



Review

Hashtag Recommendation Methods for Twitter and Sina Weibo: A Review

Areej Alsini ^{1,2,*}, Du Q. Huynh ¹ and Amitava Datta ¹

¹ Department of Computer Science and Software Engineering, The University of Western Australia, Perth 6009, Australia; du.huynh@uwa.edu.au (D.Q.H.); amitava.datta@uwa.edu.au (A.D.)

² Department of Computer Science, Umm Al-Qura University, Makkah 21995, Saudi Arabia

* Correspondence: areej.alsini@research.uwa.edu.au

Abstract: Hashtag recommendation suggests hashtags to users while they write microblogs in social media platforms. Although researchers have investigated various methods and factors that affect the performance of hashtag recommendations in Twitter and Sina Weibo, a systematic review of these methods is lacking. The objectives of this study are to present a comprehensive overview of research on hashtag recommendation for tweets and present insights from previous research papers. In this paper, we search for articles related to our research between 2010 and 2020 from CiteSeer, IEEE Xplore, Springer and ACM digital libraries. From the 61 articles included in this study, we notice that most of the research papers were focused on the textual content of tweets instead of other data. Furthermore, collaborative filtering methods are seldom used solely in hashtag recommendation. Taking this perspective, we present a taxonomy of hashtag recommendation based on the research methodologies that have been used. We provide a critical review of each of the classes in the taxonomy. We also discuss the challenges remaining in the field and outline future research directions in this area of study.



Citation: Alsini, A.; Huynh, D.Q.; Datta, A. Hashtag Recommendation Methods for Twitter and Sina Weibo: A Review. *Future Internet* **2021**, *13*, 129. <https://doi.org/10.3390/fi13050129>

Academic Editor: Mehrdad Jalali

Received: 5 April 2021
Accepted: 10 May 2021
Published: 14 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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/>).

Keywords: survey; hashtag recommendation; hashtags; Twitter; Sina Weibo

1. Introduction

Social media platforms have become fast-growing and influential media that enable people to communicate with each other easily, share information and search for exciting topics. The number of social media users was 3.6 billion in 2020, and this number is predicted to increase to 4.41 billion users in 2025 (<https://www.statista.com/statistics/278414/> (accessed on 10 May 2021)). Twitter (<https://about.twitter.com/company> (accessed on 10 May 2021)) is a microblogging social media platform that permits users to write and share short messages of 280 characters or less, including hashtags, mentions and URLs. These types of short messages are referred to as “microblogs” and “tweets” [1–7]. Founded in 2006, Twitter has quickly become an increasingly popular and powerful tool worldwide. According to the Internet Live Stat (<http://www.internetlivestats.com/twitter-statistics/> (accessed on 10 May 2021)), 500 million messages on average are posted per day by 330 million active users. In July 2020 (<https://www.statista.com/statistics/242606/> (accessed on 10 May 2021)), the United States had the largest audience size, with 62.55 million users, followed by Japan with 49.1 million users, and India ranked third with 17 million users. Sina Weibo (<https://www.statista.com/statistics/795303/china-mau-of-sina-weibo/> (accessed on 10 May 2021)), the Chinese equivalent of Twitter, had around 523 million active users in the same year.

With the information overload and increase in technology dependency, social recommendations have become a key research area. Social recommendation systems can be defined as techniques or algorithms that automatically suggest the most relevant and interesting data to social media users. Hashtag recommendation is a branch of the social recommendation systems that proposes contemporary and relevant hashtags to users as

they type tweets [5–23]. Choosing the correct hashtag has several benefits: it enables the user to quickly join a discussion and read tweets written by other users [24]. Using hashtags gives the user a chance for their tweets to be noticed and reach a wider audience [25]. Hashtags also help researchers to analyze users' behaviors or to predict the outbreaks of natural disasters and epidemics [26]; they are also helpful for companies to advertise their products and improve customer services and support through users' complaints and comments [27]. In politics, politicians can communicate with the public and advertise their campaigns [28]. Moreover, people can raise and share their voice nationally and globally [29]. For Twitter and Sina Weibo, recommending hashtags helps to enhance discussion as users are guided to use more accurate and relevant hashtags. Adopting the right hashtags helps Twitter/Sina Weibo to eliminate insignificant and noisy hashtags and reduce information overload. Automatically recommending personalized hashtags to users also helps them save time and effort in searching for relevant hashtags.

Recommending hashtags by analyzing tweets and extracting information from the Twitter/Sina Weibo hashtag universe can be a very challenging task. One of the challenges is that these tweets and hashtags are *user-generated*. Users tend to use informal language when writing their tweets; for example, users use "4U" to mean "for you", "AMA" for "ask me anything" and "BFN" for "bye for now". Spelling and grammatical mistakes are not checked or corrected. Short texts are therefore more difficult to analyze than long texts. The facts that tweets are short texts and are noisy add extra complication to the data. Furthermore, hashtags can be acronyms, shortened or misspelled words or a combination of words, numbers and punctuation marks. Thus, using hashtags as keywords does not necessarily convey the meaning of the discussion. The lack of control over the creation of hashtags has resulted in hundreds of hashtags being associated with a single discussion topic and different discussion topics being associated with a single hashtag.

Hashtag recommendation can be either general, when the suggested hashtags are obtained based on the data of all users, or personalized, when the suggested hashtags incorporate the user's preferences and data. The hashtags from all the tweets in a dataset form a space known as the "hashtag space". The suggested hashtags are said to be novel if they are not in this hashtag space (i.e., not previously used by other users). Otherwise, they are said to be predefined.

Although researchers have investigated various techniques and factors that affect the performance of hashtag recommendation methods for Twitter/Sina Weibo data [5–23], there is a lack of a systematic review of these methods. Jain et al. [30] listed only 11 methods in their survey on hashtag recommendations. Our paper aims to present a systematic review of research on hashtag recommendation for tweets and present insights from previous research papers together with their methodologies and factors to provide future perspectives on hashtag recommendation. Our paper comprises the following sections:

- Section 2 presents the methodology that we adopted for our literature search;
- Section 3 provides the results of the literature review based on the research questions defined in Section 2;
- Section 4 provides a taxonomy of the selected research papers on hashtag recommendation for tweets based on the methodologies adopted in the papers;
- Sections 5 and 6 give a critical overview of each class in the taxonomy and provide a detailed review of the selective methods in each class;
- Section 7 outlines our research limitations;
- Section 8 summarizes the whole paper and highlights future research directions.

2. Methodology

Based on this objective, we seek answers to the following research questions:

- **RQ1:** Has the number of research papers regarding hashtag recommendation for Twitter and Sina Weibo been increasing in the last decade?
- **RQ2:** What are the hashtag recommendation methods used in Twitter and Sina Weibo?
- **RQ3:** What are the techniques used in the hashtag recommendation methods?

- **RQ4:** What are the characteristics of the dataset used in each paper related to the research topic?
- **RQ5:** What are the future directions for hashtag recommendation methods for Twitter and Sina Weibo?

To address these research questions, we adopt the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) [31], which is an established methodological protocol for systematic reviews. The subsection below explains the eligibility criteria, information sources and searches, paper selection and results.

2.1. Eligibility Criteria

In this paper, we search for peer-reviewed papers in journals and proceeding conferences from four electronic databases: CiteSeer, IEEE Xplore, Springer and ACM. The selected papers met the following criteria:

- (i) Papers should be related to hashtag recommendation within the platforms of Twitter and Sina Weibo.
- (ii) Papers should include recommendation methods and techniques.
- (iii) Papers should include a dataset.

2.2. Information Sources and Search

Each of the four electronic databases listed above has different searching fields. For example, the advanced search in IEEE Xplore digital library enables searching keywords in many fields (such as full text, publication title, authors, and abstract) and time-frame for publication year.

Our search focuses on articles published in the English Language between 2010 to 2020. As for the searching field, we concentrate on the title and abstract in the first stage. Since the keywords used for the search are extracted from the research objectives and research questions, our selected keywords include “hashtag”, “recommendation”, “Twitter”, and “Sina Weibo”.

2.3. Paper Selection

The study selection procedure included two stages. In the first stage, we extracted articles that contain at least two keywords in the search fields (titles and abstracts). This is for criterion (i) mentioned in Section 2.1. Once the keywords were identified, we examined the selected articles to check if they met criteria (ii) and (iii) (see Section 2.1). During this stage, the selected articles that are unrelated to our study were discarded. In the second stage, papers that were collected from the first stage were reviewed in detail.

Figure 1 shows a flowchart of the steps used to identify the eligible publications from the four electronic databases and published between 2010 and 2020. After the keyword search step, a total of 2742 articles were identified. Then, 2585 papers were removed for being duplicated or unrelated to our research topic. Following that, the abstracts of the selected articles were skimmed to assess the paper’s eligibility for further examination. Next, the selected papers were screened for online platform qualification. From the pool chosen, 39 out of 139 papers were excluded from the study because they did not use data from Twitter or Sina Weibo in their studies. The following step eliminated 27 papers for not meeting criterion (iii) and removed 12 theoretical papers. In the final stage, 61 eligible papers remained after all the culling steps mentioned above. For each of the eligible papers, we documented information about the recommendation method, authors, publication year, recommendation, hashtag type, features, platforms and dataset (see Table 1).

Table 1. Summary of the 61 eligible papers.

