



# Identifying Covid-19 misinformation tweets and learning their spatio-temporal topic dynamics using Nonnegative Coupled Matrix Tensor Factorization

Thirunavukarasu Balasubramaniam<sup>1</sup> · Richi Nayak<sup>1</sup> · Khanh Luong<sup>1</sup> · Md. Abul Bashar<sup>1</sup>

Received: 18 November 2020 / Revised: 26 May 2021 / Accepted: 4 June 2021 / Published online: 15 June 2021  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2021

## Abstract

Social media platforms like Twitter have become an easy portal for billions of people to connect and exchange their thoughts. Unfortunately, people commonly use these platforms to share misinformation which can influence other people adversely. The spread of misinformation is unavoidable in an extraordinary situation like Covid-19, and the consequences can be dreadful. This paper proposes a two-step ranking-based misinformation detection (RMiD) technique. Firstly, a novel ranking-based approach leveraging the scalable information retrieval infrastructure is applied to detect misinformation from a huge collection of unlabelled tweets based on a related but very small labelled misinformation data set. Secondly, the identified misinformation tweets are represented as a coupled matrix tensor model and Nonnegative Coupled Matrix Tensor Factorization is applied to learn their spatio-temporal topic dynamics. The experimental analysis shows that RMiD is capable of detecting misinformation with better coverage and less noise in comparison with existing techniques. Moreover, the coupled matrix tensor representation has improved the quality of topics discovered from unlabelled data up to 4% by leveraging the semantic similarity of terms in labelled data.

**Keywords** Covid-19 · Misinformation detection · Topic modelling · Ranking · Spatio-temporal patterns · Nonnegative tensor factorization · Saturating Coordinate Descent

## 1 Introduction

Social media has become an acceptable medium of information consumption and exchange where users express them without inhibition. During the Covid-19 pandemic, users commonly rely on social media for social connectedness and news consumption.<sup>1</sup> With the increasing use of social media such as Twitter for information sharing, mining of

this abundant and freely available data can provide valuable insights such as conversion topics, community opinion, and sentiments (Bashar et al. 2020; Mohotti et al. 2019; Tenakoon et al. 2019; Chen et al. 2020).

Given the freedom of speech, there is high possibility of misinformation presence and their spread on social media sites. Especially in events like Covid-19 pandemic, people will not only spread but also react to the misinformation unintentionally. The misinformation tweets can lead to unwanted behaviour such as fear and racism which can put social well-being at risk and impose further liability to already stressed authorities (Heverin and Zach 2012; Shimizu 2020). Therefore, it becomes important to detect misinformation tweets as well as misinformation topics (Wu et al. 2019). Identifying misleading topics with their spatial and temporal context will help authorities to appropriately respond to the information on time.

There exist several supervised machine learning techniques that build an automatic detection model using the

✉ Thirunavukarasu Balasubramaniam  
thirunavukarasu.balas@qut.edu.au

Richi Nayak  
r.nayak@qut.edu.au

Khanh Luong  
khanh.luong@qut.edu.au

Md. Abul Bashar  
m1.bashar@qut.edu.au

<sup>1</sup> School of Computer Science and Centre for Data Science, Queensland University of Technology, 2 George St, Brisbane City, QLD 4000, Australia

<sup>1</sup> [https://blog.twitter.com/en\\_us/topics/company/2020/An-update-on-our-continuity-strategy-during-Covid-19.html](https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-Covid-19.html).

labelled data and its features (Yu et al. 2017; Guo et al. 2018; Wu et al. 2019; Zhang et al. 2019). However, these methods face challenges due to the variations and uncertainty on misinformation topic as well as the lack of labelled misinformation data (Shahi and Nandini 2020). This makes it difficult to apply a supervised machine learning technique on labelled data to build a classifier to detect misinformation tweets (Yang et al. 2019). Due to the continuous generation of tweets, the labelling of a huge volume of tweets is not feasible. A misinformation detection technique without depending heavily on labelled data are inevitable. Using a very small amount of labelled misinformation data, this paper proposes a novel ranking-based approach leveraging the scalable information retrieval infrastructure (Gormley and Tong 2015) to identify misinformation tweets from a large collection of tweet data.

Understanding insight on misinformation tweets can be attained by applying text mining methods, especially, topic modelling or clustering (Allcott and Gentzkow 2017; Debnath and Bardhan 2020). A challenge in applying these methods is to deal with high-dimensional short text data that normally embed unstructured words. To detect misinformation topics from a Covid-19 corpus, Latent Dirichlet Allocation (LDA), one of the most popular topic modelling techniques (Blei et al. 2003), has been used (Han et al. 2020). LDA, in general, does not work with short, highly sparse, and low word co-occurrence text which are the common characteristics of social media data (Zhao et al. 2011; Mohotti et al. 2019).

A dimensionality reduction technique can be applied to learn the low-dimensional embedded space of the data before exploring the embedded topics. Nonnegative Matrix Factorization (NMF) is known as the prominent dimensionality reduction technique for text data which returns low-rank factor matrices that uncover the underlying structures and/or patterns in the data (Lee and Seung 1999; Berry et al. 2007; Luong et al. 2018). Specifically, NMF factorizes the original high-dimensional tweet-term data matrix into two low-rank matrices that represent the intensity of tweets per topic and the intensity of terms per topic (Xu et al. 2003). These two-dimensional topic modelling approaches, NMF and LDA, are limited to represent the tweets with terms only (Sun and Axhausen 2016). They fail to associate these tweets with spatial and temporal dimensions that facilitate useful insights for misinformation topics (Balasubramaniam et al. 2020a).

Identifying these spatial-temporal patterns associated with topics uncover useful information demonstrating the occurrence of misinformation topics over time and locations. We propose to present this data with a 3-mode tensor model to represent terms in the tweets, their time-stamp and their location and fuse the tensor with more auxiliary information in matrix form. We present a coupled matrix tensor model

for modelling the misinformation tweet data and propose an efficient Nonnegative Coupled Matrix Tensor Factorization (N-CMTF) process to identify latent factors in the data to highlight misinformation topic patterns.

More specifically, the contributions of the paper are as below:

1. We frame the problem of detecting misinformation topics over time and space based on an unlabelled Covid-19 Australian Sphere Tweets (CAST) data set (more details of the data set is provided in Sect. 5).
2. We propose a novel ranking-based method to identify misinformation tweets in the CAST data set, a collection of 6.4 million tweets on Covid-19. With the unavailability of labels on this data, we utilize a small labelled misinformation data (Poynter<sup>2</sup>) and identify misinformation topics on this data. These misinformation topics are then used to identify misinformation tweets in CAST using a novel and effective ranking-based misinformation detection (RMiD) technique.
3. We model the identified misinformation data in CAST as a coupled matrix tensor where the tensor has three modes term, time, and location, and the term mode is coupled and shared with a matrix whose second mode is topics from the Poynter data. We then extend the tensor factorization method, Saturating Coordinate Descent (SaCD) (Balasubramaniam et al. 2020b), to solve the N-CMTF problem. With this high-dimensional modelling and factorization, we effectively extract latent relationships of misinformation terms/topics and simultaneously learn the time and location where the topic is discussed.
4. Finally, extensive experiments are conducted to evaluate the quality of misinformation tweet detection and their spatio-temporal topics representation.

## 2 Related works

This section reviews literature related to two main contributions made in this paper: (1) misinformation detection; and (2) topic modelling.

### 2.1 Misinformation detection techniques

The problem of misinformation detection generally falls into the group of supervised learning problem based on text features, except for only a handful of methods that use visual features (Gupta et al. 2013; Jin et al. 2017). Methods to identify misinformation posts can be based on shallow learning

<sup>2</sup> <https://www.poynter.org/ifcn-covid-19-misinformation/>.

or deep learning which include at least two steps of training and testing the detection system. These methods attempt to build a binary classification which use different types of features and auxiliary attributes in the training process to decide if a post is a misinformation.

Most shallow learning-based misinformation detection methods utilize content-based features such as lexical features, syntactic features and/or topic features, or use social context-based features to consider all user profile-related information such as interactions among features or contents (Wu et al. 2019; Guo et al. 2020). For example, authors in Castillo et al. (2011, 2013) use various types of features including tweet's content, users' posting and retweeting behaviours and citations from external sources to build an annotated data set and extract corresponding features to decide different levels of credibility of tweets.

Authors in Qazvinian et al. (2011) not only attempt to collect rumour-related tweets but also identify tweets that users trust or question. They used three different sets of features, which are extracted from tweets' content, user behaviour and memes, to identify rumours. The method builds different Bayes classifiers to learn feature importance and then learn a linear function of these Bayes classifiers to conduct the above two main tasks. Authors in Hu et al. (2014) use content, network information and users' sentiment information to build a classifier detecting spammers. The sentiment information is computed using a supervised sentiment analysis model which relies on a labelled data set. The content features are obtained as topics using NMF, the well-known topic extraction technique for high-dimensional text data.

In the last few years, deep learning-based misinformation detection methods have attracted much attention (Shu et al. 2019; Qian et al. 2018). These methods vary by the use of input data (such as post text content or social context information of users) and network architectures [such as convolutional neural networks (CNN) (Yu et al. 2017; Chen et al. 2017) or recurrent neural network (RNN) (Ma et al. 2016; Guo et al. 2018) or Bayesian deep learning model (Zhang et al. 2019)]. The ultimate goal is to identify the complex and nonlinear latent features embedded in the data to achieve an accurate data representation for detecting misinformation.

These supervised learning methods heavily depend on a large labelled data set for training a misinformation model (Yang et al. 2019). Only a handful of works have been proposed that can detect misinformation when a small or no labelled data are available (Jin et al. 2017; Le and Mikolov 2014). In cases where the labelled data consist of only one class (misinformation), misinformation detection is treated as a problem of single class classification (Manevitz and Yousef 2001) or text matching (Jin et al. 2017). A few researches focus on learning the representation of single class using auto-encoders (Padnekar et al. 2020) and use the learned representation on unlabelled data to identify data

points that match/mismatch with the learned representation. Data points matching the learned representation can be considered as inliers and those that mismatch can be considered as outliers/anomalies.

