



Review

Automatic Short Text Summarization Techniques in Social Media Platforms

Fahd A. Ghanem ^{1,2} , M. C. Padma ¹ and Ramez Alkhatib ^{3,*}

¹ Department of Computer Science & Engineering, PES College of Engineering, (Affiliated to University of Mysore), Mandya 571401, Karnataka, India; fahd.a.ghanem@gmail.com (F.A.G.); padmapes@gmail.com (M.C.P.)

² Department of Computer Science, College of Education-Zabid, Hodeidah University, Hodeidah P.O. Box 3114, Yemen

³ BMB Nord, Research Center Borstel, Parkallee 35, 23845 Borstel, Germany

* Correspondence: ralkhatib@fz-borstel.de; Tel.: +49-176-6520-0553

Abstract: The rapid expansion of social media platforms has resulted in an unprecedented surge of short text content being generated on a daily basis. Extracting valuable insights and patterns from this vast volume of textual data necessitates specialized techniques that can effectively condense information while preserving its core essence. In response to this challenge, automatic short text summarization (ASTS) techniques have emerged as a compelling solution, gaining significant importance in their development. This paper delves into the domain of summarizing short text on social media, exploring various types of short text and the associated challenges they present. It also investigates the approaches employed to generate concise and meaningful summaries. By providing a survey of the latest methods and potential avenues for future research, this paper contributes to the advancement of ASTS in the ever-evolving landscape of social media communication.

Keywords: text summarization; social media; evaluation metrics; abstractive summarization; extractive summarization; machine learning



Citation: Ghanem, F.A.; Padma, M.C.; Alkhatib, R. Automatic Short Text Summarization Techniques in Social Media Platforms. *Future Internet* **2023**, *15*, 311. <https://doi.org/10.3390/fi15090311>

Academic Editor: Massimo Cafaro

Received: 9 August 2023

Revised: 1 September 2023

Accepted: 6 September 2023

Published: 13 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background

In recent years, the landscape of communication has undergone a profound transformation with the ascendancy of social media platforms such as Twitter, Sina Weibo, and Facebook. These digital forums have swiftly emerged as primary conduits for real-time dissemination of news and updates, creating an unprecedented synergy between information and immediacy. Nevertheless, this surge in activity has engendered a proliferation of data on a staggering scale, posing a formidable challenge for users striving to unearth the precise morsels of information they seek. The sheer volume and the accelerated pace of data generation within these platforms have ushered in an era of information abundance that, paradoxically, obscures the very insights users aim to glean. At the nexus of this challenge lies the practice of manual text summarization—an art that entails extracting the vital essence from a document. This process, though imbued with the power to distill significance, exacts a toll on time and resources, especially when faced with substantial textual corpora. The indispensability of manual summarization reverberates across fields as diverse as journalism, research, and academia, where distilled and meticulous summaries stand as pillars of knowledge dissemination [1]. The tide of innovation, however, has bestowed upon us the mantle of automation. Text summarization, with its intricate web of algorithms and linguistic nuances, has burgeoned into a field of study that embodies the duality of precision and automation [2]. This evolution materializes in two principal strands: extractive techniques, adept at selecting and weaving key sentences into succinct

narratives, and abstractive techniques, proficient at birthing new sentences that encapsulate the quintessence of the original text. These methodologies, once honed, find homes in an array of applications: news condensation, social media abstraction, and scientific exposition [3]. Yet, amid these strides, a noteworthy chasm persists—a divide that underscores the prowess of human summarization over its automated counterparts. This divide emerges from the intricate choreography of comprehension, where the holistic significance of text passages dances to the tune of myriad factors. The subtleties of context, tone, and intent conspire to elevate human summarization above its algorithmic brethren. Thus emerges the clarion call—a resounding summons—to ascend the peaks of automation and engender algorithms of enhanced potency. This clarion call is resolute in its quest to bridge the gap and align the efficacy of automated short text summarization (ASTS) techniques with their human counterparts. Within the realm of social media and microblogs, ASTS unfurls as a grand enterprise akin to Multi-Document Summarization (MDS), a domain meticulously explored in the annals of information retrieval [4]. This paradigm equips users with the power to deftly distill cardinal insights from a multitude of documents, weaving them into concise narratives that encapsulate specific topics or events. Yet, this realm is riddled with the challenge of redundancy—an abundance of information that often conceals rather than reveals. The task of sieving through this labyrinth of extraneous data calls for strategic data reduction and adept summarization techniques that, when forged, yield succinct gems of insight. In the tapestry of social media and microblogging, the threads of summarization weave an invaluable tapestry. This tapestry emerges as concise summaries, capturing the essence of real-time dialogues surrounding specific topics or events. The significance of these summaries extends to individuals, enterprises, agencies, and organizations seeking to fathom the prevailing public sentiment—a virtual looking glass into the collective voice [5]. As we stand on the precipice of information saturation, the mandate for ASTS on microblogs is magnified—an indomitable necessity demanding the architecting of methodologies that distill the facets of intriguing topics or burgeoning events from the voluminous sea of social media [6].

1.2. Motivation and Objectives

A myriad of proposed automatic short text summarization (ASTS) techniques has surfaced. Nonetheless, a discernible void persists within the scholarly discourse—an absence that reverberates as the lack of a comprehensive survey and taxonomy to meticulously scrutinize and appraise the existing array of ASTS methods, complete with their inherent strengths and vulnerabilities. It is within this void that our research paper finds its impetus—a driving force rooted in the ambition to bridge this gap in comprehension. This clarion call emerges as the foundation of our survey paper, a scholarly endeavor meticulously crafted to illuminate and dissect the most recent and efficacious ASTS techniques within the dynamic context of social media. The primary tenet of this survey is to unfurl a comprehensive panorama—a sweeping overview that navigates through the labyrinthine landscape of ASTS techniques harmonized specifically for the intricate tapestry of social media platforms. To accomplish this, we have meticulously forged a robust classification framework that intricately weaves together pivotal factors, ranging from the platform itself to the techniques employed, the methodologies embraced, the domains addressed, linguistic nuances, summary types, datasets harnessed, and the evaluative metrics employed. We endeavor to unveil profound insights into the boundaries and latent potentials of these methods, propelling the readers into a realm of deeper comprehension. By distilling the distinct advantages and shortcomings inherent to various approaches, our overarching goal is to furnish researchers with a compass that steers them towards the salient research challenges resonating within the realm of short text summarization—particularly in the vibrant arena of social media. Ultimately, our work aspires to become a beacon that illuminates the contours of critical research issues, facilitating the inception of pioneering and effective ASTS methods. It is our fervent hope that this work, by streamlining the creation of high-caliber and precise summaries on social media platforms, will cascade

into ripples of progress that fuel advancements in the field and embolden the cauldron of innovation in ASTS techniques tailored to the ever-evolving and dynamic realm of social media communication.

1.3. Existing Surveys

This section offers an overview of previous surveys related to automatic short text summarization (ASTS) and compares them to our survey. Atefeh and Khreich conducted a survey on event detection-based methods for Twitter streams, categorizing the methods based on event type, detection task, and detection method while also outlining commonly used features [7]. However, their survey is restricted to event-based methods for tweet streams and lacks the comprehensiveness of our study.

Kawade and Pise conducted a survey that focused on summarization techniques for microblogs, specifically extracting situational information from disaster events on Twitter [8]. Ramachandran and Ramasubramanian also proposed a survey studying various techniques for event detection from microblogs, reviewing methods based on the utilized techniques, applications, and the detected elements (events or topics) [9].

Rudrapal et al. presented a survey on topic-based summarization techniques, primarily concentrating on current automatic evaluation methods for summarization techniques in social media [10]. Similarly, Hasan et al. surveyed event-detection methods for online Twitter data, classifying the methods based on shared common traits [11].

Additionally, Ermakova et al. presented a comprehensive survey on current metrics used for summary evaluation, highlighting their limitations and proposing an automatic approach for evaluating metrics without relying on human annotation [12].

The existing literature on ASTS has predominantly focused on Twitter, leaving other social media platforms with limited attention. However, this review paper aims to bridge this gap by presenting a comprehensive taxonomy of ASTS techniques across multiple platforms, including Twitter, Facebook, and Sina Weibo.

1.4. Structure of the Paper

The arrangement of this article unfolds in the subsequent manner: Section 2 provides a succinct delineation of types of Short Text on social media, succeeded by an exploration of social media datasets in Section 3. The Structure of ASTS is expounded upon in Section 4, while Section 5 expounds on diverse approaches to summarization. An exploration of the methods employed to evaluate ASTS is undertaken in Section 6. Moving on, Section 7 delves into an elaborate discourse on the existing ASTS techniques employed within the realm of social media. The ensuing sections, namely Sections 8 and 9, engage in a comprehensive analysis of discussions and challenges pertinent to the domain. Ultimately, Section 10 draws the curtain on the study, culminating the exploration.

2. Types of Short Text on Social Media

Understanding the different types of short text on social media is essential for developing effective summarization techniques that cater to the unique characteristics of each platform and user interaction. The following are some common types of short text found on various social media platforms.

2.1. Tweets

Tweets on Twitter are commonly used for sharing thoughts, news updates, opinions, and reactions. Twitter has emerged as a crucial platform for exploring diverse topics, ranging from global breaking news and sports to science, religion, emerging technologies [13], pandemics, and virus outbreaks [14]. However, the overwhelming volume of content on the Twitter timeline poses a complex challenge for users trying to stay updated in their areas of interest. As of 2 October 2022, Twitter witnesses an astonishing volume of activity, with approximately 6000 tweets sent per second, over 350,000 tweets per minute, 500 million

tweets per day, and nearly 200 billion tweets per year [15]. With 229 million daily activated users [16], the scale of Twitter's content is undeniably immense.

Registered Twitter users have the privilege of engaging with tweets through actions such as posting, liking, and retweeting, while unregistered users are limited to reading tweets. It is worth noting that tweets were initially constrained to 140 characters before expanding to 280 for non-CJK characters in November 2017 [17]. Although tweets can be informative in their raw form, the deluge of tweets can become overwhelming, making it challenging for users to efficiently digest and process the vast amount of information available.

2.2. Facebook Posts

While Facebook originated as a social networking platform, it has also evolved into a noteworthy channel for news and updates among its user base. According to data from Statista, Facebook boasts an excess of 2.9 billion monthly active users worldwide [18]. Facebook posts are a prominent type of short text commonly found on the Facebook social media platform [19]. Users can share their thoughts, experiences, updates, news articles, and various other content in the form of concise text snippets. These posts are typically limited to a certain number of characters, encouraging brevity in communication. Due to the character constraints for each post, users must be concise and to the point when expressing themselves. This brevity poses a challenge for readers to comprehend the complete context of a post, especially when navigating through a vast number of posts on their feed. ASTS techniques that can effectively summarize Facebook posts hold significant value for users and businesses alike. Such techniques can provide a concise and coherent overview of a user's Facebook feed, enabling efficient content consumption and knowledge extraction. By condensing lengthy posts into key insights and important information, ASTS facilitates quick comprehension and enhanced user engagement on the platform. Furthermore, summarizing Facebook posts can be valuable for social media monitoring, sentiment analysis, and opinion mining. By generating accurate and informative summaries, ASTS techniques can help organizations track public opinions, monitor brand sentiment, and identify emerging trends from the deluge of Facebook posts.

2.3. Instagram Captions

On the Instagram social media platform [19], Instagram captions play a significant role as a form of short text commonly encountered. When users share photos or videos, they often complement their visual content with brief textual descriptions known as captions. These captions serve as a means for users to provide context, express emotions, share stories, or convey messages that enhance the impact of their visual posts. Due to Instagram's visual-centric nature, captions are typically limited to a specific number of characters, encouraging users to be concise and impactful in their expressions. The brevity of captions is crucial for capturing the attention of followers and effectively conveying the intended message. Automatic short text summarization (ASTS) techniques that can adeptly summarize Instagram captions offer practical applications. For users, such techniques enable efficient browsing and comprehension of multiple posts in their feeds. For businesses and content creators, caption summarization facilitates content curation, sentiment analysis, and audience engagement, leading to more meaningful interactions.

