What's Your Problem? Automatically Summarizing Scientific Research with Open Problems

Laidlaw, Cassidy

Scientists looking to familiarize themselves with a new field are often forced to read out-of-date books or spend hours poring over research papers, slowing the research process. The purpose of this project is to ameliorate this situation by creating an automatic system that can condense information from research articles into a short, coherent summary of open problems and research directions in a particular field. To create such a system, I used a variety of algorithms from machine learning and natural language processing and also devised a few of my own. The system picks phrases out of abstracts from an online database that refer to open problems and then groups them into clusters of similar problems. The most important clusters are summarized and form the sections of the generated review paper. I tested each step of this system by comparing it to a human's performance on the same task and in general achieved very positive results; in all steps I was able to improve accuracy over existing methods. In addition, I tested the system as a whole by sending generated review papers to researchers to evaluate. While the generated papers did not score as high as human written review papers, the scientists stated that with the scarceness of human written reviews, the system seemed like "a really useful tool." This system shows great promise and could help scientists in the future spend less time learning about new areas and more time researching them, moving science forward more quickly.

Awards Won:

Fourth Award of $500