Information Fusion 51 (2019) 42–57

Contents lists available at ScienceDirect

Information Fusion

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

Full Length Article

Data fusion in cyber-physical-social systems: State-of-the-art and perspectives

Puming Wang[a], Laurence T. Yang[a,b,*], Jintao Li[a], Jinjun Chen[c], Shangqing Hu[d]

[a] School of Computer Science and Technology, Huazhong University of Science and Technology, China
[b] Department of Computer Science, St. Francis Xavier University, Antigonish, Canada
[c] School of Software and Electrical Engineering, Swinburne University of Technology, Melbourne, Australia
[d] School of International Education and Exchange, Changzhou University, Changzhou, China

ARTICLE INFO

Keywords:
CPSS
Data fusion
Data fusion framework design
Tensor

ABSTRACT

Cyber-Physical-Social systems (CPSSs) are the extension of Cyber-Physical systems (CPS), which seamlessly integrate cyber space, physical space and social space. CPSSs promote the information resource from single space to tri-space, so as to lead a revolution in data science (DS). This paper aims to provide a comprehensive review of data fusion in CPSSs for readers. We firstly analyze data collection and representation in CPSS and propose to use tensors to represent CPSS data, then a general definition of CPSS data fusion is proposed to clarify the concept of information fusion in CPSS. After that, some representative data fusion methods related to CPSS are reviewed. Furthermore, we propose a series of tensor based data fusion methods for CPSS data. Also, we review the design of data fusion frameworks and propose a comprehensive data fusion framework for CPSS. Some challenges and future works are discussed as well.

1. Introduction

With the new surging of various smart terminal and techniques, including smart phone, smart Internet of Things (IoT), mobile social network and cloud computing, human as sensor has become a promising perception paradigm [1–3]. The novel perception technologies insensibly promote the data source into a new information space paradigm, which seamlessly integrates *cyber space (CS), physical space (PS)* and *social space (SS)*, namely Cyber-Physical-Social Systems (CPSS).

CPSS can be seen as a new stage of development for ubiquitous computing, which typically consists of three stages including Cyber-Physical systems (CPS), Cyber-Social systems (CSS) and Cyber-Physical-Social Systems (CPSS). In particular, along with CPS and CSS, these two systems are becoming better and being universally connected. They are finally integrated into CPSS to enable smart interaction between cyber, physical and social spaces [4]. In nature, CPS is a system which maps the objects in physical space to the cyber space, CSS is a system which maps the objects in social space to the cyber space, CPSS combine CPS and CSS, which map the social objects and physical objects to the cyber space. Actually, the CPSS can be seen as U-systems, which cover the functions of CPS and CSS. Further introduction about these concepts can refer to [4–7].

In terms of data fusion, CPSS fully integrate the capabilities of CPS and CSS through fusing various data from cyber space, physical space and social space. Some hot topics of CPSS are emerging, to name a few, [8,9] focused on the study of CPSS big data, [10] focused on the research of CPSS security, [11] focused on the study of CPSS intelligence. And some paradigmatic CPSS have emerged [5,6,12–14]. They all consider humans' needs, interests, hobbies, emotions, social relationships as part of the data fusion system [5,15]. In this sense, humans become not only the data users, but also data providers in a data fusion system. This can provide valuable inferences which are not available from conventional physical sensors.

As a novel information space paradigm, CPSS will lead to a revolution in data collection and representation, data fusion, framework design.

Data collection and representation in CPSS. In CPSS, the data are generated by various systems, to name a few, biological, environmental, sociological, and psychological system. There are big differences in the mechanism of generating data. Especially social systems are introduced in CPSS, the data sources in CPSS have posed new features, such as large scale, multi sources, heterogeneous network, cross domains, cross media, cross language, dynamic evolution and universality [4,5,16,17]. How to collect and represent the data from CPSS is a big challenge.

Data fusion method in CPSS. Traditional data fusion usually processes data from physical space. However, CPSS data are from tri-space. These data have different distributions, densities, scales, representations and modalities. Especially in CPSS, the data from human and social net-

* Corresponding author at:
E-mail address: ltyang@gmail.com (L.T. Yang).

https://doi.org/10.1016/j.inffus.2018.11.002
Received 5 March 2018; Received in revised form 23 October 2018; Accepted 1 November 2018
Available online 2 November 2018
1566-2535/

Table 1
Some applications of CPSS data fusion.

Application domain	Literatures
Smart home	[19–22]
Smart networks	[23–27]
Transport systems	[28–32]
Medical services	[33–36]
Smart city	[37–40]
Smart Business	[41–44]

works are involved, it is necessary to take into account representation, organization, processing, indexing, retrieval, dissemination of these social data. Therefore, data fusion in CPSS requires sophisticated data processing techniques and algorithms [4–7,18]. How to fuse CPSS data is a big challenge.

