Knowledge-Based Systems 98 (2016) 130–147

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Tensor-based anomaly detection: An interdisciplinary survey

Hadi Fanaee-T [a,*], João Gama [b]

[a] Laboratory of Artificial Intelligence and Decision Support/ INESC TEC and FCUP/University of Porto, Rua Dr. Roberto Frias, Porto 4200-465, Portugal
[b] Laboratory of Artificial Intelligence and Decision Support/ INESC TEC and FEP/University of Porto, Rua Dr. Roberto Frias, Porto 4200-465, Portugal

ARTICLE INFO

Article history:
Received 12 October 2015
Revised 18 January 2016
Accepted 20 January 2016
Available online 8 February 2016

Keywords:
Anomaly detection
Tensor analysis
Multiway data
Tensor decomposition
Tensorial learning

ABSTRACT

Traditional spectral-based methods such as PCA are popular for anomaly detection in a variety of problems and domains. However, if data includes tensor (multiway) structure (e.g. space-time-measurements), some meaningful anomalies may remain invisible with these methods. Although tensor-based anomaly detection (TAD) has been applied within a variety of disciplines over the last twenty years, it is not yet recognized as a formal category in anomaly detection. This survey aims to highlight the potential of tensor-based techniques as a novel approach for detection and identification of abnormalities and failures. We survey the interdisciplinary works in which TAD is reported and characterize the learning strategies, methods and applications; extract the important open issues in TAD and provide the corresponding existing solutions according to the state-of-the-art.

1. Introduction

Those patterns in data that do not conform to expected behavior are called *anomalies* and the process of detection of such patterns is known as *anomaly detection* [1]. Anomaly detection is an essential component of many safety, monitoring and surveillance systems. The reason is that it uncovers significant and critical facts about the system's behavior that leads to prevention of further escalation and losses. Plenty of methods have been developed during the last two decades for anomaly detection in different domains, the majority of which are covered in the survey paper [1]. One group of methods that is mentioned in this survey is spectral methods. These approaches attempt to project high dimensional data onto a lower subspace in which anomalies can be identified more easily. The main assumption of these techniques is that normal and abnormal instances appear significantly different in the projected subspace [1]. However, in many real-world applications we deal with data with tensor (multiway) structure which unfortunately is widely ignored. In such circumstances, anomalies may remain invisible with the matrix-based spectral methods. Besides, ignoring the tensor structure in data can cause some problems and result in wrong results. As an example some real failure case studies of matrix-based solutions and superiority of tensor-based solutions over them are listed in Table 1 which can manifest how much tensors are required for anomaly detection.

Although authors in [1] discuss the matrix methods in their survey, they exclude tensors and their applications in anomaly detection. This is while over the last twenty years, since the work of Nomikos and MacGregor [12], research related to tensor-based anomaly detection (TAD) has been exponentially growing. Furthermore, many methods have been developed in multiple disciplines from chemometrics and environmental monitoring to signal processing and data mining. Despite the popularity of this research area (though with different terminologies), no comprehensive survey on TAD is yet available. The most probable reason is that the TAD belongs to wide scopes and spans across different research fields.

Our main objective in this survey is to bridge the gap between two popular research areas of anomaly detection and tensors. We study the literature from all major disciplines where tensors are frequently applied and classify the contributions related to TAD based on some factors such as applications, learning types, methods and evaluation metrics. Moreover, we identify and classify the important issues and proposed solutions in TAD research. We follow a motivational strategy in this survey, in the sense that we do not limit ourselves introducing only techniques that are already applied for anomaly detection. Rather, we include those methods that are used in the close applications, such as classification, regression and forecasting that may show a great potential for anomaly detection. Therefore, this survey can be regarded as a comprehensive complement for Section 9 of [1] and from the tensor point of view it can be considered as a focused complement for applications of tensors in data mining, i.e. survey paper in [13]. Our assumption is that the reader is familiar with basic concepts

* Corresponding author. Tel.: +351 222 094 000.
E-mail address: hadi.fanaee@fe.up.pt (H. Fanaee-T).

http://dx.doi.org/10.1016/j.knosys.2016.01.027

Table 1
Some empirical evidences in the literature indicating the superiority of tensor-based solutions over matrix solutions.

Study	Tensor method	Matrix method	Matrix method's reported problem
[2]	Tucker3	PCA	Difficult interpretation of score plots
[3]	PARAFAC	PCA	Difficult interpretation of score plots
[4]	Non-negative Multiway PCA	PCA	Lower classification accuracy
[5]	Incremental Tensor subpsace learning	PCA	Lower tracking performance
[6]	Multiway PCA	PCA	Higher error rate in damage detection
[7]	Multiway PCA	PCA	Lower recognition accuracy
[8]	HOSVD	SVD	Higher prediction error
[9]	Tucker3	SVD	SVD fails on modeling tensor structured data
[10]	PARAFAC	PCA	Loss of multiway linkages plus over-fitting
[11]	PARAFAC	PCA	PCA fails to identify the right variance

Table 2
Tensor-based anomaly detection examples.

Domain	Typical tensor	Application	Ref.
Process control	Batch × Measurements × Time	Detection of faulty batches	[12]
Environment	Variables × Site × Time	Detection of spatiotemporal source of pollution	[18]
Video surveillance	ImgRow × ImgCol × Time	Abnormal event/objects discovery	[19]
Network security	OriginIP × DestIP × Time	Abnormal traffic discovery	[16]
Social networks	Person × Person × Time	Event detection	[20]
Text-based systems	Actor × Keyword × Time	Event detection	[21]
Neuroscience	Frequency × Channel × Time	Seizure recognition	[22]
Remote sensing	ImgRow × ImgCol × Wavelength	Target detection	[23]
Sensors	Measurements × Location × Time	Anomaly detection	[15]
Transportation	Origin × Destination × Time	Detection of urban traffic problems	[24]
Metallurgy Eng.	Coils × PSD × Frequency	Fault detection in hot strip mill	[25]
Civil structures	Location × Time × Frequency	Detection of damages in civil structures	[26]
Mechanical systems	Experiment × Sensor × Time	Damage detection in aircraft wing flap	[6]
Power systems	Experiments × Variables × Time	Detection of voltage sags	[27]
Medical diagnosis	Medication × Patient × Diagnosis	Heart failure prediction	[28]
Epidemiology	Space × Time × Indicators	Disease outbreak prediction	[29]
Seismology	Location × Time × Frequency	Predicting earthquake ground motion	[30]
Criminology	Lng × Lat × Time × Indicators	Crime occurrence forecasting	[31]

in anomaly detection and tensor decomposition (or tensorial learning). For this reason, we omit explanation of the straightforward concepts related to tensor decomposition, anomaly detection and spectral-based anomaly detection. Instead, we refer the reader to the recent surveys about anomaly detection [1] and tensor decomposition [13,14] that adequately cover essential technical materials for understanding the current review.

The article is organized as follows. In Section 2, we introduce the history of TAD and its applications. Section 3 presents learning methods for TAD. Section 4 discusses the techniques for tensor decomposition. Section 5 outlines the issues in TAD along with the corresponding solutions. In Section 6 we discuss the evaluation metrics used in TAD and introduce the available software for tensor analysis. Section 7 concludes the survey.

2. History and applications

A tensor is a geometric object used in mathematics and physics for extension of concepts such as scalars, vectors and matrices to higher dimensions. The origin of the word "tensor" is the Latin *tendere* "to stretch" firstly appeared in anatomy in the seventeenth century to denote muscle's stretch. It was later used in mid-eighteenth-century by William Hamilton to describe some concepts in quaternion algebra. Tensor calculus, which comes closer to the word's current meaning, was introduced in 1900 by Italian mathematician Gregorio Ricci-Curbastro and his doctoral student Tullio Levi-Civita. In 1915, tensor was used by Albert Einstein in general relativity theory for explaining geometric and causal structure of space-time and definition of concepts such as distance, volume, curvature, angle, future and past. The first principles of tensor decomposition [14] were founded by American mathematician Frank Hitchcock in 1927. Complex and multiway structure of human behaviors was probably the first motivation for use of tensors in data analysis. Psychologists such as Raymond Cattell, Ledyard Tucker and Richard Harshman were pioneers in extending tensor decomposition applications in psychology during three decades from 1940s to 1970s. In 1981, tensor decomposition was introduced by Appellof and Davidson to the Chemometrics community. The first applications of tensors in anomaly detection appeared in this community almost a decade later. The work of Nomikos and MacGregor [12] about multi-way batch monitoring was a pioneer in motivating tensor (multiway) methods in the monitoring and fault detection problems. The modern application of tensors in anomaly detection appeared a decade ago in a series of articles from Jimeng Sun and colleagues [15–17] who had a major contribution to the growth of TAD research. Nowadays, TAD's application has been widespread in wider areas, including environmental monitoring, video surveillance, network security, social networks, text-based systems, neuroscience, remote sensing, engineering and other domains. In the following, some of these applications are discussed in more detail (See Table 2 for summary).

2.1. Process control

The first footprint of tensor(multiway) methods as earlier mentioned can be seen in the monitoring of batch processes. The common objective in operating batch processes is to achieve value-added products of high-quality with competitive prices. The goal of the batch process analysis is to understand the major sources of batch-to-batch variations [12], real-time detection of faulty batches and use it to improve the operation policies.

Tensors are very popular monitoring techniques in production of chemicals and other manufacturing applications. Examples are polymerization processes [32–35], semiconductor etching process

[36–38], manufacturing pharmaceutical materials [39,40], wastewater treatment [41], bioprocesses [42], fed-batch fermentation process [40,43–45], nuclear waste storage tank monitoring [46] and winemaking process [47].

In the majority of these applications, the typical tensor is a three-order tensor of I (batch) × J (measurement) × K (time) which usually is unfolded in batch or time mode. Therefore, usually the matrix of I × JK or J × IK is processed which is called respectively batch-wise and time-wise unfolded matrix. The main goal of tensor-based batch processing is to identify the abnormal batches or time instants.

2.2. Environmental monitoring

Thanks to recent advances in sensor technologies, it is feasible to analyze tens of ecological parameters through different locations and times. The need for tensor analysis has emerged in this domain, mainly due to existing spatiotemporal variations in such data. Identification of locations or time periods related to abnormal measurement is the main goal of this application. Tensors have recently been applied in water quality monitoring [2,48–52], air pollution control [18,53] and monitoring of soil quality [54,55].

The multi-way data in these applications follows a general scheme of *variables* × *sampling site* × *sampling time* where the first dimension normally includes the chemical (e.g. oxygen rate), physical (e.g. temperature) and biological parameters (e.g. faecal coliforms) measured by the sensors.

2.3. Video surveillance

Identification of time instants in video surveillance cameras is of great interest in public security for the prevention of terrorism/crime activities. Tensors are natural data models for video data and therefore can provide more accurate framework for abnormal activity discovery. Video data can be represented as a 4D tensor of *RGB color* × *image row* × *image column* × *time* or a 3D tensor of *image row* × *image column* × *time*. The most relevant works that use tensor model for anomaly detection are [19,56] which apply TAD in video surveillance cameras. [57] also model 3D video as tensor for human action recognition. A tensor-based approach is proposed in [58] for real-time tracking of moving points from infrared image sequences. Some other works [5,59,60] also use tensors for object tracking in video data so that these works are versatile enough to be adapted for anomaly detection purposes. Some methods like [61,62] exploit tensors respectively, for crowd density estimation and motion recognition that can be useful for anomaly detection as well.

2.4. Network security

Computer-based systems are at risk from various attacks and malicious activities. Anomaly detection in these networks has been for long years the center of attention by many researchers. Tensors are powerful tools for anomaly detection in these networks. The reason is that a tensor can easily model the dynamic of traffic matrices that requires extra dimension of time. Moreover, in network security application it is very difficult to obtain labels for abnormal situations. Usually only the history of normal operation is available. Therefore semi-supervised and unsupervised tools such as tensor decomposition can be adequate tools.

There is no unique tensor data structure for analysis of network data. The majority of works use the *origin* × *destination* × *time* scheme. This format is used for analyzing a wide range of network data such as TCP/IP network, emails, phone calls, IP-TV and World Wide Web (WWW). For instance, in TCP/IP network, the two most popular models used are *SourceIP* × *TargetIP* × *Time* [16,63–65,65,66] and *SourceIP* × *TargetIP* × *Port* × *Time* [16,20]. In email or phone call networks the tensor models are constructed in the scheme of *Sender* × *Recipient* × *Time* [16,63,67–69]. There exists another type of works that model the interaction of user with the system. Examples are *IP* × *URL* × *User* × *Time* [70] and *Users* × *URL* × *Time* [71] in web-access log data and *User* × *TV Program* × *Time* in IP-TV system [72]. Anomaly detection from Internet networks are also addressed in [73]. The authors propose a method based on tensor decomposition for finding the source of disturbances originated in the network elements in a large Internet network. A three-order tensor model of *VP* × *AS* × *time* is introduced where VP denotes the vantage point and AS refers to a network element. The built model is then used to track large earthquakes occurred during the network activity.

2.5. Social networks

Social networks are a special case of general networks where nodes of networks are mostly live agents (e.g. humans) and the edges show the interaction of these agents. Tensors are normally used for the detection of anomalous people, links and communities which is obtained by taking into account their long term behavior over time. The general tensor model for this task is *person* × *person* × *time*. One of the popular tensor-based practices is related to the analysis of anomalies in electronic discussion network data set such as ENRON [16,17,67–69,74,75]. Tensors are used in analysis of Facebook data [20], phone calls [63], location-based social networks (*user* × *location* × *time*) [76] and analysis of physical social networks such as face-to-face contacts of individuals [77]. Apart from the traditional model, [78] proposed new tensor models such as *nodes* × *measures* × *time* and *communities* × *measures* × *time* for dynamic social networks where measures such as betweenness and degree closeness are computed from social network in each snapshot.

2.6. Text-based systems

Tensors are used for modeling the user/topic evolution in text-based systems. The constructed models are later applied to event and anomaly detection or co-clustering. The general tensor model for textual data analysis is *user* × *keyword* × *time*. Such model is used for anomaly detection from Twitter data [21] and analysis of chatrooms [9] and bibliographic data (*author* × *keyword* × *time*) [16,17,68]. The tensor-based topic modeling techniques such as [71] also show potential regarding text-based event detection.

2.7. Neuroscience

The brain is one of the complex systems that produces a rich source of multiway data. The reason is that every occurring activity in the brain is managed via different regions of the brain during a specific period of time. Therefore, brain data is inherently spatiotemporal. The two well-known tools for capturing the brain's activities in a machine-readable format are Electroencephalography (EEG) signals or Functional Magnetic Resonance Imaging (fMRI). The data being generated from these tools is analyzed via tensor models to detect abnormal activities or patterns in the brain. For instance, tensors are used to find the responsible regions for generating the abnormal neural activity resulting in the initial seizure discharge [22,79]. The information obtained from this analysis is very helpful for the success of an epilepsy surgery. Different from the above-mentioned application, tensors are used for mental workload monitoring of operators in safety-critical applications (e.g. controlling the Unmanned Air Vehicle (UAV) [80]).

