

Data Mining and Analyzing Basic Features of the State of the Union Addresses

Trevor Bihl Kenneth Bauer Spring Ohio MAA Meeting April 2013



theguardian





e AFIT of Today is the Air Force of Tomorrow.

The state of our union is ... dumber: How the linguistic standard of the presidential address has declined

Using the Flesch-Kincaid readability test the Guardian has tracked the reading level of every state of the union









- 1. What is special about the State of the Union Address (SUA)?
 - Generally annual, constitutionally required feedback from Pres.
- 2. What is the Flesch-Kincaid (FK) Readability Metric?
 - Should be somewhat familiar to anyone that remembers MS Word's spell-check in the 90s...
- 3. Goals of this Research
 - 1. Analyze Available Data
 - 1. Using more than just FK
 - 2. See if FK reflects changes in the SUA
 - 1. Will define classes
 - 2. Apply Artificial Neural Networks and
 - 3. Feature Screening used to reduce data
- 4. Interpretation and Conclusions

State of the Union Address (SUA) AFIT

The AFIT of Today is the Air Force of Tomorrow.

- One of the few Constitutional requirements of the President
 - Annual requirement, but date not specified
 - Frequency not specified either, Washington gave 2 in 1790
 - Form not specified Jefferson to Wilson had written SUAs
- The form, medium, content, and use have changed over time
- We are considering Joint Addresses (1981, 1989, 1993, 2001, 2009) as SUAs (supported in literature)
- 227 data points should be enough to observe trends
 - Some years had none (1841, 1881, and 1933)
 - Some years had 2 (1790, 1953, and 1981)



Flesch-Kincaid



The AFIT of Today is the Air Force of Tomorrow.

- It is impossible to quantify Eloquence
 - Eloquence is subjective and individual
- The Flesch-Kincaid, a fairly common algorithm, is used to calculate grade level based on^[ref]

 $FK = 0.39 \left(\frac{total \ words}{total \ sentences} \right) + 11.8 \left(\frac{total \ syllables}{total \ words} \right) - 15.5.$

 This function is also built into MS Word, which was used for calculation – actual MS Word coefficients are unknown

[ref] DuBay, William, The principles of readability, Impact Information, 2004.

Initial Regression Analysis



The AFIT of Today is the Air Force of Tomorrow.



 $FK = -0.0456 \cdot year + 101.35$

 $R^2 = 0.757$ p-value < 0.05 Air University: The Intellectual and Leadership Center of the Air Force

Aim High...Fly - Fight - Win





Selecting Data Features



Building a Dataset



The AFIT of Today is the Air Force of Tomorrow.

- So, What Data is Available?
- <u>Content Analysis</u> Analyze words in the SUA
 - Not selected: a) Timely; b) Subjective
- <u>Non-subjective components</u> of the SUA were selected:
 - Total Words, Average Words/Sentence, Average Chars/Word
 - FK Reading Level
- Incorporate External Factors?
 - Want to avoid any political inferences and potential political fallout
 - Also want to avoid older parties that no long exist
 - So, Presidential Education Level selected



Total Number of Words







Aim High...Fly - Fight - Win



Words/Sentence









Characters/Word







Aim High...Fly - Fight - Win

V

College Type of President

AIR FORCE INSTITUTE OF TECHNOLOGY









Examining Data Features

A few regions appear to exist in the data

We'll just examine FK, but the same locations appear in all features except Pres. Ed.















Aim High...Fly - Fight - Win





Defining Classes



Defined Classes



The AFIT of Today is the Air Force of Tomorrow.

Scenario	Class 1	Class 0	Number in Class 1	Number in Class 0
Written vs Oral	Orally delivered SUA	Written SUAs	97	130
Televised vs All Others	Televised SUAs	All non-televised SUAs	65	162
Public vs Congress	All SUAs broadcast to the US public	All Written SUAs and SUAs delivered orally to only Congress	78	149
Prime-Time SUA	All televised SUAs delivered to a primetime audience	All other SUAs	48	179





Data Classification



Data Classification



The AFIT of Today is the Air Force of Tomorrow.