Category	Method	Ref.	Year	Name	Recommendation		Hashtag Type		Features				# Tweets		# Users		# Unique Hashtags				
					General	Personalized	Novel	Pre-Defined	Textual	URL	Social	Temporal	Media	Other	Twitter	Sina Weibo	Twitter	Sina Weibo	Twitter	Sina Weibo	
Tweets similarity		Zangerle et al. [8]	2011	-	*		*							3,209,281	-	-	-	510,170	-		
		Li et al. [32]	2012	-	*		*							665	-	-	-	-	-		
		Zangerle et al. [33]	2013	-	*		*							50,000,000	-	-	-	7,777,194	-		
		Sedhai and Sun [34]	2014	-	*		*	*	*					1,370,000	-	-	-	1,000,000	-		
		Li et al. [35]	2016	-	*		*			*				16,000	-	-	-	2,450,000	-		
		Otsuka et al. [36]	2016	-	*		*							8,300,000	-	-	-	-	-		
		Dey et al. [37]	2017	-	*		*							175,000	-	-	-	251,649	-		
		Ben-Lhachemi and Nfaoui [38]	2017	-	*		*							295,767	-	-	-	390,807	-		
		Ben-Lhachemi and Nfaoui [39]	2018	-	*		*							1,212,300	-	-	-	-	-		
		Kaviani and Rahmani [40]	2020	-	*		*							100,000	-	-	-	-	-		
Text-based	Probabilistic	Efron [41]	2010	-	*		*							3,414,330	-	874,892	-	50,097	-		
		Mazzia and Juett [42]	2011	-	*		*							1,318,323	-	5000	-	56	-		
		Ding et al. [1]	2012	TSTM	*		*								-	-	-	-	116,958		
		Li and Xu [43]	2016	User-IBTM	*	*		*							1,385,425	-	2,094	-	-	-	
		Tariq et al. [44]	2013	-	*		*								141,881	-	-	-	143	-	
		Godin et al. [45]	2013	-	*		*								1,800,000	-	-	-	-	-	
		Ding et al. [2]	2013	TTM	*		*								-	110,000	-	-	-	37,224	
		She and Chen [46]	2014	TOMOHA	*		*								48,651	-	183	-	637	-	
		Ma et al. [11]	2014	HPM	*		*				*				1,217,928	-	13,711	-	14,055	-	
		Zhang et al. [7]	2014	TUK-TTM	*	*		*							-	166,864	-	-	-	17,516	
Classification-based		Xu et al. [47]	2015	-	*		*							-	124,707	-	6,661	-	33,777		
		Gong et al. [3]	2015	CNHR	*		*								-	1,118,792	-	-	-	305,227	
		Lu and Lee [48]	2015	TOT-MMM	*		*			*					741,317	-	-	-	16,839	-	
		Gong et al. [4]	2016	PTTM	*		*								-	50,000	-	-	-	3,174	
		Tomar et al. [49]	2014	-	*		*								-	-	-	-	-	-	
		Chen and Kao [50]	2015	TMSHR	*		*								226,981	-	-	-	-	-	
		Ghaly et al. [51]	2016	-	*		*								627,084	-	-	-	7,961	-	
		Li et al. [52]	2016	LSTM-tweet	*		*								1,000	-	-	-	-	-	
		Li et al. [53]	2016	TAB-LSTM	*		*								42,000	-	5,015	-	20	-	
		Ma et al. [54]	2018	tSAM-LSTM	*		*								600,000	-	-	-	-	-	
Graph-based		Li et al. [55]	2019	TCAN	*		*							-	1,692,507	-	-	-	2,000		
		Peng et al. [56]	2019	AMEN	*		*								600,000	-	-	-	27,720		
		Khabiri et al. [9]	2012	-	*		*								127,846	-	2000	-	3,104		
		Ferragina et al. [57]	2015	HE-graph	*		*				*				36,558,421	-	-	-	134,522		
		Al-Dhelaan and Alhawasi [14]	2015	-	*		*								5,245	-	-	-	5,245		
		Li et al. [58]	2015	-	*		*								27,199	-	15,586	-	10,891		
		Matrix Factorisation	Badami and Nasraoui [59]	2018	-	*		*							-	74,662	-	-	-	-	
		Hybrid user-based		Diaz-Aviles et al. [60]	2012	RMFX	*	*		*						208,160	-	-	-	68,187	-
				Chen et al. [10]	2013	TeRec	*	*		*			*			35,350,508	-	413,987	-	37,297	-
				Kywe et al. [61]	2012	-	*		*								20,000,000	-	87,287	-	29,334
Xing et al. [62]	2013			-	*		*								3,534,869	-	65,410	-	449,206		
Wang et al. [63]	2014			-	*		*								-	-	-	-	-		
Zhao et al. [64]	2016			-	*		*								-	21,992	-	12,156	-		
Li et al. [17]	2017			-	*		*								1,674,789	-	14,630	-	28,526		
Wang et al. [65]	2019			-	*		*								15,947	-	20,625	-	11,347		
Kou et al. [66]	2018			-	*		*								37,533	-	22,849	-	-		
Social Collaborative Filtering				Harvey and Crestani [67]	2015	-	*		*			*				-	67,835	-	4,373	-	4061
		Kowald et al. [18]	2017	BLL	*	*		*		*				333,784	-	23,476	-	51,899	-		
		Alsini et al. [16]	2017	-	*		*							8,157,702	-	127,112	-	1,507,773			
		Alsini et al. [19]	2019	-	*		*							174,965	-	100	-	2,655			
		Javari et al. [22]	2020	PHAN	*	*		*			*				-	745,262	-	-	-		
		Alsini et al. [23]	2020	CBHR	*	*		*			*				217,965	-	23,169	-	7,023		
		Li and Zhang [68]	2013	-	*		*								262,178	-	125,708	-	-		
		Feng and Wang [13]	2014	Hybrid+	*	*		*	*	*	*				-	-	-	-	-		
		Jeon et al. [12]	2014	-	*		*								240,000	-	80	-	404		
		Hybrid	Miscellaneous	Song et al. [5]	2015	SenSim+Ac+Te	*	*		*			*			-	50	-	-	-	-
Yu and Zhu [6]	2015			Linear	*	*		*			*				-	-	-	-	-		
Zhang et al. [15]	2017			-	*		*					*			4,357,329	2,899,438	1,934,381	108,956	-		
Ma et al. [20]	2019			CoA-MN	*	*		*			*				402,782	-	-	-	3,292		
Belhadi et al. [21]	2020			PM-HRec	*	*		*			*				334,019	-	-	-	3,280		
Kumar et al. [69]	2020			-	*		*								4,000,000	-	-	-	-		
															-	-	-	-	-		
															-	-	-	-	-		
															329,369	-	-	-	85,216		

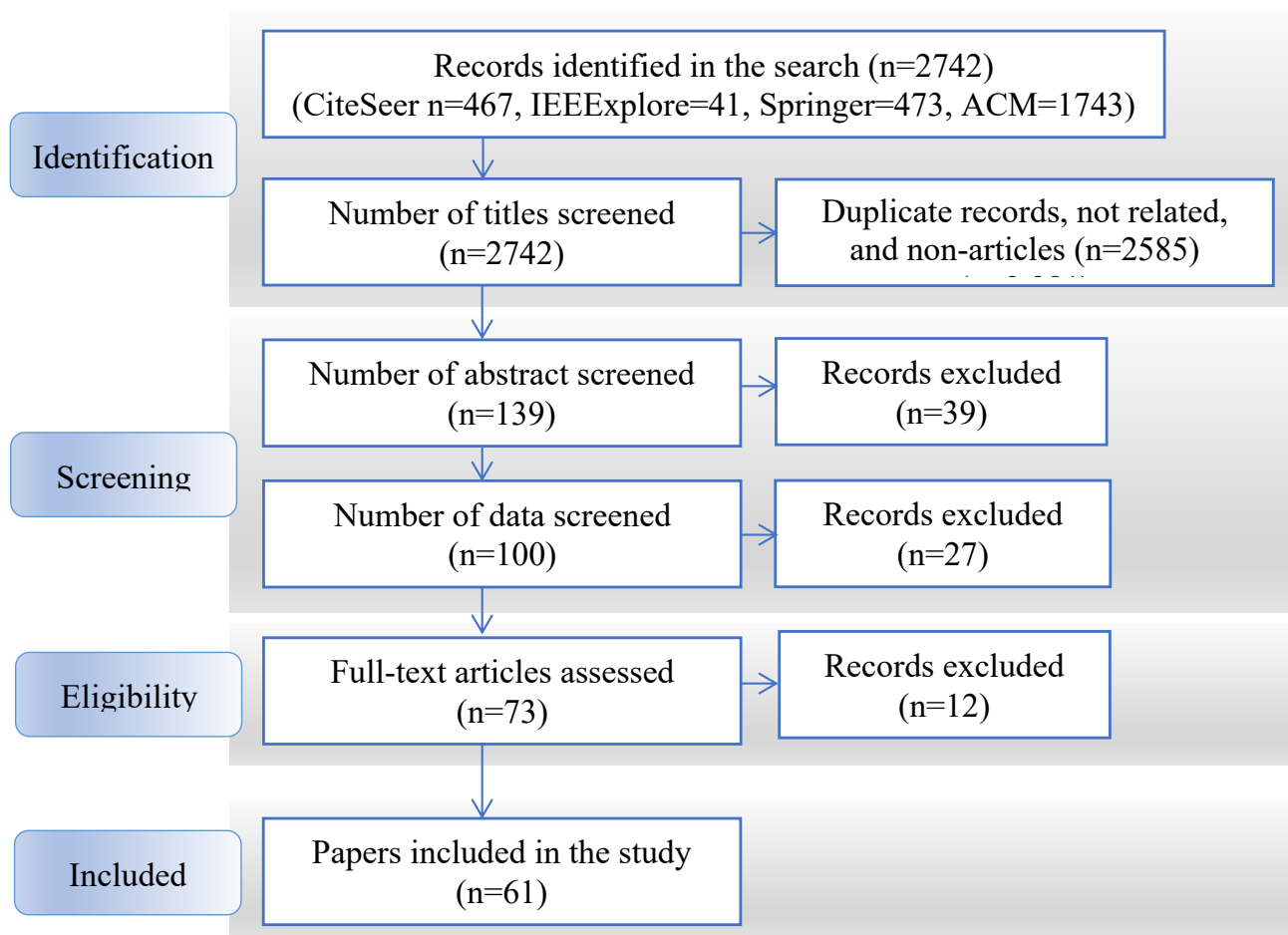


Figure 1. Flowchart of the steps used in the literature search.