Data fusion framework design. CPSS are considerably complex systems due to the heterogeneity of networks, complexity of software and the hardware entities, which lacks effective design approaches to systematically consider these issues in a unified form. Another issue is that the social part in CPSS still has not been received much attentions, the mining and analysis of social data is still in the relatively early stage, there is a lack of the mining methods for large-scale high dimensional social data, there are not feasible approaches to uniformly model cyber, physical and social space.

CPSS data fusion has some primary applications in various domain, such as smart home, smart networks, transport systems, medical service, smart city, economy. The Table 1 shows these applications.

Smart home. There are some applications of CPSS in smart home, to name a few, Gator tech Smart House [19], CASAS [20], Georgia Tech's Aware Home [21], User-configurable semantic home [22].

Medical services. In smart medical service domain, the typical applications are automatic patient monitoring and diagnosis [33], environment monitoring of the ward [34,35], dynamical association of medical and ambient information [36].

Transport systems. To enable intelligent transport systems, there are many related projects under research, development and deployment. Examples are MIT cartel project [28] that predicts the bus arrival time [29], cloud-based driving direction system [30], InternetCAR Project [31], active accident avoidance application [32].

Smart networks. There are many data fusion applications in smart networks. For example, data fusion for Software Defined Network (SDN) [23–25], big data fusion for mining and modelling the dynamic patterns of service providers in cellular data networks [26], big data fusion model for QoS in smart networks [27], which proposed a big data fusion framework for smart networks.

Smart city. Recently, researchers have started to study urban planning and management with CPSS. Such as detecting urban events [45–49], urban sensing architecture [37], Device-to-Device based mobile social networks [38], public traffic [39,40], public environment monitoring [50], Los Angeles Urban sensing [51], the Intel-Sponsored Urban Atmospheres Project [52], Harvard City Sense Project [53], etc.

Smart business. Smart business is a relatively mature field in CPSS, for example, analysing the influence factors of land price, selecting company's location, and the choice of residential location, business process management [41–44]. By analyzing a large number of user's sign in data, suggestions are provided for the location of the business [41].

Although there are some applications in CPSS, and several general [54–59] and specific [60–63] reviews of data fusion exist. However, the community currently is lack of a comprehensive overview of the state-of-the-art techniques on CPSS data fusion. To fill this gap, we survey the data collection and representation, data fusion, framework design in CPSS in this paper.

We organize the paper as follows. Section 2 briefly introduces the basic concept of CPSS and data fusion, then characterize data fusion in CPSS. In Section 3, data collection and representation in CPSS are surveyed in detail. Data fusion methods related to CPSS such as CPS data fusion and CSS data fusion, are reviewed in Section 4. In Section 5, we propose a series of tensor based fusion method for CPSS data. We survey CPSS data fusion framework design in Section 6. Subsequently, we review the future work and challenges of CPSS data fusion in Section 7. Conclusions are given in Section 8.

2. Background

2.1. Tensor and tensor eigenvalue

Tensors are higher-order generalizations of matrices, i.e., multi-way arrays, which can represent additional types of variability in their higher dimensions. Tensor factorizations are one approach to interactively analyze a set of matrices that are homogeneous in size. In fact, a tensor factorization can be interpreted as a joint matrix factorization in which the different factorizations are coupled with each other by sharing certain factors. Taking this idea one step further, the information in a set of heterogeneous data sets can be fused by factorizing each data set with a tensor decomposition and coupling these factorizations through their factors [8,9,64,65].

2.1.1. Tensor eigenvector

Tensor eigenvalue problem is a hot topic in the current data science field. In 2005, Qi [66] and Lim [67] independently introduced the notion of eigenvalue problems for tensors, namely Z-eigenvalue and Z-eigenvector. For example, given a tensor A, the value λ, vector x, they satisfy Eq. 1,

$$Ax^{m-1} = \lambda x, \tag{1}$$

then λ is named an eigenvalue of A and x an eigenvector of A associated with λ, where

$$x^T x = 1.$$

Ever since, many researchers propose the various eigenvectors and eigenvalues of tensors, and some solutions are developed too. Some early applications of tensor eigenvalue have emerged, for example high-order Markov Chain, machine learning. Due to the ability to express multi-modal or multi-aspect data, tensor plays more and more important roles in data representation, fusion and application [64,68,69].

2.1.2. Tensor eigentensor

In [70,71], a new concept, namely eigentensor was proposed as Eq. 2, which is an extension of eigenvector in modal or in order. The emergence of eigentensor makes the feature change from single-modal to multi-modal. In other word, the cubical feature is emerging. This provides a promising method for feature in a fusion format. In [70], the authors used the tensor power method (TPM) to compute the largest eigentensor for prediction system, the eigentensor shows a cubical User-Spatio-Temporal profitability distribution at one time. The experiment shows that prediction accuracy is greatly improved.