The general tensor model for EEG data is *frequency* × *channel* × *time* [22,79–82]. If measurements are recorded across different subjects or conditions, extra dimensions can be added to the simple model. These kind of higher-order data structures are mostly used for classification purposes. For instance, in [83,84], multi-subject EEG data is modeled as a fourth-order tensor of *frequency* × *channel* × *time* × *subject*. Likewise, EEG data is modeled as a fifth-order tensor *frequency* × *channel* × *time* × *subject* × *condition* [82]. Note that the tensor models does not operate directly on EEG raw signals, instead, a preprocessing step (usually via wavelet transformation) is required to transform the raw EEG signals to EEG tensors [82].

Tensors are also applied to fMRI data analysis. fMRI images can be used to detect brain regions that have been damaged by various neurodegenerative diseases such as Alzheimer and Parkinson. A typical fMRI scan image may contain 64 × 64 × 14 voxels (3D equivalent of pixels) sampled at different consecutive time instants, producing a single matrix. Multiple scans on a given subject generate a higher-order tensor of *voxel* × *time* × *runs* which is usually used in fMRI data analysis [10]. Scans can also be performed for multiple subjects resulting in *voxel* × *time* × *subjects* [10]. The tensor model can have extra dimensions such as trials (e.g. rest, finger tapping, etc.) resulting in a fourth-order tensor of *voxel* × *time* × *trials* × *runs* [85].

2.8. Remote sensing

Nowadays, with the aid of hyperspectral imaging technology we are able to capture spectral images with a different range of spectra. We can create multiple images of a scene or object via light from different parts of the spectrum. Furthermore, these hyperspectral images can be used for target and object detection and identifying materials from long distances and of course anomalies.

The simplest tensor model used for hyperspectral images is a third-order tensor of *spatial rows* × *spatial column* × *wavelength* that is used for target detection and classification [23,86,87] or for space object material identification [88]. The more advanced tensor models are those used by [89] who add two extra dimensions to the hyperspectral tensor. The new model which is called multi-feature-tensor representation is a fifth-order tensor of *spatial rows* × *spatial column* × *wavelength* × *scale* × *direction* which scale and direction are the parameters of the Gabor function, chosen as constant numbers. The Gabor function is a popular technique for texture representation and discrimination in image processing.

The other dimension that can be added to the simple model is time. The majority of remote sensing techniques are based on the assumption that the spectral signature of objects is persistent and uniform over time, which might not be true. Therefore, a new model called multi-temporal hyperspectral tensor, denoted by *spatial rows* × *spatial column* × *wavelength* × *time* is proposed in [90]. This model is obtained by combining multiple hyperspectral images obtained at different time instances. It is considered as a new generation model of soft sensors in the remote sensing community.

2.9. Sensors

One of the potential applications of tensors is anomaly detection in sensor networks which uses the same tensor model as environmental monitoring differing in the speed by which sensor gather data and are mostly used in real time monitoring. The sensor networks are modeled as third-order tensor of *measurements* × *space* × *time* in [15,17,91]. In some other circumstances, sensors may gather some information from people. The scheme of the tensor in this condition is *persons* × *measurements* × *time*. For instance, in [92] six measurements are gathered from 20 people during a period of 255 h in an office environment. Then via tensor decomposition, some meaningful events are detected which have been linked to some regular events such as lunch break or general meeting or a monthly seminar.

2.10. Engineering

Tensor decomposition has been used in civil engineering [26,93] for detection of abnormal changes in the structure vibration response. Different sensors are employed in different parts of the structures and their vibration responses are measured during a period of time. Therefore, the tensor model is represented as *space* × *time* × *frequency*.

Application of tensors in metallurgy engineering can be seen in [25] where tensor decomposition is used for fault detection in the hot strip mill, specifically for damage on the surface of coils. The data generated from ASIS (automatic surface inspection system) is modeled as a third-order tensor of *Coils* × *PSD* × *frequencies* where PSD (power spectrum densities) and frequencies are obtained via autoregressive processes of several signals modeled by Fast Fourier transform (FFT).

An example from the mechanical engineering domain can be observed in [6] where tensors are applied to detect damage in sensitive artefacts such as aircraft wing flap. The main problem in aircraft wing flap includes barely visible impacts on its surface. To deal this problem, the authors propose a new multiway model for detection of damages via monitoring multiple sensors. They suggest a tensor scheme of *experiment* × *sensor* × *time* for the analysis task.

The electrical engineering community has also used tensors for voltage sag detection in power distribution networks [27]. The tensor model of *experiments* × *variables* × *time* is proposed which later is unfolded time-wise to detect sag points.

The robotic engineers also used tensors for prediction of fall up in walking robots [94]. Inspired by the tensor-based batch process monitoring, they model the non-linear trajectory of walking robots and suggest a third-order tensor of *trajectory slices* × *scaled state variables*(*e.g. position*, *angle*) × *time* for fault detection.

2.11. Transportation systems

Traffic data (*Origin* × *Destination* matrix) is frequently used for traffic planning and management in intelligent transportation systems. Tensor decomposition has been used on the tensor *Origin* × *Destination* × *Time* for discovery of spatiotemporal traffic structure [24,95] that has important applications to urban planning and traffic jam control. Sometimes the collected data might also be abnormal due to failures in the collection process and recording systems. This problem which is known as outlier recovery is addressed in [96] with tensors. Tensors also are used for prediction of missing values in traffic tensors (known as tensor completion) [97].

2.12. Medical applications

Tensors are exploited for analysis of electronic medical records. In [98] a change detection system is developed for pain management decision making. A collection of medical forms completed at various treatment and recovery stages are modeled as a sixth-order tensor of *initial pain* × *initial infusion* × *sex* × *surgery site* × *pain* × *month* and based on that some interesting change patterns are detected. Tensor decomposition is also applied to electronic health records (EHR) for prediction of heart failure [28]. A tensor model of *Medication* × *Patient* × *Diagnosis* is used for this purpose. Tensors are also used in bio-informatics for modeling micro-array gene expression tensors (*gene* × *sample* × *time*) that can be used for diagnosing diseases [99]. Tensor decomposition has recently been

Table 3
Existing and potential learning techniques for tensor-based anomaly detection.

Model	Category	Examples
Supervised	Dimensionality reduction based	Categorical target [26,93]
		Numerical target [30]
	Classification based	Support tensor machines [89]
		Supervised tensorLearning [111]
		Tensor least square [112]
		Multilinear discriminant analysis [114]
		Factorization machines [115]
		Tensor subspace learning [116]
	Regression based	Multiway PLS (N-PLS) [108]
		Tensor ridge regression [121]
		Support tensor regression [121]
		H-MOTE [123]
		Tensor regression [122]
	Time series based	Multilinear dynamical systems [124]
		Greedy low-rank tensor learning [125]
		Tensor hidden Markov model [126]
		Tensor time series models [127,128]
		Tensor singular spectrum analysis [129]
		TriMine [71]
Semi Supervised	Monitoring of decomposition statistics (SPE, T2, etc.)	[18,25,35,39,40,44,45,94,118,119,130–132,134–136]
	Eigenspace based	[29,100]
Un-supervised	Analysis of score-plots	1D [32,76,77]
		2D [18,37,64,83]
		3D [64,100]
		Latent factors time series [48,67]
		Multivariate-SPC on multiple latent factors [95]
	Streaming residuals	Dynamic tensor analysis [16,59,98,101]
		Window-based tensor analysis [15]
		Spatio-temporal tensor streams [91]
	Histogram based	[133]

exploited in epidemiology for detection and spotting disease outbreaks [29,100]. A third order tensor of *Space* × *Time* × *Indicators* is suggested for the monitoring task.

2.13. Other applications

Many other applications from tensor-based methods have appeared in recent years, in particular during the last five years that are inherently different from the traditional applications of tensors. In [11] spectral changes of substrates and products are monitored in real time via modeling temporal evolution of enzyme activity with third-order tensor of *wavenumber* × *time* × *activity*. Tensor analysis is applied for tracking the analysis of proteins. In [101] authors use tensor analysis to model the deviations of contacts between residues and their environment with respect to each other (i.e., relative behavior) as well as with respect to time (i.e. temporal behavior). The tensor model used in this work is in scheme of *contract matrix* × *time* where the contact matrix $A_{ij}(t)$ represents the normalized value of the number of heavy atoms in residue i coming in contact with the heavy atoms in residue j at time t.

A dynamic pattern of international trades and the asymmetric relations between countries is studied in [69] which can potentially be applied for anomaly detection (e.g. economic crisis).

Tensor decomposition has applications in seismology. A third-order tensor of *space* × *time* × *frequency* is built in [30] for the prediction of ground motion after earthquake. Time-frequency components are obtained by transforming of acceleration records of earthquake ground motions with continuous wavelet transform.

Tensor decompositions are used for analysing climate tensors *climate indicator* × *grid* × *time* [102–104] which makes them capable techniques for prediction of climate changes.

Tensors are used for crime forecasting [31]. A fourth-order tensor of *longitude* × *latitude* × *time* × *measurements* is used for this purpose where measurements refer to criminal activities such as residential burglaries, construction permits, motor vehicle larceny, offender data, etc.

One of the recently emerged topics in anomaly detection is acoustic anomaly detection in which several acoustic sensors are monitored for event detection. Acoustic anomaly detection can be used, for instance, in safety monitoring of nuclear power plants [105]. Unfortunately, although tensor decomposition shows great potential, is not yet used for this purpose, whereas we can find works that model voice data as a third-order tensor of *rate* × *scale* × *frequency* [106,107] or *rate* × *time* × *frequency* [4]. These tensor models might be used for acoustic anomaly detection.

3. Tensor-based anomaly detection: existing and potential methods

Tensor methods are better known for unsupervised and semi-supervised learning. However, in recent years, many supervised tensor learning methods and tensor time series models have been developed. Some of these recent techniques are not yet used for anomaly detection, but might prove themselves useful for this purpose. Table 3 presents the summary of these methods with corresponding references. In the following, these strategies are described in more detail.

3.1. Supervised models

Perhaps we can seek the first footprint of using tensors in supervised anomaly detection in Multiway PLS models [108]. The second important role of tensors is in dimensionality reduction for classification problems. Nowadays, more supervised tensor-based learning methods are developed. Some of these techniques, in spite of their potential for anomaly detection, are not yet applied for this application. The goal of this section is to provide a structured list of existing and potential approaches.

3.1.1. Tensor decomposition for dimensionality reduction

In this category of supervised models, tensor decomposition is used as a dimensionality reduction tool for feature extraction (a more advanced alternative for matrix-based dimensionality reduction solutions such as PCA). Depending on the target value, methods can be grouped in two categories.

In the first group of methods [26,93,109], it is assumed that we have two sets, train and test, where train set contains normal samples. Tensor decomposition is applied on the normal tensor as a dimensionality reduction tool. Then, one of the factor matrices (usually time) is fed to a regular classifier (e.g. k-nearest neighbors or SVM) for making a model from the normal samples.

The goal is to predict the labels of the observations in the test set. Therefore, the built model from train set is used to predict the label (normal or abnormal) of observation in the test factor matrix. For instance, in [26,93], a PARAFAC decomposition with k number of components is applied on the $space \times time \times frequency$ tensor corresponding to the normal samples and then the derived time factor matrix is trained via k-NN (features are the latent variables). The built model is then used for classification of time points in the arriving data. In other related work, a combination of PARAFAC and self-organizing map (SOP) is used [90] for classification of signatures of multitemporal-hyperspectral images.

The second group of methods [30] follows the same procedure as the former, but instead of binary labels (abnormal/normal) a numeric target is given for prediction. Therefore, regression models are replaced with categorical classifiers. Targets can be single or multiple variables. For instance, in [30] the authors propose to train a GRNN (generalized regression neural networks) on the tensor subspace latent variables for prediction of multiple seismological variables. They used this method for prediction purposes. This kind of approaches can be easily extended for anomaly detection. A further step, however, is required. For instance, the difference of predicted and actual values can be used along with a threshold to detect anomalies.

Note that tensor decomposition is not necessarily used as dimensionality reduction tool in classification tasks. Rather, it can serve along with various other tasks such as case-based reasoning [6] and clustering [72,110].

3.1.2. Tensor classifiers

Tensor classifiers are those that adapt regular classifiers for tensorial data. In these methods, data is trained directly via tensor-based classifier and then the built model is used for prediction. A binary tensor classifier has a great ability for anomaly detection from multiway data. A good example for this category is a method presented by Zhang et al. [89] where SVM (support vector machines) is extended to STM (support tensor machines). The new tensorial classifier is trained directly with the tensorial data of specific objects and then the built model is used for target detection. In another work [111], a general framework called Supervised Tensor Learning (STL) is proposed that adapts many conventional machine learning methods to take higher order tensors as inputs. This model is successfully tested for binary classification problems which can be very useful for anomaly detection. In [112], in addition to another version of STM a new method is also presented called Tensor Least Square (TLS) which is the extension of least square classifier. A new type of STM is also presented in [113] which is applied for gait and action recognition.

Multilinear discriminant analysis (MDA) [114] is also proposed for tensor-based image classification that is an extension of Linear discriminant analysis (LDA) for tensor data. Factorization machines [115] is another method for tensor-based classification that extends SVM for tensors using PARAFAC which is motivated for SVM difficulty in collaborating filtering problems. Tensor classifiers are also known as supervised multilinear subspace learning in image processing community. The recent survey paper [116] covers the majority of advances for tensor subspace learning.

3.1.3. Tensor regression

The first tensor regression models emerged in the 1980s from the Chemometrics community as the traditional name of N-PLS or multiway PLS [117]. In these techniques which are widely used for anomaly detection [33–35,50,108,118–120] a model is built based on the relationship of the input tensor (X) to some quality measurements (Y). That model is then used for predicting the quality measurements of new tensors. Deviations of predicted target variables from the normal reference are interpreted as abnormal behavior.

Apart from the traditional multiway regression models, some novel techniques have been recently developed in different research communities. One is [121] that proposes two tensor regression models called tensor ridge regression (TRR) and support tensor regression (STR) that respectively extend vector regression models such as ridge regression (RR) and support vector regression via some properties of PARAFAC model. The authors apply these methods to facial data for human-age estimation and head/body-pose prediction. These methods can be quite interesting for a couple of problems in TAD.

Another tensor regression model is proposed in [122] which is motivated by some problems in brain imaging where observed binary diagnosis status (Y) is required to be modeled based on the fMRI images as an input tensor (X). The proposed tensor model is used to identify regions of interest in brains that are relevant to a clinical response with applications for detection of brain diseases, including Attention Deficit Hyperactivity Disorder and Alzheimer.

Moreover, Zhu et al. [123] proposes a tensor-based regression algorithm called H-MOTE that is capable to incorporate background knowledge into the model. This model is used for prediction of wafer quality in semiconductor manufacturing.