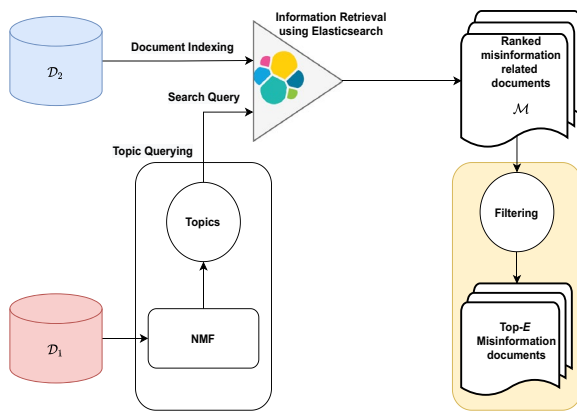
A recent work (Jin et al. 2017) compared many similarity-based text matching techniques and the term frequency-inverse document frequency (TF-IDF)-based text matching found superior. To perform the semantic text matching Doc2Vec (Le and Mikolov 2014) is used instead of TF-IDF. Authors in Yang et al. (2019) propose to extract users' opinions via their behaviour such as posting a tweet, replying, and retweeting and use this information as an umbrella for an unsupervised model based on the collapsed Gibbs sampling approach (Robert and Casella 2013). An unsupervised graph-based approach is used to detect fake news in three steps starting from labelling seeds using biclique graph, to spreading labels within bicliques and ending with having the data set labelled (Hosseinimotlagh and Papalexakis 2018; Gangireddy et al. 2020).

Distinct from these existing works, we leverage a relevant small misinformation data set and propose a ranking-based approach using an information retrieval system to identify misinformation tweets (documents) on a large collection of tweets.

## 2.2 Topic modelling approaches

LDA (Blei et al. 2003) and NMF (Lee and Seung 1999) are well-known techniques for extracting topics from the text data. They have been applied to the Twitter data in recent years (Montenegro et al. 2018; Qomariyah et al. 2019; Nugroho et al. 2015a, b). Comparatively, NMF is known to produce improved outcomes for sparse and short text data like microblogging sites due to its ability to extract the compact low-rank feature matrix that uncovers the most important features for each topics (Yan et al. 2013; Mohotti and Nayak 2018b).

Researchers have started to apply topic modelling approaches on social media posts to understand people behaviours during Covid-19 pandemic (Yin et al. 2020; Han et al. 2020). LDA was used to discover the topics from a twitter-like social media platform, Sina-Weibo and relate the topics to temporal and spatial occurrences separately (Han et al. 2020). Due to the inability of LDA to represent multi-dimensional data, these patterns were not simultaneously identified. In our recent work (Balasubramaniam et al. 2020a), we model a Covid-19 pandemic data as a tensor model and by applying Nonnegative Tensor Factorization (NTF) we show the temporal and spatial information associated with the topics. Similarly, Kassab et al. (2020) compares NMF and NTF to understand the Covid-19 twitter dynamics. However, the analysis is restricted to temporal patterns alone.



**Fig. 1** Ranking-based misinformation detection (RMiD)—overall process

While there exist many applications of NTF in discovering spatio-temporal patterns of users or other entities (Balasubramaniam et al. 2019; Sapienza et al. 2018), the applications on text or social media domain are scarce, and especially misinformation text is scarce. To our best knowledge, this is the first method that represents the social media data (i.e. misinformation tweets) with a coupled matrix tensor model to embed as much as the information it can and learn the spatio-temporal topic dynamics.

### 3 Ranking-based Misinformation Detection (RMiD)

To understand the topic dynamics of misinformation, the misinformation detection is treated as a problem of text matching. The objective is to identify existing misinformation in a huge collection of documents<sup>3</sup> that are similar to a set of known (labelled) misinformation. Let  $\mathcal{D}_1 = \{d_1, d_2, \dots, d_Q\}$  be the small set of labelled misinformation documents and  $d_i = \{w_1, w_2, \dots, w_P\}$  be the  $i$ th document with  $P$  unique terms/words. Let  $\mathcal{D}_2 = \{d_1, d_2, \dots, d_C\}$  be the huge collection of documents that may include misinformation. With the availability of scarce labelled data  $\mathcal{D}_1$ , we propose an information retrieval-based approach, RMiD, to identify a ranked list of documents that contain misinformation words/topics,  $\mathcal{M}$ , from  $\mathcal{D}_2$ . Figure 1 details the overall process that involves the following three steps: (1) document indexing; (2) querying; and (3) filtering.

<sup>3</sup> Each tweet is considered as a document in this paper. Therefore, terms "document" and "tweet" are used interchangeably throughout the paper.

### 3.1 Document indexing

Indexing facilitates storing and allows a search engine to retrieve the data quickly. The indexing will optimize and increase the run-time performance; thus finding ranked relevant information from a large data collection can be achieved within a few milliseconds (Chen et al. 2016). We use Elasticsearch (Gormley and Tong 2015), an open-source tool as a local search engine to facilitate indexing and information retrieval. Since the objective is to identify the misinformation data in  $\mathcal{D}_2$ , only the documents from  $\mathcal{D}_2$  are indexed.

### 3.2 Querying

To retrieve a ranked list of misinformation related documents from the indexed  $\mathcal{D}_2$ , each document in  $\mathcal{D}_1$  is used as a search query. For each search query, a set of relevant misinformation documents from  $\mathcal{D}_2$  are identified.

Let,

$$q_i \leftarrow d_i \tag{1}$$

be the query formed from the terms or topics of  $d_i \in \mathcal{D}_1$ . Let  $\mathcal{M}_i = \{d_1, d_2, \dots, d_E\} \in \mathcal{D}_2$  be the set of top- $E$  relevant (misinformation) documents that are identified by a search engine in response to query  $q_i$ , where  $E < C$ .

#### 3.2.1 Document ranking

A ranked list of top- $E$  documents  $\mathcal{M}_i$  from  $\mathcal{D}_2$  with the corresponding relevance score vector  $s_i$  is identified that are relevant to  $d_i$  according to the terms/topics used in  $q_i$  as per the ranking function (Sutanto and Nayak 2018; Mohotti and Nayak 2018a),

$$R_f : q_i \rightarrow \mathcal{D}_2 = \{(\mathcal{M}_i^j, s_i^j)\} \forall j \in \{1, 2, \dots, E\}, \tag{2}$$

where  $\mathcal{M}_i^j$  indicates  $j$ th document in  $\mathcal{M}_i$  and  $s_i^j \in s_i$  indicates the relevance score of  $j$ th document in  $\mathcal{M}_i$ .

The relevance score  $s_i$  is used to rank the documents. A higher relevance score indicates the high relatedness of the returning document for the given query. The relevance score of  $j$ th document in  $\mathcal{M}_i$  with  $q_i$  is calculated using term frequency \* inverse document frequency (tf\*idf) (Fuhr et al. 2012) as,

$$s_i^j = \text{relevance score}(q_i, \mathcal{M}_i^j) = \sum_{w \in q_i} \sqrt{tf_{w, \mathcal{M}_i^j} \times idf_w^2 \times \text{norm}(w, \mathcal{D}_2)}, \tag{3}$$

where  $tf_{w, \mathcal{M}_i^j}$  is the frequency of term  $w$  in document  $\mathcal{M}_i^j$ ,  $idf_w^2$  is the frequency of term  $w$  in the document collection  $\mathcal{D}_2$ ,  $\text{norm}(w, \mathcal{D}_2)$  is a normalization controlling the importance of terms based on term frequency (Gormley and Tong 2015).

### 3.2.2 Topic querying

Using terms of each document  $d_i \in \mathcal{D}_1$  explicitly as a search query  $q_i$  may not perform the search  $\mathcal{D}_2$  semantically and will focus on finding the terms present in the search query. Let us consider the following two misinformation documents example.

Drinking a bleach solution will prevent you from getting the coronavirus.  
 Gargling with bleach can cure coronavirus.

Though these two misinformation tweets do not use the same terms, they convey the same information. Searching using these two tweets as query terms will retrieve many noisy results due to the presence of unimportant terms like *getting* and *solution*. To semantically enrich a search query, we propose to apply topic modelling on the collection  $\mathcal{D}_1$  and use the terms representing a topic as a query (i.e. query terms) on  $\mathcal{D}_2$ , instead of using the terms of each document in  $\mathcal{D}_1$ . In the above example, these two documents will form a topic where only the important terms like *drinking*, *gargling*, and *bleach* will be highlighted.

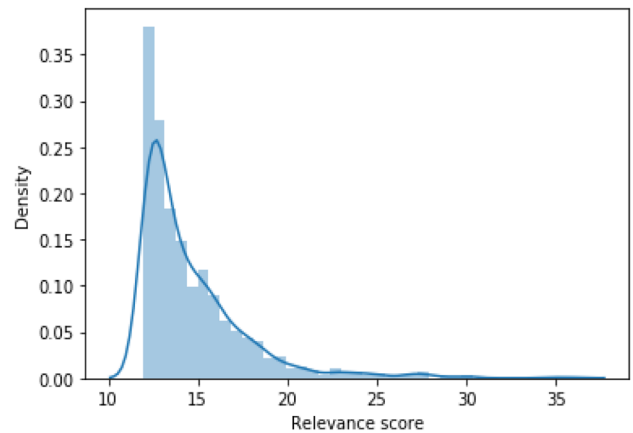
*Nonnegative Matrix Factorization*: NMF is a well-known dimensionality reduction technique that factorizes a high-dimensional matrix into two smaller, low-dimensional factor matrices, especially for sparse and short text data such as tweets (Xu et al. 2003). Let  $\mathbf{Z} \in \mathbb{R}^{(P \times Q)}$  be an input matrix that represents  $P$  rows as the number of unique terms and  $Q$  columns as the number of unique documents. The entries in this matrix are populated using term-frequency. The objective function of NMF can be formulated as (Berry et al. 2007),

$$\min_{\mathbf{W}, \mathbf{H} \geq 0} f(\mathbf{W}, \mathbf{H}) = \|\mathbf{Z} - \mathbf{WH}^T\|^2, \tag{4}$$

where  $\mathbf{W} \in \mathbb{R}^{(P \times J)}$  and  $\mathbf{H} \in \mathbb{R}^{(Q \times J)}$  are the factor matrices,  $J$  is the rank (number of topics), and  $\|\cdot\|$  indicates the Frobenius norm.

Factor matrix  $\mathbf{W}$  represents the topics in  $\mathcal{D}_1$ . Each column,  $\mathbf{W}_i \in \mathbb{R}^{P \times 1}$ , represents a topic and the magnitude of the elements in it determines the importance of the terms in that topic. Using the internal measures like Pattern Distinctiveness (Balasubramaniam et al. 2019) and Silhouette Index (Rousseeuw 1989) (more details provided in Sect. 5.1.1), the rank  $R$  (i.e. the number of topics) can be identified.