2.4. WhatsApp Messages

WhatsApp messages represent a fundamental type of short text frequently encountered within the messaging application. Users utilize WhatsApp to exchange short text messages privately or within group chats, enabling direct communication for personal conversations, professional discussions, information sharing, and expressing emotions. The instant messaging nature of WhatsApp encourages brevity in messages, prompting users to convey their thoughts concisely. Character limitations often lead to the use of abbreviations, emojis, and other text-based expressions to communicate effectively within the restricted space. The Automatic short text summarization (ASTS) techniques for What-

sApp messages pose a distinct challenge. Summarizing these short text snippets involves capturing the conversation’s core meaning while preserving essential details and context. Efficient summarization of WhatsApp messages empowers users to swiftly grasp the conversation’s main points, streamlining communication and enhancing their overall messaging experience.

2.5. YouTube Comments

YouTube comments hold considerable importance as a widespread form of short text on the video-sharing platform. Users utilize comments to express thoughts, provide feedback, share opinions, and engage in discussions with other viewers and content creators. Due to the platform’s interactive nature, YouTube comments are usually brief and focused, enabling users to convey their sentiments and ideas succinctly. The character limit for comments encourages users to be concise, often resulting in the use of abbreviations, emojis, and other shorthand expressions. The effective summarization of YouTube comments yields valuable benefits for both viewers and content creators.

2.6. Sina Weibo

Sina Weibo, often informally referred to as the “Twitter of China,” holds a pivotal position in the arena of instant information sharing. Functioning as a potent microblogging platform, Sina Weibo has emerged as a vital conduit for accessing the latest trends, news, and updates, primarily catering to the Chinese-speaking populace. Providing users with a dynamic arena for sharing thoughts, opinions, and news snippets via succinct posts, Sina Weibo captures the essence of swift information exchange. The platform’s significance is underscored by its substantial user engagement, with Statista’s 2023 data indicating an impressive monthly active user count exceeding 584 million. This robust user participation solidifies Sina Weibo’s stature as a widely embraced social media platform within the Chinese-speaking demographic. Central to Sina Weibo’s identity is its role as a primary wellspring of real-time news. Its capability to expeditiously disseminate updates regarding noteworthy events is evident, spanning a spectrum from political milestones and natural calamities to the latest developments in entertainment and cultural shifts.

3. Social Media Datasets

The utilization of social media datasets is of great significance in the development of automatic short text summarization (ASTS) algorithms. These datasets usually encompass substantial amounts of concise text data, including tweets, status updates, and comments, which serve as valuable resources for training and evaluating summarization models. In Table 1, we will provide a concise overview of the most widely used and publicly accessible datasets that are commonly employed for evaluating the efficacy of ASTS techniques in social media.

Table 1. Popular datasets utilized for evaluating ASTS in social media.

Name	Domain	Source/Language	Size	Ref.
A Large-Scale Dataset of Twitter Chatter about Online Learning during the Current COVID-19 Omicron Wave	Education	Twitter—English	52,984 tweet IDs	https://zenodo.org/record/6837118 [14] (accessed on 7 August 2023)
MonkeyPox2022Tweets	Healthcare	Twitter—English	571,831 tweets	https://doi.org/10.7910/DVN/CR7T5E [20] (accessed on 7 August 2023)
ANTIvax	COVID-19 vaccine	Twitter—Arabic and English	15 Million tweets	https://github.com/SakibShahriar95/ANTIvax [21] (accessed on 7 August 2023)
ReSoCIO	Disaster events	Twitter—French	4617 tweets	https://doi.org/10.5281/zenodo.7767294 [22] (accessed on 7 August 2023)
SNAP 2009	General	Twitter—English	467 million tweets—from 20 million users	https://snap.stanford.edu/data/twitter7.html (accessed on 7 August 2023)

Table 1. Cont.

Name	Domain	Source/Language	Size	Ref.
Twitter data 2012	Popular topics	Twitter—Chinese	50 million tweets	https://wiki.illinois.edu/wiki/display/forward/Dataset-UDI-TwitterCrawl-Aug2012 (accessed on 7 August 2023)
RepLab 2013	Automotive, banking, universities, and music/artists	Twitter—English and Spanish	more than 142,000 tweets	http://nlp.uned.es/replab2013/replab2013-dataset.tar.gz (accessed on 7 August 2023)
Andrei Olariu 2014	Important events	Twitter—English	3.4 million tweets	https://github.com/andreiolariu/online-summarizer (accessed on 7 August 2023)
UkFlood 2015	Disaster events	Twitter—English	2069 tweets	http://cse.iitkgp.ac.in/~krudra/disaster_dataset.html (accessed on 7 August 2023)
Sandyhook 2015	Disaster events	Twitter—English	2080 tweets	http://cse.iitkgp.ac.in/~krudra/disaster_dataset.html (accessed on 7 August 2023)
Hblast 2015	Disaster events	Twitter—English	1413 tweets	http://cse.iitkgp.ac.in/~krudra/disaster_dataset.html (accessed on 7 August 2023)
Hagupit 2015	Disaster events	Twitter—English	1461 tweets	http://cse.iitkgp.ac.in/~krudra/disaster_dataset.html (accessed on 7 August 2023)
Nepal Earthquake (NEQuake) 2015	Disaster events	Twitter—English	1.87 million messages	http://cse.iitkgp.ac.in/~krudra/nepal_earthquake.html (accessed on 7 August 2023)
LCSTS dataset 2016	Generic	Sina Weibo—Chinese	2,400,591 text–summary pairs, 10,666 text–summary pairs and 1106	https://paperswithcode.com/dataset/lcsts (accessed on 7 August 2023)
TwitterAAE dataset 2016		Twitter—English	60 million tweets	http://slanglab.cs.umass.edu/TwitterAAE (accessed on 7 August 2023)
CrowdFlower AI Gender dataset 2016	Gender classifier data	Twitter—English	20,000 posts with user-information	https://www.crowdfLOWER.com/data-for-everyone/ (accessed on 7 August 2023)
USPresidentialDebates 2016	Us election	Twitter—English	11 million tweets	https://www.kaggle.com/mrisdal/2016-us-presidential-debates (accessed on 7 August 2023)
Twitter events 2012–2016	different real-world events	Twitter—English	147 million tweets	https://figshare.com/articles/Twitter_event_datasets_2012-2016_/5100460 (accessed on 7 August 2023)
Customer Support on Twitter Dataset 2017	Product complaints (business)	Twitter—English	3 million tweets	https://www.kaggle.com/thoughtvector/customer-support-on-twitter (accessed on 7 August 2023)
TREC Dynamic Domain Track 2017	Ebola outbreak	Dynamic website—Chinese	90,823 tweets	http://trec-dd.org/ (accessed on 7 August 2023)
Epidemics: 1. Ebola 2. MERS 2018	Epidemics	Twitter—English	5.08 million messages, 0.215 million messages	http://cse.iitkgp.ac.in/~krudra/epidemic.html (accessed on 7 August 2023)
TSSuBERT 2018	Generic event	Twitter—English	150 million tweets	https://github.com/JoeBloggsIR/TSSuBERT (accessed on 7 August 2023)
XMUDM/IAEA 2020	Generic real-time event	Twitter—English	9,154,025 tweets	https://github.com/XMUDM/IAEA (accessed on 7 August 2023)

4. ASTS Structure

The structure of an automatic short text summarization (ASTS) system (Figure 1) can be organized into several key components, each playing a crucial role in the summarization process. It is important to note that the structure and components of an ASTS system may vary based on the specific requirements, domain, and intended use [23]. Advances in natural language processing (NLP) and machine learning techniques continue to impact the development of ASTS systems, resulting in more sophisticated and efficient summarization techniques [24]. Here is a typical structure for an ASTS system:

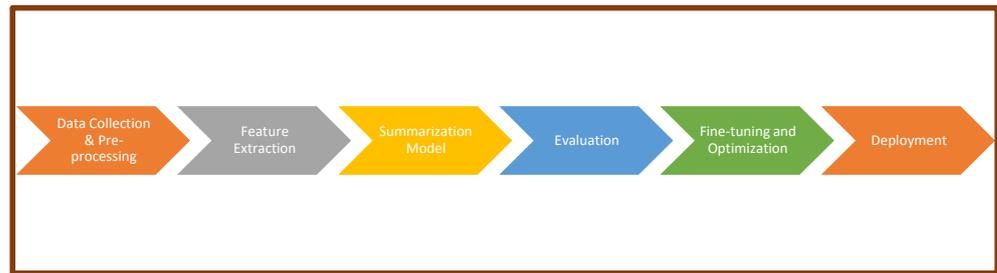


Figure 1. The structure of an automatic short text summarization system.

1. Data collection and pre-processing:
 - Data collection: Gather short text data from social media platforms or other sources, depending on the application.
 - Data pre-processing: Clean and preprocess the collected data, which includes text cleaning, tokenization, stopword removal, lemmatization, stemming, and sentence segmentation.
2. Feature extraction:
 - Extract relevant features from the preprocessed data to represent the content for the summarization model. Features can include TF-IDF vectors, word embeddings, or contextual embeddings generated using pre-trained language models like BERT or GPT.
3. Summarization model:
 - Extractive summarization: For extractive summarization, use techniques such as graph-based algorithms (e.g., TextRank) [25], attention mechanisms (e.g., Transformer-based models), or neural network-based methods (e.g., LSTM) to identify and select important sentences or phrases from the input text to form the summary.
 - Abstractive summarization: For abstractive summarization, employ neural network models, such as sequence-to-sequence architectures (e.g., LSTM with attention, Transformer-based models like BART), to generate new sentences that convey the essential information not necessarily present in the original text.
4. Evaluation:
 - Measure the quality and performance of the summarization model using evaluation metrics such as ROUGE (Recall-Oriented Understudy for Gisting Evaluation), BLEU (Bilingual Evaluation Understudy), or other domain-specific metrics. These metrics compare the generated summaries with human-written summaries or ground truth summaries.
5. Fine-tuning and optimization:
 - Fine-tune the summarization model on a domain-specific or task-specific dataset to improve its performance and adapt it to the target domain.
 - Optimize hyperparameters and model architecture through experimentation and tuning.
6. Deployment:
 - Integrate the trained ASTS model into a larger application or platform to provide automatic short text summarization functionality to users.
 - Monitor and maintain the system in production to ensure its continued accuracy and effectiveness.

5. Approaches to Summarization

Three prominent pathways emerge for short text summarization: extractive summarization, abstractive summarization, and hybrid approaches that endeavor to amalgamate

the strengths of both methods, culminating in summaries of heightened equilibrium and sophistication [7]. The choice of the summarization approach depends on various factors, including the nature of the short text data [8], the desired level of summary coherence, the intended use of the summary, and the complexity of the summarization task. Recent advances in NLP and deep learning have led to significant improvements in both extractive and abstractive summarization techniques, making short text summarization a vibrant and evolving research area [11,12]. Let us explore each approach in more detail.

5.1. Extractive Summarization

Extractive summarization involves the selection and extraction of important sentences or phrases directly from the original short text to create the summary. In this approach, the chosen sentences are taken as they appear in the original text, without any modifications. The selection process relies on ranking the sentences based on their importance and relevance to the main content of the text. One of the advantages of extractive summarization is that it preserves the exact wording of the sentences from the original text, ensuring grammatical correctness and maintaining context accuracy. Additionally, it avoids potential issues that may arise from generating new sentences that might not be contextually appropriate or could introduce factual inaccuracies. However, extractive summarization has some limitations. It may result in disjointed or redundant summaries, as the selected sentences might not flow together coherently. Moreover, it can struggle with handling overlapping information in the text, leading to certain critical content being repeated in the summary. These challenges highlight the need for careful sentence selection to create concise and coherent summaries that effectively represent the main content of the original text. Extractive summarization is widely used in various NLP applications, including automatic short text summarization (ASTS). Here are some common extractive approaches in short text summarization:

- Frequency-based methods:
Frequency-based methods assign importance scores to sentences based on the frequency of important words or phrases in the text. Sentences with a higher frequency of essential terms are considered more important and are selected for the summary.
- TF-IDF (Term Frequency–Inverse Document Frequency):
TF-IDF is a widely used technique that evaluates the importance of a word in a document relative to a corpus of documents. Sentences containing important terms with high TF-IDF scores are selected for the summary.
- TextRank algorithm:
TextRank is a graph-based ranking algorithm inspired by PageRank. It treats sentences as nodes in a graph and computes the importance of each sentence based on the similarity and co-occurrence of sentences. High-scoring sentences are chosen for the summary [25].
- Sentence embeddings:
Sentence embeddings represent sentences as dense vectors in a high-dimensional space, capturing semantic information. Similarity measures between sentences are used to rank sentences, and the most similar ones are included in the summary.
- Supervised machine learning:
In supervised approaches, models are trained on annotated data with sentence-level labels indicating whether a sentence should be included in the summary. The model then predicts the importance of sentences in new short texts and selects the most important ones for the summary [26].