3.1.4. Tensor forecasting

Tensor forecasting is an extension of vector time series models for multiway time series. The procedure for anomaly detection is the same as in univariate ones. A model is built for tensor time series and then based on that model, future tensors are predicted. In the subsequent moment, if the tensor has a considerable difference with the predicted tensor, it is marked as an anomaly. Different methods are developed for tensor forecasting. In [124] a model called Multilinear Dynamical Systems (MDS) is proposed, which is a tensorial extension of linear dynamical system (LDS). Detection of the market collapse and climate change are introduced as applications of this methodology. Another tensor forecasting method, named Greedy Low-rank Tensor Learning is proposed in [125] that is applied for forecasting tensor time series such as climate tensors.

Some time series analysis tools are also extended for tensors. For instance, a tensor-based Hidden Markov Model (HMM) approach is proposed in [126] and is used for fault detection and prediction. Some ideas in time series analysis, such as weighting and averaging are also extended for tensor analysis in [127,128]. The tensor version of singular spectrum analysis (SSA) is also presented in [129], replacing SVD with PARAFAC in regular SSA and is applied for a non-stationary source separation of seizure signals.

An innovative approach called TriMine [71] is also proposed for tensor forecasting in the context of topic modeling. In the proposed methodology, a train tensor data is decomposed as a regular tensor decomposition and then based on the obtained time factor matrix, the next factor matrix is predicted with different scales.

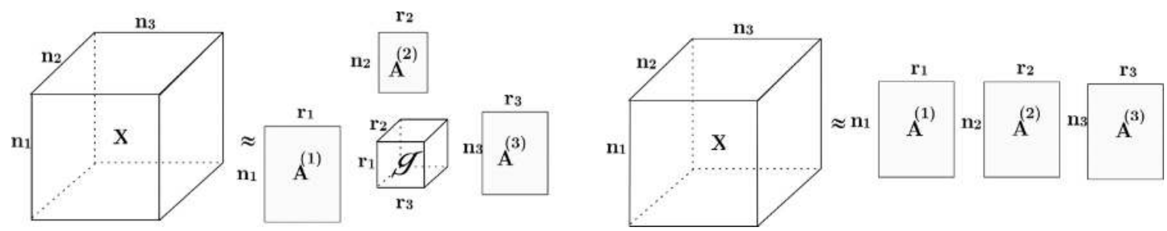

Fig. 1. Left) Tucker3 Decomposition: the third-order Tensor X is decomposed to a smaller core tensor and three factor matrices. Right) CP/PARAFAC decomposition: a third-order tensor X is decomposed to three factor matrices.

Later, the new predicted time factors are multiplied by other two dimensions to construct the tensor forecast for both short-term and long-term. This approach seems promising for multi-scale anomaly detection and prediction.

3.2. Semi-supervised models

Semi-supervised methods are twofold. The first group is originated in online fault detection from batch processes where a train tensor model corresponding to normal operation condition (NOC) is usually constructed. Then, arriving data is monitored to detect deviations from NOC model using statistics such as Squared prediction Error (SPE) or Hotelling T^2 chart [12]. Examples of this category are explained in [12,44,45,118,130–132]. There exists another group of methods that instead of the above statistics monitor the angle between Eigenvectors or Eigenvalue magnitudes in the test set in comparison with the train set. Examples of this category are explained in [29,100].

As semi-supervised models impose less human intervention, they are more desirable comparing to supervised methods. In many applications such as process control or network security, labeling data for each time instant is infeasible. Therefore, this model presents superior flexibility and simplicity.

3.3. Unsupervised models

Tensors are better known for unsupervised learning in problems such as co-clustering and anomaly detection. In this section, the popular unsupervised methods are described.

3.3.1. Score plot-based

The most traditional use of tensor decomposition in anomaly detection is with score plots obtained from the decomposition that are analyzed manually or automatically for anomaly detection or clustering. Score plots can be 1D (only one factor) [32,76,77], 2D [18,37,64,83] and 3D [64,100]. If the latent factor is time, some factors might be presented as a multivariate time series [48,67]. Sometimes this multivariate time series may also be monitored automatically with multivariate SPC methods such as Hotelling T^2 [95].

3.3.2. Streaming decomposition error-based

This group of methods is composed by those streaming decomposition methods that operate on data incrementally without the requirement for a train set. They monitor the decomposition reconstruction error for each tensor in each time instant. Anomalous time instant is the one which corresponding reconstruction error goes beyond a pre-defined threshold (e.g. twice standard deviation of errors so far). Examples are given in [5,16,17,56,59,91].

3.3.3. Histogram-based

Fanaee and Gama[133] proposed an efficient multi-aspect-streaming tensor analysis approach called MASTA based on online histograms. In this approach, the whole tensor is vectorized and is simultaneously segmented into slices in each mode. Then the distribution of each slice is compared versus the vectorized tensor using a standard metrics such as Earth Mover's Distance. The used logic is that tensor information is distributed over slices in each mode. By matching slices with the reference distribution, similar slices can be identified as well as anomalous slices.

4. Tensor decomposition

Traditional data analysis techniques, such as the PCA, clustering, regression, etc. are only able to model second-dimensional data and they do not consider the interaction between more than two dimensions. However, in several real-world phenomena, there is a mutual relationship between more than two dimensions, in particular, when the time dimension is added to the problem. Tensor (Multi-way) data analysis considers all mutual dependencies between the different dimensions and provides a compact representation of the original data in lower dimensional spaces. The most common multi-way analysis techniques are that of Tucker [137] and CP/PARAFAC [138,139], which are generalized versions of PCA or, more specifically, Singular Value Decomposition (SVD) for higher order matrices.

Among many types of tensor decomposition approaches, Tucker and PARAFAC/CP models are the most used ones. Tucker decomposition approximates a large tensor by a product of a smaller tensor with predetermined dimensions (called core tensor), multiplied by factor matrices in each dimension (See Fig. 1Left). Formally, the problem can be defined as an optimization problem [140]: Given a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_d}$, find a core tensor $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times \ldots \times r_d}$ with pre-defined integers r_i with $1 \leq r_i \leq n_i$ for i = 1,2,..., d. and factor matrices $\mathbf{A}^{(i)}$ that optimizes

$$min \ \left\| \mathcal{X} - \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \ldots \times_d \mathbf{A}^{(d)} \right\| \quad (1)$$

Subject to:
$\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times \ldots \times r_d}$,
$\mathbf{A}^{(i)} \in \mathbb{R}^{n_i \times r_i}$, $(\mathbf{A}^{(i)})^T \mathbf{A}^{(i)} = I, i = 1, 2, 3.$

In the above model, d represents the dimension of the tensor (e.g. For three-dimensional tensor, d=3) and r_1, r_2, ..., r_d ($i = 1, 2, \ldots, d$) are model input parameters (core size). The simplest algorithm for finding matrices $\mathbf{A}^{(d)}$ and $\mathcal{G}$ is a method called High-order SVD (HOSVD) [141] where firstly tensor is unfolded into lower-order matrices over all its modes(e.g. unfolding $I \times J \times K$ tensor to $I \times JK$ or $J \times IK$ or $J \times IK$ matrices) and then SVD is independently performed on each matrix (e.g. $I \times JK$ matrix) . The more sophisticated approach is high-order orthogonal iteration (HOOI) [142] that uses alternating optimization to find better projection matrices iteratively. In the HOOI algorithm, HOSVD can used for better estimation of the initial elements of $\mathbf{A}^{(d)}$ and $\mathcal{G}$.

PARAFAC/CP also is a special case of Tucker model where the core tensor is super-diagonal. Therefore, obtaining (1) for PARAFAC/CP is straightforward. Although, there exist other kinds of decomposition models, the algorithmic details of these kind of approaches is out of the scope of this survey. However, the interested

Table 4
Methods for tensor decomposition and applications to anomaly detection.

Family	Method	Anomaly detection example
Tucker	Multiway PCA (Tucker1) [46]	[25,25,27,27,39,39,44,47,94,110,134,135,147–149]
	GTucker2 [150]	[150]
	Tucker3 [137]	[2,2,3,9,48,49,53,72]
	Non-negative Tucker [151,152]	[24,83]
	HOSVD [141]	[8,29,61,153]
PARAFAC	PARAFAC [138]	[3,10,11,26,30,49,63,73,76,80,81,90,154]
	Non-negative PARAFAC [143]	[20,67,77,84,144]
	PARAFAC2 [145]	[37]
	Dynamic PARAFAC [35]	[35]
	CP-APR [146]	[70]
DEDICOM	[155]	[69,156]
Bayesian	EM-based (pTucker [157], ETF [158], InfTucker [75])	[75,92,158]
	MAP-based (ARD [159], FBCP [160])	
	Gibbs sampling (Multi-HD [161], BTA [162], BPTF [163], TriMine [71], MGP-CP [164], sp-PARAFAC [165])	[71]
LPP	TLPP [166,167]	[40,58]
	TGLPP [168]	[168]
ICA	Tucker1-based (MICA [45], MKICA [169], FS-MKICA [132])	[45,132,169]
	Tucker3-based [104]	[104]
	PARAFAC-based [85]	

readers are referred to [13,142] for more technical details about these models.

In the following we list the six main categories of methods for tensor analysis that can potentially be used for anomaly detection, including PARAFAC-based, Tucker-based, DEDICOM-based, Bayesian, Locality Preserving Projection (LPP) based and ICA-based. Table 4 demonstrates the summary of existing techniques.

4.1. PARAFAC-based

4.1.1. PARAFAC

PARAFAC/CP has been the most used decomposition model among other models. The reason is probably the similarity of implementation and interpretation, such that PARAFAC as PCA requires only one user input which is the number of components. PARAFAC model is applied in wide range of anomaly detection tasks in various domains. For example see [3,11,26,49,73,81].

4.1.2. Non-negative PARAFAC

One of the important issues in tensor decomposition is that elements in factor matrices can get negative values. These negative scores cannot be justified with the our physical knowledge (e.g. fMRI tensors). This might not be a problem when we want to work directly on eigenspace, but might be a constraint when we want to perform our analysis on the obtained components. This problem is mostly motivated by the chemometrics and neuroscience community where the output of tensor decomposition requires to be interpreted by a specialist. PARAFAC model with non-negative constraint is called non-negative PARAFAC or non-negative tensor factorization (NTF) which was presented for the first time in [143]. Nowadays, NTF has become remarkably popular due to its meaningful and physical interpretation, especially in manual score-plot based anomaly detection [20,67,77,84,144].

4.1.3. PARAFAC2

In some specific circumstances as occur in batch monitoring, a tensor with uneven-length slices appears. For instance, in batch monitoring with tensor of *batch* × *measurement* × *time*, the matrix *measurement* × *time* can be of different length for each batch due to different elapsed time for the batch. PARAFAC2 [145] which is an extension of PARAFAC provides a solution for such problems. It is used in [37] for fault detection from batch tensors with unequal time axis and its superiority over regular PARAFAC and Tucker is shown.

4.1.4. Dynamic PARAFAC

A procedure called DPARAFAC (dynamic parallel factor analysis) is introduced in [35] for online fault detection in process monitoring. This methodology includes two phases: learning and detection. In the learning phase, we are given the data of normal operation condition (NOC). Each slice of the NOC tensor (matrix *measurement* × *time*) is segmented into different equal-length windows in the time axis. Then all the segments together form a new tensor (*measurement* × *window* × *time*). PARAFAC is then applied on this tensor for each batch and loadings are obtained. The average of factor matrices for each window is obtained for all batches. Later, some statistics such as T^2 and control limits are computed for each time point. In the detection phase, when new batches of data arrives, it is arranged as the previous procedure, and is then projected onto the previous under-control subspace to assess its degree of abnormality.

4.1.5. Poisson tensor factorization

Poisson tensor decomposition (PTF) [146], also known as CANDECOMP/PARAFAC Alternating Poisson Regression (CP-APR) uses a new fitting algorithm based on Kullback–Leibler (KL) divergence instead of common ALS fitting algorithm in PARAFAC. The idea of such approaches is that count data can be better described by a Poisson distribution rather than Gaussian distribution. This model is suggested for anomaly detection from count data [70].

4.2. Tucker-based

4.2.1. Tucker1

Tucker1 or Multiway PCA (MPCA) is the first tensor model used for TAD in many applications [25,25,27,27,39,39,44,47,94,110,134,135,147–149]. Tucker1 is used when variance is only important in one dimension. Therefore, the tensor is usually unfolded through one dimension and then regular PCA is applied to the unfolded data. For instance, in batch monitoring, Tucker1 model is used on batch-wise or time-wise unfolded matrices.

4.2.2. GTucker2

Tucker2 model is barely used for anomaly detection. Only very recently a generalized version of Tucker2 called GTucker2 was proposed [150] for fault detection from tensors with unequal slice lengths. GTucker2 is equivalent to PARAFAC2, such that PARAFAC2 can be viewed as a constraint version of GTucker2. In [150] the superiority of GTucker2 is shown over Tucker1, PARAFAC, Tucker3 and PARAFAC2 on this specific problem.

4.2.3. Tucker3

The other model which promises more flexibility is known as Tucker3. This model as is presented in the previous section is normally used when there is multiway variations in all modes [2,2,3,9,48,49,53,72]. For instance, for water quality tensors, we are interested in discovering abnormal locations, time instants and measurements that are more correlated to anomalies. Therefore, Tucker3 is the preferred model [2].

4.2.4. Non-negative Tucker

There are some extensions of NTF for Tucker decomposition, called Nonnegative Tucker decomposition [151,152]. The NTF is used for modeling EEG tensors [24,83] performing better than NMF (non-negative matrix factorization) in some circumstances.

4.2.5. HOSVD

Higher-order singular value decomposition (HOSVD) is a generalization of SVD for higher-order tensors. HSOVD can be viewed as a special case of the Tucker3 model when ALS optimization is not performed, rather the tensor is unfolded across different modes and then regular SVD is applied on the unfolded matrices. Therefore, HOSVD does not provide the best approximation of a tensor, it is rather used as an initialization step in Tucker3 for reducing the number of iterations in ALS procedure [13].

4.3. ICA-based

Independent component analysis (ICA) is a popular method for decomposing a multivariate signal into additive subcomponents. The basic assumption in ICA is that subcomponents are independent, non-Gaussian signals. Extension of ICA for tensors is available for Tucker1 (MPCA) [45,132,169], Tucker3 [104], and PARAFAC [85]. All these methods except the latter one are applied for anomaly detection.

4.4. DEDICOM-based

DEDICOM (DEcomposition into DIrectional COMponents) [145,155] is a generalization of PARAFAC2 for discovering asymmetric relationships between two modes that refer to the same type of object (e.g. transactional data). This model has been found to be effective in temporal analysis of social networks [69,156]. Therefore, it can be used for event detection goals in similar scenarios.

4.5. Bayesian methods

Traditional tensor decompositions are unable to handle issues such as missing values, outliers, noises and different data types. Recently, probabilistic methods started to be taken into consideration due to their flexibility and less restrictive assumptions. They are successfully applied to anomaly detection problems [71,75,92,158] and is expected that the number of their applications be increased in near future, especially when the majority of these approaches can estimate the tensor rank during the decomposition process.