Using topics as queries instead of documents, the number of queries will be reduced from  $Q$  to  $J$  as  $J < Q$ . This reduces the redundancies in the returning documents set identified as the misinformation data set  $\mathcal{M}$ . Each of the  $J$  search queries is defined by selecting the top-10 terms based on the magnitude value from each of the  $J$  topics as,



**Fig. 2** Right-skewed distribution of relevance score ( $s_i$ ) for the search query ( $q_i$ ) “drinking lemon water prevent protects warm novel home-made masks lives”

$$q_i \leftarrow \mathbf{W}_i^{10}, \tag{5}$$

where  $\mathbf{W}_i^{10} = \{w_i^1, w_i^2, \dots, w_i^{10}\}$ . The documents retrieved based on the query defined in Equation (5) will be ranked as per Eqs. (2) and (3). Suppose, NMF outputs 25 topics, each topic in  $\mathbf{W}$  will represent  $P$  terms and their weightage. The top – 10 terms based on their weightage will make a single query. Therefore, we will have a total of 25 queries each with a length of 10 terms.

Not all topics (i.e. search queries) of  $\mathcal{D}_1$  will have equally relevant documents in  $\mathcal{D}_2$ . For instance, one topic of  $\mathcal{D}_1$  may find 100 relevant documents, whereas another may find 600 relevant documents in  $\mathcal{D}_2$ . Fixing the same number of search results to top- $E$  for all the search queries will introduce a lot of noise in the search results or will miss useful data during information retrieval. Consequently, this will affect the quality of relevant search results. Therefore, we propose a relevance score-based filtering approach that selects the appropriate number of top- $E$  relevant documents for each query.

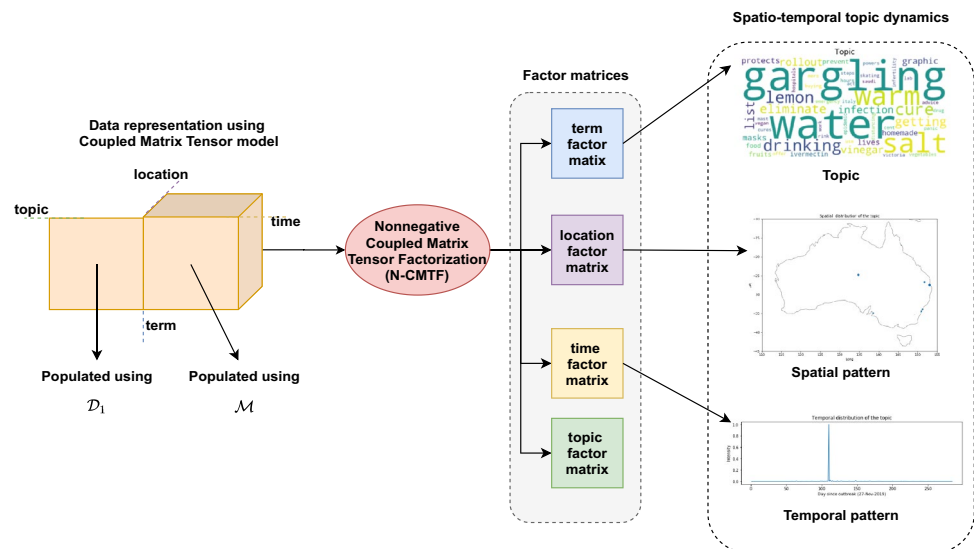
### 3.3 Filtering

The range of relevance score of identified documents varies for each query. Hence, the relevance score of documents for different queries cannot be compared (Gormley and Tong 2015). We use the relevance scores to filter the search results to include only the top- $E$  misinformation related documents and find the value of  $E$  dynamically.

Suppose using a query  $q_i$ , we retrieve  $\mathcal{M}_i$  relevant documents that are ranked according to their relevance score ( $s_i$ ).

During the investigation phase, it was noted that the search results for all the queries are right-skewed in terms of relevance score. It indicates that only a very few results are very relevant and most of the results are moderately/

**Fig. 3** Architecture: N-CMTF-based spatio-temporal topic dynamics discovery



less relevant to the search query. An example of the right-skewed relevance score distribution is shown in Fig. 2. While the query has returned a maximum relevance score of 35, most of the results are within the range of 5 to 20, returning less relevant documents to the search query.

Therefore, to retrieve only the top- $E$  misinformation documents for each query, we set a threshold automatically based on the relevance score  $s_i$ . Only the documents ( $\mathcal{M}_i$ ) with relevance score greater than the threshold will be selected. Because of the right-skewness, we set the threshold to select only the outliers towards the right side of the distribution (Howell et al. 1998) as,

$$\text{threshold} = \mu(s_i) + \sigma(s_i) \times 3, \quad (6)$$

where  $\mu(s_i)$  is the mean of  $s_i$  and  $\sigma(s_i)$  is the standard deviation of  $s_i$ . While we set this threshold which is suitable for the data set investigated, the retrieval of top- $E$  misinformation related documents can be controlled by selecting other threshold values. However, this is beyond the scope of this paper.

The above process is repeated for each search query to get the final misinformation document collection  $\mathcal{M}$ .

#### 4 N-CMTF-based spatio-temporal topic dynamics discovery

The proposed N-CMTF-based spatio-temporal topic dynamics discovery method consists of three components: (1) data model, (2) factorization, and (3) spatio-temporal topic dynamics, as shown in Fig. 3.

##### 4.1 Data representation—coupled matrix tensor model

We use a coupled matrix tensor model to represent the misinformation related tweets. Since, each *term* in  $\mathcal{M}$  is associated with *location* and *time*, a tensor representation is used to model the data by preserving the association among the modes *term*, *location*, and *time*. The time mode can be defined according to the analysis requirement. For example, it can be days, weeks, or months.

Let  $\mathcal{U} = \{u_1, u_2, \dots, u_M\}$ ,  $\mathcal{L} = \{l_1, l_2, \dots, l_N\}$  and  $\mathcal{T} = \{t_1, t_2, \dots, t_O\}$  be the set of unique terms, locations and time periods in  $\mathcal{M}$ , respectively. The traditional approach is to populate the tensor  $\mathcal{X}$  with term frequency. For example, if *term 1* is used/tweeted at *location 4* at *time 5*, and if this has occurred 3 times, the  $\mathcal{X}_{1,4,5}$ th entry of tensor will be populated with the value 3. However, in recent researches (Guacho et al. 2018; Abdali et al. 2020) binary representation has proven to capture the nuance patterns better than term frequency. Therefore, in this paper we use binary representation which gives value 1 to  $\mathcal{X}_{1,4,5}$ th if *term 1* is used/tweeted at *location 4* at *time 5* or else 0. For a more detailed empirical comparison, please refer to the supplementary material.

While a tensor model (term  $\times$  location  $\times$  time) is successful in capturing the spatio-temporal association through multi-dimensional representation, it does not guarantee to capture the semantic relatedness between the terms, especially according to their relation to misinformation topics as present in  $\mathcal{D}_1$ . A tensor model encoding (tweet  $\times$  term  $\times$  term) or (tweet  $\times$  term  $\times$  topic) can capture the co-occurrence and latent relationships between the terms (Abdali et al. 2020). However, the proposed tensor model

represented as (term × time × location) has no explicit way to capture the semantic relatedness (or co-occurrences) of terms in the tweets.

We propose to combine the tensor model (term × time × location)  $\mathcal{X}$  with a matrix that can embed the terms relatedness in  $\mathcal{M}$  and present a coupled matrix tensor model. An auxiliary matrix  $\mathbf{Y} \in \mathbb{R}^{M \times J}$  is generated where  $M$  is the number of unique terms in  $\mathcal{M}$  and  $J$  is the number of topics identified using NMF on  $\mathcal{D}_1$ .

The value of elements in  $\mathbf{Y}$  is calculated as,

$$Y_{mj} = \begin{cases} 1 & \text{if } u_m \in \mathbf{e}_j \\ 0 & \text{else,} \end{cases} \quad (7)$$

where  $\mathbf{e}_j$  is the  $j$ th topic of  $\mathcal{D}_1$  (refer to Sect. 3.2.2). The value for entry  $(m, j)$  in matrix  $\mathbf{Y}$  represents the existence of term  $u_m$  in topic  $\mathbf{e}_j$ .

Since the tensor  $\mathcal{X}$  and matrix  $\mathbf{Y}$  share some of the common terms, the fusion of  $\mathcal{X}$  and  $\mathbf{Y}$  will happen only on the term mode.

An alternative of using semantic information in the tensor model (term × location × time)  $\mathcal{X}$  is the addition of topics as 4th mode. However, this model will be meaningful when the topics are derived from the same data set  $\mathcal{M}$ , which is used to model the tensor model  $\mathcal{X}$ , by relating each term to at least one of the topics. When the topics are derived from a different data set, say  $\mathcal{D}_1$  as in this paper, there is no guarantee that all the terms in  $\mathcal{M}$  will be associated with at least one of the topics of  $\mathcal{D}_1$ .  $\mathcal{D}_1$  and  $\mathcal{M}$  will have different terms dictionaries with only some terms occurring in both dictionaries. This would pose difficulties in learning spatio-temporal topic dynamics of misinformation tweets. Firstly, the 4-mode tensor will be sparse lacking co-occurrences and present difficulty in deriving latent term dependencies. Secondly, the terms that do not belong to any topics will be left out in the 4-mode tensor representation. Consequently, it will not be possible to learn the spatio-temporal topic dynamics of the left out terms. Let us consider a toy example to understand this.

Suppose  $\mathcal{D}_1$  has 2,000 terms and  $\mathcal{M}$  has 50,000 terms with only a total of 1,000 terms occurring in both the data sets. If we model  $\mathcal{M}$  as a 3D tensor (term × location × time) and add topics derived from  $\mathcal{D}_1$  as 4th mode, only 1000 terms from  $\mathcal{M}$  will have an association with topics from  $\mathcal{D}_1$ . The majority of terms will not learn the spatio-temporal topic dynamics as they do not belong to a topic.

One possible way to overcome this issue is by introducing a surrogate topic that includes all the remaining (50,000-1,000 = 49,000) terms of  $\mathcal{M}$  that are not present in  $\mathcal{D}_1$ . By adding this additional topic, we will be able to learn the spatio-temporal topic dynamics of all the 50,000 terms of  $\mathcal{D}_1$ . However, this will be misleading as those 49,000 terms have a forced relatedness in the model that does not exist.