5.2. Abstractive Summarization

Abstractive summarization involves the generation of new sentences that capture the essential information from the original text. Unlike extractive summarization, this approach can rephrase and paraphrase the content, allowing for more coherent and concise summaries. One of the significant advantages of abstractive summarization is its ability to produce more fluent and cohesive summaries by synthesizing information in a manner similar to human language. Furthermore, it excels at handling overlapping information,

as it can consolidate similar ideas into concise statements. However, abstractive summarization does have some limitations. It requires a deeper understanding of the input text to generate contextually accurate and meaningful summaries. Additionally, it is prone to introducing factual inaccuracies or misinterpretations, especially when dealing with out-of-vocabulary words or rare entities. To mitigate these challenges, ongoing research focuses on improving the context comprehension and semantic understanding capabilities of abstractive summarization models. With advancements in NLP and deep learning, efforts are being made to enhance the accuracy and coherence of abstractive summaries while maintaining their conciseness and relevance to the original text. Here are some common abstractive approaches in short text summarization:

- **Sequence-to-sequence (Seq2Seq) models:**
Seq2Seq models, based on recurrent neural networks (RNNs) or transformers, are widely used in abstractive summarization. These models encode the input short text into a fixed-size vector and then decode it to generate a summary. The decoder generates new sentences by predicting the next word based on the context learned from the encoder [27].
- **Transformer-based models:**
Transformers, particularly pre-trained models like GPT-2 and BERT, have shown promising results in abstractive summarization. These models use self-attention mechanisms to capture contextual relationships between words and generate coherent and contextually appropriate summaries.
- **Reinforcement learning:**
Abstractive summarization can also be approached as a reinforcement learning problem. The model generates candidate summaries, and a reward mechanism is used to evaluate their quality. The model is then fine-tuned using policy gradients to optimize the summary generation process.
- **Pointer-Generator networks:**
Pointer-Generator networks combine extractive and abstractive methods. These models have the ability to copy words from the input text (extractive) while also generating new words (abstractive) to create the summary. This approach helps to handle out-of-vocabulary words and maintain the factual accuracy of the summary.

5.3. Hybrid Approaches:

Hybrid approaches in short text summarization combine elements from both extractive and abstractive summarization techniques. These approaches aim to leverage the strengths of both methods to achieve more accurate and coherent summaries. Here are some common hybrid approaches used in short text summarization:

- **Extract-then-abstract:**
This approach involves first applying an extractive summarization technique to select important sentences or phrases from the original short text. Then, an abstractive summarization model is used to rephrase and refine the extracted content into a more coherent and concise summary. This way, the final summary benefits from the factual accuracy of extractive methods and the fluency of abstractive methods.
- **Abstractive pre-processing:**
In this approach, the original short text undergoes abstractive pre-processing before being fed into an extractive summarization model. The pre-processing step involves paraphrasing and rephrasing the text to enhance its coherence and readability. The extractive model then selects sentences from the pre-processed text to form the final summary. By improving the input text's quality, this approach aims to generate more coherent and informative summaries.
- **Extract-then-cluster-then-abstract:**
This approach combines extractive summarization with clustering techniques. First, important sentences are extracted from the original text. Then, these sentences are clustered based on their similarity, grouping similar content together. Finally, an abstractive summa-

rization model is applied to each cluster to generate concise and coherent summaries for each group of related sentences. This approach helps in handling overlapping information and ensures that important content is not repeated in the final summary.

- Reinforcement learning:
Hybrid approaches can also involve the use of reinforcement learning to combine extractive and abstractive summarization [28]. The model learns to select important sentences through extractive summarization and generates new sentences using abstractive techniques. Reinforcement learning is used to optimize the summary's quality by rewarding the model for generating accurate and informative summaries while penalizing for errors.
Hybrid approaches in short text summarization aim to address the limitations of individual extractive and abstractive methods and create more effective and informative summaries. These approaches require careful design and optimization to strike the right balance between factual accuracy and fluency while maintaining the summary's coherence and relevance to the original text. Ongoing research in this area continues to explore innovative ways to combine these techniques and enhance the overall performance of short text summarization models.

6. Evaluation Metrics for Short Text Summarization

Evaluation is a crucial aspect of short text summarization, as it allows researchers and practitioners to assess the quality and effectiveness of the generated summaries. Various evaluation metrics have been proposed to measure the performance of summarization models. In this section, we discuss some of the commonly used evaluation metrics for short text summarization [29,30]:

1. ROUGE (Recall-Oriented Understudy for Gisting Evaluation):
ROUGE is one of the most widely used evaluation metrics for text summarization. It measures the overlap between the n-grams (unigrams, bigrams, trigrams, etc.) of the generated summary and the reference summary. The ROUGE score includes metrics such as ROUGE-1 (unigram overlap), ROUGE-2 (bigram overlap), and ROUGE-L (longest common subsequence) [31].
2. BLEU (Bilingual Evaluation Understudy):
Originally developed for machine translation, BLEU has been adapted for text summarization evaluation. BLEU compares the n-grams in the generated summary with those in the reference summary and calculates precision scores. It is especially useful for measuring the quality of abstractive summaries.
3. METEOR (Metric for Evaluation of Translation with Explicit Ordering):
METEOR evaluates the quality of summaries by considering exact word matches as well as paraphrased and stemmed matches. It incorporates additional features like stemming, synonym matching, and word order similarity to assess the overall quality of the generated summaries.
4. CIDEr (Consensus-based Image Description Evaluation):
Originally developed for image captioning, CIDEr has been adjusted to evaluate text summarization. It computes a consensus-based similarity score by comparing n-gram overlaps between the generated and reference summaries while also considering the diversity and quality of the generated summaries.
5. ROUGE-WE (ROUGE with word embeddings):
ROUGE-WE enhances the traditional ROUGE metrics by incorporating word embeddings to assess the semantic similarity between the generated and reference summaries. It provides a more nuanced evaluation of the generated summaries' semantic quality.
6. BertScore:
BertScore utilizes contextual embeddings from pre-trained models like BERT to measure the similarity between the generated and reference summaries. It takes into account the context of the words in the summaries, resulting in a more accurate evaluation [32].

7. Human evaluation:

In addition to automated metrics, human evaluation is essential for assessing the quality of summarization models. Human evaluators rank or rate the generated summaries based on their coherence, relevance, and informativeness compared to the reference summaries.

Each evaluation metric has its strengths and limitations. ROUGE and BLEU are widely used due to their simplicity and ease of implementation. However, they may not fully capture the semantic quality of abstractive summaries. On the other hand, newer metrics like BertScore and ROUGE-WE address some of these limitations by considering semantic similarity. A comprehensive evaluation of short text summarization models should consider multiple metrics, including both automated and human evaluations, to obtain a holistic understanding of their performance. Each metric contributes valuable insights into different aspects of summarization quality, and a combination of these metrics can provide a more reliable assessment of the summarization models.

7. ASTS Techniques for Social Media Platforms

The literature review section initiates by providing a comprehensive overview of research studies focusing on automatic short text summarization (ASTS) techniques tailored specifically for social media platforms. Within this review, we delve into the two primary categories of ASTS techniques: extractive and abstractive summarization. Additionally, we explore novel hybrid approaches that cleverly combine both extractive and abstractive methods, harnessing their respective strengths synergistically. To facilitate a better understanding and comparison, Table 2 offers a comprehensive summary with a comparative analysis of ASTS Techniques in the realm of social media. This table serves as a valuable resource for quickly assessing key metrics and outcomes, empowering researchers and practitioners to make informed decisions while selecting the most suitable ASTS techniques for their social media applications.

In their recent work, [33] introduced an innovative method that utilizes advanced tools like a state-of-the-art open-sourced search engine and a large language model to generate accurate and comprehensive summaries. By gathering information from social media and online news sources, their approach incorporates cutting-edge search strategies and leverages GPT-3 in a one-shot setting for summarizing crisis-related content. The method's evaluation took place on the TREC CrisisFACTS challenge dataset [34], which encompasses data from diverse platforms such as Twitter, Facebook, Reddit, and online news sources. The assessment involved the use of both automatic summarization metrics (ROUGE-2 and BERTScore) and manual evaluations conducted by the challenge organizers. Despite having a high redundancy ratio, the manual evaluations revealed the generated summaries to be remarkably comprehensive. Notably, the best automatic evaluation results showcased an impressive BERTScore of 0.4591 and a ROUGE-2 score of 0.0581 on the ICS 209 dataset.

In their paper, Alabid and Naseer addressed the challenge of summarizing comments related to the COVID-19 pandemic from a collection of Twitter comments [35]. Their approach involved several steps to achieve summarization. Firstly, they initiated the summarization task by clustering topics using the latent Dirichlet allocation (LDA) method and K-means clustering. This allowed them to identify relevant topics in the comments. To handle topic overlaps between multiple comments, they employed K-mean clustering to group similar phrases together. Next, they implemented the TF-IDF ranking algorithm based on context similarity to select important sentences for creating the summary. This ensured that the most relevant and significant information was included in the summarization. To determine the optimal number of topics for training LDA, they utilized the c_v topic coherence score. Moreover, they employed the Silhouette metric to measure the convergence range of data points around the center points in each cluster. For evaluation purposes, the researchers collected around 100,000 tweets between 1 March and 31 July 2022, based on predefined conditions and parameters. To assess the performance of their

proposed method, they employed the ROUGE measure for various numbers of LDA topics, obtaining the following scores:

- ROUGE-1: (F1 score: 0.06 to 0.26, recall: 0.1 to 0.21, and precision: 0.05 to 0.19)
- ROUGE-2: (F1 score: 0.04 to 0.23, recall: 0.01 to 0.19, and precision: 0.03 to 0.17)
- ROUGE-S4: (F1 score: 0.04 to 0.21, recall: 0.01 to 0.18, and precision: 0.03 to 0.15)
- ROUGE-SU4: (F1 score: 0.29 to 0.48, recall: 0.04 to 0.07, and precision: 0.07 to 0.12)

Taghandiki et al. conducted a study in which they designed and implemented a crawler to extract popular text posts from Instagram [36]. The extracted posts underwent appropriate pre-processing, and a combination of extraction and abstraction algorithms was applied to demonstrate the usefulness of each abstraction algorithm. They evaluated their proposed approach using a dataset of text posts obtained from Instagram through popular hashtags using the developed crawler. The study analyzed 820 popular text posts from Instagram, revealing an accuracy rate of 80% and a precision rate of 75% for the proposed system. The results of their automatic text summarization system showed the following values: accuracy (75%), completeness (82%), precision (80%), harmonic mean (81%), and error rates (25%).

In their work, Murshed et al. introduced a hybrid deep learning model named DEA-RNN, designed specifically to detect contextual bullying on the Twitter social media network [37]. They conducted a comprehensive evaluation using a dataset of 10,000 tweets and compared the performance of DEA-RNN to Bi-directional long short-term memory (Bi-LSTM), RNN, SVM, Multinomial Naive Bayes (MNB), and Random Forests (RF). DEA-RNN achieved an average accuracy of 90.45%, precision of 89.52%, recall of 88.98%, F1 score of 89.25%, and specificity of 90.94%.