Bayesian approaches, based on the means of used statistical inference can be divided into three categories. The first group is based on the Expectation maximization (EM) algorithm, including pTucker [157], Exponential Family Tensor Factorization (ETF) [158] and Infinite Tucker (InfTucker) [75]. The second group exploits maximum a posterior (MAP) estimation, such as Automatic Relevance Determination (ARD) [159] and Fully Bayesian CP Factorization (FBCP) [160]. Finally, the third category uses gibbs sampling as an inference engine. Examples are Multi-HD [161], Bayesian tensor analysis (BTA) [162], Bayesian Probabilistic Tensor Factorization (BPTF) [163], TriMine [71], multiplicative gamma process based CP decomposition (MGP-CP) [164] and sp-PARAFAC [165].

4.6. Locality preserving based methods

Tensor decomposition methods such as Tucker and PARAFAC do not consider the intrinsic local geometric structure of tensors. A recent group of techniques is developed for dealing with this problem on the basis of locality preserving projections (LPP). It has been shown in [40,170] that LPP-based approaches have better performance than conventional PCA-based methods which preserve only the global Euclidean structure. LPP-based approaches are more attractive when two dimensions of tensors are in a pairwise relationship (e.g. image data).

The most popular method for this family is Tensor Locality Preserving Projection (TLPP) [166,167] which is applied to detection problems [40,58]. A more sophisticated version of TLPP has been proposed very recently, called Tensor Global-Local Preserving Projections (TGLPP) and is applied for the fault detection problem in batch processes [168] which is able to capture both global and local structures of tensors simultaneously.

4.7. Tensor rank estimation

The quality of the tensor model has a direct relationship with true model selection. Although estimation of tensor rank is an NP hard problem [185], in the majority of cases, an optimal low-rank approximation is desirable. In the majority of works discussed in this survey, it is assumed that the number of components is known in advance via knowledge of the underlying phenomena. However, this might not be the case in many applications. Some approaches are developed for estimation of optimal number of ranks for both tensor decomposition approaches. Some of these approaches are listed in the below subsections (See Table 5 for summary).

4.7.1. Cumulative sum of the percentage of eigenvalues or explained variance

This is the most basic method for choosing the number of components. It is mostly used for MPCA (Tucker1) models. The number of principal components is chosen based on the cumulative percentage of eigenvalues or cumulative percentage of the explained variance. If the cumulative percentage of first k components is over a threshold (e.g. 75%), k is selected as the adequate number of components. For instance, [43] uses the eigenvalue criterion and [33,47,171,172] use cumulative variance for anomaly detection in process batch tensors.

Sometimes, instead of a threshold cut point, broken stick rule [173] is used. This approach assumes that percentage of explained variance (or eigenvalues) of a random data when is divided randomly amongst k components follows a broken-stick distribution $G_k = \frac{1}{p}\sum_{i=k}^{p}\frac{1}{i}$. Therefore, the k-*th* principal component is valuable if its value is greater than G_k (i.e. a random PC). This rule is used for model order estimation of Tucker1 for anomaly detection [33,134].

Table 5
Methods for tensor rank estimation.

Method	Common use	Fast	Auto	Application to anomaly detection
Cumulative sum of percentage of eigenvalues [43]	Tucker1	No	Yes	[43]
Cumulative sum of explained variance [33]	Tucker1	No	Yes	[33,47,171,172]
Broken stick rule [173]	Tucker1	No	Yes	[33,134]
Cross-validation [174]	Tucker1/ Tucker3/ PARAFAC	No	Yes	[33,33,35,45,135,175]
CORCONDIA [176]	PARAFAC	No	Yes	[3,18,22,26,49,77,154]
DIFFIT [177]	Tucker3	No	Yes	[83,84]
FastDIFFIT [178]	Tucker3	Yes	Yes	[95]
Multiway scree plot [179]	Tucker3	No	No	[2,37,49,51,53,72]
Split-half analysis [180]	PARAFAC	No	Yes	[3,49]
Maximum block improvement [140]	Tucker3	Yes+	Yes	[95]
Convex hull [181]	Generic	No	Yes	Is not yet applied for anomaly detection but is used for tensor rank estimation [159].
Akaikes information criterion (AIC) [182]	Generic	No	Yes	
Bayesian information criterion (BIC) [183]	Generic	No	Yes	
Automatic relevance determination (ARD) [159]	Generic	Yes	Yes	
Genetic algorithm [184]	Tucker3	No	Yes	Used for noise removal [184]

4.7.2. Cross-validation

A popular method for finding the adequate model order in component analysis is cross validation [174]. This technique is applied for fault detection problem in [33,35,45,135] for estimation of number of components in MPCA model and its extension is presented in [175] for Tucker3 and PARAFAC models. The basic idea of cross-validation is leaving out a single data element [175], a slice [186] or random half of a slice [187] at a time, perform tensor decomposition and then compute the Predictive Residual Error Sum of Squares (PRESS) = $\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(\tilde{X}_{ijk}^{PQR} - X_{ijk})$ for the elements not included in the model building. Finally, the sum of PRESS values for each principal component (p,q,r) is calculated for all eliminated parts to compute $PRESS_{pqr}$. Those (p,q,r) that give the minimum PRESS are considered a good model dimension. The more sophisticated cross-validation approaches are developed based on w-statistics [175] which use F-test strategy to determine whether an additional component is worth to adding or not.

4.7.3. CORCONDIA

Core consistency test (also known as CORCONDIA) [176] is a heuristic method used for the determination of the number of components in PARAFAC model. It is widely applied in anomaly detection from tensors [3,18,22,26,49,77,154]. Assuming P as the number of components in PARAFAC model, CORCONDIA checks the superdiagonailty of Tucker3 model with a core size of (P,P,P). If all elements in the core tensor except those with same indices $(i = j = k)$ become zero, it concludes that the PARAFAC model fits perfectly. The procedure is as follows. First, core consistency criterion is defined as the similarity percentage of Tucker3 core size with superdiagonal array T of ones and only then is PARAFAC fitted for a series of models from P=1 to F, computing core consistency for all these models. The last model in these series which corresponding Tucker3 core is similar to T is considered as the adequate number of components.

4.7.4. DIFFIT

DIFFIT (Difference in Fit) [177] is a residual-based heuristic procedure used for the estimation of the number of components in a Tucker model. It computes the Tucker decomposition for all sensible combinations of components (i,j,k) and computes the model fit as $Fit(m) = 1 - \frac{\|X-\tilde{X}\|_F}{\|X\|_F}$ for each potential model where $\|.\|$ is the Frobenius norm and $m = i + j + k$. Then the DIF(m) for m-th model is computed as $Fit(m) - Fit(m-1)$ and accordingly, DIFFIT is computed as DIFFIT(m)=DIF(m)/DIF(m+1). The model with the largest DIFFIT value is chosen as the most adequate model. The DIFFIT model has been used for estimating tensor model dimension in EEG tensors [83,84]. DIFFIT requires computing the Tucker fit for all combinations of components which is very time-consuming. [178] proposed a faster version of DIFFIT (so called Fast-DIFFIT) that requires performing a single computation of Tucker decomposition. [178] provide some evidence that this approach can be sufficient as the exact solution. Fast-DIFFIT is tested for anomaly detection purposes in [95].

4.7.5. Multiway scree plot

Multi-way score plot [188] projects Tucker3 model onto the convex hull. The most adequate model is the one on the convex hull with less complexity and better fit. This method is used in [2,37,49,51,53,72] for tensor-based monitoring and anomaly detection.

4.7.6. Split-half analysis

This technique was primarily introduced by Harshman and De Sarbo [180] for PARAFAC. The procedure splits the tensor into two (or more) parts and the model with the same number of components is built for two parts. The assumption of this method is that if the model is valid, both models on two separate sides should remain stable. A criterion called split-half stability coefficients is defined and if its value is lower than a threshold (e.g. 0.1), the model is considered stable. However, the main requirement for this method is that tensor must be splittable [188] which is restrictive for non-stochastic systems. Limited works such as [3,49] use this technique to ascertain the number of components in tensors with application to anomaly detection. Extension of this method was later proposed by Kiers and Mechelen [189].

4.7.7. Other methods

Some other approaches proposed for tensor rank estimation, which may not be used for anomaly detection applications, can be very useful to the area nonetheless. Some of these methods include convex hull [181], Akaikes information criterion (AIC) [182], Bayesian information criterion (BIC) [183] and Automatic relevance determination (ARD) [159]. These four approaches are implemented for multiway models and compared in [159] in which the superiority of ARD is concluded against the other three ones. Bayesian-based tensor decompositions may also be a good solution for tensor rank estimation since they automatically find the tensor rank in their inference procedure [160,164]. In [184] a different approach named GAHNTD is proposed based on the Genetic algorithm for finding the optimal Tucker lower rank, but no comparison is performed against other known approaches. Brockmeier et al. [190] proposed a greedy approach that builds the tensor model iteratively, and uses the BIC criterion to identify the correct number of components. A more efficient method based on maximum block improvement (MBI) is proposed in [140] that uses non-convex block optimization for finding the Tucker3 model rank. It is evident that this method outperforms DIFFIT and ARD (when the sum of dimension is predefined) both in terms of accuracy and runtime. This method is used in [95] for event detection from traffic tensors.

5. Issues

This section outlines some of the most important issues in TAD and the corresponding solutions extracted from the related works. Table 6 presents a summary of the issues and the corresponding solutions.

5.1. Data pre-processing

Data pre-processing is an important step in TAD. Tensor models are sensitive to the scale of data elements. If multiple scale data is going to be used in tensor, it must be scaled accordingly, such that all columns have the same scale [175]. This is usually done via z-score scaling [18,32]. In some cases, the input data is a continuous signal and therefore must be converted to discrete values using tools such as wavelet transform [6,22,30,82].

5.2. Processing types: offline/online/streaming

Depending on processing tensor offline or real time, TAD methods are classified into three categories: offline, online and streaming. Offline processing model [48,50,54] is usually used in score plot based unsupervised detection (Section 3.3). Online processing usually refers to semi-supervised methods (Section 3.2). A tensor model is built from a normal operation condition of the system and is then used for matching with newly observed data to identify abnormal items. Since the most expensive task is performed offline, the detection part processes only a small piece of data. For the majority of cases, the normal model does not update during the detection process. However, in some works, it is suggested to constantly update the model upon receiving a new normal batch of data [29,48]. Streaming processing uses mostly unsupervised methods that do not learn from a train set, instead they operate directly on the data [15,16,133], therefore both learning and detection are performed simultaneously. They may use some mechanisms such as forgetting to exclude the outdated data from the learning process [16].

5.3. Tensor dimensionality

The dimensionality of tensors is usually chosen based on a prior knowledge. For instance, space and time are inherent modes of the tensor when system behavior is time-changing. Furthermore, data items are subject to change according to their spatial position. Dynamic networks are also in principle three-way tensors, such that the first and second dimensions denote the interactions between nodes and the third mode models the time-changing factor. Due to higher cost of tensors comparing matrix methods, the use of tensors is justified if there is at least three-way interactions in data. Multiway ANOVA test is one of the approaches used for discovering multi-way interactions in the data. For instance, Three-way ANOVA test [189] is used in [72] for ascertaining tensor dimensionality in anomaly detection application. The other approach might be correlation or comparison of model fits in different orders (e.g. 2D vs. 3D) [10,100].

5.4. Nonlinearity

Traditional tensor decompositions are unable to model complex nonlinear interactions between entities in each mode. Nonlinearity problem in TAD is reported in some works [43,94,126,132,191]. The solution to this problem is yet limited. Some propose to eliminate nonlinearity in a preprocessing step by segmentation of tensors to different linear parts [94], while others such as [43,132,191] propose the kernelized form of existing tensor decomposition methods. The probabilistic non-parametric methods such as [75,126,157] are also suggested for dealing with this issue. A kernel non-negative Tucker decomposition is proposed in [192].

5.5. Seasonality

Most of TAD's approaches are based on the assumption that the behavior of a system is persistent and uniform over time. However, in systems that deal with human activities such as the Internet, social networks, public health, etc. this assumption is valid only for a particular temporal period. For instance, we know that during the winter, rate of flu increases. Modeling epidemic data with tensors and not incorporating seasonality, we would probably signal many false alarms for winter season.

When we apply tensors to such data, two objectives are usually pursued. One is to discover periodic patterns of an unknown system. For instance, discovering what is the seasonal pattern of water quality or land surface changes [2,3,90]. The other and more practical target is to monitor the seasonal tensor for more accurate detection of anomalies and deviations. For the latter case, knowledge of periodic patterns is necessary for building a better tensor model. Although this issue is very important, it is not commonly addressed. For instance, the authors in [18] model the indoor air pollutants into four subsets of fall, winter, spring and summer and make a tensor model for each season. They compare this strategy to a global model in which all seasons are modeled together and show that the new strategy is more accurate. The other approach is proposed in [29] where separate tensor models are built for each environmental settings (e.g. Day = weekend, Weather = cold, Flu = high, Season = winter) and then use these tensor models in a real time setting for detection of disease outbreaks.

5.6. Unequal-length slices

A tensor with uneven slices is a known problem in process monitoring. It may exist in other disciplines but is rarely taken into account in other communities. This problem arises when process duration for each batch is different and thus *measurement* × *time* matrix for each batch has an unequal-length due to different length of time axis. Four scenarios lead to such a problem [193]: 1) the majority of measurement time series have equal length, but a minority of them despite overlapping in common time part, have shorter length; 2) all measurement time series have same length but some of them have small shift due to delay or acceleration in data collection; 3) the measurement time series have the same length, but appear in different shape; and 4) The time series of measurements have different lengths and shape.

Two different groups of approaches exist for the above problems. The first group of methods suggests performing a preprocessing step on the data before performing the decomposition, which is called trajectory synchronization/alignment. In this category, the first problem is solved by treating the absent part of shorter-length series as missing values. For the second problem, synchronization is carried out with a simple shift only on the minority series. For the third and fourth scenario which are more general cases, the measurements are expressed against the other variable (known as indicator variable) other than time so that the

Table 6
Important issues and solutions in tensor-based anomaly detection.