- It appears that a relationship does exist in the data
 - Can we use it to classify the data?
- Artificial Neural Networks (ANNs) with SNR Feature Screening are explored for data classification
- ANNs are nonlinear classifiers that apply weights between connections and use activation functions in the hidden nodes to classify points
- Time was included as a feature for one of the two ANN models





ANN Properties



The AFIT of Today is the Air Force of Tomorrow.

	Settings		
Activation Function	Logarithmic Sigmoid		
Hidden Layers	1		
Decision Threshold	0.5		
Hidden Layer Nodes	4		
Maximum Training Epochs	100		
MSE Stopping Rule	0.001		
Training Method	Scaled conjugate gradient backpropagation		

Data Splitting: Training/Testing/Validation sets of 65%, 10%, and 25% 146, 23, and 56 data instances per group

Neural Network Feature Selection AFIT

The AFIT of Today is the Air Force of Tomorrow.

General Procedure:

- 1. Add a uniform (0,1) random noise parameter x_N to the original dataset
- 2. Standardize all parameters to have zero mean and variance of one
- 3. Randomly select training, test and validation sets
- 4. Train the ANN using training and test data
- 5. Apply validation set to ANN model
- 6. Compute classification accuracy
- 7. Compute SNR for each
- 8. Repeat steps 4-7 until only one non-noise parameter remains in the dataset
- 9. Apply classification accuracy results to determine an appropriate parsimonious dataset.









- ANNs are criticized for finding patterns where none exist
- As a rebuttal we examine vectors of random classes
 - Same number of 1's and 0's like the true classes
 - But, randomly chosen
 - Identical data splitting % for ANN building
- Ideally classification of these random classes should result in around 50% accuracy

SNR Feature Screening Result

The AFIT of Today is the Air Force of Tomorrow.



Results with Year as Feature

The AFIT of Today is the Air Force of Tomorrow.





Discarded Features



Iteration	Written vs Oral	Televised vs	Public vs	Prime-Time vs	
		All Others	Congress Only	All Others	
1	Words/Sent	Characters/Wor	President's	President's	
			Education	Education	
2	President's	President's	Characters/Word	Total Words	7
	Education	Education			
•					_
3	Characters/Wor	Words/Sent	Words/Sent	Flesch-Kincaid	╡.
4	Flesch-Kincaid	Total Words	Total Words	Characters/Word	_ I
5	Total Words	Flesch-Kincaid	Flesch-Kincaid	Words/Sentence	
Iteration	Written vs Oral	Televised vs	Public vs	Prime-Time vs] 1
		All Others	Congress Only	All Others	
1	President's	President's	Words/Sentence	President's	7
	Education	Education		Education	
2	Mards (Capt	Total Marda	Drosidant's	Characters (Mord	-
Z	words/sent		Fducation	Characters/ word	
3	Total Words	Characters/Wor	Total Words	Words/Sentence	
4	Characters/Wor	Words/Sent	Flesch-Kincaid	Year	ţ.
5	Year	Flesch-Kincaid	Characters/Word	Total Words	
6	Flesch-Kincaid	Year	Year	Elesch-Kincaid	for

It is safe to keep 2-3 features based on the accuracies





Conclusions







- What does this show?
 - The defined classes can be used to accurately describe the SUA
- SUA FK data doesn't appear to show anything about the US getting dumber (or smarter)
 - It appears to show self sorting as the SUA medium changes
- Others^{[ref1][ref2]} also examined types of speech, and newspapers in the 19th century
 - These found that it is common for mediums to self sort over time for efficiency
 - This research extends beyond them through using statistical methods

[[]ref1] Hallin, Daniel C., "Sound bite news: television coverage of election, 1968-1988." Journal of Communication, 1992, Issue 2, Vol. 42, pp. 5-24.

[[]ref2] Ryfe, David M. and Kemmelmeier, Markus., "Quoting practices, path dependency and the birth of modern journalism." Journalism Studies, Sept 22, 2010, pp. 1-17.





QUESTIONS





The AFIT of Today is the Air Force of Tomorrow.