In Garg et al.'s study, the researchers introduced OntoRealSumm, a novel approach capable of generating real-time tweet summaries for disasters with minimal human intervention [38]. OntoRealSumm tackles multiple challenges through a three-phase process. Experimental analysis revealed that OntoRealSumm outperforms existing research works, resulting in a 6% to 42% increase in ROUGE-N F1 scores. Furthermore, when evaluated on a dataset centered around the Harda Twin Train Derailment in India, where 31 people lost their lives and 100 were injured, OntoRealSumm demonstrated the best performance. It achieved F1 scores of 0.58, 0.29, and 0.31 for ROUGE-1, ROUGE-2, and ROUGE-L, respectively.

Ref. [39] proposed a multi-task framework (MTLTS) to obtain trustworthy summaries. They treat tweet summarization as a supervised task, enabling the automatic learning of summary-worthy features and better generalization across domains. MTLTS achieves a ROUGE1-F1 of 0.501 on the PHEME dataset.

Panchendrarajan et al. recently introduced a weekly supervised learning technique for coherent topic extraction and emotional reactions related to an event [40]. Their method provides representative microblogs and emotion distributions over time, summarizing the event effectively. The method obtains approximately a ROUGE-L of 0.4241, ROUGE-1 of 0.4787, ROUGE-2 of 0.2391, and ROUGE-3 of 0.1652 on the Mexico Earthquake dataset.

The study of Li and Zhang delved into two extractive techniques aimed at summarizing events on Twitter [41]. The first approach capitalizes on the semantic characteristics of event-related terms, scoring tweets based on computed semantic values. The second method employs a graph convolutional network constructed from a tweet relation graph, enabling the generation of hidden features to estimate tweet salience. Consequently, the most salient tweets are chosen to form the event summary. During their experiments, the researchers obtained results with ROUGE and BLEU values of 0.647 and 0.490, respectively, using a dataset of 300 evaluation events.

In the study by Saini et al., the authors introduced three versions of an innovative approach for microblog summarization, named MOOTS1, MOOTS2, and MOOTS3 [42]. These versions were developed using the principles of multi-objective optimization. The underlying optimization strategy employed was the multi-objective binary differential evolution (MOBDE). The optimization process incorporated two measures to assess tweet

similarity/dissimilarity: word mover distance and BM25. To assess the performance of the proposed approaches, the researchers evaluated them on four datasets focusing on disaster events, where only relevant tweets were included. The experimental results showcased the effectiveness of their best-proposed approach, MOOTS3, achieving a ROUGE-2 score of 0.3418 and a ROUGE-L score of 0.5009, respectively.

In another study, Saini et al. introduced two fusion models, namely the self-organizing map and granular self-organizing map (SOM+GSOM), designed to tackle the summarization challenges of social media and microblogs through clustering [43]. The models effectively retrieved the most relevant tweets from a collection, significantly reducing the number of tweets using SOM and further extracting the most pertinent tweets with GSOM. To assess their results, the researchers evaluated the models on four datasets related to disaster events: (a) the Sandy Hook elementary school shooting in the USA, (b) Uttarakhand's floods, (c) Typhoon Hangupit in the Philippines, and (d) bomb blasts in Hyderabad. The SOM+GSOM architecture demonstrated impressive performance with ROUGE-L and ROUGE-2 scores of 0.5199 and 0.3661, respectively.

In their work, Saini et al. presented an unsupervised classification method based on a neural network to identify significant tweets based on their relevance [44]. The technique utilizes the self-organizing map (SOM) and the granular self-organizing map (GSOM) to reduce the number of tweets and group them into clusters. To determine the best GSOM parameter combinations, they employed an evolutionary optimization technique. The researchers assessed the effectiveness of their proposed approach on four microblog datasets related to disasters. The results obtained for their method were 0.3458 for ROUGE-2 and 0.5000 for ROUGE-L.

Goyal et al. introduced an approach called Mythos, which serves to detect events and subevents within an event while generating abstract summaries and storylines to offer diverse perspectives on an event [45]. The Mythos system comprises three modules: an online incremental clustering algorithm is employed to identify small-scale events represented as small clusters, its event hierarchy generator generates larger events in the form of hierarchies, providing a broader context, and its summarization module uses a long short-term memory (LSTM)-based learning model to generate summaries at various levels, ranging from the most abstracted to the most detailed. These summaries at different levels are then used to create a comprehensive storyline for each event. To evaluate Mythos, the researchers conducted experimental analyses on various Twitter datasets from different domains, including the FA Cup final, Super Tuesday (ST) Primaries, US Elections (USEs), and SB. The comparison was made against existing approaches for event detection and summarization. The results of the evaluation showed that Mythos achieved the best F-measure score of 0.5644 on the USE dataset, highlighting its superior performance in event detection and summarization compared to other known approaches.

Wang and Ren introduced a novel approach called summary-aware attention for abstractive summarization of social media short texts [26]. They employed source-to-summary attention to calculate the attended summary vectors. Furthermore, they utilized dual or fusion methods to generate summary-aware attention weights, taking into account both the source text and the generated summary. The evaluation of their model was conducted on the Large Scale Chinese Short Text Summarization (LCSTS) dataset, comprising more than 2 million real text–summary pairs gathered from the Chinese microblogging website Sina Weibo. The experimental results demonstrated that their model surpassed Seq2Seq baselines. Moreover, the generated summaries effectively captured the essential ideas of the source text in a more coherent manner. The best performance achieved by their model was impressive, with a ROUGE-1 score of 40.3, a ROUGE-2 score of 27.8, and a ROUGE-L score of 38.4.

In their paper, Lin et al. introduced a real-time event summarization framework called IA EA (integrity-aware extractive–abstractive real-time event summarization) [46]. The core concept behind IA EA is the incorporation of an inconsistency detection module into a unified extractive–abstractive framework. During each update, essential new tweets

are initially extracted using the extractive module. The extraction process is then refined by explicitly detecting inconsistencies between the new tweets and previous summaries. The extractive module also captures sentence-level attention, which is later utilized by the abstractive module for acquiring word-level attention. This word-level attention is instrumental in rephrasing words to form a coherent summary. To validate their approach, the researchers carried out extensive experiments using real-world datasets. To minimize the labor involved in generating sufficient training data, they introduced automatic labeling steps, which were empirically validated for their effectiveness. Through these experiments, the IAEA framework achieved the following average ROUGE and BLEU scores: approximately 37 for ROUGE-1, 13 for ROUGE-2, 33 for ROUGE-L, 24 for BLEU-1, and 6 for BLEU-2. Additionally, the framework demonstrated an accuracy of 0.880, precision of 0.713, recall of 0.710, F1 score of 0.705, and AUC of 0.814. These results showcase the effectiveness and performance of the IAEA approach in real-time event summarization.

In their research, Ali et al. proposed a novel summarization method that revolves around sentiment topical aspect analysis, allowing for the automatic generation of comprehensive textual summaries for Twitter topics [47]. The method involves exploring the inherent sentiments associated with microblog posts and analyzing the topical aspects to enhance the summarization performance. The summarization process consists of three key stages. Initially, they adopted an approach to semantically enrich the microblog posts, extracting the sentiment and topic information linked to each post. Subsequently, for each topic-sentiment cluster of microblog posts, they constructed a Word Graph to identify different topical aspects within each specific cluster. Finally, they generated a comprehensive summary for each topic by applying various state-of-the-art summarization algorithms on each topical aspect of the Word Graph. To evaluate the effectiveness of their proposed method, the researchers performed a sequence of experiments using a Twitter dataset [48]. They evaluated the performance of their approach in comparison to existing document clustering algorithms. The results of the experiments revealed that Agglomerative clustering with Hybrid TF-IDF exhibited the most significant improvement (F-Measure delta: +0.255), followed by Bisect K-Means++ with Hybrid TF-IDF (F-Measure delta: +0.251), and then Hybrid TF-IDF alone (F-Measure delta: +0.249).

The research conducted by Lavanya et al. was centered on analyzing Twitter data related to cancer [49]. They collected data continuously over the course of one year and recorded observations comparing two sets of data. The larger dataset, however, contained more noise, which affected the performance of feature selection methods. To address this, the researchers introduced a pre-processing step and experimented with different combinations of feature selection methods used earlier. The experiments involved applying the K-means clustering technique first without feature selection and then with feature selection methods, namely variance threshold (FSVT) and Laplacian score (FSLs), along with two feature extraction methods—PCA and SVD. Both the old Twitter data (Cancer_tweets_2017) and the new, larger data (Cancer_tweets_2018) were used for experimentation. The intersection combination, utilizing the smallest feature set, exhibited better performance compared to other techniques. Additionally, SVD performed well for both datasets and various cluster sizes, requiring only five features. The researchers also explored a soft clustering algorithm, fuzzy C-means, and observed that the summary generated from the larger dataset was more meaningful than that from the smaller data, which yielded higher measures in previous experiments. This emphasized the importance of the additional pre-processing carried out. Moreover, the generated summaries from the Twitter dataset proved to be useful for the healthcare fraternity to understand the genuine feelings and opinions of patients or their relatives. This valuable insight can help improve the services and care provided to them, aligning with the holistic approach required in the healthcare industry.

Dusart et al. made a twofold contribution to their paper. Firstly, they introduced the TES 2012–2016, a sizable evaluation collection specifically designed for Twitter event summarization [50]. This valuable collection, based on Zubiaga's research [51], consists of 28 events, with an average of 3 million tweets per event. Secondly, the authors proposed

a novel neural model named TSSuBERT for extractive tweet summarization. TSSuBERT incorporates several original features, including (i) the utilization of pre-trained language models and tweet context in the form of vocabulary frequencies and (ii) the automatic adjustment of output summary size, which makes it more practical for real-world applications compared to state-of-the-art models that always output summaries of fixed size, regardless of their usefulness. (iii) TSSuBERT is an incremental method, enabling efficiency gains, as it does not require re-evaluating all documents at each time increment. To evaluate their approach, they conducted experiments using two distinct Twitter collections: TES 2012–2016 and the collection introduced by Rudra et al. [52]. The results demonstrated that on the TES 2012–2016 dataset, TSSuBERT achieved a ROUGE-1 score of 0.111, a ROUGE-2 score of 0.018, and a COS Embed score of 0.741. On the collection from Rudra et al., the model attained a ROUGE-1 score of 0.412, a ROUGE-2 score of 0.192, and a COS Embed score of 0.935.

In the paper by Dhiman and Toshniwal, they propose an unsupervised approximate graph clustering-based global event detection model for Twitter data [53]. The key contribution lies in their innovative graph-based Twitter data representation, which effectively incorporates uncertainty in tweet relationships while maintaining contextual and directional information. This approach ensures computational efficiency by keeping the runtime requirements low. To further enhance their model, they utilize component filtration of the Twitter graph, removing irrelevant tweet-pairs and edges. This focused approach allows for the selection of critical data points with a higher likelihood of representing an event. Additionally, they employ partial-k-based uncertain graph clustering to improve the model's runtime performance. The experimental performance analysis demonstrates the superiority of their proposed model compared to other competitors on different datasets. The precision, recall, and F1 score values obtained on various datasets are as follows:

RepLab2013 Twitter data: precision score of 0.7098, recall score of 0.833, and F1 score of 0.766.

Auspol Twitter data: precision score of 0.718, recall score of 0.695, and F1 score of 0.706.

Common Diseases dataset: precision score of 0.714, recall score of 0.651, and F1 score of 0.681.

Customer Support on Twitter dataset: precision score of 0.51, recall score of 0.651, and F1 score of 0.571.

These results affirm the superior performance of their proposed model in event detection across various types of datasets.

Automatic text summarization presents a multi-objective optimization (MOO) challenge, aiming to balance two conflicting objectives: retaining information from the source text while keeping the summary length concise. To address this problem, Lucky and Girsang devised a method that leverages an undirected graph to establish relationships between social media comments [54]. The approach employs the Multi-Objective Ant Colony Optimization (MOACO) algorithm to create summaries by selecting concise yet important comments from the graph, depending on the desired summary size. To evaluate the quality of their generated summaries, the researchers compared them with results from other text summarization algorithms like LexRank, TextRank, Latent Semantic Analysis, SumBasic, and KL-Sum. They used a dataset of Twitter comments related to the presidential election in Indonesia, collected using the hashtag #pilpres. The results demonstrated that MOACO was capable of producing informative and concise summaries. The average cosine distance to the source text was 0.127, indicating good retention of information, and the average number of words in the summaries was 388, which outperformed the other algorithms.