Problem	Solutions	Reference
Data pre-processing	Scaling	[18,32], [188, Chapter 6]
	Continues to discrete transformation	[6,22,30]
Processing	Offline	[48,50,54]
	Online without updating	[35,35,37,39,40,44,45,47,94,118,119,126,130–132,134,135]
	Online with updating	[29,48]
	Streaming	[15,16,133]
Tensor dimensionality	Prior knowledge	Majority of methods
	Multiway ANOVA	[72]
	Compare different ranks (e.g. 2D vs. 3D)	[10,100]
Tensor rank	See Table 5	
Nonlinearity	Eliminate nonlinearity in a preprocessing step	[94]
	Kernel tensor decomposition	[43,132,191]
Seasonality	Seasonal segmentation	[18]
	Separated tensors for each environment setting	[29]
Unequal-length slices	Treating the absent part of shorter-length series as missing values	[193]
	Dynamic time warping	[135]
	Phase division	[194]
Scalability	Sparse-optimized methods	[68,197,198]
	GPU-based	[79,199]
	Distributed and parallel approaches	[67,68,200–202,218]
Adaptivity	Incremental tensor analysis	[15,16,59,59]
	Online probabilistic	[158]
	Multi-aspect-streaming	[133]
Temporal scaling	Prior knowledge (single scale)	Majority of methods
	Multiple scale in data model	[49,50]
	Multis-scale approaches	[71]
Data fusion	Multiblock/multiway models	[193]
	Coupled matrix and tensor factorization (CMTF)	[213–216]
Noise removal	Preliminary phase removal	[188, Chapter 12]
	Two-step decomposition	[70]

shape of time series overlap for all measurements. Dynamic time warping [135] and phase division [194] techniques are also suggested for this purpose. However, these approaches are criticized in the sense that they distort the anomalous patterns and reduce the anomaly detection accuracy [193].

The second group, and more sophisticated methods are those that model uneven-length tensors in its natural form. One of the tensor models that can operate directly on uneven-length tensors is PARAFAC2. This model is proposed for fault detection [37] and its superiority has been shown over synchronization techniques. PARAFAC2 is able to directly model the original uneven-length tensor without performing further data unfolding or trajectory synchronization. However, it inherits the restrictive constraints of its equivalent model, PARAFAC. The authors in [150] propose GTucker2, a generalized version of PARAFAC2, that does not have these limitations and at the same time it can be used to model both even-length and uneven-length tensors. The authors show that GTucker2 has a better anomaly detection performance than PARAFAC2 for both even-length and uneven-length batch tensors.

5.7. Scalability

Scalability of tensor decomposition techniques is a hot and young research area in data mining, machine learning and signal processing community. The important problem is that the decomposition of big tensors is not computationally affordable by traditional techniques. Therefore, it is necessary to extend tensor methods for processing large data sets. Three major groups of solutions have been presented for this purpose, including sparse-optimized methods, GPU-based solutions and both parallel and distributed techniques.

The need for sparse-optimized methods arises from the fact that the majority of tensors in data mining applications is in principle sparse. For instance, density of Email, Web and network tensors barely exceeds 0.1%. Some works like [68,195–198] attempt to optimize the traditional tensor decomposition for large sparse tensors, in particular with operations on nonzero elements.

GPU-based techniques attempt to use new computing paradigms such as graphics processing unit(GPU) instead of CPU for speeding up the decomposition process. It is proved that GPU substantially outpaces CPU in dealing with computationally demanding and complex problems. Two examples from this category are G-PARAFAC [79] and GPUTENSOR [199].

Distributed and parallel approaches have received more attention by researchers due to the current progresses in parallel, distributed and cloud computing. The general objective of these methods is reducing the intermediate data explosion problem [68,200] and improving the runtime of tensor decomposition by splitting tensors into different sub-tensors and processing each smaller sub-tensors in a distributed, parallel or cloud environment (e.g. MapReduce). Examples of this category include GigaTensor [200], ParCube [67], PARACOMP [201] and HaTen2 [202].

5.8. Adaptivity

Standard tensor decompositions have been developed for operation in offline settings. It means that when new data is received they are unable to update the model and therefore they have to rebuild the model from scratch. Normally, due to the large volume of data in many applications, rebuilding the model is not feasible. Also keeping the whole data in memory is not possible. There exist some streaming approximation solutions for this problem for either classical tensor decomposition, subspace analysis or probabilistic tensor decomposition.

The most popular framework for incremental tensor analysis is ITA [17] consisting of three algorithms called Dynamic tensor analysis (DTA), streaming tensor analysis (STA) [16] and window-based tensor analysis (WTA) [15]. DTA decomposes the tensor incrementally by maintaining only the covariance matrix for each arriving tensor. Then, via diagonalization it outputs the principal eigenvectors of the updated covariance matrix as projection matrices. STA attempts to approximate DTA. Instead of maintaining a covariance matrix for all arriving tensors, it directly updates the principal eigenvectors using SPIRIT algorithm [203] which does not require diagonalization. The other algorithm WTA, instead of processing individual tensors uses a sliding window strategy for handling time dependency between consecutive tensors. It decomposes the sliding window with a regular Tucker or PARAFAC and then as well as DTA and STA keeps some statistics from the window in the processing of next windows.

ITA restricts the tensor growth only in time, which is a huge constraint in scalability and adaptability of other modes. In fact, ITA is only useful for large, but slender tensors. More recently a new TAD approach based on multi-aspect-streaming tensor analysis (MASTA) was proposed [133] that relaxes this constraint and allows tensor to concurrently evolve through all modes.

Incremental extensions of locality projection based methods (Section 4.6) have also been developed that are typically created for object tracking in video tensors (i.e. *spatialrow* × *spatialcolumn* × *frame*). The motivation of these methods is to model the appearance changes of objects in video data. A more recent approach from this category is DTAMU [59] that extends DTA for subspace learning. The objective of this work is to take into account the geometric structure of the image object, which is ignored in DTA. The similar ideas are used in [5,204].

Incremental version of probabilistic methods (Section 4.5)has also been presented in some works such as [158].

5.9. Multi-scale anomalies

In the discrete space, determination of the right scale (or sampling rate) for temporal dimension requires a prior knowledge about the scale of fluctuations. The sampling rate, depending on the application, can be per second [81], minute, [11,15,20,32,73,77], hour [17,18,18,64,72,91,158], k-hours [39,119], day [17,20,29], k-days [63], month [48,50,98,102,205] and year [17,69,95,133]. If there is no precise knowledge about scaling, multiple tensor model with different temporal scale may be built from data (e.g. see [96,144]). As is demonstrated in [144] the smaller scale (day) may provide a similar interpretation to a bigger scale (month), but with finer resolution. However, this might not be the case for all applications. If multiple scales have different influences on the data, a combination of more than one temporal scale may be used. For instance, in [49,50] a multi-scale scheme of *Sites* × *variables* × *year* × *month* is proposed for modeling of soil and water quality data. In this case, year and month, even though both refer to temporal dimensions, affect data in a different manner. Therefore, some meaningful patterns might be hidden if we lean to only-month or only-year scales. Recently, a multi-scale probabilistic tensor analysis framework called TriMine has been developed in [71] that accounts for several time granularities.

5.10. Data fusion

Coupled matrix and tensor factorization (CMTF) [206] are an emerging group of techniques that attempt to formulate a data fusion model based on joint factorization of matrices and higher-order tensor. In many applications, jointly analysis of an ensemble of data sets from multiple sources (also known as multi-block, multi-view, multi-set, multi-source data analysis) results in the enhancement of knowledge discovery.

The first use of data fusion based tensor and matrix decomposition in anomaly detection appeared in the work of Kourti [193] who proposed the use of multiblock/multiway PLS model for batch processes. The authors proposed that if we incorporate prior knowledge such as initial conditions for batches, raw material properties, initial ingredient charges or operation conditions in the original tensor model, the accuracy of anomaly detection will be improved.

Nowadays, the application of CMTF has been extended to wider areas such as location-based recommender systems [207,208], neuroscience [209–212], and sensory data analysis [213]. CMTF has also been used in applications related to anomaly detection such as social networks [214,215] and metabolomics [213,216]. For instance, in the metabolomics case, many heterogeneous data sets are generated via different analytical techniques for measuring biological fluids (e.g. blood). These complementary data sets if analyzed jointly may improve the understanding of the underlying biological processes corresponding to specific diseases.

A complete list of bibliography related to data fusion based on coupled matrix/tensor factorizations is gathered in [217].

5.11. Noise removal

Noise is a disturbing phenomenon in data that is disregarded by the analyst and it only negatively affects data analysis task [1]. Sometimes it can be difficult to distinguish anomalies from noises in tensor models due to their similar nature. Noise removal is usually undertaken as a preliminary phase in tensor-based modeling (See [188, Chapter 12]). However, in some works such as [48,70,110], a two-step decomposition is proposed for handling this issue. For instance, Maruhashi and Yugami [70] propose a two-step tensor decomposition framework. The first decomposition accounts for noise removal and the second decomposition that operates on the first step's output takes into account the meaningful anomalies.

6. Practical issues

In this section we introduce the fundamental tools for conducting research in TAD, mainly, software tools and evaluation metrics. The first section lists the available software and toolboxes for working with tensors and the second subsection presents a list of common evaluation metrics used in the various works.

6.1. Tensor software

Various open source toolboxes have been developed for tensor analysis in the recent decade. The most popular ones are MATLAB toolboxes like *Tensor toolbox* (http://www.sandia.gov/~tgkolda/TensorToolbox) and *N-way toolbox* (http://www.models.life.ku.dk/nwaytoolbox) which are widely used by many disciplines for tensor analysis. More recently, two toolboxes, *TensorBox* (http://www.bsp.brain.riken.jp/~phan) and *Tensorlab* (http://www.esat.kuleuven.be/sista/tensorlab) have also been developed. *TensorBox* is more focused on advanced fitting algorithms for Tucker and PARAFAC, while *Tensorlab* offers a wider range of algorithms for more complex tasks in tensor decomposition such as coupled tensor factorization, sparse and incomplete tensor decomposition, and new fitting algorithms such as quasi-Newton and nonlinear-least squares optimization, etc. *NFEA toolbox* (http://www.bsp.brain.riken.jp/~phan/nfea/nfea.html) ia s tensor toolbox specifically developed for processing EEG tensors. *CMTF toolbox* is also developed for coupled matrix and tensor factorization (http://www.models.life.ku.dk/~acare/CMTF_Toolbox). *Hierarchical Tucker toolbox* (http://anchp.epfl.ch/htucker) was developed for hierarchical Tucker decomposition. Apart from above MATLAB toolboxes, some R packages also exist for tensor decomposition, including *ThreeWay, rTensor* and *PTAk*.

6.2. Evaluation

Evaluation of TAD methods is usually similar to classical anomaly detection techniques [1]. The typical metric used includes precision/recall [70,77,91], accuracy [4,27,47,61,90,126,126] and area under ROC curve (AUC) [26,70,92]. For semi-supervised and unsupervised techniques, true and false positives (or false alarms) are also assessed [26,43,65,126]. For regression based tensor models and tensor forecasting methods, prediction error metrics such as root mean square error (RMSE) or mean absolute error (MAE) is normally used [50,59,71,90,92,124,136,154]. Detection delay has also been exploited in some works [40]. The more sophisticated metric that takes account the detection of both delay and false alarm rate is the Activity Monitoring Operating Characteristic (AMOC) curve that is used in [29]. Visual inspection of score plots with visual inspection is another evaluation method used in [11,18,20,61,64,67,76,84].

7. Conclusion

We provided the conceptual classification of many existing techniques, applications and issues for tensor-based anomaly detection. In the majority of works that we surveyed, the superiority of tensor-based methods has been shown over matrix methods. This exhibits the importance of tensors as new category in spectral-based anomaly detection. We classified the tensor-based learning into three categories of supervised, semi-supervised and unsupervised. Despite of the great ability of supervised methods, their application is not yet well-established for anomaly detection problem. We hope this survey could draw the attention of researchers to this new category of methods and their capabilities, especially tensor time series models [71,124,127,128]. Moreover, application of semi-supervised methods has been limited so far to monitoring of batch processes, although these kind of approaches can be very effective for other applications such as epidemiology and traffic data analysis.

We categorized tensor decomposition methods used in TAD into six main categories of Tucker-based, PARAFAC-based, Bayesian, LPP-based, DEDICOM-based and ICA-based and provided some examples for each branch. Among all, we found LPP-based and DEDICOM quite interesting which are unfairly less attended. DEDICOM, for instance, has a very good potential for the analysis of data in social networks and traffics (computer networks and transportation systems). LPP-based approaches are also very helpful for video and spatial data since they preserve the geometric structures in the data. Specially, the recently proposed method, TGLPP [168] seems a promising method for TAD, since it captures both local and global structure in data. Bayesian approaches are also emerging techniques with a huge contribution for anomaly detection. However, their appeal is limited due to their high computational costs. Fortunately, some new scalable methods have been proposed to deal this issue (e.g. [164,219]) and it is anticipated that we witness more works in this area in upcoming years.

We identified some important issues in TAD and suggested the possible solutions for each category, according to the state-of-the-art. We devoted a considerable portion of the survey to the problem of tensor rank estimation. Because, during the surveying of the literature we noticed that this issue has not received sufficient attention from the community. We could not find any work that studied the effect of tensor rank determination on the quality of anomaly detection or comparison of different automatic tensor rank estimation methods in accuracy of anomalies. Another problem about this issue is that the majority of methods are computationally expensive and hence infeasible for automatic purposes. Perhaps, the work of Chen et al. [140] is the most efficient method for this purpose which needs to be researched further for anomaly detection applications. However, still a need for a fast, accurate and adaptive method for tensor rank estimation is deeply felt. Probably new efforts in Bayesian tensor factorization research, for instance the recent work of Hu et al. [219] should receive more concern from researchers.

It seems that scalability, which is a quite important problem is receiving enough attention and is almost a hot topic in tensor literature. On the other side, it appears that less quantity of research is devoted to the adaptivity issue which is as important as the scalability. After the work of Sun et al. [16] we have not witnessed a serious contribution for this kind in the literature. Some recent works such as [29] propose the use of sketching techniques for coping this problem but this kind of approaches still require more research and development. Seasonality issue is also less noted in the TAD literature. In many phenomena we have the prior knowledge of seasonality that can be incorporated in TAD for more accurate anomaly detection. The recent work of Fanaee and Gama [29] might be a good starting point for further research. Data fusion based tensor approaches [217] are also predicted to be the hot topic in the near future due to the increasing number of heterogeneous data sources in modern digital systems.

Acknowledgment

This research was supported by European Commission through the project MAESTRA (Grant number ICT-750 2013–612944). The authors also acknowledge the support of the Projects NORTE-07-0124-FEDER-000059/000056 which are financed by the North Portugal Regional Operational Program (ON.2 O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF), and by national funds, through the Portuguese funding agency, Fundação para a Ciência e a Tecnologia (FCT).

