character-based model, the authors discovered that many nodes in the model corresponded to valid words despite them not being taught directly to the model. They also saw a correlation between the order of the Markov model and the perplexity, which is a measure of average predictive probabilities with smaller being better. They noted that a higher order generally resulted in lower perplexity and the infinite model always outperformed the other models.

In summary, the authors of the paper were able to estimate variable order Markov processes by extending a stick-breaking process 'vertically' over a suffix tree of hierarchical Chinese restaurant processses and make an inference on the order of words in the document.