- [1], "The Constitution of the United States." Section 3 Vol. Article 2.
- [2] Fisher, Louis., "Invoking Inherent Powers: A Primer." Presidential Studies Quarterly, March 2007, Issue 1, Vol. 37, pp. 1-22.
- [3] Binkley, Wilfred E., "The President as Chief Legislator." Annals of the American Academy of Political and Social Sciences, 1956, Vol. 307, pp. 92-105.
- [4] Hoffman, Donna R. and Howard, Alison D., Addressing the State of the Union: The Evolution and Impact of the President's Big Speech. Boulder, CO : Lynne Rienner Publishers, 2006.
- [5] Murray, Charles., "AEIdeas." *Don't Watch. It Only Encourages Them.* [Online] Jan. 26, 2011. [Cited: Nov. 10, 2012.] http://www.aei-ideas.org/2011/01/dont-watch-it-only-encourages-them/.
- [6] Kessel, John H., "The Parameters of Presidential Politics." Social Science Quarterly, 1974, pp. 8-24.
- [7] Lim, Elvin T., "Five Trends in Presidential Rhetoric: An Analysis of Rhetoric from George Washington to Bill Clinton." Presidential Studies Quarterly, Jun. 2002, Issue 2, Vol. 32, pp. 328-366.
- [8] Coe, Kevin and Neumann, Rico., "The Major Address of Modern Presidents: Parameters of a Data Set." Presidential Studies Quarterly, December 2011, Issue 4, Vol. 41, pp. 727-751.
- [9] Ostermeier, Eric., "My Message is Simple": Obama's SOTU Written at 8th Grade Level for Third Straight Year. Smart Politics. [Online] Jan 25, 2012. [Cited: Jan 25, 2012.] http://blog.lib.umn.edu/cspg/smartpolitics/2012/01/my_message_is_simple_obamas_so.php.
- [10] —. Keeping it Simple: Obama Records 2nd Lowest Flesch-Kincaid SOTU Grade Level Score Since FDR. Smart Politics/Humphrey School of Public Affairs. [Online] Jan. 27, 2011. [Cited: Feb. 1, 2011.] http://blog.lib.umn.edu/cspg/smartpolitics/2011/01/keeping_it_simple_obama_record.php.
- [11] —. 'Professor' Obama? President's State of the Union Address Notches 4th Lowest Flesch-Kincaid Grade Level Score Since FDR. Smart Politics/Humphrey School of Public Affairs. [Online] Jan. 29, 2010. [Cited: Feb. 5, 2011.] http://blog.lib.umn.edu/cspg/smartpolitics/2010/01/professor_obama_presidents_sta.php.



R7

References



The AFIT of Today is the Air Force of Tomorrow.

- [12] Hemingway, Mollie., "To Obama, We're Children. To Daniels, We're Adults." *Ricochet.* [Online] Jan. 25, 2012. [Cited: Jun. 15., 2012.] http://ricochet.com/main-feed/To-Obama-We-re-Children.-To-Daniels-We-re-Adults.
- [13] Schneiderman, Stuart., "Obama Treats the American People Like 8th Graders." Stuart Schneiderman. [Online] Jan. 25, 2012. [Cited: Jun. 15, 2012.] http://stuartschneiderman.blogspot.com/2012/01/obama-treats-americanpeople-like-8th.html.
- [14] Borevitz, Brad., *State of the Union.* [Online] January 24, 2012. [Cited: July 2, 2012.] http://stateoftheunion.onetwothree.net/index.shtml.
- [15] Mailloux, Stephen L., et al., "How reliable is computerized assessment of readability?" Computers in Nursing, 1995, Issue 5, Vol. 13, pp. 221-225.
- [16] Hayes, Donald P., Wolfer, Loreen T. and Wolfe, Michael F., "Schoolbook simplification and its relation to the decline in SAT-verbal scores." American Educational Research Journal, 1996, Issue 2, Vol. 33, pp. 489-508.
- [17] Wickman, Forrest., "Obama Might Write For 8th Graders, But He Doesn't Write Like Them." Slate. [Online] Jan. 25, 2012. [Cited: Jan. 30, 2012.]

http://www.slate.com/blogs/browbeat/2012/01/25/state_of_the_union_and_flesch_kincaid_obama_might_write_for _8th_graders_but_he_doesn_t_write_like_them.html.

