Abstract

We develop computational methods for analyzing historical documents to identify events of potential historical significance. Significant events are characterized by interactions between entities (e.g., countries, organizations, individuals) that deviate from typical interaction patterns. When studying historical events, historians and political scientists commonly read large quantities of text to construct an accurate picture of who, what, when, and where; a necessary precursor to answering the more nuanced question, Why? Our methods help historians identify possible events from the texts of historical communication. Specifically, we build on topic modeling to distinguish between topics that describe business-as-usual and topics that deviate from these patterns, where deviations are also indicated by particular entities interacting during particular periods of time. To demonstrate our methods, we analyze a corpus of 2 million State Department cables from 1973 to 1977. For example, we show that we are able to detect and characterize the Fall of Saigon.

This is joint work with Hanna Wallach and David Blei.

Bio

Allison Chaney is a PhD candidate in the Computer Science Department at Princeton University. Her research interests center around building scalable probabilistic models for large human-centered applications, ranging from recommendation systems to topic modeling of historical text. She received a B.A. in Computer Science and a B.S. in Engineering from Swarthmore College in 2008, and has worked for Pixar Animation Studios and the Yorba Foundation.

For more details, visit http://ajbc.io/.
Faculty Host: Tina Eliassi-Rad