References

[1] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: a survey, ACM Comput. Surv. 41 (3) (2009) 15.
[2] J.-D. Dong, Y.-Y. Zhang, S. Zhang, Y.-S. Wang, Z.-H. Yang, M.-L. Wu, Identification of temporal and spatial variations of water quality in Sanya Bay, China by three-way principal component analysis, Environ. Earth Sci. 60 (8) (2010) 1673–1682.
[3] F.D. Cid, R.I. Antón, R. Pardo, M. Vega, E. Caviedes-Vidal, Modelling spatial and temporal variations in the water quality of an artificial water reservoir in the semiarid midwest of argentina, Anal. Chim. Acta 705 (1) (2011) 243–252.
[4] Y. Panagakis, C. Kotropoulos, G.R. Arce, Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification, IEEE Trans. Audio Speech Lang. Process. 18 (3) (2010) 576–588.
[5] W. Hu, X. Li, X. Zhang, X. Shi, S. Maybank, Z. Zhang, Incremental tensor subspace learning and its applications to foreground segmentation and tracking, Int. J. Comput. Vis. 91 (3) (2011) 303–327.
[6] L.E. Mujica, J. Vehi, M. Ruiz, M. Verleysen, W. Staszewski, K. Worden, Multivariate statistics process control for dimensionality reduction in structural assessment, Mech. Syst. Signal Process. 22 (1) (2008) 155–171.
[7] J. Wang, Y. Chen, M. Adjouadi, A comparative study of multilinear principal component analysis for face recognition, in: Proceedings of the 37th IEEE Applied Imagery Pattern Recognition Workshop, 2008, AIPR'08, IEEE, 2008, pp. 1–6.
[8] R. Costantini, L. Sbaiz, S. Susstrunk, Higher order svd analysis for dynamic texture synthesis, IEEE Trans. Image Process. 17 (1) (2008) 42–52.

[9] E. Acar, S.A. Camtepe, M.S. Krishnamoorthy, B. Yener, Modeling and multi-way analysis of chatroom tensors, in: Intelligence and Security Informatics, Springer, 2005, pp. 256–268.
[10] A.H. Andersen, W.S. Rayens, Structure-seeking multilinear methods for the analysis of fmri data, NeuroImage 22 (2) (2004) 728–739.
[11] A. Baum, A.S. Meyer, J.L. Garcia, M. Egebo, P.W. Hansen, J.D. Mikkelsen, Enzyme activity measurement via spectral evolution profiling and parafac, Anal. Chim. Acta 778 (2013) 1–8.
[12] P. Nomikos, J.F. MacGregor, Monitoring batch processes using multiway principal component analysis, AIChE J. 40 (8) (1994) 1361–1375.
[13] M. Mørup, Applications of tensor (multiway array) factorizations and decompositions in data mining, Data Min. Knowl. Discov. 1 (1) (2011) 24–40.
[14] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, SIAM Rev. 51 (3) (2009) 455–500.
[15] J. Sun, S. Papadimitriou, S.Y. Philip, Window-based tensor analysis on high-dimensional and multi-aspect streams, in: Proceedings of International Conference on Data Mining (ICDM), vol. 6, 2006, pp. 1076–1080.
[16] J. Sun, D. Tao, C. Faloutsos, Beyond streams and graphs: dynamic tensor analysis, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2006, pp. 374–383.
[17] J. Sun, D. Tao, S. Papadimitriou, P.S. Yu, C. Faloutsos, Incremental tensor analysis: theory and applications, ACM Trans. Knowl. Discov. Data 2 (3) (2008) 11.
[18] S. Lee, H. Liu, M. Kim, J.T. Kim, C. Yoo, Online monitoring and interpretation of periodic diurnal and seasonal variations of indoor air pollutants in a subway station using parallel factor analysis (parafac), Energy Build. 68 (2014) 87–98.
[19] L. Tran, C. Navasca, J. Luo, Video detection anomaly via low-rank and sparse decompositions, in: Proceedings of Image Processing Workshop (WNYIPW), 2012, IEEE, Western New York, 2012, pp. 17–20.
[20] D. Koutra, E.E. Papalexakis, C. Faloutsos, Tensorsplat: spotting latent anomalies in time, in: Proceedings of the 16th Panhellenic Conference on Informatics (PCI), 2012, IEEE, 2012, pp. 144–149.
[21] A. Panisson, L. Gauvin, M. Quaggiotto, C. Cattuto, Mining concurrent topical activity in microblog streams, in: Proceedings of the Workshop on Making Sense of Microposts Co-located with the International World Wide Web Conference, 2014, pp. 3–10.
[22] E. Acar, C.A. Bingol, H. Bingol, R. Bro, B. Yener, Seizure recognition on epilepsy feature tensor, in: Proceedings of the 29th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, 2007. EMBS 2007, IEEE, 2007, pp. 4273–4276.
[23] N. Renard, S. Bourennane, Improvement of target detection methods by multiway filtering, IEEE Trans. Geosci. Remote Sens. 46 (8) (2008) 2407–2417.
[24] J. Wang, F. Gao, P. Cui, C. Li, Z. Xiong, Discovering urban spatio-temporal structure from time-evolving traffic networks, in: Web Technologies and Applications, Springer, 2014, pp. 93–104.
[25] W.-L. Chuang, C.-H. Chen, J.-Y. Yen, Y.-L. Hsu, Using MPCA of spectra model for fault detection in a hot strip mill, J. Mater. Process. Technol. 209 (8) (2009) 4162–4168.
[26] M.A. Prada, J. Toivola, J. Kullaa, J. Hollmén, Three-way analysis of structural health monitoring data, Neurocomputing 80 (2012) 119–128.
[27] A. Khosravi, J. Melendez, J. Colomer, J. Sanchez, Multiway principal component analysis (mpca) for upstream/downstream classification of voltage sags gathered in distribution substations, in: Advances of Computational Intelligence in Industrial Systems, Springer, 2008, pp. 297–312.
[28] J.C. Ho, J. Ghosh, S.R. Steinhubl, W.F. Stewart, J.C. Denny, B.A. Malin, J. Sun, Limestone: high-throughput candidate phenotype generation via tensor factorization, J. Biomed. Inform. 52 (2014) 199–211.
[29] H. Fanaee-T, J. Gama, Eigenevent: an algorithm for event detection from complex data streams in syndromic surveillance, Intell. Data Anal. 19 (3) (2015).
[30] Y. Bai, J. Tezcan, Q. Cheng, J. Cheng, A multiway model for predicting earthquake ground motion, in: Proceedings of the 2013 14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), IEEE, 2013, pp. 219–224.
[31] Y. Mu, W. Ding, M. Morabito, D. Tao, Empirical discriminative tensor analysis for crime forecasting, in: Knowledge Science, Engineering and Management, Springer, 2011, pp. 293–304.
[32] K. Kosanovich, M. Piovoso, K. Dahl, J. MacGregor, P. Nomikos, Multi-way PCA applied to an industrial batch process, in: Proceedings of American Control Conference, 1994, vol. 2, IEEE, 1994, pp. 1294–1298.
[33] P. Nomikos, J.F. MacGregor, Multivariate spc charts for monitoring batch processes, Technometrics 37 (1) (1995) 41–59.
[34] T. Kourti, P. Nomikos, J.F. MacGregor, Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS, J. Process Control 5 (4) (1995) 277–284.
[35] J. Chen, J.-H. Yen, Three-way data analysis with time lagged window for on-line batch process monitoring, Korean J. Chem. Eng. 20 (6) (2003) 1000–1011.
[36] J. GUO, H. CHEN, Y. LI, Mpca fault detection method based on multiblock statistics for uneven-length batch processes, J. Comput. Inf. Syst. 9 (18) (2013) 7181–7190.
[37] B.M. Wise, N.B. Gallagher, E.B. Martin, Application of parafac2 to fault detection and diagnosis in semiconductor etch, J. Chemom. 15 (4) (2001) 285–298.
[38] B.M. Wise, N.B. Gallagher, S.W. Butler, D.D. White, G.G. Barna, A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process, J. Chemom. 13 (3–4) (1999) 379–396.
[39] W. Zhifeng, Y. Jingqi, Online supervision of penicillin cultivations based on rolling mpca, Chin. J. Chem. Eng. 15 (1) (2007) 92–96.
[40] K. Hu, J. Yuan, Batch process monitoring with tensor factorization, J. Process Control 19 (2) (2009) 288–296.
[41] K.P. Singh, N. Basant, A. Malik, S. Sinha, G. Jain, Multi-way modeling of wastewater data for performance evaluation of sewage treatment plant-a case study, Chemom. Intell. Lab. Syst. 95 (1) (2009) 18–30.
[42] J.M. Amigo, A. Surribas, J. Coello, J.L. Montesinos, S. Maspoch, F. Valero, On-line parallel factor analysis. a step forward in the monitoring of bioprocesses in real time, Chemom. Intell. Lab. Syst. 92 (1) (2008) 44–52.
[43] J. Mori, J. Yu, Quality relevant nonlinear batch process performance monitoring using a kernel based multiway non-Gaussian latent subspace projection approach, J. Process Control 24 (1) (2014) 57–71.
[44] J.-M. Lee, C. Yoo, I.-B. Lee, On-line batch process monitoring using a consecutively updated multiway principal component analysis model, Comput. Chem. Eng. 27 (12) (2003) 1903–1912.
[45] C.K. Yoo, J.-M. Lee, P.A. Vanrolleghem, I.-B. Lee, On-line monitoring of batch processes using multiway independent component analysis, Chemom. Intell. Lab. Syst. 71 (2) (2004) 151–163.
[46] N.B. Gallagher, B.M. Wise, C.W. Stewart, Application of multi-way principal components analysis to nuclear waste storage tank monitoring, Comput. Chem. Eng. 20 (1996) S739–S744.
[47] A. Urtubia, G. Hernández, J. Roger, Detection of abnormal fermentations in wine process by multivariate statistics and pattern recognition techniques, J. Biotechnol. 159 (4) (2012) 336–341.
[48] P. Barbieri, G. Adami, S. Piselli, F. Gemiti, E. Reisenhofer, A three-way principal factor analysis for assessing the time variability of freshwaters related to a municipal water supply, Chemom. Intell. Lab. Syst. 62 (1) (2002) 89–100.
[49] K.P. Singh, A. Malik, V.K. Singh, N. Basant, S. Sinha, Multi-way modeling of hydro-chemical data of an alluvial river system-a case study, Anal. Chim. Acta 571 (2) (2006) 248–259.
[50] K.P. Singh, A. Malik, N. Basant, P. Saxena, Multi-way partial least squares modeling of water quality data, Anal. Chim. Acta 584 (2) (2007) 385–396.
[51] M.A. Engle, M. Gallo, K.T. Schroeder, N.J. Geboy, J.W. Zupancic, Three-way compositional analysis of water quality monitoring data, Environ. Ecol. Stat. 21 (3) (2014) 565–581.
[52] K.P. Singh, A. Malik, S. Sinha, D. Mohan, V.K. Singh, Exploring groundwater hydrochemistry of alluvial aquifers using multi-way modeling, Anal. Chim. Acta 596 (1) (2007) 171–182.
[53] I. Stanimirova, V. Simeonov, Modeling of environmental four-way data from air quality control, Chemom. Intell. Lab. Syst. 77 (1) (2005) 115–121.
[54] K.P. Singh, A. Malik, V.K. Singh, S. Sinha, Multi-way data analysis of soils irrigated with wastewater–a case study, Chemom. Intell. Lab. Syst. 83 (1) (2006) 1–12.
[55] J. Andrade, M. Kubista, A. Carlosena, D. Prada, 3-way characterization of soils by procrustes rotation, matrix-augmented principal components analysis and parallel factor analysis, Anal. Chim. Acta 603 (1) (2007) 20–29.
[56] J. Li, G. Han, J. Wen, X. Gao, Robust tensor subspace learning for anomaly detection, Int. J. Mach. Learn. Cybern. 2 (2) (2011) 89–98.
[57] Q. Zhao, G. Zhou, T. Adali, L. Zhang, A. Cichocki, Kernelization of tensor-based models for multiway data analysis: processing of multidimensional structured data, IEEE Signal Process. Mag. 30 (4) (2013) 137–148.
[58] H. Li, Y. Wei, L. Li, Y.Y. Tang, Infrared moving target detection and tracking based on tensor locality preserving projection, Infrared Phys. Technol. 53 (2) (2010) 77–83.
[59] X. Zhang, X. Shi, W. Hu, X. Li, S. Maybank, Visual tracking via dynamic tensor analysis with mean update, Neurocomputing 74 (17) (2011) 3277–3285.
[60] P. Li, Q. Sun, Tensor-based covariance matrices for object tracking, in: Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), IEEE, 2011, pp. 1681–1688.
[61] B. Zhou, F. Zhang, L. Peng, Higher-order SVD analysis for crowd density estimation, Comput. Vis. Image Underst. 116 (9) (2012) 1014–1021.
[62] T. Kobayashi, N. Otsu, Three-way auto-correlation approach to motion recognition, Pattern Recognit. Lett. 30 (3) (2009) 212–221.
[63] M. Araujo, S. Papadimitriou, S. Günnemann, C. Faloutsos, P. Basu, A. Swami, E.E. Papalexakis, D. Koutra, Com2: fast automatic discovery of temporal (comet) communities, in: Advances in Knowledge Discovery and Data Mining, Springer, 2014, pp. 271–283.
[64] H.-H. Mao, C.-J. Wu, E.E. Papalexakis, C. Faloutsos, K.-C. Lee, T.-C. Kao, Malspot: multi2 malicious network behavior patterns analysis, in: Advances in Knowledge Discovery and Data Mining, Springer, 2014, pp. 1–14.
[65] H. Kim, S. Lee, X. Ma, C. Wang, Higher-order PCA for anomaly detection in large-scale networks, in: Proceedings of the 2009 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), IEEE, 2009, pp. 85–88.
[66] K. Maruhashi, F. Guo, C. Faloutsos, Multiaspectforensics: pattern mining on large-scale heterogeneous networks with tensor analysis, in: Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2011, pp. 203–210.
[67] E.E. Papalexakis, C. Faloutsos, N.D. Sidiropoulos, Parcube: sparse parallelizable tensor decompositions, in: Machine Learning and Knowledge Discovery in Databases, Springer, 2012, pp. 521–536.
[68] T.G. Kolda, J. Sun, Scalable tensor decompositions for multi-aspect data mining, in: Proceedings of the Eighth IEEE International Conference onData Mining, ICDM'08, IEEE, 2008, pp. 363–372.