- [18] Heffernan, Virginia., "Who's dumber: Congress or Martin Luther King Jr.? The dumb report on congressional dumbness." Yahoo! News | Machine Politics. [Online] May 24, 2012. [Cited: May 24, 2012.] http://news.yahoo.com/who%E2%80%99s-dumber--congress-or-martin-luther-king-jr---the-dumb-report-oncongressional-dumbness.html.
- [19] G., R. L., "Political Language Professorial?" *Economist.* [Online] Jun. 17, 2010. [Cited: Jun. 15, 2012.] http://www.economist.com/blogs/johnson/2010/06/political language.
- [20] Austin., "Economist." *Readability Talking Down to America.* [Online] Jan. 27, 2012. [Cited: Jun. 15, 2012.] http://www.economist.com/blogs/johnson/2012/01/readability.
- [21] Kohn, Alfie., "The Dangerous Myth of Grade Inflation." Chronicle of Higher Education. Nov 8, 2002, Vol. 49, 11, p.





The AFIT of Today is the Air Force of Tomorrow.

- [22] Manhire, Brian., "Grade Inflation, Ethics and Engineering Education." 2004. American Society for Engineering Education Annual Conference & Exposition. Vol. 2560.
- [23] Kozol, Jonathan., Illiterate America. s.l. : Plume/Penguin, 1986.
- [24] Hirsch, E. D., "How to stop the drop in verbal scores." New York Times. Sept. 18, 2011.
- [25] Jain, Anil K., Duin, Robert P. W. and Mao, Jianchang., "Statistical Pattern Recognition: A Review." IEEE Transactions on Pattern Analysis and Machine Intelligence, Jan. 2000, Issue 1, Vol. 22, pp. 4-37.
- [26] Beck, Nathaniel, King, Gary and Zeng, Langche., "Theory and Evidence in International Conflict: A Response to de Marchi, Gelpi, and Grynaviski." American Political Science Review, May 2004, Issue 2, Vol. 98, pp. 379-389.
- [27] Verikas, A. and Bacauskiene, M., "Feature selection with neural networks." Pattern Recognition Letters, 2002, Vol. 23, pp. 1323-1335.
- [28] Dewdney, A. K., "The Apprentice Builds a Brain: Misled by Metaphors." Yes, We Have No Neutrons. s.l. : John Wiley & Sons, 1997.
- [29] de Marchi, Scott, Gelpi, Christopher and Grynaviski, Jeffrey D., "Untangling Neural Nets." American Political Science Review, May 2004, Issue 2, Vol. 92, pp. 371-378.
- [30] Young, Williams A., et al., "Modeling net ecosystem metabolism with an artificial neural network and Bayesian belief network." Environmental Modelling & Software, 2011, Vol. 26, pp. 1199-1210.
- [31] Young, William A. and Weckman, Gary R., "Evaluating the effects of aging for profesional football players in combine events using performance-aging curves." International Journal of Sports Science and Engineering, 2008, Issue 3, Vol. 2, pp. 131-143.
- [32] Loeffelholz, Bernard, Bednar, Earl and Bauer, Kenneth W., "Predicting NBA Games Using Neural Networks." Journal of Quantitative Analysis in Sports, 2009, Issue 1, Vol. 5, pp. 1-15.
- [33] Kuo, R. J., Chen, C. H. and Hwang, Y. C., "An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network." Fuzzy Sets and Systems, 2001, Vol. 118, pp. 21-45.





The AFIT of Today is the Air Force of Tomorrow.

