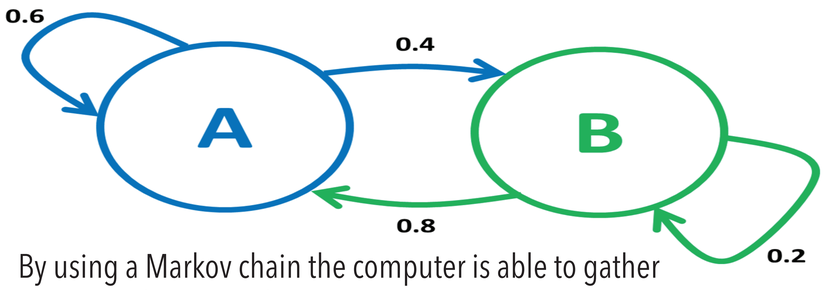
MACHINE LEARNING
WITH NATURAL LANGUAGE PROCESSING USING BAYESIAN NETWORK MODELS

By Logan Jastremski

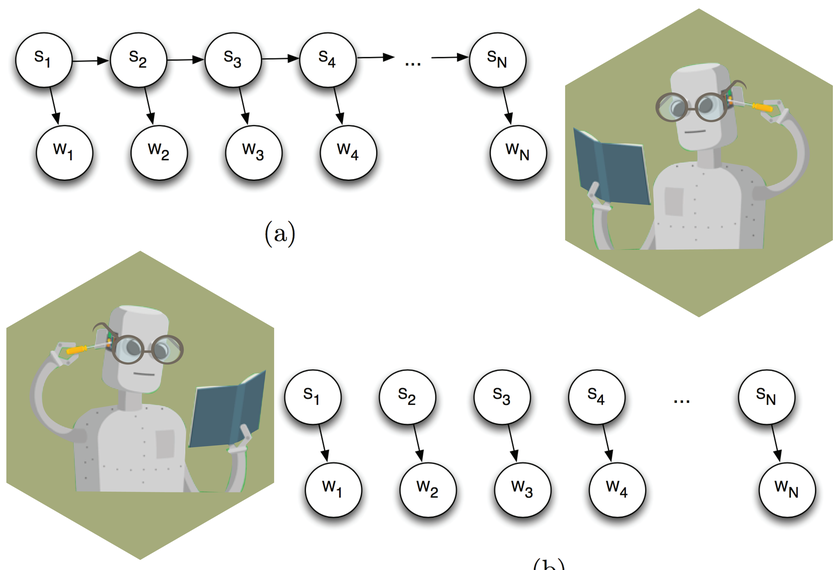
ABSRACT

Through out the history of Artificial intelligence Natural language processing (NLP) has been at the forefront. Dating back to the 1950's NLP algorithms have been refined and improved to focus more on statistical and probabilistic data sets, instead of the prior linguistic approach. In this research we compared and contrasted two algorithms: Naïve and Viterbi, both working with Bayesian networks. Each algorithm learns from a data set with part-of-speech (noun, verb, adjective, etc.) tagging Ex. "Cheerful ADJ way NOUN to PRT spend VERB an DET evening NOUN" after learning from the dataset the computer with be able to accurately predict part of speech in single words and full sentences.

BACKGROUND



By using a Markov chain the computer is able to gather enough information from the dataset to predict the part-of-speech remarkable well. In the graph above it shows how a system changes from one state to another over time. For example: If state A was a NOUN what would the probability that state B would be a VERB or vice versa. In this model the system in state A is used to help the system transition to state B in order to predict the next part-of-speech in the sentence . Using this statistical and probabilistic approach the computer is able to mathematically equate the probability of the part-of-speech in sentence structures.



(a)

(b)

METHODOLOGY

DATASETS

Each algorithm is trained on a file with over 44,204 sentences in the total data set. After the file has learned from the training set it is tested on 2,000 sentences with 29,442 words.

NAIVE

In example (b) the Naive algorithm only looks at the single word to predict It its part of speech. If it hasn't seen the word before in the dataset it makes a high probability guess on what it could be.

VITERBI

In example (a) the Viterbi algorithm looks to the previous state in the Markov Chain. With the ability to look at the last state in the sentence structure this algorithm has a higher probability of accurately predicting the part of speech.

RESULTS

When looking at the results from the dataset with can see the Viterbi is on average about 10% better at predicting the part of speech in sentence structure than the Naive algorithm. This is inline with the hypothesis that Viterbi would perform better because the algorithm is structured in such a way that the model had the ability to learn and predict more accurately based on the pervious word in the sentence structure.

Sentence :		it's late and you said they'd be here by dawn
Ground truth	(-1.43)	prt adv conj pron verb prt verb adv adp noun
1. Naive	(-0.29)	prt adj conj pron verb prt verb adv adp noun
2. Viterbi	(-0.29)	prt adj conj pron verb prt verb adv adp noun

	Words correct	Sentences correct
Ground truth	100%	100%
1. Naive	93.95%	47.5%
2. Viterbi	95.71%	58.05%

So far scored 2000 sentences with 29442 words.

CONCLUSION

Natural language processing (NLP) will continue to be at the forefront of the Artificial intelligence community for some time. While the Ultimate goal of having an algorithm be able to fully understand human speech has been continuing to improve since the 1950's we still have a ways to go. By moving to statistical and probabilistic approach we are able to more accurately able to teach computers to pick the correct part-of-speech in single words and in full sentence structures. Improving part-of-speech recognition is vital to continuing to advance these natural language processing algorithms. Hopefully in the near future we'll be able to have intelligent, intuitive, and easy to use applications that allow humans and computer to effective communicate.