Definition 1 Eigentensor: Similar to the definition of the eigenvector, the eigentensor can be described as follows:

Denoting $A \in R^{I_1 \times I_2 \times \cdots I_M \times I_1 \times I_2 \times \cdots I_M}$, $X = [x_{11\cdots1}, x_{11\cdots2}, \cdots x_{I_1 I_2 \cdots I_M}] \in R^{I_1 \times I_2 \times \cdots I_M}$,

$$A *_M X = \lambda *_M X. \tag{2}$$

If $\lambda \in R$ and $X = [x_{11\cdots1}, x_{11\cdots2}, \cdots x_{I_1 I_2 \cdots I_M}] \in R^{I_1 \times I_2 \times \cdots I_M}$ satisfy Eq. 2, we call λ an eigenvalue of A, and X an eigentensor of A, associated with the eigenvalue λ. Where the $^*{}_M$ denotes the multimode product.

2.2. Data fusion

Data fusion is a technology which combines information or data from multiple sources so as to form a unified decision. In nature, data fusion is similar to human's or animal's obtaining information with various senses, understanding the objective world through a comprehensive analysis of the brain.

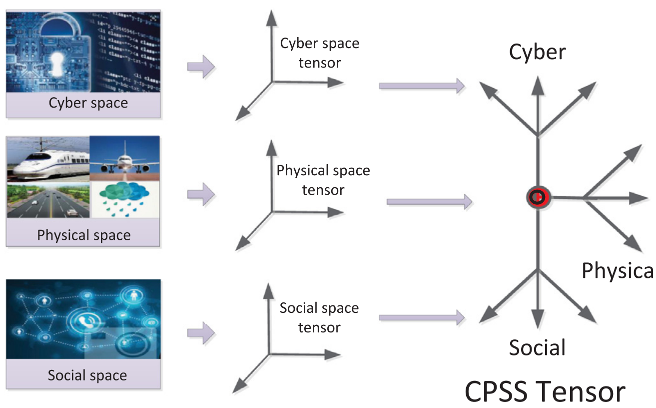


Fig. 1. The definition of CPSS Data Fusion.

Data fusion is an interdisciplinary research field, many terms are often used interchangeably in the literatures, such as information fusion, data integration. About the concept of data fusion, the researchers propose a variety of definitions or explanations from different fields or viewpoints [16,54,72–76], but there is still no generally accepted definition.

The pioneering data fusion definition is proposed by Joint Directors of Laboratories (JDL) [72] as merging the data or information from multi sources to make multilevel processing. The definition is very popular, but it is too restrictive to military applications. Klein [73] generalized this definition for a broad field applications. In [74], the researchers surveyed data fusion definitions. Another viewpoint of data fusion is multiple data sets fusion [16,54,74–76], this data fusion is called modern data fusion. Considering the context of this paper, we define the data fusion in CPSS as follows.

Definition 2 CPSS Data Fusion: CPSS Data Fusion (CDF) is the integration of data from Cyber-Physical-Social spaces, such that the corresponding data sets from tri-spaces can interact and inform each other to provide services for CPSS space. The Fig. 1 shows the definition of CDF.

2.3. Data fusion framework

Data fusion framework mainly refers to the architecture of data fusion. Until now, researchers have proposed a variety of information fusion framework, such as JDL framework, OODA framework, Omnibus framework, STDF framework [77,78]. Among them, JDL framework and its variants are most popular. The JDL framework was originally proposed for military applications and consisted of four levels: i) target assessment, ii) situation assessment, iii) threat assessment, iv) process optimization. Fig. 2 shows the architecture of the JDL fusion framework. Subsequently, the researchers added zeroth level and carried out different interpretation about the function of the level, such as the information source, signal level optimization, and signal estimation. In Section 6, we will focus on the design of the CPSS data fusion architecture.

3. Acquisition and representation

The data collection and representation are discussed in this section. Firstly, we present a comprehensive overview about CPSS data collection, then we summarize data representation and propose to use tensor to represent CPSS data.

3.1. Acquisition

In traditional data science, data acquisitions mainly focus on physical space, the sensing devices are physical sensors. Compared with traditional sensing, CPSS data acquisition has characteristics as follows.

Inter-space sensing. The inter-space sensing has another definitions such as Collaborative Sensing or Augmented Sensing, which is the most important feature of data acquisition in CPSS. CPSS have three sensing resources, various sources have distinct sensing or perception abilities, hence, it is necessary to orchestrate the different perceptions to achieve inter space sensing. From the view of signal and system, the inter-space sensing consists of two parts, input and output. Among them, the input is Tri-space sensing sources, the output is associated data [1,79]. This sensing helps to find the relation of the same physical object in different space, such as social relation in microblogs, video sharing in Youku, posts and comments in online forum, social network in email, the record data from e-commerce sites, which are associated by the same user entities. Another example is that GPS mobile data, check-in data and delivery data are associated through the same location entity, which can be used in location based service (LBS).

Human-centred and social networking sensing. Humans as sensors is a new concept (usually called Crowd Sensing). For example, when the disaster occurs, some humans will release news or upload photos on social media, these humans in fact help people perceive in their surroundings; credit card behavior when a user access to the subway station, indirectly help people perceive the travel rules and people crowding subway system; humans with smart mobile devices can help to perceive the city in the event of emergency [79,80].

