

Deep Learning Techniques to Understand Gender Disparities in Debate Speeches & Evaluations?

Huyen Nguyen

h.nguyen@ese.eur.nl

Erasmus University Rotterdam & Tinbergen Institute

Motivation

- Discrepancies between career success of men and women exist in many important areas of public life, including business, academia and politics. This raises the question of whether such discrepancies can be partially explained by differences in the persuasion tactics and evaluations across gender.
- Lack of comprehensive real-life debate datasets from well-defined settings of debate competitions to understand speech dynamics and evaluations.
- This project: Collect and analyze links between words, arguments and various speech performance measures to unravel any existence and magnitude of systematic gender differences in speeches and evaluations.

Research questions

- Is there a difference in speech delivery (i.e verbal and non-verbal content) and evaluations between men and women?
- How does the composition of opponent teams and judges affect speech performance?
- How does sentiment dynamics evolve across genders for different topics?
- What types of arguments are likely to win in gender-related topics?



Dataset

- Novel dataset of 1930 verbatim transcribed debate speeches (≈ 3.2 million words) from 250 debates in World Universities Debating Championship (WUDC), European Debating Championship (EUDC) and HWS Round Robin Champion League (HWS) on diverse topic pools.
- Currently $N_{In} = 1236$ speeches with full ranking (i.e preliminary rounds), $N_{Out} = 694$ speeches with partial ranking (i.e elimination rounds).
- Roughly 60% of the speeches have full administrative information about speakers and judges.
- Grading scale is 50 to 100, graded by 3-5 trained judges, based on constructive materials and responses.

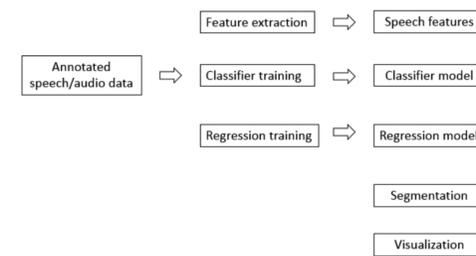
Competition	Gender	Total	Max	Min	Mean	Median	SD
WUDC	Male	247	91	58	79.80	80	4.59
	Female	103	89	66	79.24	79	4.35
EUDC	Male	277	87	66	77.17	77	4.17
	Female	148	90	63	76.92	76	5.05
HWS	Male	264	86	75	80.10	80	2.44
	Female	96	86	73	79.77	80	2.37

Speech	Description	Max	Min	Mean	SD
Duration	Entire speech duration (in seconds)	461	399	426.45	11.58
Score	Speech score (scale 50-100)	86	75	79.4	2.8
Word Count	Total number of words/speech	1910	1085	1600.42	178.23
Sentence Count	Total number of sentences/speech	101	27	56.52	17.68

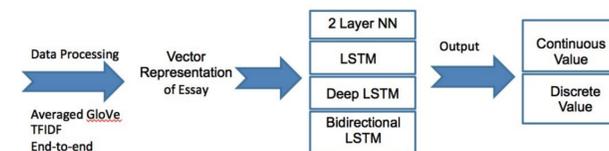
Challenges

- Comprehensiveness and interplay across determinants of speech quality, speaker and judge characteristics
- Non-linearity relationship between persuasiveness and speech elements
- Differences between "debate" and real-life persuasion

Analysis approach



- Use of NLTK and WordNet to extract various features of speech, including sentence classification, hedge cue detection, sequence labelling, text segmentation, discourse relation classification and semantic textual similarity
- Speech quality based on how well argument diagrams can be constructed [1] [2]
- Use of different neural networks, for instance, bidirectional LSTM-based model, similar to [3] and [4]



- Regression analysis of all identified speech characteristics and markers

$$S_{it} = \sum_{j=1}^n \alpha_j X_{ijt} + \sum_{j=1}^n \beta_j X_{ijt} \mathbb{I}_{F=1} + \sum_{j=1}^n \gamma_j X_{ijt} J_{ijt} + \sum_{j=1}^n \delta_j X_{ijt} O_{ijt} + \sum_{j=1}^n \eta_j X_{ijt} P_{ijt} + \epsilon_{it}$$

Ongoing research

- Finalize classification of raw speech and administrative data into the SQL system.
- Tests for prediction power of various verbal and non-verbal speech markers
- Counter-factual analysis on speaker style and scores (Oxaca-Blinder decomposition)
- Comparison of different deep learning techniques

References

- Andreas Peldszus and Manfred Stede. From argument diagrams to argumentation mining in texts: A survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 7(1):1–31, 2013.
- John Lawrence and Chris Reed. Combining argument mining techniques. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 127–136, 2015.
- H. Nguyen and L. Dery. Neural networks for automated essay grading.
- Mayer T. Petukhova, V., A. Malchanau, and H. Bunt. Virtual debate coach design: assessing multimodal argumentation performance. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 41–50. ACM, 2017.
- M. Lippi and P. Torroni. Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology (TOIT)*, 16(2):10, 2016.
- Matthew Gentzkow, Bryan T Kelly, and Matt Taddy. Text as data. Technical report, National Bureau of Economic Research, 2017.
- Naoki Egami, Christian J Fong, Justin Grimmer, Margaret E Roberts, and Brandon M Stewart. How to make causal inferences using texts. *arXiv preprint arXiv:1802.02163*, 2018.

