Abstract

People write everyday — articles, blogs, emails — with a purpose and an audience in mind. Politicians adapt their speeches to convince their audience, news media slant stories for their market, and teenagers on social media seek social status among their peers through their posts. Hence, language is *purposeful* and *strategic*. In this thesis, we introduce a framework for text analysis that make explicit the purposefulness of the author and develop methods that consider the interaction between the author, her text, and her audience's responses. We frame the authoring process as a decision theoretic problem — the observed *text* is the result of an author maximizing her utility.

We will explore this perspective by developing a set of novel statistical models that characterize authors' strategic behaviors through their utility functions. We consider three particular domains — political campaigns, the scientific community, and the judiciary — using our models and develop the necessary tools to evaluate our assumptions and hypotheses. In each of these domains, our models yield better response prediction accuracy and provide an interpretable means of investigating the underlying processes. Together, they exemplify our approach to text modeling and data exploration.

Throughout this thesis, we will illustrate how our models can be used as tools for in-depth exploration of text data and hypothesis generation.

Acknowledgements

This thesis marks an important milestone in my journey. A journey that began sixteen years ago with me typing `PRINT "Hello world!"` into the QBasic interpreter.[1] Along the way, I am very very fortunate to have the support of many people and it is my pleasure to thank those who made this possible.

First and foremost, it is difficult to overstate my gratitude to my advisor, Noah Smith, for countless hours of helpful and inspiring advice both in science and in life. Throughout graduate school, he provided invaluable guidance and patience without which I would have gotten lost. I could not have imagined a better advisor and mentor.

Besides my advisor, I would like to thank the rest of my thesis committee: Eduard Hovy, Jing Jiang, Daniel Neill, and Philip Resnik for their insightful comments from which this thesis has greatly benefited. I thank my undergraduate advisor at UIUC, Julia Hockenmaier, for introducing me to natural language processing and patiently teaching me the expectation-maximization algorithm.[2]

In my daily work, I have been blessed with the smartest, friendliest, and most amazing group of ARK-mates: Waleed Ammar, David Bamman, Dallas Card, Elizabeth Clark, Brendan O'Connor, Dipanjan Das, Jesse Dodge, Chris Dyer, Jeff Flanigan, Kevin Gimpel, Swapna Gottipati, Lingpeng Kong[3], Lucy Lin, Fei Liu, Nelson Liu, Kelvin Luu, Bill Mc-

---

[1]This statement merely sets a lower bound for my true age.

[2]Julia also encouraged us to attend Noah's talk at UIUC on text driven forecasting. I did and the rest is history.

[3]Special thanks for keeping my Mandarin ability in working condition and "making sure" that I have Chinese food more than once a week.

Dowell, George Mulcaire, Minghui Qiu, Rohan Ramanath, Bryan Routledge, Maarten Sap, Nathan Schneider, Swabha Swayamdipta, Sam Thomson, Tae Yano, Dani Yogatama, and Xiaote Zhu. There is no better group of people to be stuck in an ark with than *yinz*. Of course, I am also indebted to my many brilliant student colleagues at CMU and UW for providing a stimulating and fun environment in which to learn and grow.

I thank the Agency for Science, Technology and Research (A*STAR) for the scholarship that supported me over the last 9 years and gave me the incredible opportunity to meet with and learn from the greatest minds in the world. To my fellow friends from home who ended up with me in Pittsburgh over the years: Kawa Cheung, Weiling Neo, Junyang Ng, Jiaqi Tan, Chinyen Tee, Lionel Wong, and Yingying Wu, each of you made Pittsburgh my *kampung Singapura*.

Last but not least, I want to thank my family. Joy Sim, while in my heart you will always be my baby sis, I thank you for being my built-in best friend, for your advice that is wise beyond your years, for listening to my complaints over the last few years, and for being my sister. Mom and dad, thank you for unconditionally supporting me from afar, for raising me, for loving me, and for letting me know that there is always a home for me. 妈妈，爸爸，多谢你们多年的无私付出和养育之恩。这篇论文和我今日的成就一切因你们而有意义。

To my family I dedicate this thesis.