Heterogeneity. Heterogeneity is a main feature of data from CPSS, which arises from expression difference (i.e., modalities or views) and source difference (i.e.,cyber space, physical space, social space). In modern data fusion perspective, heterogeneity is in the diversity of order and dimensionality, namely expression difference, for example, some data are higher-order tensors, some data are matrices. Some factors, including the construction of data space, the implementation of data management system, technology and other artificial factors, lead to the heterogeneity [4]. For example, the point of interest (POI) are spatial data, the moving trajectory data are spatio-temporal data, traffic flow is the stream data, social web users posted information is text or image data.

Heterogeneous information fusion is also known as multi-modal information fusion. Compared with homogeneous data, heterogeneous data can provide information for users with more complementarity. The integration of a variety of heterogeneous and incomplete information can form a consistent understanding of the perceived scene or situation. But managing and representing large scale heterogeneous data is currently a big challenge.

3.2. Representation

Data representation is fundamental to data mining and machine learning. Both disciplines aim to discover knowledge and rules from raw data. In recent years, many works about data representation have been described and discussed.

Graph representation. Graph representation, presented by [81], is a type of data representation that have a natural representation in matrix form. A graph consists of a collection of points called nodes, some of which are connected with lines called edges. A matrix that associates its rows and columns with the nodes of a graph and whose elements equal the number of edges between any pair of these nodes is called an adjacency matrix [64].

For example, in the web graph each node is a web page and each edge represents a hyperlink from one page to another. Google[1] uses the adjacency matrix of the web graph to rank web sites in their search engine results.

In a social graph, nodes are people and edges represent relationships between them. Facebook[2] is currently a world's large social network

[1] https://www.google.com
[2] http://www.facebook.com/

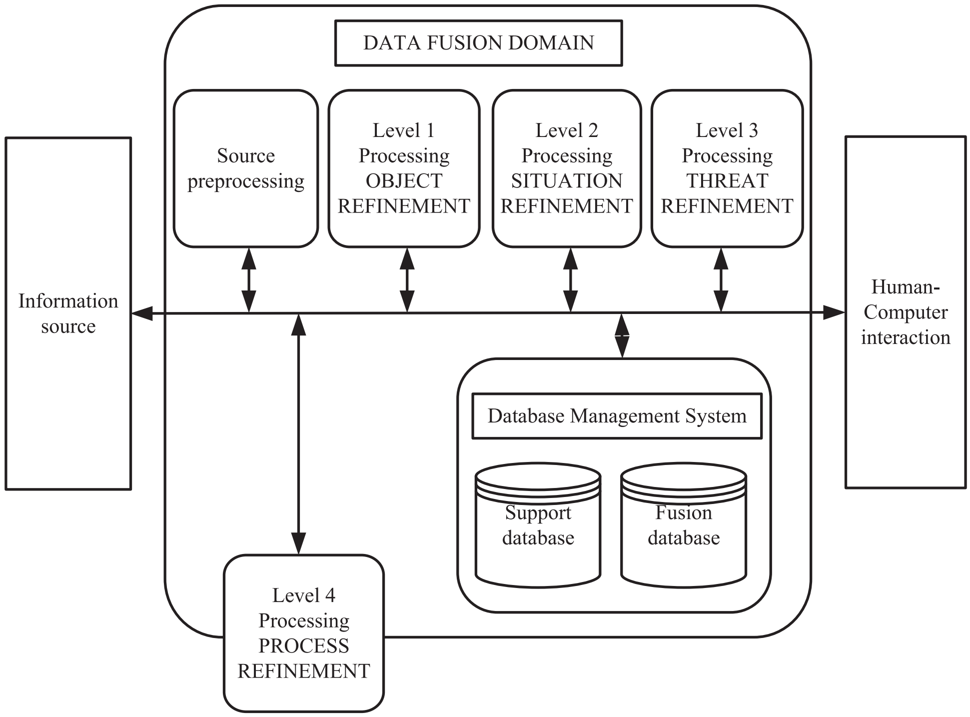


Fig. 2. The architecture of the JDL fusion framework [72].

and has a social graph of more than a billion active users. Although the graph representation method is widely utilized in data organization, it has some limitations. For example, its restricted application domains narrow its applicability in CPSS.

