ABSTRACT

Automatic Text Summarization (ATS) is an extensively researched problem with more recent approaches relying heavily on sophisticated deep learning techniques. The majority of state-of-the-art neural ATS systems are however designed to excel in summarizing open-domain news or Wikipedia articles, the characteristics of which, for example, *lead bias*, are not shared by documents from specialized domains such as law, finance, novels, etc. This motivates the need to design ***domain-specific*** ATS methods that can exploit the unique characteristics or structures of documents from respective genres to more accurately identify relevant information. Additionally, the existing summary evaluation protocol is dominated by ROUGE, that not only correlates weakly with human judgments but has also been shown of being incapable of addressing domain-specific evaluation requisites. This thesis investigates the unique modeling requirements and evaluation metrics for summarizing text in three different domains with varying shapes and content.

First, we look into the literature on ***financial*** document summarization and find out that annotated datasets are scarcely available to make any significant research progress in financial ATS. To address this, we present **ECTSum**, the first benchmark dataset for abstractive summarization of long financial documents. Interestingly, both the documents (company earnings call transcripts or ECTs), as well as the reference summaries (experts-written *Reuters* articles), are collected from the public domain, thereby making the dataset extensible and economically scalable. ECTs, apart from being lengthy and unstructured, contain a lot of facts, figures, and numbers, which if not properly captured in the summaries will render them useless for any downstream application. Therefore, summaries of financial disclosures in general, and ECTs in particular need to be ***factually consistent*** with the source

text. Accordingly, we propose **ECT-BPS**, an extract-then-paraphrase approach to produce factually consistent bullet point summaries from long ECTs. Extensive evaluation shows that *ECT-BPS* achieves state-of-the-art performance on *ECTSum* while outperforming different genres of competitive summarization baselines.

Factual consistency however does not always guarantee factual correctness. This becomes evident when we next tackle the task of summarizing ***disaster-related tweets***, where the source text itself may not be factually correct. Tweets posted during the unfolding of catastrophic events, despite carrying valuable updates, often contain unverified or rumorous information. To present a summarized account of the available data for aiding concerned stakeholders, we shift our focus from factual consistency to factual correctness or ***trustworthiness*** to derive actionable summaries. We present **MTLTS**, the first end-to-end solution that jointly determines the credibility and summary-worthiness of tweets to produce trustworthy summaries from large volumes of disaster-related microblogs in near real-time. Trustworthiness is evaluated using ***verified ratio***, a metric that tells us what proportion of summary tweets are non-rumorous (verified). Experiments on the *PHEME* dataset establish the better efficacy of *MTLTS* over existing disaster-specific summarizers.

The length of documents in the previous two works stands at two extremes, with ECTs ranging around 3000 tokens whereas tweets containing only around 280 characters. Next, we look into the task of multi-document opinion summarization of ***tourist*** reviews, where the document length lies somewhere in between. The diversity of opinions expressed in these reviews towards various ***aspects*** of a tourist destination, if succinctly summarized, can help prospective travelers and service providers to make informed decisions about their strategies. However, we argue that the notion of salience in these reviews largely depends on the aspects a user is interested in, and hence a generic summary may not satisfy the needs of all users.

Accordingly, we design an extractive summarizer to generate **personalized** aspect-based opinion summaries of tourist reviews by allowing readers to control several summary parameters, such as the length and users' aspects of interest. We evaluate ***aspect-coverage*** subjectively, given the personalized nature of the task.

The first and foremost step in opinion summarization is to identify the aspects mentioned in the reviews along with the opinions expressed towards them, a classic Aspect-based Sentiment Analysis (ABSA) task. While the aspect-discovery step in the previous work is unsupervised in nature, next we propose **PASTE**, a supervised end-to-end tagging-free solution for the task of Aspect Sentiment Triplet Extraction (ASTE) that deals with extracting (aspect, sentiment, opinion) triplets to provide the complete fine-grained sentiment information. Finally, we note that almost all existing research on ABSA or opinion summarization is limited to reviews of products or services. We however hypothesize and validate that these reviews have very different characteristics from tourist reviews. Driven by these observations, we propose **TACOS**, a new benchmark dataset to promote opinion mining research in tourism.

Keywords: domain-specific summarization, rouge limitations, evaluation protocol, finance, dataset, long document summarization, factual consistency evaluation, disaster response, trustworthy summarization, multi-task learning, rumor detection, tourism, multi-document summarization, opinion summarization, personalization, controllable summarization, aspect-based sentiment analysis, opinion mining.