



CALIFORNIA STATE SCIENCE FAIR 2011 PROJECT SUMMARY

Name(s) Galina H. Meyer	Project Number 31951
Project Title The Application of Bayesian Networks for Speech Classification	
Objectives/Goals Computers that can process the large amounts of data statistics requires have made available unprecedented means of understanding the word, resulting in radical advancements in everything from quantum physics (modeling the behavior of subatomic particles using statistics) to environmental science (modeling changes in weather) to stock market analysis (modeling micro- and macro-economy trends). One of the final frontiers of science is understanding the human mind, and how it can communicate ideas. In this project, I wanted to explore applying a specific field of probability, called Bayesian networks, to identify what abstract idea some body of text is conveying-- here, whether a politician is advocating Democratic or Republican ideas. Abstract Computers that can process the large amounts of data statistics requires have made available unprecedented means of understanding the word, resulting in radical advancements in everything from quantum physics (modeling the behavior of subatomic particles using statistics) to environmental science (modeling changes in weather) to stock market analysis (modeling micro- and macro-economy trends). One of the final frontiers of science is understanding the human mind, and how it can communicate ideas. In this project, I wanted to explore applying a specific field of probability, called Bayesian networks, to identify what abstract idea some body of text is conveying-- here, whether a politician is advocating Democratic or Republican ideas. Methods/Materials I typed my program in LISP (dialect: Racket), using the IDE called DrRacket. My corpus of speeches, called Corps, came from Foundation Bruno Kessler and was generously supplied to me by Mr. Guerini and Mr. Strapparava. I mostly used Artificial Intelligence: A Modern Approach by Stuart J. Russell and Peter Norvig for my research. My paper was formatted in LaTeX, with the IDE TeXworks. I created a Bayesian network, with values determined from training the computer with the corpus, and used Bayes Theorem to derive the probability that a speech was from a certain party given the words that were in it. Results The program works, returning surprisingly accurate results. For extremest politicians such as Huckabee, it returned 99.8% accuracy, and for more moderate politicians, such as Ronald Reagan right after he changed to the Republican party, it returned 75.4% Republican. These trends continue fairly accurately over a wide variety of politicians, even those from overseas. Conclusions/Discussion The success of this program signals that computers really can be used to extract abstract ideas from a list of words, partially understanding natural languages by observing the trends of human speech. Humans learn in a manner similar to this, by listening to years worth of conversation, so this program actually follows how humans learn a language to a certain extent. This program is not limited to a Republican versus Democrat categorization: it can also include more categories, and more abstract. One possible option is identifying a speech that promotes violence or is	
Summary Statement I categorize political speeches with a Bayesian network.	
Help Received Mr. Anderson gave me the AI: Modern Approach book	