3. Results

We analyse the eligible papers based on their years of publication, methods, techniques and dataset sizes.

3.1. Year of Publication

Figure 2 shows the distribution of the 61 papers, obtained from the paper selection procedure described above, based on their years of publication. We can see a trend of the increasing number of publications in the last decade from our list of papers, and this should be a good representative of the growing number of publications in the field in all journals.

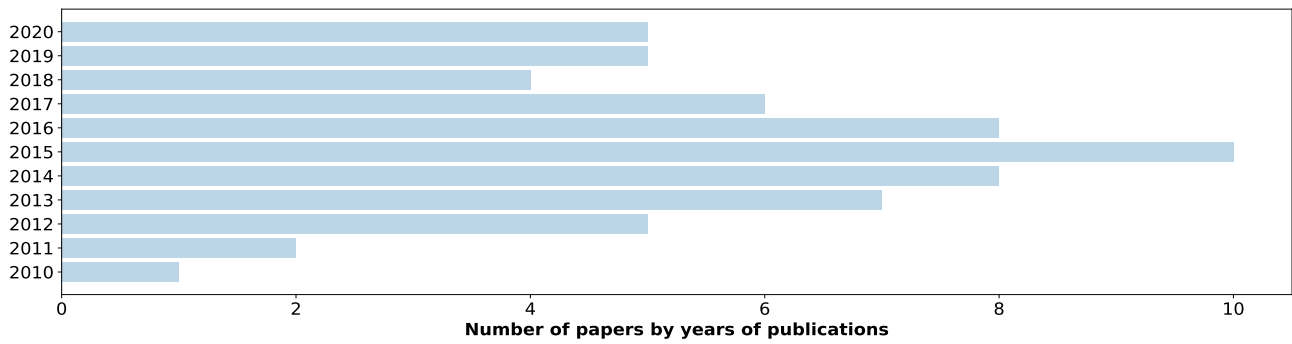


Figure 2. Distribution of papers from 2010 to 2020.

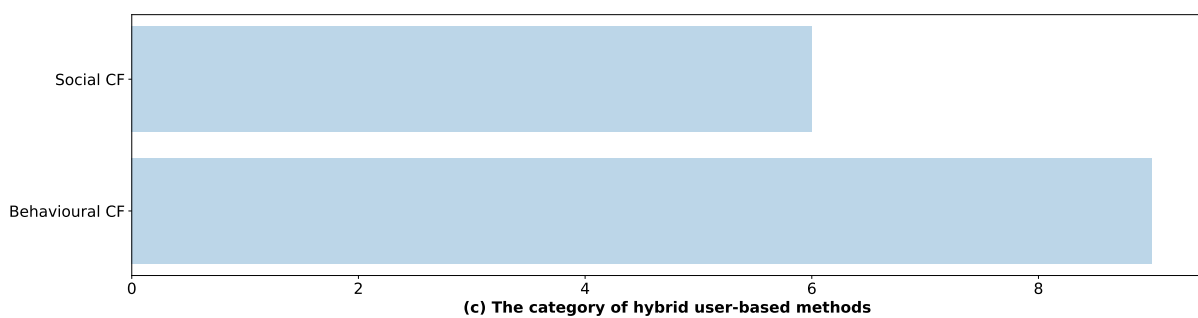
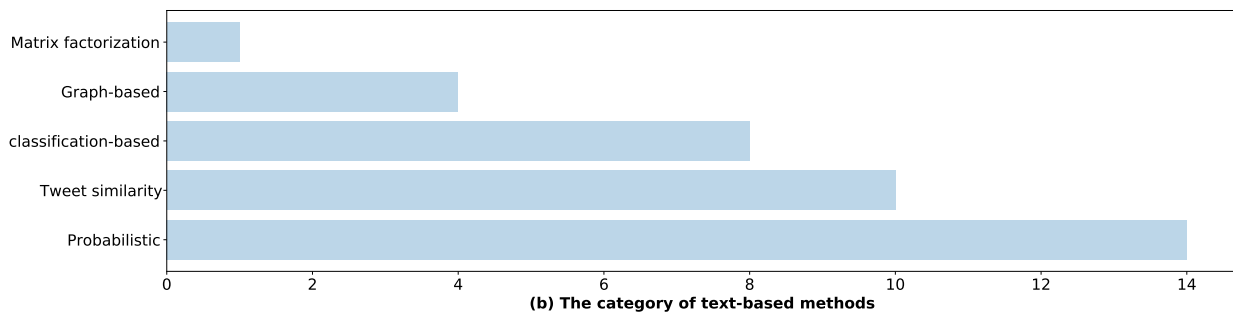
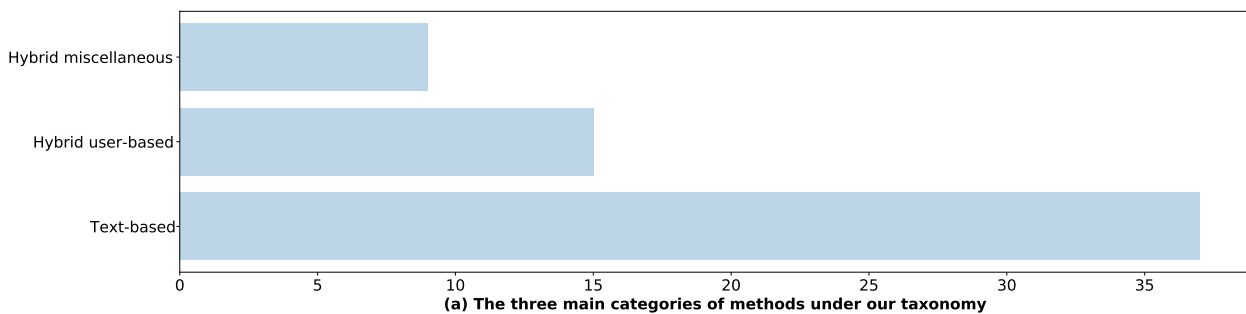


Figure 3. Distribution of the eligible papers on hashtag recommendation from 2010 to 2020 with respect to: (a) the three main categories of methods under our taxonomy; (b) the category of text-based methods; and (c) the category of hybrid user-based methods.

3.2. A Taxonomy of Hashtag Recommendation Methods

This subsection addresses RQ2 and RQ3 research questions. There is a lack of a clear taxonomy in the literature that classifies hashtag recommendation methods for tweets. From our literature search, we notice that most of the research papers were focused on the textual content of tweets instead of other data. Moreover, although collaborative filtering methods are the most popular category in the traditional/social recommendation systems, they are seldom used solely in hashtag recommendation. Taking this perspective, we propose an alternative way of classifying hashtag recommendation methods. Figure 3 shows the distributions of the eligible hashtag recommendation papers from 2010 to 2020 with respect to the methods classified under our taxonomy, which comprises three main categories (see Figure 4):

- text-based methods,
- hybrid user-based methods; and
- hybrid miscellaneous methods.

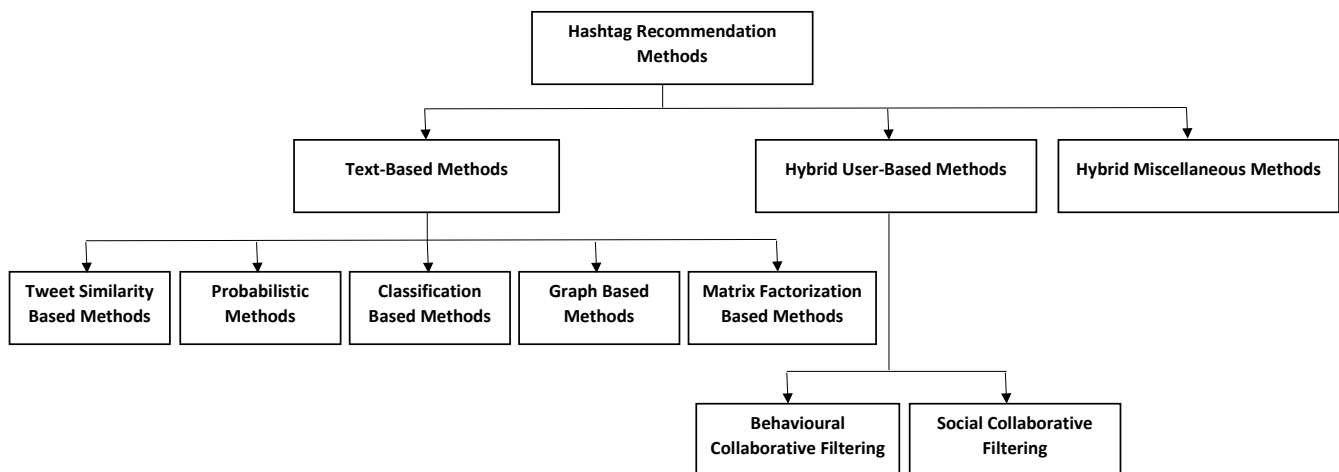


Figure 4. Our proposed taxonomy of the hashtag recommendation methods.

Text-based methods mainly depend on the textual content of the tweets, such as words, hashtags, URLs, and mentions of all users. This category is further classified into tweet similarity methods, probabilistic methods, classification based methods, graph-based methods and matrix factorization based methods. Hybrid user-based methods involve those that group related users either by their behavioral similarity or social relations to recommend hashtags. Hybrid miscellaneous methods include methods that integrate multiple modalities and factors. The included papers are listed in Table 1. Some methods incorporate more than a feature and so they have more than one * symbol in the table.

