Abstract

Email management has a fundamental role in modern work productivity. In this thesis we present evidence that email management can be potentially improved by the effective use of machine learning techniques to model different aspects of user intention. We initially propose a taxonomy of user intentions in terms of Speech Acts applied to email communication, or "email acts", and show that email act classification can be largely automated, potentially leading to better email prioritization and management.

We then describe how machine learning can be used to reduce the chances of costly email addressing errors. One type of costly error is an "email leak", i.e., mistakenly sending a message to an unintended recipient—a widespread problem that can severely harm individuals and corporations. Another type of addressing error is forgetting to add an intended collaborator as recipient, a likely source of costly misunderstandings and communication delays that can be potentially addressed with intelligent recipient recommendation. We propose several different approaches to address these problems, and show very positive experimental results in a large email collection. In addition, we describe a 4-week long user study based on the implementation of some of the proposed models in a popular email client (Mozilla Thunderbird). More than 15% of the human subjects reported that it prevented real email leaks, and more than 47% of them utilized recipient recommendations. Overall the study shows that recipient recommendation and email leak detection can be valuable additions to real email clients, with more than 80% of the subjects reporting that they would permanently use these models if a few interface/optimization changes were implemented.

Finally, we introduce a new robust rank learning algorithm to further improved recipient recommendation predictions. The algorithm is essentially a non-convex optimization procedure over a sigmoidal loss function, in which any linear baseline ranking model can be used as starting point. This new learning method provides substantial rank performance improvements on recipient recommendation tasks, outperforming all previously introduced models, including well-known state-of-the-art ranking algorithms.