Ontology representation. Ontology is defined as a collection of Synsets of concepts which in turn is a collection of Hyperyms/Hyponyms and Holonyms/Metonyms [82]. Ontology has four characteristics, namely sharing, reusability, hierarchy, reasoning. These characteristics empower the Ontology to represent an object as the conceptual model on the semantic level. The construction of Ontology consists of two steps, i) mapping the real world into a set of concepts (such as entities, attributes and processes etc.); ii) extracting the relationship between concepts. For example, Ontology in medicine domain has the concepts such as disease, symptom, treatment and relationship, such as symptoms of disease, treatment of disease. Ontology simplifies knowledge transformation, Ontology representation method was widely used in many domains [83]. In [84], some challenges of Ontology representation in IoT are discussed, such as representing incomplete or inaccurate meta data, managing large scale devices, describing unknown topology structure. Image Ontology has become an important method of data representation. And some Ontology have been established, for example natural language Ontology Wordnet[3], social Ontology SWETO[4], context Ontology Context-OWL [85]. However, Ontology is an explicit description of concepts in a special domain, the Ontology representation method is not suit for comprehensive CPSS applications.

Matrix Representation. Matrices are also called two-way arrays, expressing variability in time and space (sensors) along the two dimensions. In fact, multi-channel signals are naturally represented in a matrix, where each row of the matrix contains each sensor or channel, each column contains the measure time, and the elements of matrix denote signal values. In data mining and machine learning, a rectangular array usually represent samples or observations with some attributes or properties, each row corresponds to a sample or observation, and each column corresponds to a number of attributes or properties associated with that sample. For example, in a forest fire scene, each row in a data matrix can correspond to an observed fire point, while the columns describe the forest fire's properties such as its location, the size of the burned area, the temperature and wind speed.

Tensor Representation. A tensor is the extension of a vector in order. Tensors are essentially multidimensional arrays, and each element of the tensor has multiple index, each index represents a model or a order. Tensors are very powerful and versatile tools that can represent a wide variety of heterogeneous, multi-aspect data. For example, in [86], the electronic nose data are represented as a 3th-order tensor which consists of three different sources of information from sensor signals, gas and time axis. In [87], web link data can be defined as a 3th-order tensor consisting of three different groups of information, web pages, links and anchor texts. In [88], gait video data are represented as a 4th-order tensor consisting of four sets of information, including pixels, angles, movements, and objects.

CPSS data have large diversity, such as structured data (such as tables, lists etc.), unstructured data (such as text, pictures, video, etc.), semi-structured (such as HTML document, relational database, etc.), can also be knowledge (such as rules, models, mechanisms, etc.), these data not only coexist in a variety of forms, but also appear in various space. From the above discussion, tensors are suitable for representing CPSS data. Therefore, in Section 6, our proposed framework for CPSS data fusion uses tensors as the data representation.

4. Data fusion method

The CPSS Data Fusion (CDF) is of great challenge due to the inherent complexity, which closely relates to multidisciplinary technologies. Because CPSS are still in their infancy, little works have been focused on the fusion method. However, there exist some fusion methods about CPS or CSS, which can provide significant value for designing CPSS data fusion method. Therefore, we will survey them in following subsections. Our review of the related literatures is organized corresponding to the taxonomy from Fig. 3.

[3] https://www.wordnet.princeton.edu/wordnet/
[4] http://archive.knoesis.org/library/ontologies/sweto

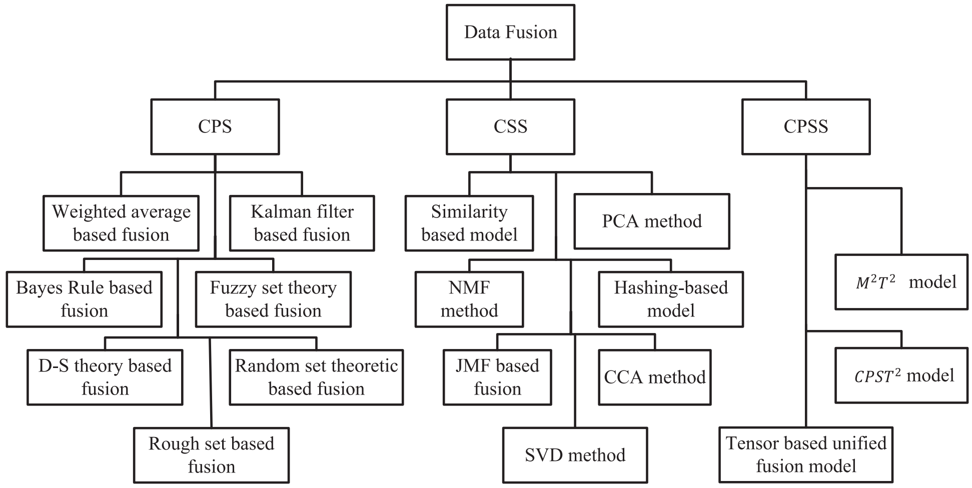


Fig. 3. Taxonomy of data fusion methodologies: different data fusion algorithms can be roughly categorized based on data space, namely, Cyber-Physical space fusion, Cyber-Social space fusion, Cyber-Physical-Social space fusion.