3.3. Dataset

This subsection addresses research question RQ4. There is an absence of benchmark datasets in this area of study. Each research study in the previously mentioned literature crawls their data. Moreover, the characteristics of collecting these data are also different. These characteristics include the number of tweets, the number of users and overall/unique hashtags. For example, the number of tweets in the dataset used in tweet-similarity-based methods ranges between 665 and 50 million tweets. The number of tweets in the dataset used in graph-based methods ranges between 5245 and 36,558,421 tweets. Similarly, the number of users varies between 80 and 1,934,381 users. Furthermore, the seed users that were used for collecting the data were also random. With the lack of benchmark datasets

and unified data collection characteristics, it is not easy to compare the performance of different algorithms.

4. Text-Based Methods

Text-based methods form the largest category, comprising five subcategories, as shown in Figure 1. These five subcategories are detailed in the subsections below.

4.1. Tweet Similarity Based Methods

Tweet similarity methods used in hashtag recommendation find similar tweets to a query tweet using similarity measures [8,32,33]. The tweet similarity method was first developed in hashtag recommendation by Zangerle et al. [8], who used the TF-IDF approach to represent tweets as one-hot vectors. For a query tweet, the set of similar tweets was retrieved using the TFIDF score, where the TFIDF score was calculated for a tweet by summing up the weights of its terms. Their research found that the hashtag extracted from the most similar tweet is the best candidate for recommendation. Zangerle et al. [33] further improved their method and found that the choice of the similarity measure affected the recommendation's quality, with the cosine similarity measure outperforming the Dice coefficient, Jaccard coefficient and Levenshtein distance.

Other works used the same technique but with different tweet representations and similarity measures. Li et al.'s work [32] was based on the intuition that *tweets can be treated as points in a high-dimensional space and that the correlation between them can be calculated mathematically*. They transformed tweets into a high-dimensional space and constructed a semantic matrix that weights words based on their correlation. They used the Euclidean distance to measure the similarity between tweets. Otsuka et al. [36] proposed a variation of the TFIDF that considered the hashtag relevancy with terms based on two maps: term-to-hashtag frequency map (THFM) and its reverse map of a hashtag-to-term frequency (HFM). This method took into consideration only the hashtag and word co-occurrences while ignoring their meanings.

More recent work has integrated word and sentence embedding techniques in hashtag recommendation. Ben-Lhachemi et al. [39] encoded the tweets using the method proposed in [70] and clustered similar tweets using DBSCAN. For a query tweet, they calculated the distances between a query tweet and the centroids of the tweet clusters and extracted candidate hashtags from the closest cluster and put forward recommendations based on the hashtags' popularity values. With the increase in popularity of BERT, it has been used by Kaviani et al. [40] in generating the embedding of tweets.

While most previous research leverages words and hashtags, one paper leverages the *hyperlinked tweets* which are tweets containing URLs. Sedhai et al. [34] recommended hashtags to hyperlinked tweets through utilizing the content of tweets and the content of the web pages using various schemes: similarity of tweets; similarity of the linked web pages; similarity of the link domain; and similarity of the named entity in the linked web pages. They applied TFIDF to encode the tweets and web pages and used cosine similarity to calculate the similarity scores. Hashtags were then aggregated and ranked using RankSVM, a learning-to-rank method.

Some other researchers calculate the similarity between a tweet and similar hashtags [35,37,38]. Ben-Lhachemi et al. [38] computed the semantic similarity between a tweet and existing hashtags using the spreading activation technique from multiple semantic networks such as WordNet, DBpedia and Wikipedia. This method expected the words of the tweets and hashtags to be in a formal English form. EmTagger [37] is another hashtag recommendation model that derived every hashtag's embeddings by learning the embeddings of words associated with these hashtags. For a given query tweet, the MOWE was used to encode its embeddings. Then, the cosine similarity was used to measure the similarity between a query tweet and hashtags. The tweets with the most similar hashtags were retrieved, and their hashtags were extracted and ranked based on their embeddings score for recommendation.

Li and Shah et al. [35] learnt the topical word embedding of words in a tweet using Topic Enhanced Word Embedding (TEWE). The embedding of each word generated using TEWE was the concatenation of its feature vector generated using LDA and the feature vector generated using skip-gram of the Word2Vec model. Then, the cosine similarity was used to measure the distance between a tweet with the closest hashtags. To weight their candidate hashtags, they estimated whether there was a shared entity between terms or sub-terms in tweets and hyperlinks. Hashtags with shared entities were ranked based on their temporal popularity. It was found that the entity-matching part did not contribute much to improving the performance of hashtag recommendation.

The main drawback with tweet-similarity methods is that a query tweet has to be compared with every other tweet in the repository to retrieve the most similar tweets. This can be expensive and time-consuming. As for the tweet-hashtags similarity methods, the major challenge is computing the embeddings of hashtags to measure their similarity with tweets.

4.2. Probabilistic Methods

Naive Bayes is a machine learning technique based on the Bayes Theorem. Mazzia et al. [42] proposed one of the earliest general hashtag recommendation models using Naive Bayes to calculate the conditional probability of a hashtag given a set of terms in the tweet as: $P(h|t_1, t_2, \dots, t_n) = P(t_1, t_2, \dots, t_n|h)P(h)/P(t_1, t_2, \dots, t_n) = P(h|t_1)P(h|t_2) \dots P(h|t_n)$. The result is a value between 0 and 1. The hashtags with the highest probabilities were recommended. However, some terms and hashtags never co-occur, leading to a joint conditional probability of 0 even if some words have a high probability score.

Topic modelling methods are statistical methods that infer the topics of documents. The inference of topics is derived from the hidden patterns in the corpus. Inspired by the LDA's popularity and its success in deriving topics from large texts, topic-based hashtag recommendation methods manipulate the standard LDA under various assumptions. Godin et al. [45] assumed that *the topics associated with tweets are sufficient to recommend novel hashtags*. Thus, they proposed a general hashtag recommendation model that determined the topics of a tweet, considering the nouns and adjectives, and chose the top topic-terms for recommendation. Although LDA has proven to be a valuable technique for large corpora [71], Yan et al. [72] and Yin et al. [73] have proven that using the standard LDA is not efficient on short texts such as tweets [45]. The terms in tweets are noisy, and word co-occurrence patterns are very sparse [72].

To enhance the quality of the topics generated using LDA and to overcome the problem of short texts in tweets, different schemes that aggregate tweets have been proposed [74,75]. Usually, such a scheme is applied at the pre-processing phase before training the model. The aggregation schemes that have been investigated in the literature are:

- The *all-tweets scheme* [75,76], which aggregates all tweets in a single document and then uses LDA to extract topics;
- The *author-based scheme*, which aggregates all tweets posted by a single user and treats them as a single document [75,77];
- The *burst-score wise scheme* [75], which aggregates tweets that contain trending terms;
- The *temporal scheme* [75], which aggregates tweets posted at a specific time;
- The *term scheme* [77], which aggregates tweets that share a word—it is applied to every word in the training set;
- The *conversation scheme* [74], which aggregates tweets based on conversations between users, e.g., replies to a tweet sent by other users (co-authors) or by the primary author.

According to Alvarez-Melis et al. [74], the best topic quality was generated when the LDA model was trained using the author-based scheme followed by the hashtag-based scheme and then the conversation scheme. Following these findings, Alsini et al. [19] adopted the author-based scheme in their hashtag recommendation model and reported good results.

In contrast, She et al. [46] tackled the noisiness of words in tweets. They assumed that *every tweet is a single local topic, and that there is a global topic for the corpus*. They

optimized Twitter-LDA [78], which generates topics for tweets, and proposed TOMOHA. This supervised topic-based hashtag recommendation model considers hashtags as labels of topics and calculates a hashtag's probability based on its local and global words. They also proposed TOMOHA-follow, which extracts the relationship between topics of a tweet, hashtags and the followees topics, assuming that *users follow the topics of their followee*. TOMOHA and TOMOHA-follow achieved similar performances and outperformed the standard LDA.

It is difficult for topic-based hashtag recommendation methods to suggest the correct hashtags because of the so-called *vocabulary gap problem*, which is usually caused by the diversified words in tweets and hashtags. Thus, machine translation models were adopted in hashtag recommendation mainly to solve this problem. Most of the proposed hashtag recommendation methods based on machine-translation worked under the assumption that *each tweet and the associated hashtags have parallel descriptions but are written in different languages* [2,79]. Ding et al. [1] introduced the topic-specific translation model (TSTM), which integrated the use of the LDA and a *word alignment technique*. Their paper's main idea was to promote the essential words in each tweet to trigger the corresponding hashtags and assess the topic-specific word alignment probabilities between a tweet and a hashtag. Ding et al. [2] later improved their original model and called it TTM. In their previous model, TSTM, they assumed that there are k topics. In their optimized model, TTM, they believed that there are $k + 1$ topics, where the additional topic includes the *background words* that are mostly used in most of the tweets but do not belong to a specific topic. Their optimized model outperformed the original model.

Rather than using the LDA, Gong et al. [3] adopted a Dirichlet Process Mixture Model (DPMM) and implanted a hidden variable that represented the hashtag type. Their model DPMM applied the Hierarchical Dirichlet Process (HDP) to automatically capture an infinite number of topics from the data instead of specifying it in advance as in the standard LDA.

The hashtag recommendation models mentioned above are general. They ignore user information, preferences, and global trends such as temporal information. TUK-TTM [7] incorporated personal and temporal factors as hidden variables. The personal variable was calculated by computing the probabilities of the topics that the user was interested in. The temporal aspect was calculated by computing the probabilities of the global trend at a certain period. TUK-TTM was trained on the optimized LDA on aggregated tweets where tweets posted in a month were considered a single document. They assigned a hidden variable to denote the topic's probability for a tweet and used the translation model to get the hashtags. They found that topics vary over time.

