

From Text to Feedback: Leveraging Data Mining to Build Educational Technologies

Danielle S. McNamara, Arizona State University

Over the past decade, our research teams have been using data mining techniques primarily for the purpose of analyzing text and writing corpora. These analyses emerged within three projects, Coh-Metrix, iSTART, and the Writing-Pal. Coh-Metrix is a text analysis tool that provides hundreds of linguistic and semantic indices on text. Within the context of the Coh-Metrix project, we have analyzed thousands of texts with the goal of better understanding the nature of text. Ultimately, our goal has been to provide educators with multidimensional information about the difficulty of text. iSTART is a reading strategy tutoring system that provides students with automated feedback on the quality of their self-explanations of text. The Writing Pal is a writing strategy tutoring system that provides strategy instruction to adolescent readers as well as practice writing essays with feedback. Within the context of the iSTART and Writing Pal projects, we have used Coh-Metrix and other text analysis tools to analyze self-explanations and essays. Our goal has been to better understand the linguistic features contributing to their quality as well as to develop and improve our automated scoring and feedback systems. All of these endeavors have called upon a variety of data mining techniques that serve to analyze the data and in turn drive the feedback algorithms that undergird the tutoring systems. This talk will describe how data mining supports the development of educational technologies at various levels.