The paper by Liang et al. introduces a novel selective reinforced sequence-to-sequence attention model for abstractive social media text summarization [55]. The proposed model incorporates a hidden layer before the encoder module and utilizes a selective gate to determine which information should be retained or discarded. Moreover, the model combines cross-entropy and reinforcement learning policy to directly optimize the ROUGE score. The experiments were conducted on the LCSTS dataset, which was obtained from

Sina Weibo. This dataset is composed of three parts: PART I with 2,400,591 text–summary pairs, PART II with 10,666 text–summary pairs, and PART III with 1106 pairs. The evaluation results demonstrated the effectiveness of the proposed model, with F1 scores of 38.2 for ROUGE-1, 25.2 for ROUGE-2, and 35.5 for ROUGE-L.

In the paper authored by Dehghani and Asadpour, a novel graph-based framework was introduced to generate storylines on Twitter [56]. This innovative approach takes into account both the semantic content and social information of tweets. The storyline is represented as a directed tree, showcasing socially salient events that unfold over time. In this representation, nodes signify main events, while edges capture the semantic relationships between related events. To identify significant sub-events, they employed a community detection technique on a temporal and semantic similarity graph. Extensive experiments were conducted on real-world events in both English and Persian datasets extracted from Twitter. The performance of their event summarization module was evaluated using ROUGE 1 and ROUGE 2 metrics, with achieved values of 0.6239 and 0.5476, respectively.

In their work, Rudrapal et al. presented a two-phase summarization approach to create abstract summaries for Twitter events [57]. Initially, the approach extracts key sentences from a pool of tweets relevant to the event, effectively reducing redundant information by leveraging Partial Textual Entailment (PTE) relations between sentences. Subsequently, an abstract summary is generated using the least redundant key sentences. The proposed approach's performance was evaluated through experiments on various datasets. The data encompassed English tweets related to 25 trending events collected from Twitter4j during the period from January to October 2017. These events covered diverse topics such as natural disasters, politics, sports, entertainment, and technology. The evaluation results were measured using ROUGE-1, ROUGE-2, ROUGE-L, and ROUGE-SU4 metrics, yielding scores of 0.49, 0.38, 0.45, and 0.39, respectively.

Dutta et al. proposed a clustering approach for micro-blogs that combines feature selection techniques to group similar messages [58]. The approach effectively reduces information overload and addresses the problem of less content and noisy messages. The best performance of the proposed methodology according to each metric was as follows: 43.4864 for the Calinski–Harabasz (CH) index, 0.795 for the DB (DB) index, 1.52 for the Dunn (D) index, 0.1425 for the I-Index (I), 0.2708 for the Silhouette (S) index, and 0.1319 for the Xie–Beni (XB) on Hagupit dataset.

In the study conducted by Dutta et al., they presented a graph-based approach to summarize tweets [59]. The first step involved constructing a tweet similarity graph, followed by the application of community detection methodology to group together similar tweets. To represent the tweet summarization, important tweets were identified from each cluster using various graph centrality measures, including degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, and PageRank. To perform their experiments, the researchers collected a dataset consisting of 2921 tweets from Twitter API using a crawler. The performance of their proposed methodology was evaluated using precision, recall, and F-measure metrics. The obtained results showed a precision of 0.957, recall of 0.937, and F-measure of 0.931, demonstrating the effectiveness of their approach in tweet summarization.

Ref. [60] presented a regularization approach for the sequence-to-sequence model, specifically tailored for the Chinese social media summarization task. Their method involved leveraging the model's learned knowledge to apply regularization to the learning objective, effectively mitigating summarization-related challenges. To enhance the evaluation process and ensure semantic consistency, they introduced a practical human evaluation approach, addressing the limitations of existing automatic evaluation methods. For their experiments, the researchers utilized the LCSTS dataset. This dataset comprises over 2.4 million text–summary pairs extracted from the popular Chinese social media microblogging service, Weibo. The automatic evaluation of the proposed method yielded the following results: ROUGE-1 score of 36.2, ROUGE-2 score of 24.3, and ROUGE-L score of 33.8. These results demonstrate the effectiveness of their regularization approach in the context of Chinese social media summarization.

In their work, Wang et al. introduced an approach for microblog summarization, which leverages the Paragraph Vector and semantic structure. The method comprises several key steps [61]. Firstly, they constructed a sentence similarity matrix, capturing the contextual information of microblogs, to identify sub-topics using the Paragraph Vector. Next, they analyzed the sentences using the Chinese Sentential Semantic Model (CSM) to extract semantic features. Additionally, relation features were derived from the combination of the similarity matrix and the aforementioned semantic features. Lastly, the most informative sentences were accurately selected from microblogs belonging to the same sub-topics based on both semantic and relation features. To evaluate their method, the researchers created an evaluation dataset consisting of 16 microblog topics. These topics were collected from the 2013 NLP&CC (CCF Conference on Natural Language Processing & Chinese Computing) conference database and Chinese microblog data gathered online. The proposed method achieved ROUGE scores as follows: 0.531706 for ROUGE-1, 0.265799 for ROUGE-2, and 0.266479 for ROUGE-SU.

Chakraborty et al. presented a method to summarize tweets related to news articles [62]. The approach revolves around capturing diverse opinions expressed in the tweets, which is achieved through the creation of a unique TSG (Twitter Sentiment Graph). By utilizing a community detection technique on this TSG, the researchers successfully identify tweets that represent these diverse opinions. Their findings demonstrate that the initial capturing of diverse opinions significantly enhances the accuracy of identifying the relevant set of tweets. To evaluate the effectiveness of their approach, the researchers utilized a dataset comprising political news articles sourced from the New York Post, along with corresponding Twitter data. The evaluation of their proposed method was carried out using ROUGE-I and ROUGE-II metrics, with the following scores obtained: ROUGE-I—precision of 0.67, recall of 0.66, and F1 score of 0.664; ROUGE-II—precision of 0.56, recall of 0.358, and F1 score of 0.548.

Rudra et al. put forth a classification-based summarization approach, catering to both English and Hindi tweets [63]. Their method efficiently extracts and summarizes tweets containing situational information. These studies contribute to the advancement of automatic short text summarization in the context of social media, particularly for disaster-related content. The outcomes of their research were as follows: the ROUGE-1 F-score is 0.5770 for the event NEquake and 0.6602 for HDerail.

Rudra et al. introduced a classification-based method for summarizing microblogs during outbreaks [52]. They used Twitter data from two recent outbreaks (Ebola and MERS) to build an automatic classification approach useful for categorizing tweets into different disease-related categories. Their approach efficiently categorizes raw tweets using low-level lexical class-specific features to extract essential information. Their study yielded a ROUGE-1 recall of 0.4961 and an F-score of 0.4980 for Ebola and a ROUGE-1 recall of 0.3862 and an F-score of 0.3758 for MERS.

Nguyen et al. proposed an approach that generates high-quality summaries by considering the social context of user-generated content and third-party sources [64]. They utilize learning to rank (L2R), utilizing Ranking SVM, with features such as sentences, user-generated content, and third-party sources to score and select sentences for the summaries, which can be retrieved through a score-based or voting mechanism. The model achieved a ROUGE-1 of 0.230 on comments of the SoLSCSum dataset.

Huang et al. introduced a generic social event summarization approach for Twitter streams [65]. The method includes participant detection, sub-event detection, and tweet summary extraction. They used an online clustering method for participant detection and an online temporal content mixture model Expectation Maximization (EM) algorithm for sub-event detection, making the approach applicable to natural tweet streams. The best evaluation results were 0.3970 of the new evaluation metric ROUGET-1 F-1 score, a modified evaluation metric of ROUGE-N, on the play-by-play data (Celtics vs. Heat) from ESPN.

In their study, He and Duan introduced a novel approach called Social Networks with Sparse Reconstruction (SNSR) to summarize Twitter data, especially in short and noisy situations [66]. To achieve this, they formulated tweet relationships as social regularization and integrated them into a group sparse optimization framework. The researchers created gold-standard Twitter summary datasets covering 12 different topics for evaluation. These datasets were built using publicly available Twitter data collected by the University of Illinois as the raw data. The experimental results on these datasets demonstrated the effectiveness of their approach, with a ROUGE-1 score of 0.44887, ROUGE-2 score of 0.13882, and ROUGE-SU4 score of 0.18147.

The research conducted by Madichetty and Muthukumarasamy tackles the task of detecting situational tweets using various deep learning architectures, including Convolutional Neural Network (CNN), long short-term memory (LSTM), bi-directional long short-term memory (BLSTM), and bi-directional long short-term memory with attention (BLSTM attention) [67]. Notably, these deep learning models are applied to both English and Hindi language tweets to identify situational information during disasters. In countries like India, during disasters, some tweets are posted in Hindi, and their information may not be available in English. To address this multilingual scenario, experiments are carried out on diverse disaster datasets, such as the Hyderabad bomb blast, Hagupit cyclone, Nepal Earthquake, Harda rail accident, and Sandy Hook shooting, considering both in-domain and cross-domain situations. The best performance results obtained for the deep learning models were as follows: On the TypHagupit dataset, the model achieved 98.72% recall, 74.37% F1 score, and 65.98% accuracy. On the SanHShoot dataset, it attained 95.56% recall, 79.53% F1 score, and 75.41% accuracy. These findings demonstrate the effectiveness of the proposed deep learning architectures for detecting situational tweets during disasters in both the English and Hindi languages.

Andy et al. utilized the distinctive attributes of pre-scheduled events in tweet streams to extract the key highlights and proposed a method for summarizing these highlights [68]. To assess the effectiveness of their algorithm, they conducted evaluations using tweets related to two episodes of the renowned TV show *Game of Thrones*, Season 7. The evaluation results for tweets centered on the highlights from Episode 3 were impressive, with scores of 38 for ROUGE-1, 10 for ROUGE-2, and 31.5 for ROUGE-L.

In their work, Phan et al. presented a tweet integration model based on identifying the shortest paths on a word graph [1]. To achieve this, they employed a topic modeling technique that automatically determines the number of topics within the tweets. Additionally, sentiment analysis was used to assess the percentage of opinions categorized as agreeing, disagreeing, or neutral for each issue. The outcome of this integration process is the creation of weighted directed graphs, where each graph corresponds to a topic and represents the shortest path on the word graph. To assess the model's performance, the researchers gathered 90,000 raw tweets from 20 Twitter users using the Streams API for Python. After pre-processing the tweets, they obtained a subset of 70,000 tweets, which was used for evaluation. The experimental results of the proposed method yielded the following scores: For ROUGE_N (F_measure: 0.30, precision: 0.31, and recall: 0.29), For ROUGE_L: (F_measure: 0.20, precision: 0.20, and recall: 0.19)

Ref. [69] proposed an innovative model that utilizes an autoencoder as a supervisor for the sequence-to-sequence model, enhancing the internal representation for abstractive summarization. To further enhance the autoencoder's supervision, they introduced an adversarial learning approach. To evaluate the effectiveness of their model, they conducted experiments on the Large Scale Chinese Social Media Text Summarization Dataset (LCSTS). This dataset comprises more than 2,400,000 text–summary pairs extracted from the Sina Weibo website. The experimental results were highly promising, achieving the following ROUGE scores: (ROUGE-1: 39.2, ROUGE-2: 26.0, and ROUGE-L: 36.2).

Table 2. Previous studies of ASTS techniques in social media.