Unlike the above word-based topical translation models, which treat tweets on the word level, Gong et al. [4] proposed a phrase-based topical translation model to perform the hashtag recommendation task. They refer to their method as PTTM (abbreviated for Phrase Topical Translation Model). They assumed that *a phrase is a natural unit that conveys a complete topic and contains real meaning that a human being wishes to express*. Under this assumption, they aligned phrases in a tweet with hashtags. They adopted a phrase extraction technique that retrieves frequent phrases from the corpus and merges the words in a given tweet into phrases. They treated each tweet as a *bag of phrases* rather than a *bag of words* as used in most of the LDA based models and assigned a single topic to each tweet. This method outperformed all previously mentioned models with the number of topics ranging from 10 to 30.

In contrast, Ma et al. [11] studied the social factor's impact through user mentions under the assumption that *users who mention each other more often are most likely to adopt similar hashtags*. Their PLSA-style model was regularized to study the importance of the social factor. They proposed the Content-Pivoted Model and the Hashtag-Pivoted Model. The former model worked under the assumption that *the user figures out the hashtag while he/she writes a tweet*, whereas the latter assumed that *the user has a pre-selected hashtag in mind*. The Content-Pivoted Model generated the topics of the tweet considering the user topic

distribution and the time topic distribution. The Hashtag-Pivoted Model generated the topics of the tweet considering an additional factor, which was the hashtag distribution. It was found that the Hashtag-Pivoted Model outperformed the Content-Pivoted Model. In both models, existing hashtags were recommended. They also highlighted that the topic models could be infeasible in determining the common topics between users who mention each other. This method also under-performed the collaborative filtering method [61], which integrates hashtags from the set of similar users.

Topics-based models have some drawbacks. The limited number of words in tweets, the noisiness in them and the diversification between words in tweets and hashtags prevent topic models from generating meaningful topics. Thus, aggregating short texts as a document before the training can provide topics with high quality. Moreover, aligning tweets and hashtags using topical-translation models can remedy the words gap problem. However, the topics in most of the aforementioned topic-models are determined in advance. In reality, the underlying topics discussed in tweets can greatly vary. In addition, topic models do not include the effect of social factors such as relations and behavior similarity. Finally, topic-based models can be computationally expensive with large datasets [49].

4.3. Classification Based Methods

Classification is a supervised learning method that trains a classifier to predict a class label given an instance. While binary classifiers tackle only two categories, multi-class classifiers tackle multiple categories. Hashtag recommendation is commonly tackled as a multi-class classification problem of hashtags [49,52,53,56], where every hashtag is considered as a distinct class label. The intuition of the classification based hashtag recommendation is that *the abundance in posts and hashtags equip classifiers with an immense amount of labelled data to learn strong representations* [80]. Classification-based hashtag recommendation requires less *task-specific assumption and engineering* in comparison with topic-based hashtag recommendation [80].

Earlier research has applied neural networks and deep learning techniques to predict hashtags. Most of the studies in this area focus only on the internal structure of the words in the tweets to automatically learn their representations and then measure their relevance to hashtags. Tomar et al. [49] learned the word distributions using the pre-trained skip-gram model word2vec on the Google News data set. The mean of the 300-dimensional feature vectors for all the words in a tweet was used to encode the tweet feature vector, which was used as an input to train deep feedforward neural networks. The Mean Squared Error (MSE) was used as an objective function to generate the corresponding hashtag feature vectors. Li and Liu et al. [53] proposed the topical attention-based LSTM model (TAB-LSTM). They generated local representations of tweets using LSTM and combined them with global representations obtained from the LDA model through an attention mechanism. A softmax classifier was attached at the top of the model to predict the hashtags. Li et al. [52] proposed a hashtag recommendation method using a Long-Short Term Memory Recurrent Neural Network (LSTM-RNN). The tweets feature vectors were represented using CNN on words pre-trained using Word2Vec skip-gram model. The generated tweets' feature vectors were then fed into their LSTM-RNN to classify hashtags. The MSE was used as a cost function between the target and the predicted hashtags. Although this method achieved the highest performance over the other techniques mentioned earlier for the top-10 hashtag recommendations, the number of hashtags was limited to 20, which restricted the applicability of the method. Despite all previous methods, Peng et al. [56] proposed the Adaptive neural MEory Network (AMEN), which models the past tweet history of the user. AMEN encoded tweets using a convolutional layer and hashtags using a recurrent neural network.

A classification based hashtag recommendation can be an impractical choice of methodology. Hashtags are noisy labels for legitimate classification; i.e., a misspelled hashtag is considered a different hashtag. Hashtag distribution follows the long tail distribution, where most hashtags are low in popularity, and very few hashtags are high in

popularity. By using every hashtag as a label, the classifier is more biased toward the more popular hashtags. Therefore, the result of the classification will be inaccurate. In all classification-based hashtag recommendation mentioned earlier, the number of hashtag labels is restricted, which contradicts the nature of hashtag generation in social media. In addition, none of these classifiers was personalized.

4.4. Graph Based Methods

Graph-based methods structure data as a network of vertices and edges, where vertices represent entities and edges represent the relationship between these entities. The edges can be weighted to determine the strength of the relationship. In graph-based hashtag recommendation, the entities can be hashtags, terms of tweets, and users. The assumption of the research adopting graph-based methods in hashtag recommendation is that *terms in tweets and hashtags have implicit relationships, forming a semantic graph; using this graph can relate terms to hashtags even if they never co-occur* [9].

As such, Ferragina et al. [57] built the HE-graph, a heterogeneous graph of hashtags and Wikipedia terms drawn from tweets. HE-graphs draw directed hashtag–term edges and undirected term–term edges derived from Wikipedia graph structure. Each hashtag is linked to a set of terms that represents its semantic. Using TFIDF and various similarity measures, the relevance scores between hashtags were determined. Al-Dhelaan et al. [14] added another dimension and built a heterogeneous graph of hashtags, tweets and users. This graph was further summarized into a homogeneous graph of hashtags. In their paper, two methods were used to calculate the relevance between hashtags: Random Walk with Restart method (RWR) and cosine similarity. The final ranking of candidate hashtags combined the RWR weight and the cosine similarity score. Khabiri et al. [9] built a hashtag prediction that can find the most significant terms in the tweet, weigh their closeness to each hashtag, and use a temporal sliding window to reflect the terms and hashtags within a time frame. They used IDF and entropy methods to find significant words. They found that IDF and entropy methods contributed similarly to the final recommendation performance. They also found that a week’s time frame generated the best results over an hour, four hours and a day for the temporal sliding window.

4.5. Matrix-Factorization-Based Methods

Matrix factorization is a group of algorithms that decompose a matrix into the product of two smaller matrices with reduced dimensions. In hashtag recommendation, matrices have been constructed following the assumption that *although hashtags and words are both made of terms, hashtags are typically used with a different intention and serve a different purpose* [59]. Thus, Badami et al. [59] constructed their matrix based on the hashtag–terms of tweets. They proposed a hashtag recommendation model using non-negative matrix factorization (NMF) to generate tweets’ latent features. To find hashtags relevant to a given tweet, tweets containing these two hashtags were retrieved, and their average cosine similarity score was calculated.

5. Hybrid User-Based Methods

Collaborative filtering is a user-based method that recommends items based on other users’ collaborative behaviors, the similarity of their interests, shared topics or social relations. The set of similar users are also called like-minded users. Although collaborative filtering is a large category in the traditional recommendation systems, it was rarely used as a sole method for recommending hashtags due to data sparseness and the cold start problem. Data sparseness is caused by the free style of writing hashtags and the increasing number of users. The cold start problem is usually caused if the user has no historical activities, making it becomes difficult to find like-minded users. Thus, hybrid collaborative filtering methods integrate collaborative filtering methods with other modalities in hashtag recommendation. It can be divided into two categories: *behavioral collaborative filtering* and *social collaborative filtering*.

5.1. Behavioural Collaborative Filtering

Some researchers have assumed that *if the users shared similar hashtags/topics in the past, most likely they will share similar hashtags/topics in the future*. Following this assumption, Diaz-Aviles et al. [60] and Chen et al. [10] proposed personalized hashtag recommendation models in the following manner: they created a user-hashtag matrix and assigned 1 if a user has adopted a specific hashtag in the past. Using an optimization method, they generated a hashtag adoption score of every user to every possible hashtag. To address the data sparseness problem, which happens when the numbers of hashtags and users increase, they updated the rates of using hashtags from random tweets sampled at certain times. They, however, used different sampling techniques. The previously mentioned methods suffer from the cold-start problem.

Content-based methods can remedy the cold-start problem that arose from the collaborative filtering methods. Kywe et al. [61] proposed a personalized hashtag recommendation model that involved integrating content similarity and collaborative filtering to extract candidate hashtags. This method profiles every user according to his/her historical hashtag usage using TFIDF. Given a user, hashtags extracted from like-minded users and similar tweets were retrieved, ranked, and recommended. Kou et al. [66] recommended hashtags based on the weight of combining three features: content similarity, collaborative filtering of users with similar hashtag usage and topical interests.

To overcome the cold-start problem, in Wang et al.'s paper [63], the recommended hashtag for a given tweet was extracted from the tweets of users sharing similar topics. They generated the topic representation of a tweet by training the LDA model on aggregated tweets containing the same hashtag. Zhao et al. [64] adopted a similar approach and combined it with users using similar historical hashtags if present in the database. They found that a smaller number of users is required to get a high hashtag recommendation performance.

5.2. Social Collaborative Filtering

Some other researchers assumed that *there is a shared interest of topics between a user and users whom he/she follows*. Harvey et al. [67] integrated the candidate hashtags extracted from the content similarity method [33] with candidate hashtags of the most similar users from the set of followees. He also incorporated a temporal weight to rank the hashtags based on the entropy score and the hashtags' age. They found that hashtags with a small entropy and age values are less likely to be reused and therefore should be penalized. Kowald et al. [18] put forward *Base Level Learning* (BLL), a method that recommended the most recent and frequent hashtags placed in the user's tweets and his/her followees. In their experiments, they found that the social factor improves the recommendation performance.