4.1. Cyber-Physical space (CPS) fusion

4.1.1. Weighted average based fusion

The weighted average method [89] is the simplest data fusion method which is a direct fusion method for data source at signal level. It is suitable for fusing homogenous data from the isomorphic sensors, these data are described about the same object. If a target is measured by K sensors, its average value is defined as Eq. 3,

$$\overline{x} = \sum_{i=1}^{k} w_i x_i, \ \sum_{i=1}^{k} w_i = 1, \tag{3}$$

where the w_i denoted the i-th sensor's weight. This method's advantages are simplicity, intuition. But this method has poor performance in data fusion domain, which usually is used in simple Internet of Things (IoT).

4.1.2. Kalman filter based fusion

Compared with weighted average based fusion, Kalman Filter (KF) model is based on minimum variance estimate instead of average value. This method was firstly proposed in the 1960s [90]. Ever since it became the most common technique in target tracking and navigation system. Sequently, to cope with bearings-only passive sonar data, an improved KF method named Extended Kalman Filter (EKF) [91] was proposed, due to its linearization, the EFK is not the optimal. Both KF and EKF were originally used on the data from a single sensor. To form a global estimate, Willner et al. [92] firstly developed the idea of combining information from local sensors at a central fusion node. Since then, the KF method was gradually used in the Internet of Things (IoT). But this kind of method has the drawback that each local sensor requires the global estimate, which required two-way communication, this negates some of the advantages of parallelization. KF method is not suitable for large scale data fusion from CPSS.

4.1.3. Bayes rule based fusion

The Bayes' rule [93] is an early method used in IoT for uncertainty information, which set the prior probability firstly, then use the Bayes' rule to calculate the posterior probability. According to the posterior probability, the researchers can deal with uncertain problems.

Bayes' rule provides a principled means which combines the observed information with objects' prior confidences. But there are some barriers in the Bayes' rule method applications. Among them, the main one is finding the data probability distribution, especially when the data comes from a low level sensor. In addition, it is difficult to know the prior probability in practice, when the prior probability is not consistent with the actual situation, the inference result is very poor.

Bayes' rule are based on statistical rules, there are some similar methods, such as voting method and user-defined rules method. These methods have the advantages of simple model and a lower computational complexity. The Bayes rule has gradually become a main method of multi-sensor data fusion in cyber-physical systems (CPS) for incomplete data fusion [93].

4.1.4. Fuzzy set theory based fusion

Fuzzy set theory is a local theory based on classification, which was first proposed in 1960s [94]. The fuzzy theory relaxes the constraints about the definition of probability theory, so that data fusion can be modeled in loose way so as to solve the conflict between information and decision [95,96]. Fuzzy theory usually combines with other methods [97] to improve the accuracy of information fusion. Fuzzy theory uses membership function to deal with the fuzziness, but the degree of membership is given by experience or domain expert, it is quite subjective. So fuzzy set theory is not a popular fusion method. Fuzzy theory always is used in some specific IoT scenario where degree of membership can be easily got.

4.1.5. Rough set based fusion

Rough set is firstly proposed by Pawlak [98] to represent imprecise or incomplete data. In recent years, rough set theory has been successfully applied in the fields of knowledge representation, data mining, machine learning, knowledge discovery. Rough set based fusion is described as follows,

- The collected information are transformed into attribute data, including condition attributes and conclusion attributes, then a convenient analysis table is made according to condition attributes and conclusion attributes.
- The core idea of the rough set is classification, the continuous attribute values of data are dispersed, then these data are clustered in terms of discrete attribute values.
- On the basis of these classifications, the redundant or repeated information is deleted with reduction and kernel concept, and the information table is simplified.
- Classify the reduced data according to different attributes, and find the kernel.
- Listing the possibility decision table based on the nuclear and the original samples.
- The minimum rule is summed up to get the fastest fusion algorithm.

Compared to fuzzy set theory, rough sets are based on data itself and do not need any prior information. They have more abilities in data fusion, to name a few, processing fuzzy data from physical space, reducing redundant information, compressing data [99,100]. The fuzzy set theory is mainly used in the Internet of Things (IoT).

4.1.6. D-S Theory based fusion

Evidential theory [101] is a mathematical tool that is used to fuse uncertain data. The evidence theory is the extension of probability theory, which was firstly proposed by Dempster to construct general framework of uncertain reasoning model, Shafer further develop the theory [102]. Since evidence theory is also named Dempster–Shafer evidence theory [103,104], or D-S theory [105,106]. The D-S theory describes the uncertain information with interval estimation instead of point estimation, hence it is more flexible in data fusion. However D-S theory's computational complexity increases exponentially with the number of sensors [107]. Accordingly evidential theory is unfit for large scale CPSS data fusion.

4.1.7. Random set theoretic based fusion

Random set theory was firstly proposed to study integral geometry in 1970s [108]. In nature, random set theory is another extension of probability theory. Random set is a kind of random element which value is random set, and it is a generalization of the random variable (or random vector) in probability theory. Its highlight is the unifying capability which has been proved by some works. Among them, the most typical job is proposed by Mahler et al. [109,110]. They promoted the random set theory to be a unified data fusion framework. In his book [110], the random set theory and its applications were introduced in detail, readers can refer to the book. Some researchers have proved that random sets theory, evidence theory, and fuzzy set theory can be mutually transformed with some constrains [111].

