

Natural Language Generation with Markov Chains and Grammar

Sam Zhang
Computer Systems Lab 2009-2010

Abstract

Language can be generated stochastically using Markov Chain databases. In first order Markov Chain generation, a test corpus is analyzed for the probabilistic weights between each word and the word following it. Then a text is generated randomly by following these weights. This project explores the use of a *second-order* Markov Chain in conjunction with a semantic similarity model to create a de facto grammar such that the output has similar diction, syntax, and subjectively perceived tone as the source text, yet still different.

Background and Introduction

To create a comprehensible text, this project uses a corpus-based method to glean the basic grammatical rules and vocabulary, then using those, to piece together a new text (Figures 1 and 2).

Discussion

As can be seen, the two inaugural addresses (Figures 1 and 2) reflect the semiotic landscape of the time. Bush uses “power”, “burns”, “faiths”, “democratic”, “free”, and “idealistic”. Washington uses “nor party animosities”, reflecting his anti-partisan views; “peculiarly conscious”, “surely pledges”, and other phrases uncommon in contemporary political discourse.

Results and Conclusions

The project successfully regenerates texts using a second-order Markov Chain. This has applicability in the field of human-computer interaction, automated responses, and computational linguistics.

Figure 1: 2005 Bush's inaugural address:

In the truths of Sinai , the questions that come to those who feel its power , it burns those who feel its power , it burns those who feel its power , and the varied faiths of our soldiers . You have seen together . For as long as whole regions of the United States to seek and support the growth of democratic movements and institutions in every generation by reaffirming all that is the concentrated work of helping raise up free governments K the idealistic work of generations . The great objective of ending tyranny is the...

Figure 2: 1789 Washington's inaugural address:

Than side no local prejudices or attachments , no separate views nor party animosities , will misdirect the comprehensive and equal eye which ought to be peculiarly conscious of his own deficiencies . In these honorable qualifications I behold the surest pledges that as on one side no local prejudices or attachments , no separate views nor party animosities , will misdirect the comprehensive and equal eye which ought to be suppressed ...