Javari et al. [22] looked into the hashtag recommendation from a different angle. They built PHAN, a graph-based model of representative users and hashtags. They built their model based on the notion that *when a node user becomes popular, the hashtag represents a topic of interest*. As such, a represented node user was selected if his/her followers exceeds a given threshold. Accordingly, the representative users and hashtags were projected into latent space using Generalized Matrix Factorization (GMF), a Neural Collaborative Filtering method. GMF is a multi-layer neural network that learns user-item relationships for recommendation. They optimized GMF by setting an attention weight that learned the influence of a representative node on a user with respect to a hashtag. With the absence of labels for the representative users, they trained their model based on weak supervision of estimated labels to every representative node.

Although several studies [18,22,67] have highlighted the importance of user relations (behavioural and social), the degree of their influence is not clear. Alsini et al. [16,19,23] present an overview of the community detection algorithms as techniques used to group like-minded users. The influence of four relations based on the hashtag usage, topics, followees and mentions on the performance of hashtag recommendation over communities

have been investigated [23]. It was found that the level of social relations affects the performance of hashtag recommendation.

6. Hybrid Miscellaneous Methods

In this category of research, multi-modalities and multi-factors are incorporated to recommend hashtags under various assumptions. Joen et al. [12] assumed that *users frequently tweet about topics they are interested in*. Thus, their hashtag recommendation model determined the relationship between the significant terms and hashtags by their shared topics. With 16 predefined class labels (art and design, books, business, etc.), they classified significant terms in a tweet into one of these class labels using Naive Bayes. To personalize their model, they extracted the significant terms from every user and classified them into one of the predefined class categories. They ranked candidates' hashtags based on the similarity of the terms to the hashtags, user interests, and popularity. They found that the performance of the hashtag recommendation on sports tweets outperformed the performance on tweets in other categories, and the performance on news tweets was the poorest compared to all other categories.

SenSim+Ac+Te is a hashtag recommendation model [5] that incorporated three factors: similarity of a hashtag to a tweet, user acceptance degree and the development tendency of a hashtag. They used a modified version of LDA called JGibbLDA as the base of their hashtag recommendation model. They first generated the topical feature vectors of tweets and hashtags. Then, they calculated their similarity. Their research observed that given two hashtags that were different but relevant in the sense that they were about the same topic or event, people usually tended to use the shorter hashtag. Under this observation, they gave the shorter hashtag a higher probability and thus a higher ranking score. They calculated the user acceptance score for a hashtag h as $\log(N_h)$, where N_h was the number of tweets containing the hashtag h . As for the development tendency, they used Polynomial Spline Estimator and Sigmoid function to detect the shift in hashtag trends over seven days. Hashtags with a similar development tendency and higher user acceptance had higher ranks for the recommendation.

The linear model and Topic-STG are two personalized hashtag recommendation models introduced by Yu et al. [6]. These two models were based on the assumption that *user behaviors on hashtags present the same meaning as the operations on tweets containing these hashtags*. These operations included the number of 'retweet', 'create', 'comment', 'favorite' and 'add friend' actions. They also assumed that *tweets and hashtags with a larger number of operations are popular* and that this popularity changes over time. According to these two assumptions, in the linear model, they calculated the content similarity of the user interest with the content of hashtags using the negative symmetric Kullback–Leibler divergence where the user's interest is generated using LDA. They also classified hashtags into *time-sensitive* and *time-insensitive* and calculated the popularity scores for a hashtag. Based on these three variables, they recommended the top- k hashtags. In the Topic-STG model, they assumed that the user's interest varies over time. On that basis, they proposed a directed and weighted graph-based model of users, topic, hashtags and session time. Using the session time, they determined the user's short- and long-term preferences. Their Topic-STG model outperformed the linear model but has a higher complexity with respect to an increase in the number of topics. It was also difficult to incorporate additional features.

Kumar et al. [69] incorporated external knowledge from Wikipedia and news articles extracted from the web. Feng and Wang [13] proposed Hybrid+. This optimization-based framework includes features related to tweets (terms, URLs and mentions), users (ID, location and social relation), and hashtags (length, frequency, uptrend and time). They found that a user's interest can change within a week. Moreover, links, mentions, and location were a weak indication for hashtags. However, the integration of all features improved the results.

Media of images were also investigated in the context of hashtag recommendation [15,20]. Zhang et al. [15] noticed that a *hashtag is only related to a specific region of*

an input image. Thus, they proposed a co-attention network that incorporated textual and visual information to recommend hashtags. They trained their network as a multi-class classification problem and used LSTM to extract the textual features and the VGG network pre-trained on the 1000 class objects to extract the visual features. It was found that the visual and textual factors influence the recommendation. However, the images in the dataset were diverse and mostly graffiti, selfies or smartphone shots. Ma et al. [20] also proposed a co-attention memory network (CoA-MN). However, they incorporated the historical tweets of hashtags to represent these hashtags. CoA-MN achieved results better than classification-based methods.

7. Limitations

As we restricted our search of papers from only four electronic databases (CiteSeer, IEEE Explore, Springer and ACM), we expect many more papers on hashtag recommendation have been published in other journals.

Due to the absence of benchmark datasets, hashtag recommendation methods' performance cannot be compared in the same way as in other research areas. This problem is attributed to the legal terms set by the social media companies that do not permit data sharing. Thus, this paper has not identified the superiority of methods over others but has given only critical discussions about hashtag recommendation methodologies.

8. Conclusions and Future Research Directions

This paper discusses how hashtag recommendation systems for tweets have evolved within the field of online social networks. It also presents a new taxonomy for hashtag recommendation of tweets based on their methodologies. The taxonomy classifies hashtag recommendation methods for tweets into three main categories: text-based, hybrid user-based and hybrid miscellaneous methods. Text-based methods find hashtags similar to what a user intends to adopt based on the textual information. This category is further classified into tweet-similarity-based methods, probabilistic methods, classification based methods, graph-based methods and matrix factorization based methods. Since methods of collaborative filtering suffer from the cold-start problem, they are integrated with other methods. Hybrid user-based hashtag recommendation methods recommend hashtags based on the similarity of the users' behavior, interests or relations. This category is further classified into behavioral and social collaborative filtering methods. Hybrid miscellaneous hashtag recommendation take advantage of multi-modalities and multi-factors to recommend the hashtags. Regardless of the specific techniques employed, it has become clear that the best outcome can be achieved using the hybrid methods (user-based or miscellaneous) for their ability to overcome problems occurring with content-based and collaborative filtering methods. It was noticed that understanding various factors that affect the performance of hashtag recommendation and the underlying assumptions have a significant impact on the algorithmic approach that should be considered.

This subsection also addresses the research question RQ5. We highlight some open challenges, which can be considered future research directions. These challenges are as follows:

- Despite the advancement of the current methods, further improvements are required to propose more effective methods that are less expensive in terms of time and computation and provide a personalized recommendation that covers a broader range of pre-defined and novel hashtags with higher accuracy. Furthermore, most of the previous research was tested offline. Recommending personalized hashtags in real-time is more difficult where the recommended hashtags need to be accurate and given instantly.
- As an extension to work presented in Alsini et al.'s paper [23], the association of the four networks and their combined effect on the performance of hashtag recommendation can be examined. In addition, rather than considering the mutual tie relationships

between users, weighted relationships can be used to construct the networks and detect communities.

- It is challenging to compare newly proposed methods with baseline methods due to the variance in the size of the datasets (i.e., number of tweets, users, and hashtags). It is recommended for future research papers to set a minimum size of the dataset for evaluation.
- Accuracy-based metrics were the primary measures of evaluation for a long time. In recent years, *concepts of evaluation*, which are metrics beyond accuracy, have been studied to evaluate the value of the traditional recommendations. For example, *diversity* is concerned with the variety of items recommended by the system, and *novelty* is concerned with how the recommended items are new to users [81,82]. However, concepts of the evaluation were rarely used to evaluate hashtag recommendation methods. The value of the recommendations also needs to be studied in terms of user satisfaction and expectation.
- With the dynamic nature of social media platforms, studies of hashtag recommendation should focus more on the automatic update of the data on the recommendation.

Author Contributions: Conceptualization, A.A. and D.H.; methodology, A.A.; formal analysis, A.A.; data curation, A.A.; writing—original draft preparation, A.A.; writing—review and editing, A.A., D.H. and A.D.; supervision, A.D. and D.H.; project administration, A.D.; 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, the study does not report any data.

Acknowledgments: We would like to show our thanks to Dawood Amin.