[69] B.W. Bader, R.A. Harshman, T.G. Kolda, Temporal analysis of semantic graphs using asalsan, in: Proceedings of the Seventh IEEE International Conference on Data Mining, ICDM 2007, IEEE, 2007, pp. 33–42.
[70] K. Maruhashi, N. Yugami, Multiaspectspotting: spotting anomalous behavior within count data using tensor, in: Advances in Knowledge Discovery and Data Mining, Springer, 2014, pp. 474–485.
[71] Y. Matsubara, Y. Sakurai, C. Faloutsos, T. Iwata, M. Yoshikawa, Fast mining and forecasting of complex time-stamped events, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2012, pp. 271–279.
[72] H. Fanaee-T, M. Oliveira, J. Gama, S. Malinowski, R. Morla, Event and anomaly detection using tucker3 decomposition, in: Proceedings of European Conference on Artificial Intelligence – Ubiquitous Data Mining Workshop (UDM 2012), 2012, pp. 8–12.
[73] K. Glass, R. Colbaugh, M. Planck, Automatically identifying the sources of large internet events, in: Proceedings of the 2010 IEEE International Conference on Intelligence and Security Informatics (ISI), IEEE, 2010, pp. 108–113.
[74] W. Peng, T. Li, Temporal relation co-clustering on directional social network and author-topic evolution, Knowl. Inf. Syst. 26 (3) (2011) 467–486.
[75] Z. Xu, F. Yan, Y. Qi, Bayesian nonparametric models for multiway data analysis, IEEE Trans. Pattern Anal. Mach. Intell. 37 (2) (2015) 475–487, doi:10.1109/TPAMI.2013.201.
[76] E. Papalexakis, K. Pelechrinis, C. Faloutsos, Spotting misbehaviors in location-based social networks using tensors, in: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, 2014, pp. 551–552.
[77] L. Gauvin, A. Panisson, C. Cattuto, Detecting the community structure and activity patterns of temporal networks: a non-negative tensor factorization approach, PLoS One 9 (1) (2014) e86028.
[78] M. Oliveira, J. Gama, Visualization of evolving social networks using actor-level and community-level trajectories, Expert Syst. 30 (4) (2013) 306–319.
[79] D. Chen, X. Li, L. Wang, S. Khan, J. Wang, K. Zeng, C. Cai, Fast and scalable multi-way analysis of massive neural data, IEEE Trans. Comput. 64 (3) (2015) 707–719, doi:10.1109/TC.2013.2295806.
[80] R. Rosipal, L.J. Trejo, P.L. Nunez, Application of multi-way EEG decomposition for cognitive workload monitoring, in: Proceedings of the 6th International Conference on Partial Least Squares and Related Methods, Beijing, China, 2009, pp. 145–149.
[81] F. Miwakeichi, P.A. Valdes-Sosa, E. Aubert-Vazquez, J.B. Bayard, J. Watanabe, H. Mizuhara, Y. Yamaguchi, Decomposing EEG data into space-time-frequency components using parallel factor analysis and its relation with cerebral blood flow, in: Neural Information Processing, Springer, 2008, pp. 802–810.
[82] M. Mørup, L.K. Hansen, C.S. Herrmann, J. Parnas, S.M. Arnfred, Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG, NeuroImage 29 (3) (2006) 938–947.
[83] F. Cong, A.-H. Phan, P. Astikainen, Q. Zhao, Q. Wu, J.K. Hietanen, T. Ristaniemi, A. Cichocki, Multi-domain feature extraction for small event-related potentials through nonnegative multi-way array decomposition from low dense array EEG, Int. J. Neural Syst. 23 (02) (2013).
[84] F. Cong, A.H. Phan, Q. Zhao, T. Huttunen-Scott, J. Kaartinen, T. Ristaniemi, H. Lyytinen, A. Cichocki, Benefits of multi-domain feature of mismatch negativity extracted by non-negative tensor factorization from eeg collected by low-density array, Int. J. Neural Syst. 22 (06) (2012).
[85] C.F. Beckmann, S.M. Smith, Tensorial extensions of independent component analysis for multisubject FMRI analysis, Neuroimage 25 (1) (2005) 294–311.
[86] S. Bourennane, C. Fossati, A. Cailly, Improvement of classification for hyperspectral images based on tensor modeling, IEEE Geosci. Remote Sens. Lett. 7 (4) (2010) 801–805.
[87] N. Renard, S. Bourennane, Dimensionality reduction based on tensor modeling for classification methods, IEEE Trans. Geosci. Remote Sens. 47 (4) (2009) 1123–1131.
[88] Q. Zhang, H. Wang, R.J. Plemmons, V. Pauca, Tensor methods for hyperspectral data analysis: a space object material identification study, J. Opt. Soc. Am. A 25 (12) (2008) 3001–3012.
[89] L. Zhang, L. Zhang, D. Tao, X. Huang, A multifeature tensor for remote-sensing target recognition, IEEE Geosci. Remote Sens. Lett. 8 (2) (2011) 374–378.
[90] S. Hemissi, I.R. Farah, K. Saheb Ettabaa, B. Solaiman, Multi-spectro-temporal analysis of hyperspectral imagery based on 3-d spectral modeling and multilinear algebra, IEEE Trans. Geosci. Remote Sens. 51 (1) (2013) 199–216.
[91] L. Shi, A. Gangopadhyay, V.P. Janeja, Stensr: spatio-temporal tensor streams for anomaly detection and pattern discovery, Knowl. Inf. Syst. (2014) 1–21.
[92] K. Hayashi, T. Takenouchi, T. Shibata, Y. Kamiya, D. Kato, K. Kunieda, K. Yamada, K. Ikeda, Exponential family tensor factorization for missing-values prediction and anomaly detection, in: Proceedings of the 2010 IEEE 10th International Conference on Data Mining (ICDM), IEEE, 2010, pp. 216–225.
[93] M.A. Prada, M. Dominguez, P. Barrientos, S. Garcia, Dimensionality reduction for damage detection in engineering structures, Int. J. Mod. Phys. B 26 (25) (2012).
[94] J.D. Karssen, M. Wisse, Fall detection in walking robots by multi-way principal component analysis, Robotica 27 (2) (2009) 249–257.
[95] H. Fanaee-T, J. Gama, Event detectionfrom traffic tensors: a hybrid model, Neurocomputing (2015). https://fe.up.pt/~dma10087/Neurocomputing2016.pdf
[96] H. Tan, J. Feng, G. Feng, W. Wang, Y.-J. Zhang, Traffic volume data outlier recovery via tensor model, Math. Probl. Eng. 2013 (2013).
[97] H. Tan, G. Feng, J. Feng, W. Wang, Y.-J. Zhang, F. Li, A tensor-based method for missing traffic data completion, Transp. Res. Part C: Emerg. Technol. 28 (2013) 15–27.
[98] A. Hall, G. Qu, I.K. Sethi, C. Hartrick, Tensor-based temporal behavior analysis in pain medicine, in: Proceedings of the 2012 11th International Conference on Machine Learning and Applications (ICMLA), vol. 1, IEEE, 2012, pp. 626–629.
[99] Y. Li, A. Ngom, Non-negative matrix and tensor factorization based classification of clinical microarray gene expression data, in: Proceedings of the 2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2010, pp. 438–443.
[100] H. Fanaee-T, J. Gama, An eigenvector-based hotspot detection, in: Proceedings of the 16th Portuguese Conference on Artificial Intelligence (EPIA 2013), 2013, pp. 251–260.
[101] A. Ramanathan, P.K. Agarwal, C.J. Langmead, Using Tensor Analysis to Characterize Contact-map Dynamics of Proteins, Department of Computer Science, School of Computer Science, Carnegie Mellon University, 2008.Technical Report 1-2008
[102] D.G. Leibovici, Spatio-temporal multiway decomposition using principal tensor analysis on k-modes: the r package ptak, J. Stat. Softw. 34 (10) (2010) 1–34.
[103] D. Leibovici, G. Quillevere, J.-C. Desconnets, A method to classify ecoclimatic arid and semiarid zones in circum-Saharan Africa using monthly dynamics of multiple indicators, IEEE Trans. Geosci. Remote Sens. 45 (12) (2007) 4000–4007.
[104] S. Unkel, A. Hannachi, N.T. Trendafilov, I.T. Jolliffe, Independent component analysis for three-way data with an application from atmospheric science, J. Agric. Biol. Environ. Stat. 16 (3) (2011) 319–338.
[105] A.R. Marklund, J. Dufek, Development and comparison of spectral methods for passive acoustic anomaly detection in nuclear power plants, Appl. Acoust. 83 (2014) 100–107.
[106] N. Mesgarani, S.V. David, J.B. Fritz, S.A. Shamma, Mechanisms of noise robust representation of speech in primary auditory cortex, Proc. Natl. Acad. Sci. USA 111 (18) (2014) 6792–6797.
[107] N. Mesgarani, M. Slaney, S.A. Shamma, Discrimination of speech from non-speech based on multiscale spectro-temporal modulations, IEEE Trans. Audio Speech Lang. Process. 14 (3) (2006) 920–930.
[108] P. Nomikos, J.F. MacGregor, Multi-way partial least squares in monitoring batch processes, Chemom. Intell. Lab. Syst. 30 (1) (1995) 97–108.
[109] K. Nazarpour, S. Sanei, L. Shoker, J.A. Chambers, Parallel space-time-frequency decomposition of EEG signals for brain computer interfacing, in: Proceedings of EUSIPCO06, 2006.
[110] K. Villez, M. Ruiz, G. Sin, J. Colomer, C. Rosen, P.A. Vanrolleghem, Combining multiway principal component analysis (MPCA) and clustering for efficient data mining of historical data sets of sbr processes., Water Sci. Technol. 57 (10) (2008).
[111] D. Tao, X. Li, W. Hu, S. Maybank, X. Wu, Supervised tensor learning, in: Proceedings of the Fifth IEEE International Conference on Data Mining, IEEE, 2005, pp. 8–pp.
[112] D. Cai, X. He, J. Han, Learning with Tensor Representation, Technical Report, Department of Computer Science, University of Illinois, 2006.UIUCDCS-R-2006-2716
[113] I. Kotsia, W. Guo, I. Patras, Higher rank support tensor machines for visual recognition, Pattern Recognit. 45 (12) (2012) 4192–4203.
[114] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, H.-J. Zhang, Multilinear discriminant analysis for face recognition, IEEE Trans. Image Process. 16 (1) (2007) 212–220.
[115] S. Rendle, Factorization machines, in: Proceedings of the 2010 IEEE 10th International Conference on Data Mining (ICDM), IEEE, 2010, pp. 995–1000.
[116] H. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, A survey of multilinear subspace learning for tensor data, Pattern Recognit. 44 (7) (2011) 1540–1551.
[117] S. Wold, P. Geladi, K. Esbensen, J. Öhman, Multi-way principal components-and PLS-analysis, J. Chemom. 1 (1) (1987) 41–56.
[118] J. Chen, K.-C. Liu, On-line batch process monitoring using dynamic PCA and dynamic PLS models, Chem. Eng. Sci. 57 (1) (2002) 63–75.
[119] O. Marjanovic, B. Lennox, D. Sandoz, K. Smith, M. Crofts, Real-time monitoring of an industrial batch process, Comput. Chem. Eng. 30 (10) (2006) 1476–1481.
[120] Y. Li, Z. Wang, J. Yuan, On-line fault detection using svm-based dynamic MPLS for batch processes, Chin. J. Chem. Eng. 14 (6) (2006) 754–758.
[121] W. Guo, I. Kotsia, I. Patras, Tensor learning for regression, IEEE Trans. Image Process 21 (2) (2012) 816–827.
[122] H. Zhou, L. Li, H. Zhu, Tensor regression with applications in neuroimaging data analysis, J. Am. Stat. Assoc. 108 (502) (2013) 540–552.
[123] Y. Zhu, J. He, R.D. Lawrence, A general framework for predictive tensor modeling with domain knowledge, Data Min. Knowl. Discov. (2014) 1–24.
[124] M. Rogers, L. Li, S. Russell, Multilinear dynamical systems for tensor time series, in: Advances in Neural Information Processing Systems, 2013, pp. 2634–2642.
[125] M.T. Bahadori, Q.R. Yu, Y. Liu, Fast multivariate spatio-temporal analysis via low rank tensor learning, in: Advances in Neural Information Processing Systems, 2014, pp. 3491–3499.
[126] J. Yu, Multiway discrete hidden markov model-based approach for dynamic batch process monitoring and fault classification, AIChE J. 58 (9) (2012) 2714–2725.

