CrowdSummary: Crowdsourced Abstractive Summary Generation with an Intelligent Interface

Shih-Wen Huang
Department of Computer Science
University of Illinois at Urbana-Champaign
shuang51@illinois.edu

1 Introduction

A well-written abstractive summary can save readers much time and effort in comprehending the information in an extensive document. However, abstractive summary generation is a task that is difficult for both humans and computers. For humans, it takes ample time for one to digest all of the information in a document and select important sentences that can be used to generate the abstract. On the other hand, computers can easily select important sentences in a document using statistical natural language processing algorithms. However, it is still not possible for computers to combine the important sentences in a document and generate a concise and coherent abstract, which is a task that is relatively easy for humans. In this project, we aim at building a crowd-computer hybrid system that first locates important sentences using statistical natural language processing and then recruits crowd workers to combine these sentences to generate high-quality abstractive summaries for users in a cheap and affordable manner.

2 Related Work

There are two branches of works that are highly related to the current project:

2.1 Crowd-powered systems:

The idea of incorporating crowds into a computation system was first proposed by Bernstein et al. [2]. They developed Soylent, a word processor that recruited online workers from Amazon Mechanical Turk (AMT) to help users edit their writing (e.g., text shortening and proofreading). Bigham et al. [3] built VizWiz, a tool that can provide nearly real-time answers to blind peoples visual queries by queuing crowd workers before the query. In addition, crowd-powered systems also are used in capturing important moments in video [1], image searching [9], answering database queries [6], and itinerary planning [10]. These examples have demonstrated that crowd-powered systems can be used to successfully solve problems that are difficult for digital computers.

2.2 Automatic summarization:

Many natural language processing algorithms have been proposed for automatic summarization [8]. In general, a summary can be classified into one of two categories: extractive summary and abstractive summary. Extractive summaries consist of sentences that are directly retrieved from the original text. On the other hand, abstractive summary can contain sentences that are newly written by the summarizer. Research has shown that state-of-the-art statistical approaches can be used to generate informative extractive summaries [7]. However, research also shows that it is still not possible for computers to compress or combine these sentences to generate high-quality abstractive summaries [8]. As a result, summaries generated by automatic summarization systems usually contain quite redundant and incoherent information.
3 System Design

In this project, we intend to design a system that utilizes both the strength of humans and computers to generate abstractive summaries. The system provides an interface for users to upload files or enter texts that they want to summarize. After getting the text from users, the system performs log-likelihood ratio test [5] to select important sentences from the text. Then, the system automatically generates Human Intelligence Tasks (HITs) on AMT to recruit crowd workers to generate the summaries. When a worker accepts the HIT, he/she is redirected to an interface that shows the text uploaded by the users with the important sentences highlighted. The machine-generated summary is also presented in a textbox. The worker can directly edit the text in the textbox to improve the machine-generated summary. (Figure 1) This intelligent interface allows workers to generate summaries without reading the whole document, which saves them much time and effort. In addition, the system also utilizes crowd workers to make machine-generated summaries more readable and concise.

4 Evaluation

To evaluate our proposed system, we will ask two editors from oDesk to rate the summaries generated by our system and the baseline systems. This expert evaluation has shown to be more reliable than crowd-evaluation [4]. Specifically, we will compare the quality of the summaries generated by our system to the machine-generated and expert summaries. We hope to see that the summaries generated by our system are much better than the machine-generated summaries and are similar in quality to the expert summary. Moreover, we will also compare the summary generation time between the CrowdSummary system with and without highlighting important sentences. This can help us see if the intelligent interface in our system helps crowd workers to generate summaries faster.

References