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

References

1. Ding, Z.; Zhang, Q.; Huang, X. Automatic Hashtag Recommendation for Microblogs using Topic-Specific Translation Model. In Proceedings of the COLING 2012: Posters, The COLING 2012 Organizing Committee, Mumbai, India, 8–15 December 2012; pp. 265–274.
2. Ding, Z.; Qiu, X.; Zhang, Q.; Huang, X. Learning Topical Translation Model for Microblog Hashtag Suggestion. In Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, IJCAI '13, Beijing, China, 3–9 August 2013; pp. 2078–2084.
3. Gong, Y.; Zhang, Q.; Huang, X. Hashtag Recommendation Using Dirichlet Process Mixture Models Incorporating Types of Hashtags. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Lisbon, Portugal, 17–21 September 2015; pp. 401–410. [CrossRef]
4. Gong, Y.; Zhang, Q.; Han, X.; Huang, X. Phrase-based hashtag recommendation for microblog posts. *Sci. China Inform. Sci.* **2016**, *60*, 012109. [CrossRef]
5. Song, S.; Meng, Y.; Zheng, Z. Recommending Hashtags to Forthcoming Tweets in Microblogging. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, China, 9–12 October 2015; pp. 1998–2003. [CrossRef]
6. Yu, J.; Zhu, T. Combining long-term and short-term user interest for personalized hashtag recommendation. *Front. Comput. Sci.* **2015**, *9*, 608–622. [CrossRef]
7. Zhang, Q.; Gong, Y.; Sun, X.; Huang, X. Time-aware Personalized Hashtag Recommendation on Social Media. In Proceedings of the COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, Dublin City University and Association for Computational Linguistics, Dublin, Ireland, 23–29 August 2014; pp. 203–212.
8. Zangerle, E.; Gassler, W.; Specht, G. Recommending#-Tags in Twitter. In Proceedings of the Workshop on Semantic Adaptive Social Web (SASWeb 2011), CEUR Workshop Proceedings, Girona, Spain, 22–26 June 2011; pp. 67–78.
9. Khabiri, E.; Caverlee, J.; Kamath, K.Y. Predicting Semantic Annotations on the Real-Time Web. In Proceedings of the 23rd ACM Conference on Hypertext and Social Media, HT '12, Association for Computing Machinery, New York, NY, USA, 25–28 June 2012; pp. 219–228.
10. Chen, C.; Yin, H.; Yao, J.; Cui, B. TeRec: A Temporal Recommender System over Tweet Stream. *Proc. VLDB Endow.* **2013**, *6*, 1254–1257. [CrossRef]
11. Ma, Z.; Sun, A.; Yuan, Q.; Cong, G. Tagging Your Tweets: A Probabilistic Modeling of Hashtag Annotation in Twitter. In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, Shanghai, China, 1 November 2014; pp. 999–1008.

12. Jeon, M.; Jun, S.; Hwang, E. *Hashtag Recommendation Based on User Tweet and Hashtag Classification on Twitter*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 325–336.
13. Feng, W.; Wang, J. We can learn your #hashtags: Connecting tweets to explicit topics. In Proceedings of the 2014 IEEE 30th International Conference on Data Engineering, Chicago, IL, USA, 31 March–4 April 2014; pp. 856–867. [CrossRef]
14. Al-Dhelaan, M.; Alhawasi, H. Graph Summarization for Hashtag Recommendation. In Proceedings of the 2015 3rd International Conference on Future Internet of Things and Cloud, Rome, Italy, 24–26 August 2015; pp. 698–702. [CrossRef]
15. Zhang, Q.; Wang, J.; Huang, H.; Huang, X.; Gong, Y. Hashtag Recommendation for Multimodal Microblog Using Co-Attention Network. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, Melbourne, Australia, 19–25 August 2017; pp. 3420–3426. [CrossRef]
16. Alsini, A.; Datta, A.; Li, J.; Huynh, D. Empirical Analysis of Factors Influencing Twitter Hashtag Recommendation on Detected Communities. In Proceedings of the Advanced Data Mining and Applications—13th International Conference, ADMA 2017, Singapore, 5–6 November 2017; pp. 119–131. [CrossRef]
17. Li, Y.; Jiang, J.; Liu, T.; Qiu, M.; Sun, X. Personalized Microtopic Recommendation on Microblogs. *ACM Trans. Intell. Syst. Technol.* **2017**, *8*. [CrossRef]
18. Kowald, D.; Pujari, S.C.; Lex, E. Temporal Effects on Hashtag Reuse in Twitter: A Cognitive-Inspired Hashtag Recommendation Approach. In Proceedings of the 26th International Conference on WWW, International World Wide Web Conferences Steering Committee: Geneva, Switzerland, 1 April 2017; pp. 1401–1410.
19. Alsini, A.; Datta, A.; Huynh, D.Q.; Li, J. Community Aware Personalized Hashtag Recommendation in Social Networks. In *Data Mining*; Islam, R., Koh, Y.S., Zhao, Y., Warwick, G., Stirling, D., Li, C.T., Islam, Z., Eds.; Springer: Singapore, 2019; pp. 216–227.
20. Ma, R.; Qiu, X.; Zhang, Q.; Hu, X.; Jiang, Y.G.; Huang, X. Co-attention Memory Network for Multimodal Microblog’s Hashtag Recommendation. *IEEE Trans. Know. Data Eng.* **2019**. [CrossRef]
21. Belhadi, A.; Djenouri, Y.; Lin, C.W.; Cano, A. A Data-Driven Approach for Twitter Hashtag Recommendation. *IEEE Access* **2020**, *8*, 79182–79191. [CrossRef]
22. Javari, A.; He, Z.; Huang, Z.; Jeetu, R.; Chen-Chuan Chang, K. Weakly Supervised Attention for Hashtag Recommendation Using Graph Data. In Proceedings of the Web Conference 2020, WWW ’20, Taipei, Taiwan, 20–24 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1038–1048. [CrossRef]
23. Alsini, A.; Datta, A.; Huynh, D.Q. On Utilizing Communities Detected From Social Networks in Hashtag Recommendation. *IEEE Trans. Comput. Soc. Syst.* **2020**, *7*, 971–982. [CrossRef]
24. DeMasi, O.; Mason, D.; Ma, J. Understanding Communities via Hashtag Engagement: A Clustering Based Approach Authors. In Proceedings of the International AAAI Conference on Web and Social Media, Cologne, Germany, 17–20 May 2016.
25. Laniado, D.; Mika, P. Making Sense of Twitter. In *The Semantic Web—ISWC 2010*; Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 470–485.
26. Chowdhury, J.R.; Caragea, C.; Caragea, D. On identifying hashtags in disaster twitter data. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020. [CrossRef]
27. Xiao, F.; Noro, T.; Tokuda, T. News-Topic Oriented Hashtag Recommendation in Twitter Based on Characteristic Co-occurrence Word Detection. In *Web Engineering*; Brambilla, M., Tokuda, T., Toksodorf, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 16–30.
28. Jungherr, A.; Schoen, H.; Jürgens, P. The Mediation of Politics through Twitter: An Analysis of Messages posted during the Campaign for the German Federal Election 2013. *J. Comput. Med. Commun.* **2015**, *21*, 50–68.10.1111/jcc4.12143. [CrossRef]
29. Ince, J.; Rojas, F.; Davis, C.A. The social media response to Black Lives Matter: How Twitter users interact with Black Lives Matter through hashtag use. *Ethn. Rac. Stud.* **2017**, *40*, 1814–1830. [CrossRef]
30. Jain, K.; Jindal, R. A Survey on Hashtag Recommendations. In Proceeding of the 27th Conference of Fruct Association, Trento, Italy, 7–9 September 2020; Volume 27.
31. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* **2009**. [CrossRef]
32. Li, T.; Wu, Y.; Zhang, Y. Twitter Hash Tag Prediction Algorithm. Available online: https://wayneyuwu.com/paper/icomp_hashtag.pdf (accessed on 10 May 2021).
33. Zangerle, E.; Gassler, W.; Specht, G. On the Impact of Text Similarity Functions on Hashtag Recommendations in Microblogging Environments. *Soc. Netw. Anal. Min.* **2013**, *3*, 889–898. [CrossRef]
34. Sedhai, S.; Sun, A. Hashtag Recommendation for Hyperlinked Tweets. In Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, Gold Coast, QLD, Australia, 6–11 July 2014.
35. Li, Q.; Shah, S.; Nourbakhsh, A.; Liu, X.; Fang, R. Hashtag Recommendation Based on Topic Enhanced Embedding, Tweet Entity Data and Learning to Rank. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, Indianapolis, IN, USA, 24–28 October 2016.
36. Otsuka, E.; Wallace, S.A.; Chiu, D. A hashtag recommendation system for twitter data streams. *Comput. Soc. Netw.* **2016**, *3*, 3. [CrossRef]
37. Dey, K.; Shrivastava, R.; Kaushik, S.; Subramaniam, L.V. EmTagger: A Word Embedding Based Novel Method for Hashtag Recommendation on Twitter. In Proceedings of the 2017 IEEE International Conference on Data Mining Workshops, ICDM Workshops 2017, New Orleans, LA, USA, 18–21 November 2017; pp. 1025–1032. [CrossRef]