**Table 1** Table of symbols

Symbol	Definition
$\mathcal{X}$	Tensor (Euler script letter)
$U, L, T$	3 modes of $\mathcal{X}$
$U, V$	2 modes of $\mathbf{Y}$
$M, N, O, J$	Length of mode $U, L, T, V$ , respectively
$\mathbf{U}$	Matrix (upper case. bold letter)
$\mathbf{u}$	Vector (lower case. bold letter)
$u$	Scalar (lower case, italic letter)/element
$\mathbf{X}_n$	Mode- $n$ matricization of tensor
$R$	Rank of tensor
$\otimes$	Kronecher product
$\odot$	Khatri-Rao product
$*$	Hadamard product
$\circ$	outer product
$\ \cdot\ $	Frobenius norm

On the other hand, the proposed coupled matrix tensor model has a natural capability to represent the topics from a different data set. The term × topic matrix  $\mathbf{Y}$  that is coupled with tensor  $\mathcal{X}$  along the shared term mode will have entries populated with 1 if a term does appear in a topic as per Equation (7). For terms that do not belong to any topic, the entire row of  $\mathbf{Y}$  will be populated with 0. Since topic-term association is handled independently in this coupled model, the term’s association with other modes (i.e. time and location) of the tensor will remain unaffected. Hence, learning spatio-temporal topic dynamics of terms using the coupled matrix tensor model will be complete for all terms present in  $\mathcal{M}$ .

Additionally, this is an innovative use of transfer learning from one data set ( $\mathcal{D}_1$ ) to another data set ( $\mathcal{M}$ ). We learned the topics on  $\mathcal{D}_1$  (a smaller data set with misinformation data) and used them in  $\mathcal{M}$  (a large data set with thousands of tweets) to transfer semantic information. The data represented in this fused model will preserve the closeness of the terms based on the topics learned from  $\mathcal{D}_1$  leading to a better topic learning in  $\mathcal{M}$  (Bahargam and Papalexakis 2018).

### 4.2 Nonnegative Coupled Matrix Tensor Factorization (N-CMTF)

N-CMTF is the fusion of NMF and NTF on any shared dimension.

For matrix  $\mathbf{Y} \in \mathbb{R}^{(M \times J)}$ , the objective function of NMF can be formulated as,

$$\min_{\mathbf{U}, \mathbf{V} \geq 0} f_1(\mathbf{U}, \mathbf{V}) = \|\mathbf{Y} - \mathbf{UV}^T\|^2, \quad (8)$$

where  $\mathbf{U} \in \mathbb{R}^{(M \times R)}$  and  $\mathbf{V} \in \mathbb{R}^{(J \times R)}$  are the factor matrices,  $R$  is the rank, and  $\|\cdot\|$  indicates the Frobenius norm (Table 1).

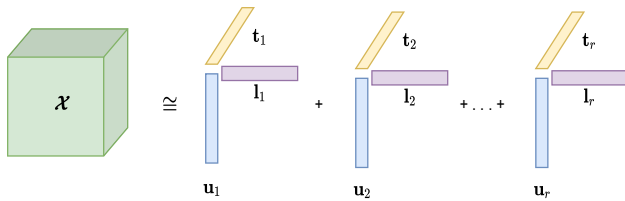


Fig. 4 CP Tensor Factorization

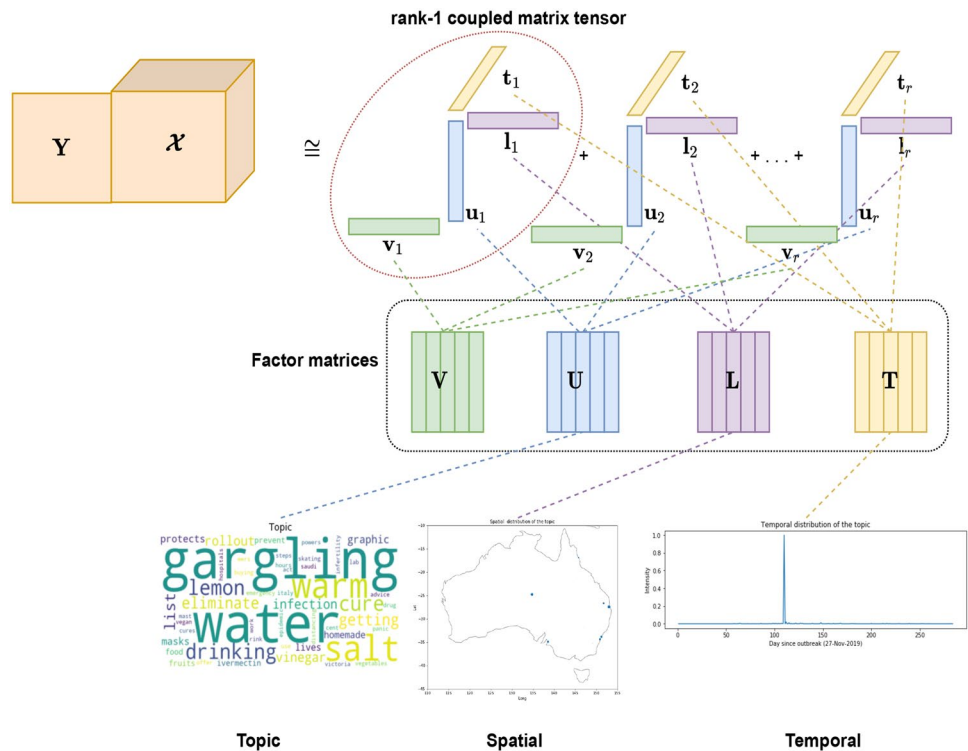
For a third-order tensor  $\mathcal{X} \in \mathbb{R}^{(M \times N \times O)}$ , Candecomp/Parafac (CP) (Carroll and Chang 1970) factorization decomposes the tensor into multiple rank-1 tensors as shown in Fig. 4. Each rank-1 tensor comprises of three components representing each mode of the tensor. The mode-wise components are grouped to form the factor matrices representing each mode of the tensor. CP factorization is formulated as,

$$\mathcal{X} \cong [\mathbf{U}, \mathbf{L}, \mathbf{T}] = \sum_{r=1}^R \mathbf{u}_r \circ \mathbf{l}_r \circ \mathbf{t}_r, \tag{9}$$

where  $\mathbf{U} \in \mathbb{R}^{(M \times R)}$ ,  $\mathbf{L} \in \mathbb{R}^{(N \times R)}$ , and  $\mathbf{T} \in \mathbb{R}^{(O \times R)}$  are the factor matrices representing three modes. In our case, it is *term*, *location*, and *time*.

$R$  is the rank of the tensor, indicating the hidden features in the lower dimension. For *term* mode, it represents the topics. For *location* and *time* mode, it represents the spatial and temporal patterns, respectively.

Fig. 5 Spatio-temporal topic dynamics using N-CMTF. Each column in a factor matrix (i.e. middle row) represents a pattern



NTF imposes nonnegativity constraint on the factor matrices, and the values for factor matrices are derived by solving the optimization minimization problem defined as,

$$\min_{\mathbf{U}, \mathbf{L}, \mathbf{T} \geq 0} f_2(\mathbf{U}, \mathbf{L}, \mathbf{T}) = \|\mathcal{X} - [\mathbf{U}, \mathbf{L}, \mathbf{T}]\|^2. \tag{10}$$

Factorization algorithms like alternating least square (ALS) (Cichocki et al. 2009) is usually used to solve Equation (10) and derive the values for factor matrices.

It can be noted from Eqs. (8) and (10) that the factor matrix  $\mathbf{U}$  is shared with both NMF and NTF. Therefore, the fused objective function can be defined as,

$$f = f_1 + f_2. \tag{11}$$

By substituting Eqs. (8) and (10) in (11), the objective function of N-CMTF becomes,

$$\min_{\mathbf{U}, \mathbf{V}, \mathbf{L}, \mathbf{T} \geq 0} f(\mathbf{U}, \mathbf{V}, \mathbf{L}, \mathbf{T}) = \|\mathbf{Y} - \mathbf{UV}\|^2 + \|\mathcal{X} - [\mathbf{U}, \mathbf{L}, \mathbf{T}]\|^2. \tag{12}$$

Because of this fused objective function, the factor matrix  $\mathbf{U}$  will learn more cohesive topics in  $\mathcal{M}$ , taking advantage of the similarity of the existing terms based on the topics in  $\mathcal{D}_1$ .

### 4.3 Spatio-temporal topic dynamics

The misinformation subset  $\mathcal{M}$  is obtained from  $\mathcal{D}_2$  after applying RMiD using small misinformation labelled data  $\mathcal{D}_1$ . We represent  $\mathcal{M}$  as a coupled matrix tensor (CMT) model



and decompose it to obtain factor matrices that learn latent dependencies in the data. Algorithm 1 details the overall process of the SaCD algorithm for learning spatio-temporal topic dynamics in  $\mathcal{M}$  using N-CMTF.

The factorization of  $\mathcal{M}$  with N-CMTF using SaCD will result in  $R$  rank-1 coupled matrix tensors as shown in Fig. 5. Each of the rank-1 coupled matrix tensors consists of four vectors (or components), where each component represents a mode (i.e. topic, location, time). Each vector is a feature/pattern which reveals a distinct characteristic of the respective mode. Because of the rank-1 components, the features are learned with the dependencies among them preserved. These rank-1 coupled tensors are used to understand the spatio-temporal topic dynamics.

## 5 Experimental analysis

Experiments are conducted to answer the following questions:

1. How effectively RMiD can detect Covid-19 misinformation tweets from a huge collection of unlabelled data using a very small set of labelled data?
2. How good the spatio-temporal topic dynamics of Covid-19 misinformation are learned using N-CMTF with SaCD? What are those spatio-temporal topic dynamics?

*data sets:* Two Covid-19 related data sets are used to understand spatio-temporal misinformation topic dynamics.

1. An online fact checking organization, Poynter has collected posts from different social media platforms (e.g., Facebook, YouTube, and Instagram) and performed fact checks on their content. These posts are then annotated with the label for misinformation (True or False). It also collects the location where the misinformation is originated. We filtered this labelled data to include only the misinformation that originated within Australia as our focus is to discover misinformation topic dynamics in Australian Twitter-sphere during Covid-19. This has led to the selection of 119 facts labelled “True” for misinformation. Applying NMF as the topic modelling approach on this data set, a total of 49 topics are identified. A close investigation of 119 misinformation posts and 49 topics reveal that most of the misinformation posts are unique and there exists less association among each of misinformation. This resulted in the formation of a high number of topics.