In the framework of random set theory, the uncertain information and fuzzy information can be uniformly described and expressed. This provides a prerequisite for handling heterogeneous data. Therefore, random set theory plays a very important role in the integration of heterogeneous information in CPS.

4.2. Cyber-social space (CSS) fusion

With the development of online social networks, humans can easily share information with each other and create richer social information [112]. Human gradually become producers and consumers of data. Meanwhile, information space gradually considers social factors so as to reach a new stage, viz. CSS. CSS data fusion paradigm gives full play to human's role in fusion process [113], and can make up for the deficiency of physical space in semantic analysis and logic analysis.

Preliminary researches in CSS have been studied. Some works focus on social data fusion framework [114–117], and others focus on human data fusion methodology [118–120]. Also in [112], a comprehensive review about CSS data fusion was provided, the paper summarize motivations, advantages, approaches and challenges of data fusion in CSS.

4.2.1. PCA

The explosive popularity of social media produces massive data at an unprecedented speed. For example, 250 million tweets are posted per day; 3000 photos are uploaded per minute to Flickr; and Facebook users had increased from 100 million in 2008 to 800 million in 2012. The massive and high dimensional social media data challenges traditional data mining tasks such as classification and clustering due to the curse of dimensionality. Principal Component Analysis (PCA) is mainly used in CSS for data compression and dimensionality reduction. PCA extracts the main features from the linear space constructed by CSS data, then CSS data are processed in a low dimensional space constructed by the main features, so as to solve the bottleneck of the high dimension data from social media [121,122].

4.2.2. SVD

Singular Value Decomposition (SVD) is an important matrix factorization in linear algebra. In general, for any matrix $A \in R^{m\times n}$, its singular value decomposition is defined as follows,

$$A = U\Sigma V^T, \tag{4}$$

$U \in R^{m\times m}$ and $V \in R^{n\times n}$ are orthogonal matrices, Σ is a diagonal matrix.

Recommender system is a classic CSS system, which fully studies humans' interest, then recommends the suitable things (commodities or information) to meet the humans' need. SVD is mainly used in recommender systems to address information overload. With the development of electronic commerce systems, the magnitudes of users and items grow rapidly, result in the extreme sparsity of user rating data set. The SVD-based recommendation systems alleviate the sparsity problems of the user-item rating matrix, so as to provide better recommendation than traditional collaborative filtering algorithms.

4.2.3. NMF

Lee et al. [123,124] firstly added nonnegative constraints on matrix factorization, and proposed the concept of nonnegative matrix factorization (NMF). NMF represents data as a linear combination of some base vectors, and these combination coefficients and base vectors are nonnegative. Given a matrix A, its NMF method can be described as follows,

$$A = UV, \tag{5}$$

where U and V are non negative matrixes.

NMF is in line with person's psychology [125], NMF is widely used in social media domain, for example, learning evolving and emerging topics [126] from social media, trust-link prediction [127] in social networks, social event detection [128].

4.2.4. Hashing-based model

Generally hashing algorithm maps heterogeneous data into a common space for data fusion, the common space always is a binary space (such as Hamming space). The binary representations help to find relevant objects fast and reduce storage complexity. Zhang et al. [129] proposed a composite hashing model for information retrieval, unfortunately the method focused on single domain. In order to cope with the problem, a classification-based hashing was developed for cross-space heterogeneous data in [130]. Zhen et al. [131] improved the above algorithms and proposed a probabilistic model to learn hash functions.

Hashing-based algorithm is a popular solution strategy for heterogeneous data in CSS. For example, Liu et al. [132] developed a hash-based method to fuse heterogeneous social data for improving accuracy of image retrieval. Slutsky et al. [133] combined hashing algorithm with latent dirichlet allocation (LDA) for topic modeling in social streams. In [134], based on locality-sensitive hashing method, an incremental community detection method was proposed for social tagging systems. Shalita et al. [135] proposed Social Hash (SH) framework to solve assignment problems on social networks.

4.2.5. Similarity based model

Similarity-based data fusion approach is another popular method for heterogeneous CSS data fusion. This kind of method firstly computes the similarity of objects to construct the similarity matrix, then complete the heterogeneous data fusion based on the similarity matrix. The fusion method can be carried out in the data level, also can be extended to decision-making level.