38. Ben-Lhachemi, N.; Nfaoui, E.H. An extended spreading activation technique for hashtag recommendation in microblogging platforms. In Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics, WIMS 2017, Amantea, Italy, 19–22 June 2017; Akerkar, R., Cuzzocrea, A., Cao, J., Hacid, M., Eds.; ACM: New York, NY, USA, 2017, pp. 16:1–16:8. [CrossRef]
39. Ben-Lhachemi, N.; Nfaoui, E.H. Using Tweets Embeddings For Hashtag Recommendation in Twitter. *Proc. Comput. Sci.* **2018**, *127*, 7–15. [CrossRef]
40. Kaviani, M.; Rahmani, H. EmHash: Hashtag Recommendation using Neural Network based on BERT Embedding. In Proceedings of the 6th International Conference on Web Research (ICWR), Tehran, Iran, 22–23 April 2020; pp. 113–118. [CrossRef]
41. Efron, M. Hashtag Retrieval in a Microblogging Environment. In Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Geneva, Switzerland, 19–23 July 2010; Association for Computing Machinery: New York, NY, USA, 2010; pp. 787–788. [CrossRef]
42. Mazzia, A.; Juett, J. Suggesting Hashtags on Twitter. EECS 545 Project, Winter Term. 2011. Available online: <http://www-personal.umich.edu/~amazzia/pubs/545-final.pdf> (accessed on 10 May 2021).
43. Li, J.; Xu, H. User-IBTM: An Online Framework for Hashtag Suggestion in Twitter. In *Web-Age Information Management*; Cui, B., Zhang, N., Xu, J., Lian, X., Liu, D., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 279–290.
44. Tariq, A.; Karim, A.; Gomez, F.; Foroosh, H. Exploiting Topical Perceptions over Multi-Lingual Text for Hashtag Suggestion on Twitter. In Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2013, St. Pete Beach, FL, USA, 22–24 May 2013.
45. Godin, F.; Slavkovikj, V.; De Neve, W.; Schrauwen, B.; Van de Walle, R. Using Topic Models for Twitter Hashtag Recommendation. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013.
46. She, J.; Chen, L. TOMOHA: TOPic Model-based Hashtag Recommendation on Twitter. In Proceedings of the 23rd International Conference on WWW, Seoul Korea, 7 April 2014.
47. Xu, J.; Zhang, Q.; Huang, X. Personalized Hashtag Suggestion for Microblogs. In *Social Media Processing*; Zhang, X., Sun, M., Wang, Z., Huang, X., Eds.; Springer: Singapore, 2015; pp. 38–50.
48. Lu, H.M.; Lee, C.H. A Twitter Hashtag Recommendation Model that Accommodates for Temporal Clustering Effects. *IEEE Intell. Syst.* **2015**, *30*, 18–25. [CrossRef]
49. Tomar, A.; Godin, F.; Vandersmissen, B.; De Neve, W.; Van de Walle, R. Towards Twitter Hashtag Recommendation Using Distributed Word Representations and a Deep Feed Forward Neural Network. In Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, Bangalore, India, 19–22 September 2014; pp. 362–368.
50. Chen, J.; Kao, H. LDA based semi-supervised learning from streaming short text. In Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015, Paris, France, 19–21 October 2015.
51. Ghaly, R.S.; Elabd, E.; Mostafa, M.A. Tweets classification, hashtags suggestion and tweets linking in social semantic web. In Proceedings of the 2016 SAI Computing Conference (SAI), London, UK, 13–15 July 2016; pp. 1140–1146.
52. Li, J.; Xu, H.; He, X.; Deng, J.; Sun, X. Tweet modelling with LSTM recurrent neural networks for hashtag recommendation. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN) IEEE, Vancouver, BC, Canada, 24–29 July 2016; pp. 1570–1577.
53. Li, Y.; Liu, T.; Jiang, J.; Zhang, L. Hashtag Recommendation with Topical Attention-Based LSTM. In *Coling*; Calzolari, N., Matsumoto, Y., Prasad, R., Eds.; ACL: Stroudsburg, PA, USA, 2016; pp. 3019–3029.
54. Ma, J.; Feng, C.; Shi, G.; Shi, X.; Huang, H. Temporal enhanced sentence-level attention model for hashtag recommendation. *CAAI Trans. Intell. Technol.* **2018**, *3*, 95–100. [CrossRef]
55. Li, Y.; Liu, T.; Hu, J.; Jiang, J. Topical Co-Attention Networks for hashtag recommendation on microblogs. *Neurocomputing* **2019**, *331*, 356–365. [CrossRef]
56. Peng, M.; Lin, Y.; Zeng, L.; Gui, T.; Zhang, Q. Modeling the Long-Term Post History for Personalized Hashtag Recommendation. In *Chinese Computational Linguistics*; Sun, M., Huang, X., Ji, H., Liu, Z., Liu, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 495–507.
57. Ferragina, P.; Piccinno, F.; Santoro, R. On Analyzing Hashtags in Twitter. In Proceedings of the International AAAI Conference on Web and Social Media Ninth International AAAI Conference on Web and Social Media, Oxford, UK, 26–29 May 2015; pp. 110–119.
58. Li, L.; Su, C.; Sun, Y.; Xiong, S.; Xu, G. Hashtag Biased Ranking for Keyword Extraction from Microblog Posts. In *Knowledge Science, Engineering and Management*; Zhang, S., Wirsing, M., Zhang, Z., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 348–359.
59. Badami, M.; Nasraoui, O. Cross-Domain Hashtag Recommendation and Story Revelation in Social Media. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 4294–4303.
60. Diaz-Aviles, E.; Drumond, L.; Schmidt-Thieme, L.; Nejdl, W. Real-Time Top-n Recommendation in Social Streams. In Proceedings of the Sixth ACM Conference on Recommender Systems, RecSys '12, Dublin, Ireland, 9–13 September 2012.
61. Kywe, S.M.; Hoang, T.; Lim, E.; Zhu, F. On Recommending Hashtags in Twitter Networks. In *SocInfo*; Aberer, K., Flache, A., Jager, W., Liu, L., Tang, J., Guéret, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7710, pp. 337–350.
62. Xing, X.; Zhang, W.; Zhang, X.; Xu, N. SOCITEMREC: A Framework for Item Recommendation in Social Networks. Available online: <http://www.jatit.org/volumes/Vol48No3/25Vol48No3.pdf> (accessed on 10 May 2021).

63. Wang, Y.; Qu, J.; Liu, J.; Chen, J.; Huang, Y. What to Tag Your Microblog: Hashtag Recommendation Based on Topic Analysis and Collaborative Filtering. In *Web Technologies and Applications*; Chen, L., Jia, Y., Sellis, T., Liu, G., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 610–618.
64. Zhao, F.; Zhu, Y.; Jin, H.; Yang, L.T. A Personalized Hashtag Recommendation Approach using LDA-based Topic Model in Microblog Environment. *Fut. Generat. Comput. Syst.* **2016**, *65*, 196–206. [[CrossRef](#)]
65. Wang, C.; Sun, Z.; Zhao, Y.; Chi, C.H.; van den Heuvel, W.J.; Lam, K.Y.; Stantic, B. Top-N Hashtag Prediction via Coupling Social Influence and Homophily. In *Advanced Data Mining and Applications*; Li, J., Wang, S., Qin, S., Li, X., Wang, S., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 343–358.
66. Kou, F.; Du, J.; Yang, C.; Shi, Y.; Cui, W.; Liang, M.; Geng, Y. Hashtag Recommendation Based on Multi-Features of Microblogs. *J. Comput. Sci. Technol.* **2018**, *33*, 711–726. [[CrossRef](#)]
67. Harvey, M.; Crestani, F. Long Time, No Tweets! Time-aware Personalised Hashtag Suggestion. In Proceedings of the Advances in Information Retrieval: 37th European Conference on IR Research, ECIR 2015, Vienna, Austria, 29 March–2 April 2015; Springer International Publishing: Cham, Switzerland, 2015; pp. 581–592. [[CrossRef](#)]
68. Li, R.; Zhang, Y. Social-Correlation Based Mutual Reinforcement for Short Text Classification and User Interest Tagging. In *Advanced Data Mining and Applications*; Motoda, H., Wu, Z., Cao, L., Zaiane, O., Yao, M., Wang, W., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 444–455.
69. Kumar, N.; Baskaran, E.; Konjengbam, A.; Singh, M. Hashtag recommendation for short social media texts using word-embeddings and external knowledge. *Know. Inform. Syst.* **2020**. [[CrossRef](#)]
70. Arora, S.; Liang, Y.; Ma, T. A Simple but Tough-to-Beat Baseline for Sentence Embeddings. In Proceedings of the ICLR, Toulon, France, 24–26 April 2017.
71. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
72. Yan, X.; Guo, J.; Lan, Y.; Cheng, X. A Bitern Topic Model for Short Texts. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013.
73. Yin, J.; Wang, J. A Dirichlet Multinomial Mixture Model-based Approach for Short Text Clustering. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014.
74. Alvarez-Melis, D.; Saveski, M. Topic Modeling in Twitter: Aggregating Tweets by Conversations. AAAI Press: Cambridge, MA, USA, 2016; pp. 519–522.
75. Mehrotra, R.; Sanner, S.; Buntine, W.L.; Xie, L. Improving LDA Topic Models for Microblogs via Tweet Pooling and Automatic Labeling. In Proceedings of the SIGIR, Dublin, Ireland, 28 July–1 August 2013; pp. 889–892.
76. Weng, J.; Lim, E.P.; Jiang, J.; He, Q. TwitterRank: Finding Topic-sensitive Influential Twitterers. In Proceedings of the Third ACM International Conference on Web Search and Data Mining, Houston, TX, USA, 4–6 February 2010; ACM: New York, NY, USA, 2010; pp. 261–270. [[CrossRef](#)]
77. Hong, L.; Davison, B.D. Empirical Study of Topic Modeling in Twitter. In Proceedings of the First Workshop on Social Media Analytics, Washington, DC, USA, 25 July 2010; Association for Computing Machinery: New York, NY, USA, 2010; pp. 80–88. [[CrossRef](#)]
78. Zhao, W.X.; Jiang, J.; Weng, J.; He, J.; Lim, E.P.; Yan, H.; Li, X. Comparing Twitter and Traditional Media Using Topic Models. In *Advances in Information Retrieval*; Clough, P., Foley, C., Gurrin, C., Jones, G.J.F., Kraaij, W., Lee, H., Mudoch, V., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 338–349.
79. Liu, Z.; Chen, X.; Sun, M. A Simple Word Trigger Method for Social Tag Suggestion. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Edinburgh, UK, 27–31 July 2011; pp. 1577–1588.
80. Weston, J.; Chopra, S.; Adams, K. #TagSpace: Semantic Embeddings from Hashtags; EMNLP; Moschitti, A., Pang, B., Daelemans, W., Eds.; ACL: Stroudsburg, PA, USA, 2014; pp. 1822–1827.
81. Silveira, T.; Zhang, M.; Lin, X.; Liu, Y.; Ma, S. How good your recommender system is? A survey on evaluations in recommendation. *Int. J. Mach. Learn. Cybernet.* **2017**, *10*. [[CrossRef](#)]
82. Ziegler, C.N.; McNee, S.M.; Konstan, J.A.; Lausen, G. Improving Recommendation Lists through Topic Diversification. In Proceedings of the 14th International Conference on World Wide Web, Chiba, Japan, 10–14 May 2010; Association for Computing Machinery: New York, NY, USA, 2005; pp. 22–32. [[CrossRef](#)]