Ref.	Objective	Method	Domain	Dataset	Evaluation Metrics	Observation
[46]	Integrity-aware, inconsistency, readability, quality, and efficiency	Deep Neural Network	Event	XMUDM/IAEA (https://github.com/XMUDM/IAEA) (accessed on 7 August 2023)	BLEU-1 = 24 BLEU-2 = 6 ROUGE-1 = 37 ROUGE-2 = 13 ROUGE-L = 33 Accuracy = 0.880 Precision = 0.713 Recall = 0.710 F1 score = 0.705 AUC = 0.814	This approach is a novel real-time event summarization framework, but it needs to study the efficiency issue in real-time event summarization systems.
[69]	Accuracy	Supervised learning	Generic	LCSTS (http://icrc.hitsz.edu.cn/Article/show/139.html) (accessed on 7 August 2023) dataset	ROUGE-1 = 39.2 ROUGE-2 = 26.0 ROUGE-L = 36.2	This model outperforms the sequence-to-sequence baseline by a large margin and achieves state-of-the-art performances on a Chinese social media dataset.
[60]	Accuracy, semantic consistency, and fluency Relatedness Faithfulness	Cross-entropy	Generic	LCSTS dataset	Human ROUGE-1 = 36.2 ROUGE-2 = 24.3 ROUGE-L = 33.8	This approach improves the semantic consistency by 4% in terms of human evaluation.
[55]	To optimize the ROUGE score	Unsupervised cross-entropy and reinforcement	Generic	LCSTS dataset	F1 scores ROUGE-1 = 38.2 ROUGE-2 = 25.2 ROUGE-L = 35.5	This model decides what information should be retained or removed during summarization, but cannot combine visual information and textual information for social media event summarization.
[57]	Redundancy and sentence identification	Ranking	Generic	Collected tweets	ROUGE-1 = 0.49 ROUGE-2 = 0.38 ROUGE-L = 0.45 ROUGE-SU4 = 0.39	This approach outperforms the baseline approach, but it needs to measure the trustworthiness of tweet content to be included in the summary.
[52]	Time-aware	Integer Linear Programming	Epidemic	Two datasets (Ebola and MERS) (http://cse.iitkgp.ac.in/~krudra/epidemic.html) (accessed on 7 August 2023)	ROUGE-1 for Ebola F-scores = 0.4980 Recall = 0.4961 ROUGE-1 for MERS F-scores = 0.4980 Recall = 0.4961	Using CatE addresses the problem of overlapping keywords across classes.
[64]	Social context and high-quality summary	Ranking and Selection SVM	Generic	SoLSCSum, USA Today-CNN, and VSoLSCSum	ROUGE-1 = 0.230 on SoLSCSum	To enhance semantics, more deep analyses should consider LSTMs or CNNs.
[63]	Quality and similarity	Classification ILP	Specific disaster events	Disaster datasets	ROUGE-1 F-score = 0.5770 (NEquake) F-score = 0.6602 (HDerail)	There is no consideration for the semantic similarity between two non-English words; this helps reduce errors.
[59]	Degree centrality, closeness centrality, and betweenness centrality	Graph-based	Disaster events	Collection of 2921 tweets	Precision = 0.957 Recall = 0.937 F-measure = 0.931	This is a simple and effective method for Tweet summarization, but as future work, other algorithms can be developed that can help predict certain patterns and trends on Twitter.
[58]	Quality and similarity	Clustering and Feature selection	Specific (disaster events)	Disaster datasets	CH index = 43.4864 DB index = 0.795 D index = 1.52 I-Index = 0.1425 S index = 0.2708 XB = 0.1319 on Hagupit dataset	Simplicity and effectiveness satisfy real-time processing needs.
[68]	Identify highlights and characters	LSTM	Event	Collected tweets and around two episodes of GOTS7	ROUGE-1 = 38 ROUGE-2 = 10 ROUGE-L = 31.5	No consideration for diversity.
[67]	Accuracy	Deep learning	Disaster events	Disaster datasets, Hindi tweets, and dataset	TypHagupit dataset Recall = 98.72% F1 score = 74.37% Accuracy = 65.98% SanHShoot dataset Recall = 95.56% F1 score = 79.53% Accuracy = 75.41%	Deep learning models perform badly on Hindi tweets due to the lack of enough Tweets available.
[49]	Similarity and efficiency	Selection methods, clustering, and extraction	Health care	More than one million tweets related to cancer were collected	Similarity and efficiency	The generated summary from the larger-scale dataset was found to be more meaningful than the smaller dataset.
[43]	Performance, similarity, and dissimilarity	Unsupervised classification, tf-idf scores	Specific (disaster events)	Disaster datasets	ROUGE-2 = 0.3661 ROUGE-L = 0.5199	Reported improvement in ROUGE-2 and ROUGE-L scores, but executing time is high.

Table 2. *Cont.*

Ref.	Objective	Method	Domain	Dataset	Evaluation Metrics	Observation
[44]	Quality	SOM, GSOM, and GA	Disaster events	Disaster events UkFlood, Sandyhook, Hblast, and Hagupit	ROUGE-2 = 0.3458 ROUGE-L = 0.5000	Showed improvement in ROUGE-2 and ROUGE-L score.
[40]	Diversity	Supervised learning	Event	Collected tweets and event-based dataset	ROUGE-1 = 0.4787 ROUGE-2 = 0.2391 ROUGE-3 = 0.1652 ROUGE-L = 0.4241	Reported better topic coverage and conveyed more diverse user emotions.
[39]	Credibility, worthiness, accuracy	Supervised	Disaster events	PHEME dataset	ROUGE1-F1 = 0.501	Extractive technique for summarizing tweets with the task rumor detection.
[66]	Coverage and sparsity, diversity, and redundancy	Social Network and Sparse Reconstruction (SNSR)	Generic	UDI-TwitterCrawl-Aug2012 (https://wiki.illinois.edu/wiki/display/forward/Dataset-UDI-TwitterCrawl-Aug2012) (accessed on 7 August 2023)	ROUGE-1 = 0.44887 ROUGE-2 = 0.13882 ROUGE-SU4 = 0.18147	Social relations help in optimizing the summarization process.
[54]	Summary size reduction	Graph-based, Multi-Objective Ant Colony Optimization	Election event	Presidential election in Indonesia Twitter dataset	Cosine distance = 0.127 Word count = 388	Reliable for producing concise and informative summaries from social media comments.
[42]	Similarity and redundancy	Multi-objective binary differential evolution	Specific (disaster events)	Disaster events (UkFlood, Sandyhook, Hblast, and Hagupit)	Best result for MOOTS3 ROUGE-2 = 0.3418 ROUGE-L = 0.5009	Using BM25 similarity measure and SOM as a genetic operator score showed improvement in the summarization process.
[1]	Accuracy	Maximal association rules	Generic	Collected Twitter dataset	ROUGE_N F_measure = 0.30 Precision = 0.31 Recall = 0.29 ROUGE_L F_measure = 0.20 Precision = 0.20 Recall = 0.19	Improves the accuracy of a summary.
[53]	Contextual relationships and computational cost	Approximate model, graph-based, and clustering	Generic	RepLab2013 (http://nlp.uned.es/replab2013/replab2013-dataset.tar.gz) (accessed on 7 August 2023) and Customer Support on Twitter Dataset (https://www.kaggle.com/thoughtvector/customer-support-on-twitter) (accessed on 7 August 2023)	RepLab2013 Twitter Precision = 0.7098 Recall = 0.833 F1 score = 0.766 Auspul Twitter Precision = 0.718 Recall = 0.695 F1 score = 0.706 Common Diseases Precision = 0.714 Recall = 0.651 F1 score = 0.681 Customer Support on Twitter Precision = 0.51 Recall = 0.651 F1 score = 0.571	Improved run-time performance for event detection but no consideration for location.
[38]	Coverage and diversity	Ontology	Disaster events	10 disaster datasets	F1 scores ROUGE-1 = 0.58 ROUGE-2 = 0.29 ROUGE-L = 0.31	Relies on existing ontology; cannot determine the category of tweet.
[62]	Diversity, coverage, and relevance	Graph partitioning	News	News articles and Twitter dataset related to articles	ROUGE-I precision = 0.67 Recall = 0.66 F1 score = 0.664 ROUGE-II precision = 0.56 recall = 0.358 F1 score = 0.548.	This method of summarizing news article tweets.
[65]	Summary quality	Clustering	Sporting events	NBA basketball games	ROUGET-1 F-1 = 0.3970	Achieves improvement in performance similar to counterpart.
[47]	Content score and performance	LDA and WorkGraph	Generic	[48]	Agglomerative clustering with Hybrid TF-IDF F-Measure delta = +0.255 Bisect K-Means++ with Hybrid TF-IDF F-Measure delta = +0.251 Hybrid TF-IDF alone F-Measure delta = +0.249	This method takes both topic sentiments and topic aspects into account together.

Table 2. Cont.

Ref.	Objective	Method	Domain	Dataset	Evaluation Metrics	Observation
[41]	Saliency estimation	Semantic types and graph convolutional network	Generic	Collected tweets for 1000 events	ROUGE-2 = 0.647 BLEU = 0.490	Two extractive strategies were investigated. The first ranks tweets based on semantic terms; the second generates hidden features.
[56]	Redundancy and accuracy	Graph-based	Specific events	Iran Election, IranNuclearProgram, and USPresidentialDebates	ROUGE-1 = 0.6239 ROUGE-2 = 0.5476	This method identifies the significant events, summarizes them, and generates a coherent storyline of their evolution with a reasonable computational cost for large datasets.
[61]	Quality	Multi-objective DE	Specific Topic	Dataset collected from NLP&CC	ROUGE-1 = 0.531706 ROUGE-2 = 0.265799 ROUGE-SU = 0.266479	This method suggests a summary technique that takes the Paragraph Vector and semantic structure into account.
[45]	Detection, Fast Happening Subevents, and hierarchy	LSTM	Specific events	FA CUP final, ST, Primaries, and USEs SB (http://www.socialsensor.eu/results/datasets/72-twitter-tdt-dataset) (accessed on 7 August 2023)	Precision, recall, F-measure, and ROUGE-1 Best F-measure = 0.5644 on the USE dataset	This method generates a storyline for an event using summaries at different levels, but it needs a separate deep learning module that can be trained to combat noise in tweets.
[26]	Time complexity, fluency, and adequacy	summary-aware attention	Generic	LCSTS dataset (http://icrc.hitsz.edu.cn/Article/show/139.html) (accessed on 7 August 2023)	Automatic and human ROUGE-1 = 40.3 ROUGE-2 = 27.8 ROUGE-L = 38.4	This model outperforms Seq2Seq baselines and generates summaries in a coherent manner, but increases the computational cost.
[50]	Predict the cosine score	Neural model	Specific event	TSSuBERT (https://github.com/JoeBloggsIR/TSSuBERT) (accessed on 7 August 2023) and Rudra et al. [52]	TSSuBERT ROUGE-1 = 0.111 ROUGE-2 = 0.018 COS Embed = 0.741 Rudra et al. dataset ROUGE-1 = 0.412 ROUGE-2 = 0.192 COS Embed = 0.935	A neural model automatically summarizes enormous Twitter streams, but it needs to learn the thresholds used in the filtering step.
[37]	Detecting contextual bullying	hybrid deep learning	Psychology (cyber-bullying)	Collected Twitter dataset	Accuracy = 90.45% Precision = 89.52% Recall = 88.98% F1 score = 89.25% Specificity = 90.94%	They could not perform the analysis in relation to the users' behavior.
[33]	Accuracy and redundancy	open-sourced search engine and a large language model	Emergency events	TREC and CrisisFACTS	Manual BERTScore = 0.4591 ROUGE-2 = 0.0581	It did not evaluate the reduction in redundancy.
[35]	Time, similarity, and efficiency	Latent Dirichlet allocation (LDA) method and K-means clustering	COVID19Vaccine	Collected Twitter dataset	ROUGE-1: (F1 score: 0.06 to 0.26, recall: 0.1 to 0.21, and precision: 0.05 to 0.19) ROUGE-2: (F1 score: 0.04 to 0.23, recall: 0.01 to 0.19, and precision: 0.03 to 0.17) ROUGE-S4: (F1 score: 0.04 to 0.21, recall: 0.01 to 0.18, and precision: 0.03 to 0.15) ROUGE-SU4: (F1 score: 0.29 to 0.48, recall: 0.04 to 0.07, and precision: 0.07 to 0.12)	The ROUGE may not be the most suitable evaluation matrix for assessing extractive summaries' effectiveness.
[36]	Accuracy and relevance	Combination extraction and abstraction algorithms	Events	Collected Instagram posts	Accuracy = 75% Completeness = 82% Precision = 80% Harmonic mean = 81% Error rates = 25%	The accuracy and efficiency of the summarization process can be enhanced.