2. The Covid-19 Australian Sphere Tweets (CAST) data set, collected with the help of QUT Digital Observatory,<sup>4</sup> consists of Covid-19 related tweets collected based on the keywords like Covid, Covid-19, Covid19, Coronavirusoutbreak, Coronavirus, for the period between 27th November 2019 and 7th September 2020. It is a huge collection of 6.4 million tweets posted by Australian Twitter users along with their location and time.

The data sets are pre-processed using standard social media data pre-processing library (ekphrasis) (Baziotis et al. 2017).

### 5.1 Performance analysis of misinformation identification

Using the posts and topics in Poynter data set, the objective is to detect misinformation in CAST data set by applying the proposed RMiD technique. These results are compared with the results obtained by three state-of-the-art misinformation detection techniques.

#### 5.1.1 Benchmarks and experimental settings

We mainly select the benchmarks that are closely associated with the problem of detecting misinformation from a huge collection of unlabelled data using a small set of labelled data. Since all the samples in labelled data belong to the same class indicating misinformation, single class classification and similarity-based approaches are identified as suitable benchmarks.

1. Since the Poynter data set consists of facts belonging to only one class, One-Class Support Vector Machine (OCSVM) (Manevitz and Yousef 2001) is trained on it and the trained model is used to predict the misinformation in CAST data set.
2. A TF-IDF (Jin et al. 2017) similarity-based misinformation detection technique is implemented that detects misinformation in the CAST data set by the pair-wise similarity comparison. Any tweet in CAST that has the similarity score  $\geq 0.5$  with a post in Poynter can be considered as a misinformation tweet.
3. Doc2Vec (Le and Mikolov 2014) is also a similarity-based misinformation detection technique that attempts to find the similarity based on the semantic closeness of terms. A Doc2Vec model is built on Poynter data set and used to represent each tweet in CAST data set as a numeric vector. Then, cosine similarity is applied to the CAST data set detecting the misinformation tweets. Any

<sup>4</sup> <https://www.qut.edu.au/institute-for-future-environments/facilities/digital-observatory>.

tweet in CAST that has the similarity score  $\geq 0.4$  with any post in Poynter can be considered as a misinformation tweet.

4. AE-LSTM and AE-DNN (Padnekar et al. 2020) are autoencoders build with LSTM and DNN layers, respectively. Firstly an autoencoder is pre-trained using the Poynter data set to capture the representation of misinformation. Then, the CAST data set is passed into the autoencoder. The final loss value indicates how much each tweets deviates from the learned representation. A threshold of  $\leq 0.5$  is set to select the tweets that highly matched with the misinformation.

To evaluate the effectiveness of topic querying and relevance score-based filtering in RMiD, we introduce four variations of RMiD.

1. RMiD-DQ—Terms in each misinformation post from Poynter are directly used as the search query terms to detect misinformation in the CAST data set. Top-1000 misinformation tweets are retrieved for each search query.
2. RMiD-DQF—Filtering is applied to the misinformation tweets detected using the RMiD-DQ data set. This will select the most relevant misinformation tweets from the CAST data set.
3. RMiD-TQ—The total of 49 topics obtained from the Poynter data set is used as the search query. However, instead of identifying top- $E$  misinformation tweets in the CAST data set, the top-1000 misinformation tweets are retrieved for each query.
4. RMiD-TQF - Filtering is applied to the misinformation tweets detected using RMiD-TQ.

The misinformation identified using OCSVM, TFIDF, Doc2Vec, RMiD-DQ, RMiD-DQF are represented in a tensor model  $\mathcal{X}$  and the topics are learned using NTF SaCD. On the other hand, misinformation identified using RMiD-TQ and RMiD-TQF are represented using coupled matrix tensor model because of the availability of auxiliary matrix  $\mathbf{Y}$  and the topics are learned using N-CMTF SaCD.

The number of topics (rank of NTF/N-CMTF) is identified based on the evaluation measures like Pattern Distinctiveness (Balasubramaniam et al. 2019) and Silhouette Index (Rousseeuw 1989). The Silhouette Index (SI) (Rousseeuw 1989) is calculated as the ratio of inter-cluster to intra-cluster distances to measure the separability and compactness of clustering and similarly with topic modelling. The higher the value, the better the topic modelling. Pattern Distinctiveness (PD) (Balasubramaniam et al. 2019) is generally used to evaluate the quality of patterns learned i.e. clusters or topics. It computes the similarity among patterns, therefore, the lower the value, the more unique the patterns and is the goal.

For each technique, the number of topics (or rank in the tensor) ranging from 2 to 49 was tested. The number of topics for which the SI is at the maximum and PD is at the minimum is selected as the optimal number of topics. Figure 6 shows the sensitivity analysis on the choice of rank/optimal number of topics. For CMTF variants, since matrix  $\mathbf{Y}$  and tensor  $\mathcal{X}$  shares same terms mode, both  $\mathbf{Y}$  and  $\mathcal{X}$  are set the same rank. For example, if the rank of tensor is set to 5, the rank of matrix is also set to 5.

The details of misinformation tweets identified in the CAST data set using these techniques are summarized in Table 2.

### 5.1.2 Evaluation measures

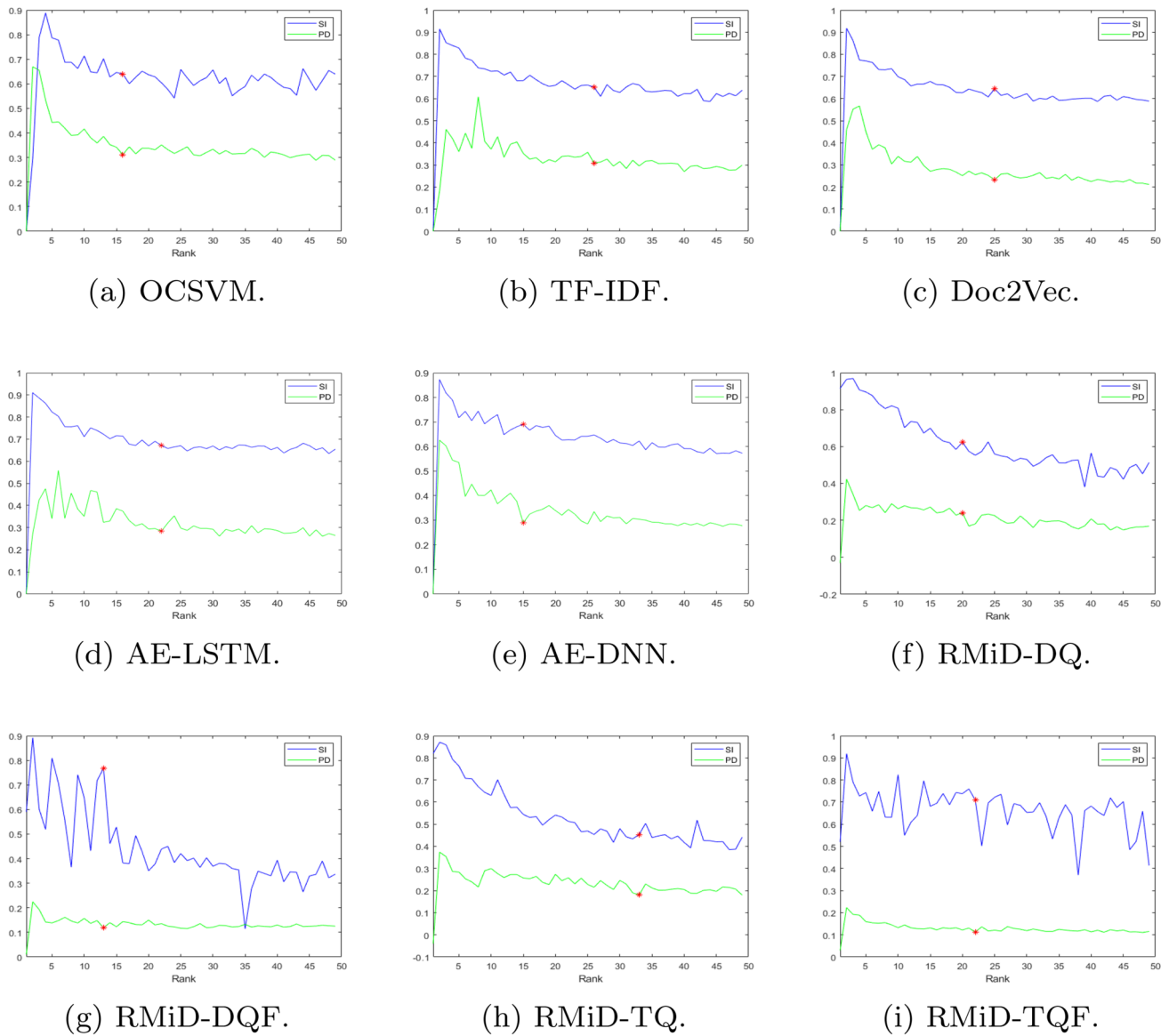
Five evaluation measures are used to assess the quality of misinformation detection and to answer Question 1. Since the CAST data set does not contain labelled tweets, the following intrinsic measures are used.

1. **Terms coverage** measures the percentage of unique terms present in the Poynter data that are also present in the misinformation tweets identified in the CAST data.
2. **Potential Noise** measures the percentage of unique terms present in the misinformation data that are not present in the Poynter data.
3. **Topic-terms coverage** measures the percentage of unique terms present in the topics of misinformation data are also present in the Poynter data set.
4. **Topic-terms avg. similarity** measures the average cosine similarity between the topics of misinformation data and the misinformation posts in the Poynter data set. For each topic of misinformation data, a most similar post in Poynter is found and the average similarity score of all the topics is calculated.
5. **Topics with  $\geq 50\%$  similarity** measures the percentage of topics in misinformation data that have more than 50% cosine similarity with any of the misinformation post in Poynter data set.

While terms coverage and potential noise are used to evaluate the quality of identified misinformation tweets in the CAST data set, topic-terms coverage, topic-terms avg. similarity, and topics with  $\geq 50\%$  similarity evaluate the quality of topics learned from the identified misinformation tweets from CAST data sets.