- [34] Burke, H. B., et al., "Artificial neural networks improve the accuracy of cancer survival prediction." Cancer, 1997, Vol. 79, pp. 857-862.
- [35] Beck, Nathaniel, King, Gary and Zeng, Langche., "Improving Quantitative Studies of International Conflict: A Conjecture." American Political Science Review, March 2000, Issue 1, Vol. 94, pp. 21-35.
- [36] Zeng, Langche., "Prediction and Classification with Neural Network Models." Sociological Methods and Research, 1999, Issue 4, Vol. 27, pp. 499-524.
- [37] Zeng, Langche., "Neural Network Models and Political Data Analysis." [book auth.] Diana Richards. *Political Complexity: Nonlinear Models of Politics.* Ann Arbor : University of Michigan Press, 2000, pp. 239-68.
- [38] Schrodt, Philip., *Patterns, Rules, and Learning: Computational Models of International Behavior.* Vinland, Kansas : Parus Analytical Systems, 2004.
- [39] Zhang, Guoqiang Peter., "Neural Networks for Classification: A Survey." IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews, Nov 2000, Issue 4, Vol. 30, pp. 451-462.
- [40] Johnson, Jerald B. and Omland, Kristian S., "Model selection in ecology and evolution." TRENDS in Ecology and Evolution, Feb. 2004, Issue 2, Vol. 19, pp. 101-108.
- [41] Bauer, Kenneth W., Alsing, Stephen G. and Greene, Kelly A., "Feature screening using signal-to-noise ratios." Neurocomputing, 2000, Vol. 31, pp. 29-44.
- [42] Reagan, Ronald., "Address Before a Joint Session of the Congress on the State of the Union." [Speech]. Washington, DC : s.n., Jan 26, 1982.
- [43] Leuchtenburg, Willaim E., "The Twentieth-Century Presidency." [book auth.] Harvard Sitkoff. *Perspectives on Modern America.* s.l. : Oxford University Press, 2001, pp. 9-32.
- [44] Teten, Ryan L., "Evolution of the Modern Rhetorical Presidency: Presidential Presentation and Development of the State of the Union Address." Presidential Studies Quarterly, June 2003, Issue 2, Vol. 33, pp. 333-346.
- [45] Kessel, John H., "The Seasons of Presidential Politics." Social Science Quarterly, 1977, Vol. 58, pp. 418-435.
- [46] Druckman, James N. and Holmes, Justin W., "Does Presidential Rhetoric Matter? Priming and Presidential Approval." Presidential Studies Quarterly, December 2004, Issue 4, Vol. 34, pp. 755-778.





The AFIT of Today is the Air Force of Tomorrow.

- [47] Neuendorf, Kimberly A., The Content Analysis Guidebook. s.l. : Sage Publications, 2002.
- [48] Woolley, John and Peters, Gerhard., "The American Presidency Project." State of the Union Addresses and Messages: research notes by Gerhard Peters. [Online] 2011. [Cited: Feb 5, 2011.] http://www.presidency.ucsb.edu/sou.php#axzz1dnvQDCCz.
- [49] DuBay, William H., The Principles of Readability. Costa Mesa, CA : Impact Information, 2004.
- [50] Mills, C. Wright., The Power Elite. 1956.
- [51] Japkowicz, Nathalie and Stephen, Shaju., "The class imbalance problem: a systematic study." Intelligent Data Analysis, Oct. 2002, Issue 5, Vol. 6.
- [52] Japkowicz, Nathalie., *Learning from imbalanced data sets: a comparison of various strategies.* AAAI Technical Report: WS-00-05. 2000.
- [53] Kutner, Michael H., et al., Applied Linear Statistical Models. Fifth Edition. New York : McGraw-Hill, 2005.
- [54] Dillon, William R. and Goldstein, Matthew., "A measure of fit." *Multivariate Analysis Methods and Applications.* New York : John Wiley & Sons, 1984, pp. 220-225.
- [55] Kurkova, Vera., "Kolmogorov's Theorem and Multilayer Neural Networks." Neural Networks, 1992, Vol. 5, pp. 501-506.
- [56] Looney, Carl G., Pattern Recognition Using Neural Networks. 1997 : Oxford University Press.
- [57] Atkinson, John Maxwell., Lend me your ears. s.l. : Oxford University Press, 2005.
- [58] Hallin, Daniel C., "Sound bite news: television coverage of election, 1968-1988." Journal of Communication, 1992, Issue 2, Vol. 42, pp. 5-24.
- [59] Orwell, George., "Politics and the English Language." Horizon, Apr. 1946.
- [60] Gold, Kevin., "An Information Pipeline Model of Human-Robot Interaction." Proceedings on Human-Robot Interaction, 2009, pp. 85-92.
- [61] Ryfe, David M. and Kemmelmeier, Markus., "Quoting practices, path dependency and the birth of modern journalism." Journalism Studies, Sept 22, 2010, pp. 1-17.
- [62] Swing, Edward L., et al., "Television and video game exposure and the development of attention problems." Pediatrics, 2010, Vol. 126, pp. 214-221.

• Air University: The Intellectual and Leadership Center of the Air Force

Aim High...Fly - Fight - Win