8. Discussion

Despite the extensive research efforts aimed at addressing automatic short text summarization (ASTS) problems, the distinctive characteristics of short text summarization present unresolved challenges within the ASTS task. As a result, there remains ample room

for further exploration in this area. Some topics of ASTS require greater attention in future research endeavors:

- Extractive-based ASTS techniques encompass a range of methodologies, including machine learning and optimization-based, statistical-based, graph-based, and fuzzy-logic-based approaches. Within the machine learning realm, both supervised and unsupervised techniques aim to enhance accuracy [67] quality [52,58,63,65], and similarity [42,59]. Yet, a comprehensive evaluation often requires a broader focus, encompassing diversity, relevancy, completeness, readability, and redundancy.
- Supervised classification techniques, although effective, demand substantial training datasets, leading to challenges in their creation. Unsupervised methods, while promising, encounter the complexities inherent in the diversity and noise of tweets. Optimization-based techniques seek to optimize redundancy [44,54,66], diversity, coverage [67], and other objectives. However, their potential to enhance summary quality, accuracy, similarity, sparsity, and computational efficiency deserves deeper exploration. Resolving conflicting objectives should involve multi-objective optimization algorithms.
- Considerations in statistical and graph-based ASTS techniques
 - The landscape of automatic short text summarization (ASTS) techniques is enriched by the contributions of both statistical-based and graph-based approaches. However, a nuanced evaluation reveals distinct considerations and priorities within each of these methodologies. Statistical-based techniques emphasize summary quality and relevancy yet often overlook diversity, accuracy, readability, time constraints, and computational costs. In contrast, graph-based methods prioritize diversity, accuracy [1], and performance [12], with less emphasis on quality, similarity, completeness, redundancy, coverage, readability, time, and computational costs. As ASTS techniques continue to evolve, a synthesis of statistical-based and graph-based methodologies holds the potential to elevate summarization outcomes to new heights. By integrating the strengths of both approaches, researchers could create a more versatile and balanced framework that not only prioritizes quality, diversity, and accuracy but also addresses other critical dimensions. A subset of ASTS techniques dabbles in the application of fuzzy logic; however, the extent of its effectiveness remains uncertain. Further exploration is warranted to ascertain its potential contribution to the advancement of ASTS methodologies.
- Balancing act for optimal summarization
 - In the pursuit of effective summarization, the choice between extractive, abstractive, or hybrid approaches hinges on the nature of the input data, the desired level of summary coherence, the intended application, and the complexity of the summarization task. Extractive methods often excel when preserving source text information is crucial, while abstractive methods shine when generating concise and engaging summaries is paramount. However, it is the hybrid approach that embodies the quest for a harmonious equilibrium. By integrating the strengths of both extractive and abstractive techniques, hybrids aspire to deliver summaries that capture the core essence of the source while infusing them with creative and contextual innovations. The realm of ASTS is not about choosing one approach over the other, but rather about finding the ideal equilibrium that aligns with the summarization goals. This dynamic interplay ensures that the art of summarization continues to evolve, offering versatile solutions that adapt to the nuances of diverse text inputs and user preferences.
- Dataset limitations and multi-lingual evaluation
 - The efficacy of automatic short text summarization (ASTS) techniques is inherently tied to the quality and diversity of the datasets used for evaluation. However, it is important to acknowledge the limitations present within exist-

ing datasets, which are summarized in Table 1, as these limitations can impact the generalizability and reliability of the generated summaries. One of the key limitations lies in the scope and representativeness of the datasets. Many existing datasets are primarily centered around a single platform, such as Twitter, which might not fully capture the intricacies and variations present in other social media platforms like Facebook, Instagram, or Sina Weibo. This platform-centric nature of datasets can lead to biased evaluations, where the performance of ASTS techniques might not translate seamlessly across different platforms. Furthermore, the majority of these datasets are predominantly designed for English text, leaving a significant gap in evaluation capabilities for other languages. This language bias restricts the broader applicability of ASTS techniques across multilingual social media platforms. Languages like Arabic, Hindi, Chinese, and others are underrepresented, which hinders a comprehensive evaluation of techniques' performance on a global scale. A robust evaluation framework should encompass a spectrum of languages, thereby enabling researchers to gauge the effectiveness of their techniques across different linguistic nuances and structures.

- By incorporating multiple languages in evaluation, ASTS techniques can be assessed for their adaptability, effectiveness, and challenges in accommodating varying grammatical rules, idiomatic expressions, and cultural contexts. A multi-lingual evaluation not only highlights the strengths and weaknesses of techniques across diverse languages but also fuels advancements that are more universally applicable.
- In addition to expanding the range of languages, evaluating ASTS techniques on different social media platforms is equally crucial. This broader scope ensures that techniques can effectively handle the idiosyncrasies of various platforms and tailor their summarization outputs accordingly.

9. Challenges

Summarizing short text on social media poses several unique and significant challenges due to the dynamic and diverse nature of content found on these platforms. Addressing these challenges requires continuous advancements in natural language processing techniques and domain-specific knowledge. By overcoming these hurdles, ASTS techniques can play a vital role in enhancing information retrieval, trend analysis, sentiment monitoring, and understanding public opinion on social media platforms. The following section discusses some of the key challenges faced by automatic short text summarization (ASTS) techniques in effectively summarizing short text on social media.

- **Noisy and informal language:** Social media platforms are known for their casual and conversational style of communication. Users often employ abbreviations, emojis, slang, and colloquial language, making the text noisy and informal. Deciphering the meaning and extracting valuable insights from such language can be challenging for summarization algorithms.
- **Abbreviations and acronyms:** The prevalence of abbreviations and acronyms used by users further complicates summarization. These shorthand expressions may not always be universally understood, and different communities develop their own unique set of abbreviations, adding to the difficulty of accurately capturing the content's essence.
- **Contextual understanding:** Short text on social media often lacks explicit context, leading to ambiguity in meaning. ASTS techniques must be capable of interpreting the implicit context surrounding posts, comments, and tweets to generate coherent and relevant summaries.
- **Multi-modality:** Social media content frequently includes multimedia elements such as images, videos, and emojis, which complement the textual information. Sum-

marization must consider these multi-modal aspects to provide comprehensive and informative summaries.

- Dynamic language and trends: The language and trends on social media evolve rapidly, with new terms, hashtags, and trending phrases emerging constantly. Keeping up with the ever-changing language requires ASTS techniques to adapt and update regularly.
- Nested conversations and replies: Social media conversations often involve nested replies and discussions, requiring summarization techniques to capture the broader context while maintaining conciseness.
- Sentiment and emotional analysis: Emojis, punctuation, and sentiment-laden language are common in social media content, indicating the users' emotions and attitudes. Properly conveying the sentiment and emotional tone in summaries is vital for comprehensive content understanding.
- Data volume and real-time processing: Social media platforms generate an enormous volume of short text data every second. Summarization techniques must handle this vast amount of data efficiently, considering the real-time nature of social media communication.
- Data sparsity and imbalanced datasets: Some topics or events on social media may attract an overwhelming amount of attention, while others receive limited engagement. Summarization algorithms must account for data sparsity and imbalanced datasets to generate unbiased and comprehensive summaries.
- Ethical and biased content: Social media content may include biased or harmful information. ASTS techniques should be sensitive to ethical concerns and avoid promoting or amplifying harmful narratives in summaries.

10. Conclusions

This paper represents a comprehensive endeavor to explore automatic short text summarization techniques, illuminating both the progress achieved and the obstacles confronted in this captivating realm. The extensive body of research underscores the significance and intricacy of the ASTS task. The primary objective of this paper is to provide readers with a foundational understanding of automatic short text summarization and acquaint them with the diverse categories of ASTS systems in meticulous detail. The paper encompasses a research survey covering extractive, abstractive, and hybrid text summarization methods. Throughout this survey, the significance of inventive approaches that intricately balance succinctness and comprehensive content coverage has been underscored. Upon deeper scrutiny, it becomes apparent that the distinctive attributes of short text summarization give rise to specific challenges demanding comprehensive solutions. The availability of databases for specific languages also emerges as a concern and a novel aspect for future consideration in this domain. Looking ahead, we anticipate continuous advancements in machine learning that will further enhance the refinement of automatic short text summarization systems. This evolution holds the potential to unlock efficient information extraction and comprehension from succinct textual content, heralding a paradigm shift in our interaction with vast data volumes and the derivation of insightful understanding.

Author Contributions: Conceptualization, F.A.G. and R.A.; methodology, F.A.G., M.C.P. and R.A.; formal analysis, F.A.G., M.C.P. and R.A.; investigation, R.A. and M.C.P.; resources, F.A.G. and R.A.; data curation, F.A.G. and R.A.; writing—original draft preparation, F.A.G., R.A. and M.C.P.; writing—review and editing, F.A.G. and R.A.; visualization, F.A.G. and R.A.; supervision, M.C.P. and R.A.; project administration, M.C.P.; funding acquisition, R.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Phan, H.T.; Hoang, D.T.; Nguyen, N.T.; Hwang, D. Tweet Integration by Finding the Shortest Paths on a Word Graph. In *Modern Approaches for Intelligent Information and Database Systems*; Springer: Cham, Switzerland, 2018; pp. 87–97.
2. Nenkova, A.; McKeown, K. Automatic summarization. *Found. Trends® Inf. Retr.* **2011**, *5*, 103–233. [CrossRef]
3. Nallapati, R.; Zhai, F.; Zhou, B. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In Proceedings of the AAAI Conference on Artificial Intelligence Palo, Alto, San Francisco, CA, USA, 4–9 February 2017; p. 31.
4. Liu, Y.; Lapata, M. Text Summarization with Pretrained Encoders. In Proceedings of the 2019 Conference on Empirical Methods in Natural, Hong Kong, China, November 2019; pp. 3728–3738.
5. Jones, K.S. Automatic summarising: The state of the art. *Inf. Process. Manag.* **2007**, *43*, 1449–1481. [CrossRef]
6. Sharifi, B.P.; Inouye, D.I.; Kalita, J.K. Summarization of twitter microblogs. *Comput. J.* **2014**, *57*, 378–402. [CrossRef]
7. Atefeh, F.; Khreich, W. A survey of techniques for event detection in twitter. *Comput. Intell.* **2015**, *31*, 132–164. [CrossRef]
8. Kawade, P.B.; Pise, N.N. Summarization Approach Of Situational Information from Microblogs During Disaster Events: A Survey. *Int. J. Latest Trends Eng. Technol.* **2016**, *7*, 108–112.
9. Ramachandran, D.; Ramasubramanian, P. Event detection from Twitter—A survey. *Int. J. Web Inf. Syst.* **2018**, *14*, 262–280. [CrossRef]
10. Rudrapal, D.; Das, A.; Bhattacharya, B. A survey on automatic Twitter event summarization. *J. Inf. Process. Syst.* **2018**, *14*, 79–100.
11. Hasan, M.; Orgun, M.A.; Schwitler, R. A survey on real-time event detection from the Twitter data stream. *J. Inf. Sci.* **2018**, *44*, 443–463. [CrossRef]
12. Ermakova, L.; Cossu, J.V.; Mothe, J. A survey on evaluation of summarization methods. *Inf. Process. Manag.* **2019**, *56*, 1794–1814. [CrossRef]
13. Taecharungroj, V. “What Can ChatGPT Do?” Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. *Big Data Cogn. Comput.* **2023**, *7*, 35. [CrossRef]
14. Thakur, N. A Large-Scale Dataset of Twitter Chatter about Online Learning during. *Data* **2022**, *7*, 109. [CrossRef]
15. Anon. Twitter Usage Statistics. 2022. Available online: <https://www.internetlivestats.com/twitter-statistics/> (accessed on 7 August 2023).
16. Anon. Twitter User Count. 2022. Available online: <https://expandedramblings.com/index.php/twitter-stats-facts/> (accessed on 7 August 2023).
17. Anon. Counting Character. 2022. Available online: <https://developer.twitter.com/en/docs/counting-characters> (accessed on 7 August 2023).
18. Dixon, S.J. Global Social Networks Ranked by Number of Users. 2023. Available online: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> (accessed on 27 August 2023).
19. Kaplan, A.M.; Haenlein, M. Users of the world, unite! The challenges and opportunities of Social Media. *Bus. Horiz.* **2010**, *53*, 59–68. [CrossRef]
20. Thakur, N. MonkeyPox2022Tweets: A Large-Scale Twitter Dataset on the 2022 Monkeypox Outbreak, Findings from Analysis of Tweets, and Open Research Questions. *Infect. Dis. Rep.* **2022**, *14*, 855–883. [CrossRef] [PubMed]
21. Hayawi, K.; Shahriar, S.; Serhani, M.A.; Taleb, I.; Mathew, S.S. ANTi-Vax: A novel Twitter dataset for COVID-19 vaccine misinformation detection. *Public Health* **2022**, *203*, 23–30. [CrossRef] [PubMed]
22. Caillaud. French Entity-Linking Dataset between Annotated. 2023. Available online: <https://zenodo.org/record/7767294> (accessed on 9 August 2023).
23. Pano, T.; Kashef, R. A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19. *Big Data Cogn. Comput.* **2020**, *4*, 33. [CrossRef]
24. Abdulateef, S.; Khan, N.A.; Chen, B.; Shang, X. Multidocument Arabic Text Summarization Based on Clustering and Word2Vec to Reduce Redundancy. *Information* **2020**, *11*, 59. [CrossRef]
25. Mihalcea, R.; Tarau, P. Textrank: Bringing order into text. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Barcelona, Spain, 1 July 2004; pp. 404–411.
26. Wang, Q.; Ren, J. Summary-aware attention for social media short text abstractive summarization. *Neurocomputing* **2021**, *425*, 290–299. [CrossRef]
27. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. In *Advances in Neural Information Processing Systems 27: Annual Conference*; Curran Associates, Inc.: Red Hook, NY, USA, 2014; pp. 3104–3112.
28. Hariardi, W.; Latief, N.; Febryanto, D.; Suhartono, D. Automatic summarization from Indonesian hashtag on Twitter using TF-IDF and phrase reinforcement algorithm. In Proceedings of the 6th International Computer Science and Engineering, Vienna, Austria, 21–22 May 2016; pp. 17–19.
29. Moratanch, N.; Chitrakala, S. A survey on extractive text summarization. In Proceedings of the 2017 International Conference on Computer, Communication and Signal Processing (ICCCSP), Chennai, India, 10–11 January 2017; pp. 1–6.
30. Lloret, E.; Plaza, L.; Aker, A. The challenging task of summary evaluation: An overview. *Lang. Resour. Eval.* **2018**, *52*, 101–148. [CrossRef]
31. Lin, C.-Y. Rouge: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*; Association for Computational Linguistics: Barcelona, Spain, 2004; pp. 74–81.