### 5.1.3 Results and discussion

A key point to note in Table 2 is sparsity of the tensor model  $\mathcal{X}$  representing misinformation tweets with their time and location. The representation of misinformation data detected using OCSVM, TF-IDF, Doc2Vec, AE-LSTM, and



**Fig. 6** Sensitivity analysis on setting ranks (i.e. the number of topics). Red point indicates the optimal number of topics where SI is high and PD is low

**Table 2** Insight on the misinformation tweets identified in the CAST data set

	OCSVM	TF-IDF	Doc2Vec	AE-LSTM	AE-DNN	RMiD-DQ	RMiD-DQF	RMiD-TQ	RMiD-TQF
# Query	–	–	–	–	–	119	119	49	49
# Misinfo. tweets	80149	177498	6940	8943	23787	119000	2263	49000	974
Avg. s	–	–	–	–	–	17.20	27.30	15.48	25.14
# Unique terms	77000	108503	22254	23277	77838	32120	4184	23038	2476
# Unique days	239	246	233	214	232	245	223	243	188
# Unique locations	7566	12882	2104	2365	4001	9051	715	5916	373
# Observations in $\mathcal{X}$	1259485	1659810	103263	84928	578587	1657422	31115	666796	13769
Sparsity of $\mathcal{X}$ (%)	0.999991	0.999995	0.999991	0.999993	0.999992	0.999984	0.999961	0.999984	0.999931
Matrix ( $\mathbf{Y}$ ) size	–	–	–	–	–	–	–	23038 × 49	2476 × 49
Rank (# topics)	16	26	25	26	15	20	13	33	22



**Table 4** Performance analysis: spatio-temporal topic dynamics using NTF and N-CMTF

Evaluation	Method	RMiD-DQ	RMiD-DQF	RMiD-TQ	RMiD-TQF
FS	NTF ALS	0.05	0.07	0.06	0.15
	NTF SaCD	0.05	0.07	0.06	0.15
	NTF 4D	–	–	0.04	0.08
	ACMTF	–	–	0.05	0.06
	N-CMTF SaCD	–	–	0.06	0.14
SI	NTF ALS	0.32	0.65	– 0.02	0.58
	NTF SaCD	<b>0.62</b>	<b>0.76</b>	0.57	0.67
	NTF 4D	–	–	0.28	0.63
	ACMTF	–	–	0.04	0.03
	N-CMTF SaCD	–	–	<b>0.58</b>	<b>0.71</b>
CH	NTF ALS	148.87	<b>83.59</b>	49.32	47.42
	NTF SaCD	<b>190.50</b>	80.79	112.25	51.66
	NTF 4D	–	–	109.49	48.39
	ACMTF	–	–	46.45	41.48
	N-CMTF SaCD	–	–	<b>125.08</b>	<b>55.50</b>
DV	NTF ALS	2.83	2.08	3.17	1.95
	NTF SaCD	<b>2.34</b>	<b>1.97</b>	2.34	1.97
	NTF 4D	–	–	<b>2.06</b>	1.99
	ACMTF	–	–	4.01	2.66
	N-CMTF SaCD	–	–	2.33	<b>1.87</b>
PD	NTF ALS	0.31	0.14	0.51	0.13
	NTF SaCD	<b>0.24</b>	<b>0.12</b>	0.20	0.11
	NTF 4D	–	–	0.51	<b>0.08</b>
	ACMTF	–	–	0.43	0.13
	N-CMTF SaCD	–	–	<b>0.18</b>	0.11
Runtime (secs)	NTF ALS	67.96	<b>0.45</b>	<b>31.70</b>	<b>0.49</b>
	NTF SaCD	<b>66.73</b>	1.08	53.05	1.27
	NTF 4D	–	–	127.76	1.46
	ACMTF	–	–	515.68	11.62
	N-CMTF SaCD	–	–	55.89	1.28

Bold value indicates the best result

from the Poynter data set can be included in the tensor model populated with the CAST data. The NTF 4D version in this comparison includes a surrogate topic to include all CAST terms that do not belong to any Poynter topics. Please refer to supplemental material for a detailed comparison between 4D NTF with and without the additional topic holding all nontopic inclusive tweets. We also compare N-CMTF SaCD with ACMTF (Acar et al. 2017) which is an advanced CMTF without nonnegative constraint. This would help to evaluate the performance of SaCD and N-CMTF in discovering spatio-temporal topic dynamics.

In addition to SI and PD, we use three measurement criteria for evaluating topic modelling task, including Fit Score (FS), Calinski–Harabasz (CH), and Davies–Bouldin (DV).

Fit Score (FS) (Bro et al. 1999) is a general-purpose evaluation measure, assessing the quality of factorization approximation. The higher the FS, the better is the quality of factor matrices to recreate the input tensor/matrix.

Calinski–Harabasz (CH) index (Caliński and Harabasz 1974) evaluates the different number of clusters and topics. It is computed as the distances between clusters and centroids.

The Davies–Bouldin (DV) index (Davies and Bouldin 1979) is normally used as a clustering measurement criterion, which aims to maximize between-cluster distance and to minimize the distance between centroids of clusters and the other data points. For the topic extraction task, it is used to evaluate the cohesiveness of terms in each topic and the separation among topics that are learned. Lower the value, the better is the result.

## 5.2.2 Results and discussion

Table 4 presents the performance of the proposed N-CMTF and other benchmarks for understanding the topic dynamics of the misinformation detected using RMiD variations.



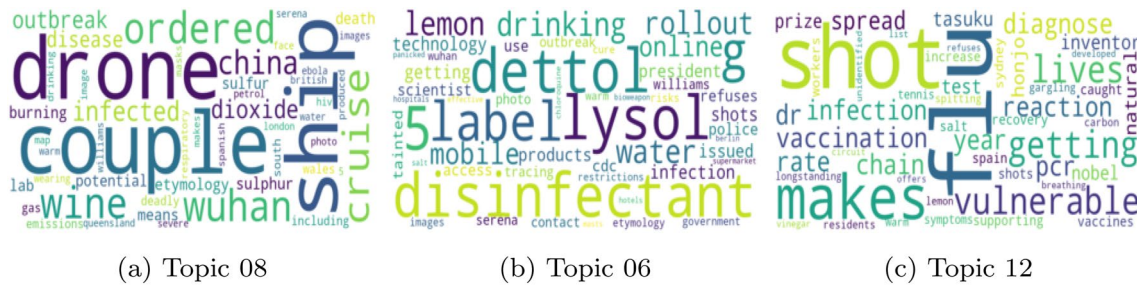


Fig. 11 RMiD-DQF. Top-3 Topics. The average similarity score is 0.69

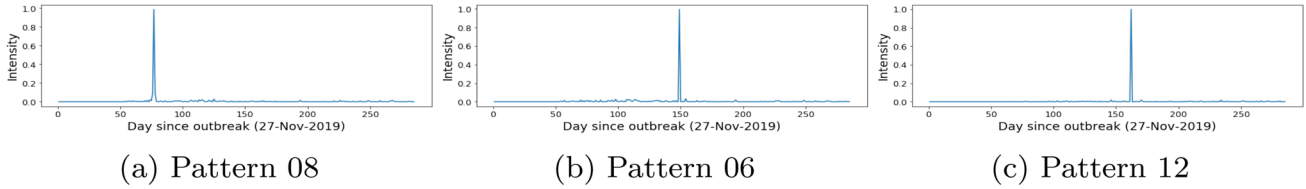


Fig. 12 RMiD-DQF. Temporal distribution of the topic

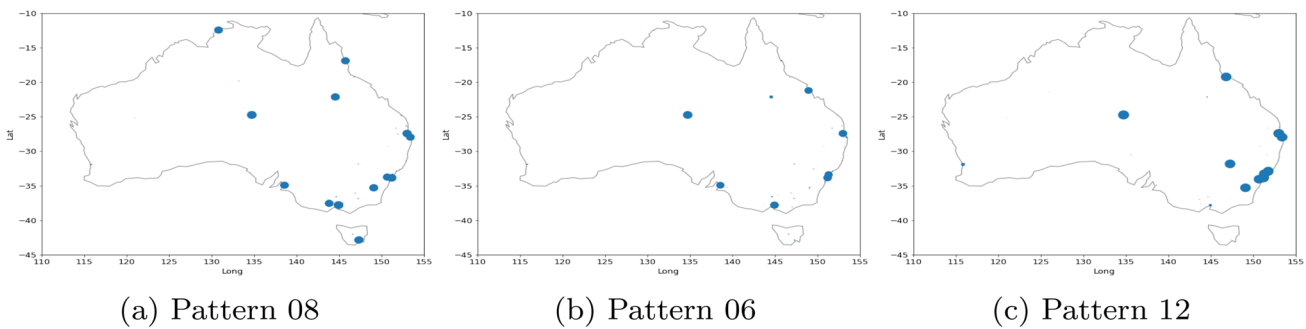


Fig. 13 RMiD-DQF. Spatial distribution of the topic



Fig. 14 RMiD-TQ. Top-3 Topics. The average similarity score is 0.65

dynamics learned by N-CMTF SaCD are of improved quality than of NTF 4D. An increase of sparsity caused by the addition of fourth dimension in NTF 4D could be a reason behind this performance. Please refer to the supplemental material for the sparsity details of NTF 4D.