[127] N. Thai-Nghe, T. Horváth, L. Schmidt-Thieme, Factorization models for forecasting student performance, in: Proceedings of International Conference on Educational Data Mining (EDM), 2011, pp. 11–20.
[128] N. Thai-Nghe, Predicting Student Performance in an Intelligent Tutoring System, Department of Computer Science, University of Hildesheim, Hildesheim, Germany, 2011 (Ph.D. thesis).
[129] S. Kouchaki, S. Sanei, Tensor based singular spectrum analysis for nonstationary source separation, in: Proceedings of the 2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, 2013, pp. 1–5.
[130] D.S. Lee, P.A. Vanrolleghem, Monitoring of a sequencing batch reactor using adaptive multiblock principal component analysis, Biotechnol. Bioeng. 82 (4) (2003) 489–497.
[131] C.K. Yoo, D.S. Lee, P.A. Vanrolleghem, Application of multiway ICA for on-line process monitoring of a sequencing batch reactor, Water Res. 38 (7) (2004) 1715–1732.
[132] X. Tian, X. Zhang, X. Deng, S. Chen, Multiway kernel independent component analysis based on feature samples for batch process monitoring, Neurocomputing 72 (7) (2009) 1584–1596.
[133] H. Fanaee-T, J. Gama, Multi-aspect-streaming tensor analysis, Knowl.-Based Syst. 89 (2015) 332–345.
[134] N.A.A. Majid, M.P. Taylor, J.J. Chen, M.A. Stam, A. Mulder, B.R. Young, Aluminium process fault detection by multiway principal component analysis, Control Eng. Pract. 19 (4) (2011) 367–379.
[135] X. Gao, On-line Monitoring of Batch Process with Multiway PCA/ICA, in: Dr. P Sanguansat (Ed.), Principal Component Analysis, 2012, pp. 239–262, doi:10.5772/38386. Chapter 13.
[136] Z.I. Stefanov, K.A. Hoo, Hierarchical multivariate analysis of cockle phenomena, J. Chemom. 17 (10) (2003) 550–568.
[137] L.R. Tucker, Some mathematical notes on three-mode factor analysis, Psychometrika 31 (3) (1966) 279–311.
[138] R.A. Harshman, Foundations of the PARAFAC Procedure: Models and Conditions for an "Explanatory" Multi-Modal Factor Analysis, UCLA Working Papers in Phonetics, vol. 16(1), 1970, p. 84.
[139] J.D. Carroll, J.-J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition, Psychometrika 35 (3) (1970) 283–319.
[140] B. Chen, Z. Li, S. Zhang, On optimal low rank tucker approximation for tensors: the case for an adjustable core size, J. Global Optim. (2014) 1–22, doi:10.1007/s10898-014-0231-x.
[141] L. De Lathauwer, B. De Moor, J. Vandewalle, A multilinear singular value decomposition, SIAM J. Matrix Anal Appl. 21 (4) (2000) 1253–1278.
[142] T.G. Kolda, B.W. Baderz, Tensor decompositions and application, SIAM Rev. 51 (2009) 455–500.
[143] J.D. Carroll, G. De Soete, S. Pruzansky, Fitting of the latent class model via iteratively reweighted least squares candecomp with nonnegativity constraints, in: Multiway Data Analysis, North-Holland Publishing Co., 1989, pp. 463–472.
[144] B.W. Bader, M.W. Berry, M. Browne, Discussion tracking in enron email using parafac, in: Survey of Text Mining II, Springer, 2008, pp. 147–163.
[145] H.A. Kiers, An alternating least squares algorithm for parafac2 and three-way dedicom, Comput. Stat. Data Anal. 16 (1) (1993) 103–118.
[146] E.C. Chi, T.G. Kolda, On tensors, sparsity, and nonnegative factorizations, SIAM J. Matrix Anal. Appl. 33 (4) (2012) 1272–1299.
[147] H. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, Multilinear principal component analysis of tensor objects for recognition, in: Proceedings of the 18th International Conference on Pattern Recognition, ICPR 2006, vol. 2, IEEE, 2006, pp. 776–779.
[148] N. Lv, G.Y. Bai, L.L. Jiang, Fault detection for batch processes based on segmentation mpca, in: Advanced Materials Research, vol. 1030, Trans Tech Publ, 2014, pp. 1701–1708.
[149] J. Peng, H. Liu, Y. Hu, J. Xi, H. Chen, Ascs online fault detection and isolation based on an improved mpca, Chin. J. Mech. Eng. 27 (5) (2014) 1047–1056.
[150] L. Luo, S. Bao, Z. Gao, J. Yuan, Batch process monitoring with gtucker2 model, Ind. Eng. Chem. Res. 53 (39) (2014) 15101–15110.
[151] Y.-D. Kim, S. Choi, Nonnegative tucker decomposition, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR'07, IEEE, 2007, pp. 1–8.
[152] M. Mørup, L.K. Hansen, S.M. Arnfred, Algorithms for sparse nonnegative tucker decompositions, Neural Comput. 20 (8) (2008) 2112–2131.
[153] S. Süsstrunk, L. Sbaiz, R. Costantini, et al., Dynamic texture analysis and synthesis using tensor decomposition, in: Proceedings of International Symposium on Visual Computing, vol. 4292, Springer, 2006, pp. 1161–1170.Lecture Notes in Computer Science
[154] E. Selli, C. Zaccaria, F. Sena, G. Tomasi, G. Bidoglio, Application of multi-way models to the time-resolved fluorescence of polycyclic aromatic hydrocarbons mixtures in water, Water Res. 38 (9) (2004) 2269–2276.
[155] R.A. Harshman, Models for analysis of asymmetrical relationships among n objects or stimuli, in: Proceedings of First Joint Meeting of the Psychometric Society and the Society for Mathematical Psychology, McMaster University, Hamilton, Ontario, 1978.
[156] B.W. Bader, R.A. Harshman, T.G. Kolda, Temporal Analysis of Social Networks Using Three-way Dedicom, vol. 119, Sandia National Laboratories, 2006.TR SAND2006-2161.
[157] W. Chu, Z. Ghahramani, Probabilistic models for incomplete multi-dimensional arrays, in: Proceedings of JMLR Workshop and Conference Proceedings, AISTATS 2009, vol. 5, Microtome Publishing (paper), 2009, pp. 89–96.J. Mach. Learn. Res.
[158] K. Hayashi, T. Takenouchi, T. Shibata, Y. Kamiya, D. Kato, K. Kunieda, K. Yamada, K. Ikeda, Exponential family tensor factorization: an online extension and applications, Knowl. Inf. Syst. 33 (1) (2012) 57–88.
[159] M. Mørup, L.K. Hansen, Automatic relevance determination for multi-way models, J. Chemom. 23 (7–8) (2009) 352–363.
[160] Q. Zhao, L. Zhang, A. Cichocki, Bayesian cp factorization of incomplete tensors with automatic rank determination, IEEE Trans Pattern Anal. Mach. Intell. PP (99) (2015).1–1 doi: 10.1109/TPAMI.2015.2392756.
[161] I. Porteous, E. Bart, M. Welling, Multi-hdp: a nonparametric Bayesian model for tensor factorization, in: Proceedings of the 23rd National Conference on Artificial intelligence vol. 3, AAAI Press, 2008, pp. 1487–1490.
[162] D. Tao, M. Song, X. Li, J. Shen, J. Sun, X. Wu, C. Faloutsos, S.J. Maybank, Bayesian tensor approach for 3-d face modeling, IEEE Trans. Circuits Syst. Video Technol. 18 (10) (2008) 1397–1410.
[163] L. Xiong, X. Chen, T.-K. Huang, J.G. Schneider, J.G. Carbonell, Temporal collaborative filtering with Bayesian probabilistic tensor factorization., in: Proceedings of SIAM International Conference on Data Mining, SDM, vol. 10, SIAM, 2010, pp. 211–222.
[164] P. Rai, Y. Wang, S. Guo, G. Chen, D. Dunson, L. Carin, Scalable Bayesian low-rank decomposition of incomplete multiway tensors, in: Proceedings of the 31st International Conference on Machine Learning (ICML-14), 2014, pp. 1800–1808.
[165] J. Zhou, A. Bhattacharya, A.H. Herring, D.B. Dunson, Bayesian factorizations of big sparse tensors, J. Am. Stat. Assoc. 110 (512) (2015) 1562–1576.
[166] X. He, D. Cai, P. Niyogi, Tensor subspace analysis, in: Advances in Neural Information Processing Systems, 2005, pp. 499–506.
[167] G. Dai, D.-Y. Yeung, Tensor embedding methods, in: Proceedings of the National Conference on Artificial Intelligence, vol. 21, AAAI Press, Menlo Park, CA, 2006, p. 330.Cambridge, MA; London/MIT Press.
[168] L. Luo, S. Bao, Z. Gao, J. Yuan, Tensor global-local preserving projections for batch process monitoring, Ind. Eng. Chem. Res. 53 (24) (2014) 10166–10176, doi:10.1021/ie403973w.
[169] Y. Zhang, S.J. Qin, Fault detection of nonlinear processes using multiway kernel independent component analysis, Ind. Eng. Chem. Res. 46 (23) (2007) 7780–7787.
[170] K. Hu, J. Yuan, Multivariate statistical process control based on multiway locality preserving projections, J. Process Control 18 (7) (2008) 797–807.
[171] J.A. Westerhuis, T. Kourti, J.F. MacGregor, Comparing alternative approaches for multivariate statistical analysis of batch process data, J. Chemom. 13 (3-4) (1999) 397–413.
[172] J.F. MacGregor, M.H. Bharati, H. Yu, Multivariate image analysis for process monitoring and control, in: Intelligent Systems and Smart Manufacturing, International Society for Optics and Photonics, 2001, pp. 17–26.
[173] D.A. Jackson, Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches, Ecology (1993) 2204–2214.
[174] B. Efron, Estimating the error rate of a prediction rule: improvement on cross-validation, J. Am. Stat. Assoc. 78 (382) (1983) 316–331.
[175] D. Louwerse, A.K. Smilde, H.A. Kiers, Cross-validation of multiway component models, J. Chemom. 13 (5) (1999) 491–510.
[176] R. Bro, H.A. Kiers, A new efficient method for determining the number of components in parafac models, J Chemom. 17 (5) (2003) 274–286.
[177] M.E. Timmerman, H.A. Kiers, Three-mode principal components analysis: choosing the numbers of components and sensitivity to local optima, Br. J. Math. Stat. Psychol. 53 (1) (2000) 1–16.
[178] H.A. Kiers, A. Kinderen, A fast method for choosing the numbers of components in tucker3 analysis, Br. J. Math. Stat. Psychol. 56 (1) (2003) 119–125.
[179] C.A. Andersson, R. Bro, The n-way toolbox for matlab, Chemom. Intell. Lab. Syst. 52 (1) (2000) 1–4.
[180] R.A. Harshman, W.S. De Sarbo, An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques, in: H.G. Law, C.W. Snyder, Jr. J. Hattie, R.P. McDonald (Eds.), Research Methods for Multimode Data Analysis, Praeger, New York, 1984, pp. 602–642.
[181] E. Ceulemans, H.A. Kiers, Selecting among three-mode principal component models of different types and complexities: a numerical convex hull based method, Br. J. Math. Stat. Psychol. 59 (1) (2006) 133–150.
[182] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control 19 (6) (1974) 716–723.
[183] G. Schwarz, et al., Estimating the dimension of a model, Ann. Stat. 6 (2) (1978) 461–464.
[184] A. Karami, M. Yazdi, A.Z. Asli, Best rank-r tensor selection using genetic algorithm for better noise reduction and compression of hyperspectral images, in: Proceedings of the 2010 Fifth International Conference on Digital Information Management (ICDIM), IEEE, 2010, pp. 169–173.
[185] J. Håstad, Tensor rank is np-complete, J. Algorithms 11 (4) (1990) 644–654.
[186] J. Riu, R. Bro, Jack-knife technique for outlier detection and estimation of standard errors in parafac models, Chemom. Intell. Lab. Syst. 65 (1) (2003) 35–49.
[187] P.M. Kroonenberg, et al., Three-mode Component and Scaling Models. Wiley StatsRef: Statistics Reference Online. (2015), 1–17. URL: http://dx.doi.org/10.1002/9781118445112.stat06459.pub2.
[188] P.M. Kroonenberg, Applied Multiway Data Analysis, John Wiley & Sons, 2008.
[189] H.A. Kiers, I.V. Mechelen, Three-way component analysis: principles and illustrative application, Psychol. Methods 6 (1) (2001) 84.

[190] A.J. Brockmeier, J.C. Principe, A.H. Phan, A. Cichocki, A greedy algorithm for model selection of tensor decompositions., in: Proceedings of the International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2013, pp. 6113–6117.
[191] M.M. Rashid, J. Yu, Nonlinear and non-Gaussian dynamic batch process monitoring using a new multiway kernel independent component analysis and multidimensional mutual information based dissimilarity approach, Ind. Eng. Chem. Res. 51 (33) (2012) 10910–10920.
[192] A. Karami, M. Yazdi, A. Zolghadre Asli, Noise reduction of hyperspectral images using kernel non-negative tucker decomposition, IEEE J. Sel. Top. Signal Process. 5 (3) (2011) 487–493.
[193] T. Kourti, Abnormal situation detection, three-way data and projection methods; robust data archiving and modeling for industrial applications, Annu. Rev. Control 27 (2) (2003) 131–139.
[194] N.-y. Lu, F.-r. Gao, Y. Yang, F. Wang, PCA-based modeling and on-line monitoring strategy for uneven-length batch processes, Ind. Eng. Chem. Res. 43 (13) (2004) 3343–3352.
[195] T.G. Kolda, B.W. Bader, J.P. Kenny, Higher-order web link analysis using multilinear algebra, in: Proceedings of the Fifth IEEE International Conference on Data Mining, IEEE, 2005, p. 8.
[196] B.W. Bader, T.G. Kolda, Efficient matlab computations with sparse and factored tensors, SIAM J. Sci. Comput. 30 (1) (2007) 205–231.
[197] G. Allen, Sparse higher-order principal components analysis, in: Proceedings of the International Conference on Artificial Intelligence and Statistics, 2012, pp. 27–36.
[198] M. Baskaran, B. Meister, N. Vasilache, R. Lethin, Efficient and scalable computations with sparse tensors, in: Proceedings of the 2012 IEEE Conference on High Performance Extreme Computing (HPEC), IEEE, 2012, pp. 1–6.
[199] B. Zou, C. Li, L. Tan, H. Chen, Gputensor: efficient tensor factorization for context-aware recommendations, Inf. Sci. 299 (2015) 159–177.
[200] U. Kang, E. Papalexakis, A. Harpale, C. Faloutsos, Gigatensor: scaling tensor analysis up by 100 times-algorithms and discoveries, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2012, pp. 316–324.
[201] N. Sidiropoulos, E. Papalexakis, C. Faloutsos, Parallel randomly compressed cubes: a scalable distributed architecture for big tensor decomposition, IEEE Signal Process Mag. 31 (5) (2014) 57–70.
[202] J. Inah, E.E. Papalexakis, U. Kang, C. Faloutsos, Haten2: Billion-scale tensor decompositions, in: Proceedings of the 2015 IEEE International Conference on Data Engineering, IEEE, 2015.
[203] S. Papadimitriou, P. Yu, Optimal multi-scale patterns in time series streams, in: Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data, ACM, 2006, pp. 647–658.
[204] X. Li, W. Hu, Z. Zhang, X. Zhang, G. Luo, Robust visual tracking based on incremental tensor subspace learning, in: Proceedings of Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, IEEE, 2007, pp. 1–8.
[205] B.W. Bader, A.A. Puretskiy, M.W. Berry, Scenario discovery using nonnegative tensor factorization, in: Progress in Pattern Recognition, Image Analysis and Applications, Springer, 2008, pp. 791–805.
[206] E. Acar, T.G. Kolda, D.M. Dunlavy, All-at-once Optimization for Coupled Matrix and Tensor Factorizations (2011). http://arxiv.org/abs/1105.3422.
[207] V.W. Zheng, Y. Zheng, X. Xie, Q. Yang, Towards mobile intelligence: learning from GPS history data for collaborative recommendation, Artif. Intell. 184 (2012) 17–37.
[208] B. Ermiş, E. Acar, A.T. Cemgil, Link prediction in heterogeneous data via generalized coupled tensor factorization, Data Min. Knowl. Discov. 29 (1) (2015) 203–236.
[209] H. Becker, P. Comon, L. Albera, Tensor-based preprocessing of combined eeg/meg data, in: Proceedings of the 20th European Signal Processing Conference (EUSIPCO), IEEE, 2012, pp. 275–279.
[210] A. Cichocki, Tensor Decompositions: A New Concept in Brain Data Analysis? (2013). http://arxiv.org/abs/1305.0395.
[211] W. Swinnen, B. Hunyadi, E. Acar, S.V. Huffe, M. De Vos, Incorporating higher dimensionality in joint decomposition of EEG and FMRI, in: Proceedings of the 22nd European Signal Processing Conference (EUSIPCO), IEEE, 2014, pp. 121–125.
[212] S.A. Khan, S. Kaski, Bayesian multi-view tensor factorization, in: Machine Learning and Knowledge Discovery in Databases, Springer, 2014, pp. 656–671.
[213] E. Acar, M.A. Rasmussen, F. Savorani, T. Næs, R. Bro, Understanding data fusion within the framework of coupled matrix and tensor factorizations, Chemom. Intell. Lab. Syst. 129 (2013) 53–63.
[214] Y.-R. Lin, J. Sun, P. Castro, R. Konuru, H. Sundaram, A. Kelliher, Metafac: community discovery via relational hypergraph factorization, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2009, pp. 527–536.
[215] S.-H. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng, H. Zha, Like like alike: joint friendship and interest propagation in social networks, in: Proceedings of the 20th International Conference on World Wide Web, ACM, 2011, pp. 537–546.
[216] E. Acar, A.J. Lawaetz, M.A. Rasmussen, R. Bro, Structure-revealing data fusion model with applications in metabolomics, in: Proceedings of the 35th IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBC), IEEE, 2013, pp. 6023–6026.
[217] E. Acar, Data Fusion Based on Coupled Matrix/Tensor Factorizations, 2015, (http://www.models.life.ku.dk/~acare/DataFusion).
[218] A.L. de Almeida, A.Y. Kibangou, Distributed large-scale tensor decomposition, in: Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2014, pp. 26–30.
[219] C. Hu, P. Rai, C. Chen, M. Harding, L. Carin, Scalable Bayesian non-negative tensor factorization for massive count data, in: Machine Learning and Knowledge Discovery in Databases, Springer, 2015, pp. 53–70.