32. Zhang, R.; Li, W.; Gao, D.; Ouyang, Y. Automatic twitter topic summarization with speech acts. *IEEE Trans. Audio Speech Lang. Process.* **2012**, *21*, 649–658. [[CrossRef](#)]
33. Pereira, J.; Fidalgo, R.; Lotufo, R.; Nogueira, R. Crisis Event Social Media Summarization with GPT-3 and Neural Reranking. In Proceedings of the 20th ISCRAM Conference, Omaha, NE, USA, May 2023; Volume 8.
34. Track, T.C. CrisisFACTS. 2023. Available online: <https://crisisfacts.github.io/#datasets> (accessed on 27 August 2023).
35. Alabid, N.; Naseer, Z. Summarizing twitter posts regarding COVID-19 based on n-grams. *Indones. J. Electr. Eng. Comput. Sci.* **2023**, *8*, 1008–1015. [[CrossRef](#)]
36. Taghandiki, K.; Ahmadi, M.H.; Ehsan, E.R. Automatic summarisation of Instagram social network posts Combining semantic and statistical approaches. *arXiv* **2023**, arXiv:2303.07957v1.
37. Murshed, B.A.H.; Abawajy, J.; Mallappa, S.; Saif, M.A.N.; Al-Ariki, H.D.E. DEA-RNN: A hybrid deep learning approach for cyberbullying detection in Twitter social media platform. *IEEE Access* **2022**, *10*, 25857–25871. [[CrossRef](#)]
38. Garg, P.K.; Chakraborty, R.; Dandapat, S.K. Ontorealsumm: Ontology based real-time tweet summarization. *arXiv* **2022**, arXiv:2201.06545.
39. Mukherjee, R.; Vishnu, U.; Peruri, H.C.; Bhattacharya, S.; Rudra, K.; Goyal, P.; Ganguly, N. Mtlts: A multi-task framework to obtain trustworthy summaries from crisis-related microblogs. In Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, Tempe, AZ, USA, 21–25 February 2022; pp. 755–763.
40. Panchendrarajan, R.; Hsu, W.; Li Lee, M. Emotion-Aware Event Summarization in Microblogs. In Proceedings of the Companion Proceedings of the Web Conference 2021, New York, NY, USA, April 2021; pp. 486–494.
41. Li, Q.; Zhang, Q. Twitter event summarization by exploiting semantic terms and graph network. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 15347–15354. [[CrossRef](#)]
42. Saini, N.; Saha, S.; Bhattacharyya, P. Microblog summarization using self-adaptive multi-objective binary differential evolution. *Appl. Intell.* **2022**, *52*, 1686–1702. [[CrossRef](#)]
43. Saini, N.; Saha, S.; Mansoori, S.; Bhattacharyya, P. Fusion of self-organizing map and granular self-organizing map for microblog summarization. *Soft Comput.* **2020**, *24*, 18699–18711. [[CrossRef](#)]
44. Saini, N.; Saha, S.; Mansoori, S.; Bhattacharyya, P. Automatic parameter selection of granular self-organizing map for microblog summarization. In *Neural Information Processing, Proceedings of the 27th International Conference, ICONIP 2020, Bangkok, Thailand, 23–27 November 2020*; Springer: Cham, Switzerland, 2020; pp. 680–692. [[CrossRef](#)]
45. Goyal, P.; Kaushik, P.; Gupta, P.; Vashisth, D.; Agarwal, S.; Goyal, N. Multilevel event detection, storyline generation, and summarization for tweet streams. *IEEE Trans. Comput. Soc. Syst.* **2019**, *7*, 8–23. [[CrossRef](#)]
46. Lin, C.; Ouyang, Z.; Wang, X.; Li, H.; Huang, Z. Preserve integrity in realtime event summarization. *ACM Trans. Knowl. Discov. Data (TKDD)* **2021**, *15*, 1–29. [[CrossRef](#)]
47. Ali, S.M.; Noorian, Z.; Bagheri, E.; Ding, C.; Al-Obeidat, F. Topic and sentiment aware microblog summarization for twitter. *J. Intell. Inf. Syst.* **2020**, *54*, 129–156. [[CrossRef](#)]
48. Abel, F.; Gao, Q.; Houben, G.-J.; Tao, K. Analyzing user modeling on twitter for personalized news recommendations. In *User Modeling, Adaption and Personalization, Proceedings of the 19th International Conference, UMAP 2011, Girona, Spain, 11–15 July 2011*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 1–12.
49. Lavanya, P.G.; Kouser, K.; Suresha, M. Efficient pre-processing and feature selection for clustering of cancer tweets. In Proceedings of the Intelligent Systems, Technologies and Applications: Proceedings of ISTA 2018, Bangalore, India, 19–22 September 2020; pp. 17–37.
50. Dusart, A.; Pinel-Sauvagnat, K.; Hubert, G. Tssubert: Tweet stream summarization using bert. *arXiv* **2021**, arXiv:2106.08770.
51. Zubiaga, A. A longitudinal assessment of the persistence of twitter datasets. *J. Assoc. Inf. Technol.* **2018**, *69*, 974–984. [[CrossRef](#)]
52. Rudra, K.; Sharma, A.; Ganguly, N.; Imran, M. Classifying and summarizing information from microblogs during epidemics. *Inf. Syst. Front.* **2018**, *20*, 933–948. [[CrossRef](#)] [[PubMed](#)]
53. Dhiman, A.; Toshniwal, D. An approximate model for event detection from twitter data. *IEEE Access* **2020**, *8*, 122168–122184. [[CrossRef](#)]
54. Lucky; Girsang, A.S. Multi-Objective Ant Colony Optimization for Automatic Social Media Comments Summarization. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*. [[CrossRef](#)]
55. Liang, Z.; Du, J.; Li, C. Abstractive social media text summarization using selective reinforced Seq2Seq attention model. *Neurocomputing* **2020**, *410*, 432–440. [[CrossRef](#)]
56. Dehghani, N.; Asadpour, M. SGSG: Semantic graph-based storyline generation in Twitter. *J. Inf. Sci.* **2019**, *45*, 304–321. [[CrossRef](#)]
57. Rudrapal, D.; Das, A.; Bhattacharya, B. A new approach for twitter event summarization based on sentence identification and partial textual entailment. *Comput. Syst.* **2019**, *23*, 1065–1078. [[CrossRef](#)]
58. Dutta, S.; Ghatak, S.; Das, A.K.; Gupta, M.; Dasgupta, S. Feature selection-based clustering on micro-blogging data. In *Computational Intelligence in Data Mining: Proceedings of the International Conference on CIDM 2017*; Springer: Singapore, 2019; pp. 885–895.
59. Dutta, S.; Das, A.K.; Bhattacharya, A. Community detection based tweet summarization. In *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2018*; Springer: Singapore, 2019; Volume 2, pp. 797–808.
60. Wei, B.; Ren, X.; Zhang, Y.; Cai, X.; Su, Q.; Sun, X. Regularizing output distribution of abstractive chinese social media text summarization for improved semantic consistency. *ACM Trans. Asian Low-Resour. Lang. Inf. Process. (TALLIP)* **2019**, *18*, 1–15. [[CrossRef](#)]

61. Wang, R.; Luo, S.; Pan, L.; Wu, Z.; Yuan, Y.; Chen, Q. Microblog summarization using paragraph vector and semantic structure. *Comput. Speech Lang.* **2019**, *57*, 1–19. [[CrossRef](#)]
62. Chakraborty, R.; Bhavsar, M.; Dandapat, S.K.; Chandra, J. Tweet summarization of news articles: An objective ordering-based perspective. *IEEE Trans. Comput. Soc. Syst.* **2019**, *6*, 761–777. [[CrossRef](#)]
63. Rudra, K.; Ganguly, N.; Goyal, P.; Ghosh, S. Extracting and summarizing situational information from the twitter social media during disasters. *ACM Trans. Web (TWEB)* **2018**, *12*, 1–35. [[CrossRef](#)]
64. Nguyen, M.-T.; Tran, D.-V.; Nguyen, L.-M. Social context summarization using user-generated content and third-party sources. *Knowl.-Based Syst.* **2018**, *144*, 51–64. [[CrossRef](#)]
65. Huang, Y.; Shen, C.; Li, T. Event summarization for sports games using twitter streams. *World Wide Web* **2018**, *21*, 609–627. [[CrossRef](#)]
66. He, R.; Duan, X. Twitter summarization based on social network and sparse reconstruction. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; p. 32.
67. Madichetty, S.; Muthukumarasamy, S. Detection of situational information from Twitter during disaster using deep learning models. *Sādhanā* **2020**, *45*, 1–13. [[CrossRef](#)]
68. Andy, A.; Wijaya, D.T.; Callison-Burch, C. Winter is here: Summarizing twitter streams related to pre-scheduled events. In Proceedings of the Second Workshop on Storytelling, Florence, Italy, August 2019; Association for Computational Linguistics: Barcelona, Spain, 2019; pp. 112–116.
69. Ma, S.; Sun, X.; Lin, J.; Wang, H. Autoencoder as assistant supervisor: Improving text representation for chinese social media text summarization. *arXiv* **2018**, arXiv:1805.04869.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.