Figures 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 and 19 present the top-3 topics based on the topic-terms similarity score, along with spatio-temporal patterns for all the RMiD variations (refer to respective figure caption for the average similarity score of top-3 topics). Table 5 maps

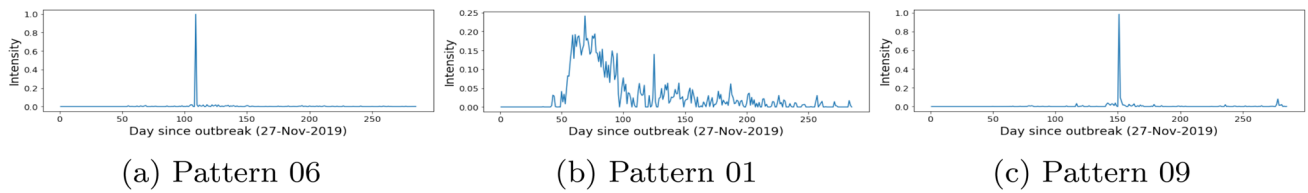


Fig. 15 RMiD-TQ. Temporal distribution of the topic

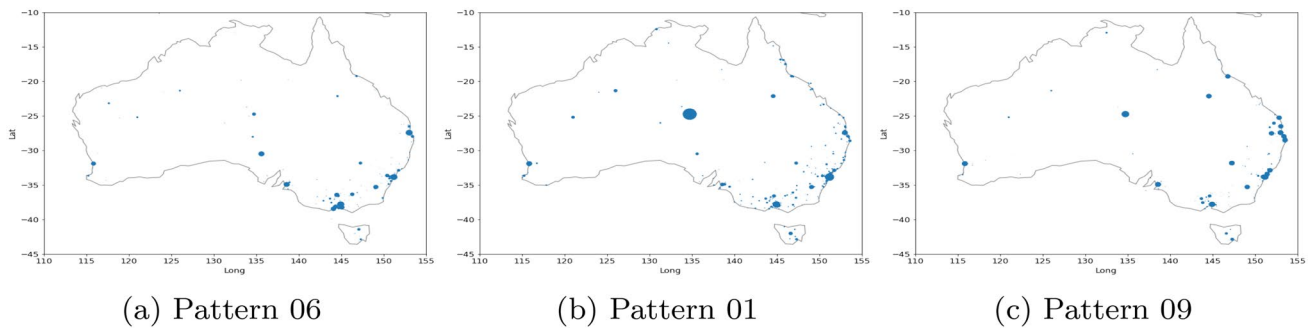


Fig. 16 RMiD-TQ. Spatial distribution of the topic

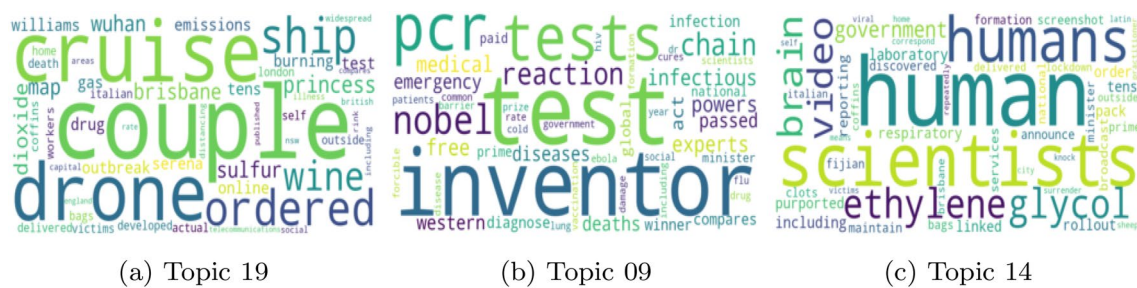


Fig. 17 RMiD-TQF. Top-3 Topics. The average similarity score is 0.76

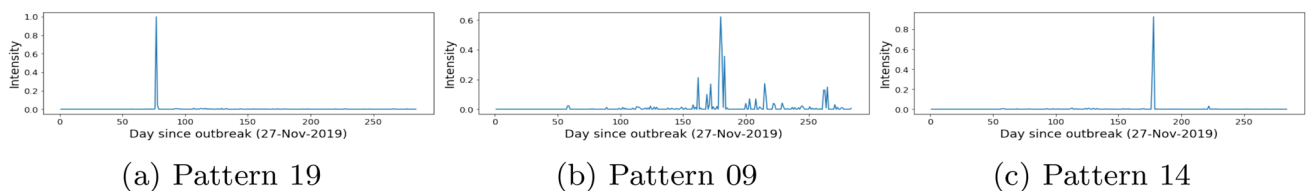


Fig. 18 RMiD-TQF. Temporal distribution of the topic

the misinformation topics of RMiD variations to the most similar misinformation from Poynter data set. While RMiD-TQF succeeds to find a match for all the topics, other RMiD variations struggle to find a misinformation topic. This is because of the identification of common topics, where

irrelevant terms gets high weightage than the terms related to misinformation.

In some cases, the misinformation from Poynter is referring to a video or an image, where the analysis of text becomes insignificant. For example, the misinformation “This is a video of panicked shoppers storming a supermarket amid the coronavirus pandemic”. sounds



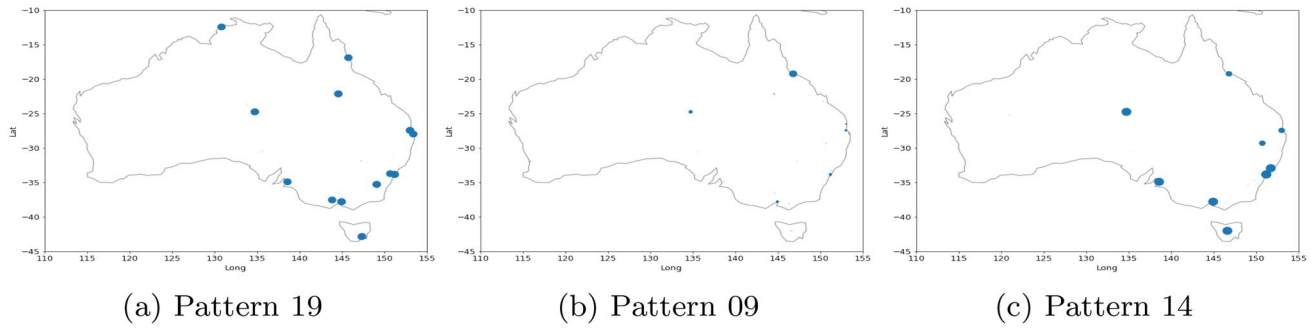


Fig. 19 RMiD-TQF. Spatial distribution of the topic

Table 5 Similarity comparison of topics identified to the original misinformation source (Poynter)

RMiD variations	Topic #	Most similar misinformation from Poynter data set
RMiD-DQ	11 (Fig. 8a)	Common topic
	10 (Fig. 8b)	Common topic
	03 (Fig. 8c)	Multiple videos make the claim that information labels on Lysol disinfectant products prove that the “human coronavirus” is not a new phenomenon
RMiD-DQF	08 (Fig. 11a)	An Australian couple ordered wine by drone on a Princess Cruise ship
	06 (Fig. 11b)	Multiple videos make the claim that information labels on Lysol disinfectant products prove that the “human coronavirus” is not a new phenomenon
	12 (Fig. 11c)	Common topic
RMiD-TQ	06 (Fig. 14a)	A Twitter post claimed Cristiano Ronaldo is transforming his string of hotels to hospitals to aid in the fight against COVID-19
	01 (Fig. 14b)	Common topic
	09 (Fig. 14c)	Android users in Australia have been automatically signed up to COVIDSafe app
RMiD-TQF	19 (Fig. 17a)	An Australian couple ordered wine by drone on a Princess Cruise ship
	09 (Fig. 17b)	A Facebook video claims the PCR test inventor said the tests do not effectively diagnose COVID-19
	14 (Fig. 17c)	A video post claims that a study by Australian scientists found Covid-19 was “human made”

to be true unless the video referred in it says otherwise. Therefore, the identification of misinformation based on the proposed and bench-marked approaches will fail. In such cases, video or image source should also be included in the process. This is a potential future direction of this paper.

## 6 Conclusion

Misinformation about highly sensitive events like Covid-19 can have many impacts and sometimes they can be life-threatening. The continued discussion of such misinformation over time and location will worsen the situation. Therefore, identification of misinformation tweets and topics in the Covid-19 situation is crucial which can help governments and organizations to control the spread and mitigate the effects. The RMiD approach proposed in this paper will help to identify misinformation tweets present in a huge collection of tweets. Moreover, the N-CMTF technique will

help to extract and understand the spatio-temporal topic dynamics of misinformation. The framework proposed in this paper reduces human effort in identifying misinformation tweets and learning their spatio-temporal topic dynamics, thus making it a feasible automated task to perform.

**Supplementary Information** The online version supplementary material available at <https://doi.org/10.1007/s13278-021-00767-7>.

## References

Abdali S, Shah N, Papalexakis EE (2020) HiJoD: Semi-supervised multi-aspect detection of misinformation using hierarchical joint decomposition. arXiv preprint, [arXiv:2005.04310](https://arxiv.org/abs/2005.04310)

Acar E, Levin-Schwartz Y, Calhoun VD, Adali T (2017) ACMTF for fusion of multi-modal neuroimaging data and identification of biomarkers. In: 2017 25th European signal processing conference (EUSIPCO). IEEE, pp 643–647

Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. *J Econ Perspect* 31(2):211–36

- Bahargam S, Papalexakis EE (2018) Constrained coupled matrix-tensor factorization and its application in pattern and topic detection. In: 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE, pp 91–94
- Balasubramaniam T, Nayak R, Yuen C (2019) Sparsity constraint nonnegative tensor factorization for mobility pattern mining. In: Pacific Rim international conference on artificial intelligence (PRICAI). Springer, Berlin, pp 582–594
- Balasubramaniam T, Nayak R, Bashar MA (2020a) Understanding the spatio-temporal topic dynamics of covid-19 using nonnegative tensor factorization: a case study. In: 2020 IEEE symposium series on computational intelligence (SSCI)
- Balasubramaniam T, Nayak R, Yuen C (2020b) Efficient nonnegative tensor factorization via saturating coordinate descent. In: ACM Transactions on Knowledge Discovery from Data (TKDD)
- Bashar MA, Nayak R, Balasubramaniam T (2020) Topic, sentiment and impact analysis: Covid19 information seeking on social media. arXiv preprint, [arXiv:2008.12435](https://arxiv.org/abs/2008.12435)
- Baziotis C, Pelekis N, Doukleridis C (2017) Dastories at semeval-2017 task 4: Deep lstm with attention for message-level and topic-based sentiment analysis. In: Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017). Association for Computational Linguistics, Vancouver, pp 747–754
- Berry MW, Browne M, Langville AN, Pauca VP, Plemmons RJ (2007) Algorithms and applications for approximate nonnegative matrix factorization. *Comput Stat Data Anal* 52(1):155–173
- Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *J Mach Learn Res* 3:993–1022
- Bro R, Andersson CA, Kiers HA (1999) Parafac2—part II. Modeling chromatographic data with retention time shifts. *J Chemom Soc* 13(3–4):295–309
- Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. *Commun Stat Theory Methods* 3(1):1–27
- Carroll JD, Chang J-J (1970) Analysis of individual differences in multidimensional scaling via an n-way generalization of “eckart-young” decomposition. *Psychometrika* 35(3):283–319
- Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter. In: Proceedings of the 20th international conference on world wide web, ser. WWW '11. Association for Computing Machinery, New York, pp 675–684
- Castillo C, Mendoza M, Poblete B (2013) Predicting information credibility in time-sensitive social media. *Internet Res* 23:560–588
- Chen Y-C, Liu Z-Y, Kao H-Y (2017) IKM at SemEval-2017 task 8: convolutional neural networks for stance detection and rumor verification. In: Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017). Association for Computational Linguistics, Vancouver, pp 465–469
- Chen D, Chen Y, Brownlow BN, Kanjamala PP, Arredondo CAG, Radszpiler BL, Raveling MA (2016) Real-time or near real-time persisting daily healthcare data into HDFs and elasticsearch index inside a big data platform. *IEEE Trans Ind Inf* 13(2):595–606
- Chen E, Lerman K, Ferrara E (2020) Tracking social media discourse about the covid-19 pandemic: development of a public coronavirus twitter data set. *JMIR Public Health Surveill* 6(2):e19273
- Cichocki A, Zdunek R, Phan AH, Amari S-i (2009) Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation. Wiley, New York
- Davies DL, Bouldin DW (1979) A cluster separation measure. *IEEE Trans Pattern Anal Mach Intell* 2:224–227
- Debnath R, Bardhan R (2020) India nudges to contain covid-19 pandemic: a reactive public policy analysis using machine-learning based topic modelling. arXiv preprint, [arXiv:2005.06619](https://arxiv.org/abs/2005.06619)
- Fuhr N, Lechtenfeld M, Stein B, Gollub T (2012) The optimum clustering framework: implementing the cluster hypothesis. *Inf Retrieval* 15(2):93–115
- Gangireddy SCR, Long DP, C, Chakraborty T (2020) Unsupervised fake news detection: a graph-based approach. In: Proceedings of the 31st ACM conference on hypertext and social media, ser. HT '20. Association for Computing Machinery, New York, pp 75–83
- Gormley C, Tong Z (2015) Elasticsearch: the definitive guide: a distributed real-time search and analytics engine. O'Reilly Media, Inc, New York
- Guacho GB, Abdali S, Shah N, Papalexakis EE (2018) Semi-supervised content-based detection of misinformation via tensor embeddings. In: 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE, pp 322–325
- Guo H, Cao J, Zhang Y, Guo J, Li J (2018) Rumor detection with hierarchical social attention network. In: Proceedings of the 27th ACM international conference on information and knowledge management, ser. CIKM '18. Association for Computing Machinery, New York, pp 943–951
- Guo B, Ding Y, Yao L, Liang Y, Yu Z (2020) The future of false information detection on social media: new perspectives and trends. *ACM Comput Surv* 53(4):1–36
- Gupta A, Lamba H, Kumaraguru P, Joshi A (2013) Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy. In: Proceedings of the 22nd international conference on world wide web, ser. WWW '13 Companion. Association for Computing Machinery, New York, pp 729–736
- Han X, Wang J, Zhang M, Wang X (2020) Using social media to mine and analyze public opinion related to Covid-19 in China. *Int J Environ Res Public Health* 17(8):2788
- Heverin T, Zach L (2012) 2 - Law enforcement agency adoption and use of Twitter as a crisis communication tool. In: Crisis information management. Elsevier, pp 25–42. <https://www.sciencedirect.com/science/article/pii/B9781843346470500023>
- Hosseinimotlagh S, Papalexakis EE (2018) Unsupervised content-based identification of fake news articles with tensor decomposition ensembles. In: Proceedings of the workshop on misinformation and misbehavior mining on the web (MIS2)
- Howell D, Rogier M, Yzerbyt V, Bestgen Y (1998) Statistical methods in human sciences, vol 721. Wadsworth, New York
- Hu X, Tang J, Gao H, Liu H, (2014) Social spammer detection with sentiment information. In: 2014 IEEE international conference on data mining, pp 180–189
- Jin Z, Cao J, Zhang Y, Zhou J, Tian Q (2017) Novel visual and statistical image features for microblogs news verification. *IEEE Trans Multimedia* 19(3):598–608
- Jin Z, Cao J, Guo H, Zhang Y, Wang Y, Luo J (2017) Detection and analysis of 2016 us presidential election related rumors on twitter. In: International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation. Springer, Berlin, pp 14–24
- Kassab L, Kryshchenko A, Lyu H, Molitor D, Needell D, Rebrova E (2020) On nonnegative matrix and tensor decompositions for covid-19 twitter dynamics. arXiv preprint, [arXiv:2010.01600](https://arxiv.org/abs/2010.01600)
- Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. *Nature* 401(6755):788
- Lee D, Seung HS (2000) Algorithms for non-negative matrix factorization. *Adv Neural Inf Process Syst* 13:556–562
- Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: International conference on machine learning, pp 1188–1196
- Luong K, Balasubramaniam T, Nayak R (2018) A novel technique of using coupled matrix and greedy coordinate descent for multi-view data representation. In: International conference on web

- information systems engineering (WISE). Springer, Berlin, pp 285–300
- Ma J, Gao W, Mitra P, Kwon S, Jansen BJ, Wong K-F, Cha M (2016) Detecting rumors from microblogs with recurrent neural networks. In: Proceedings of the twenty-fifth international joint conference on artificial intelligence, ser. IJCAI'16. AAAI Press, pp 3818–3824
- Manevitz LM, Yousef M (2001) One-class SVMs for document classification. *J Mach Learn Res* 2:139–154
- Mohotti WA, Nayak R (2018a) An efficient ranking-centered density-based document clustering method. In: Pacific-Asia conference on knowledge discovery and data mining. Springer, Berlin, pp 439–451
- Mohotti WA, Nayak R (2018b) Corpus-based augmented media posts with density-based clustering for community detection. In: 2018 IEEE 30th international conference on tools with artificial intelligence (ICTAI), pp. 379–386
- Mohotti WA, Lukas DC, Nayak R (2019) Concept mining in online forums using self-corpus-based augmented text clustering. In: Pacific Rim international conference on artificial intelligence. Springer, Berlin, pp 397–402
- Montenegro C, Ligutom C, Orio JV, Ramacho DAM (2018) Using latent dirichlet allocation for topic modeling and document clustering of dumaguete city twitter data set. In: Proceedings of the 2018 international conference on computing and data engineering, ser. ICCDE 2018. Association for Computing Machinery, New York, pp 1–5
- Nugroho R, Yang J, Zhong Y, Paris C, Nepal S (2015) Deriving topics in twitter by exploiting tweet interactions. In: 2015 IEEE international congress on big data, pp 87–94
- Nugroho R, Zhong Y, Yang J, Paris C, Nepal S (2015) Matrix inter-joint factorization—a new approach for topic derivation in twitter. In: 2015 IEEE international congress on big data, pp 79–86
- Padnekar SM, Kumar GS, Deepak P (2020) Bilstm-autoencoder architecture for stance prediction. In: 2020 international conference on data science and engineering (ICDSE). IEEE, pp 1–5
- Qazvinian V, Rosengren E, Radev DR, Mei Q (2011) Rumor has it: identifying misinformation in microblogs. In: Proceedings of the 2011 conference on empirical methods in natural language processing. Association for Computational Linguistics, Edinburgh, Scotland, pp 1589–1599. <https://www.aclweb.org/anthology/D11-1147>
- Qian F, Gong C, Sharma K, Liu Y (2018) Neural user response generator: fake news detection with collective user intelligence. *IJCAI* 18:3834–3840
- Qomariyah S, Iriawan N, Fithriasari K (2019) Topic modeling twitter data using latent dirichlet allocation and latent semantic analysis. In: AIP conference proceedings, vol 2194, no 1. AIP Publishing LLC, p 020093
- Robert C, Casella G (2013) Monte Carlo statistical methods. Springer, Berlin
- Rousseeuw J (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65. <https://www.sciencedirect.com/science/article/pii/0377042787901257>
- Sapienza A, Bessi A, Ferrara E (2018) Non-negative tensor factorization for human behavioral pattern mining in online games. *Information* 9(3):66
- Shahi GK, Nandini D (2020) Fakecovid—a multilingual cross-domain fact check news data set for covid-19. arXiv preprint, [arXiv:2006.11343](https://arxiv.org/abs/2006.11343)
- Shimizu K (2020) 2019-ncov, fake news, and racism. *Lancet* 395(10225):685–686
- Shu K, Cui L, Wang S, Lee D, Liu H (2019) defend: Explainable fake news detection. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp 395–405
- Sun L, Axhausen KW (2016) Understanding urban mobility patterns with a probabilistic tensor factorization framework. *Transp Res Part B Methodol* 91:511–524
- Sutanto T, Nayak R (2018) Fine-grained document clustering via ranking and its application to social media analytics. *Soc Netw Anal Min* 8(1):29
- Tennakoon TMG, Luong K, Mohotti W, Chakravarthy S, Nayak R (2019) Multi-type relational data clustering for community detection by exploiting content and structure information in social networks. In: Pacific Rim international conference on artificial intelligence. Springer, Berlin, pp 541–554
- Wu L, Morstatter F, Carley KM, Liu H (2019) Misinformation in social media: definition, manipulation, and detection. *SIGKDD Explor News* 21(2):80–90
- Xu W, Liu X, Gong Y (2003) Document clustering based on non-negative matrix factorization. In: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, pp 267–273
- Yang S, Shu K, Wang S, Gu R, Wu F, Liu H (2019) Unsupervised fake news detection on social media: a generative approach. *Proc AAAI Conf Artif Intell* 33:5644–5651
- Yan X, Guo J, Liu S, Cheng X, Wang Y (2013) Learning topics in short texts by non-negative matrix factorization on term correlation matrix. In: Proceedings of the 2013 SIAM international conference on data mining. SIAM, pp 749–757
- Yin H, Yang S, Li J (2020) Detecting topic and sentiment dynamics due to covid-19 pandemic using social media. arXiv preprint, [arXiv:2007.02304](https://arxiv.org/abs/2007.02304)
- Yu F, Liu Q, Wu S, Wang L, Tan T (2017) A convolutional approach for misinformation identification. In: Proceedings of the twenty-sixth international joint conference on artificial intelligence, IJCAI-17, pp 3901–3907
- Zhang Q, Lipani A, Liang S, Yilmaz E (2019) Reply-aided detection of misinformation via bayesian deep learning. In: The world wide web conference, ser. WWW '19. Association for Computing Machinery, New York, NY, pp 2333–2343
- Zhao WX, Jiang J, Weng J, He J, Lim E-P, Yan H, Li X (2011) Comparing twitter and traditional media using topic models. In: European conference on information retrieval. Springer, Berlin, pp 338–349

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.