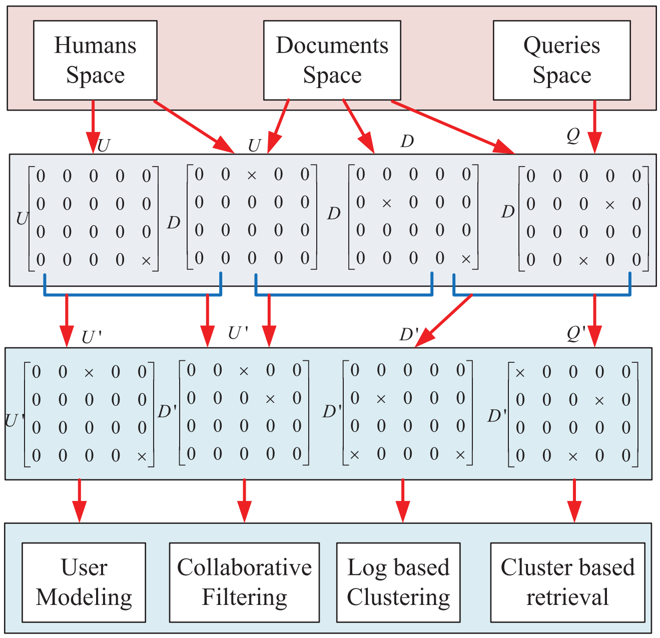


Fig. 4. Visualization of the similarity based method from [137], the data from CSS are fused by URM and SimFusion, completing a cross space searching.

The similarity-based data fusion approaches have been widely used, for example location based service [136] in CSS, search service in CSS [137]. Among them, the typical work is [137]. The work mainly consists of two parts, the first is unified relationship matrix (URM), which represents a set of heterogeneous data objects and their interrelationships, the second is SimFusion, which completes the similarity of inter-space objects. Based on URM and SimFusion, this work designs an inter information searching system as Fig. 4.

4.2.6. CCA Method

Like PCA, Canonical Correlation Analysis (CCA) [138] is a multivariate statistical analysis technology, which extracts the correlation between the two data sets. The essence of CCA is a dimensionality reduction method, this method is suitable for cross space data. CCA maps the correlative data from two data sets into a low dimension space. The solution of the low dimension space can be transformed to an optimization problem as Eq. 6,

$$\max_{x_i \neq 0, x_t \neq 0} \frac{x_i^T \sum_{IT} x_t}{\sqrt{x_i^T \sum_{II} x_i}\sqrt{x_t^T \sum_{II} x_t}}. \tag{6}$$

Ever since CCA algorithm was proposed, some improved CCA methods have been developed for various applications. For example, Kettenring [139] developed a general CCA method to cope with multiple matrices. Zhu et al. [140] used Kernel CCA (KCCA) to cope with the dimensionality reduction in nonlinear space. CCA algorithm has the characteristic of cross space fusion, so it is widely used in CSS data fusion, such as sentiment analysis [141], semantics mining [142].

4.2.7. Joint matrix factorization based fusion

Joint factorization of multiple matrices is a research hotspot of CSS data fusion. In order to analyze multi-relational data from CSS space, Singh et al. [143] proposed a novel framework to simultaneously factorize multiple matrices which shared the same mode, this is the pioneer of the collective matrix factorization (CMF). Given matrices $X \in R^{I\times J}$ and $Y \in R^{I\times K}$, they shared the common mode-1, their CMF is defined as Eq. 7,

$$f(A, B, W) = \left\|X - AB^T\right\|^2 + \left\|Y - AW^T\right\|^2, \tag{7}$$

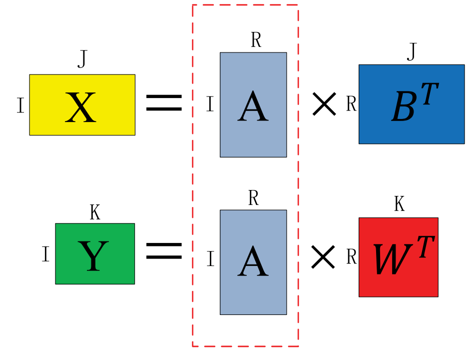


Fig. 5. Visualization of joint matrix factorization. $X \in R^{I\times J}$ and $Y \in R^{I\times K}$, share the common mode-1, $A \in R^{I\times R}$ is the shared factor matrix, $B \in R^{J\times R}$ and $W \in R^{K\times R}$ are the independent factor matrices.

where $A \in R^{I\times R}$ is the shared factor matrix corresponding to the common mode; $B \in R^{J\times R}$ and $W \in R^{K\times R}$ are the independent factor matrices which respectively arise from the other mode of X and Y; and R is the number of factors. Fig. 5 shows the main idea of the joint matrix factorization. Joint factorization of multiple matrices fusion methods have widely been used in CSS data fusion. He et al. [144] used joint matrix factorization to fuse link information and content information for mining topic community existing on online social network (OSN). Cui et al. [145] proposed a joint matrix factorization method for unsupervised action categorization.

The existing data fusion methods are mainly application specific. In particular, the methods such as Section 4.1 mainly were used in IoT or CPS, the methods form Section 4.2 mainly were used in CSS domain. These methods can't effectively represent the multi-model data or multi-relation data, which are not suit for the CPSS data.

5. Proposed fusion methodology for cyber-physical-social space

As Section 4 describes, there are some data fusion methods in various fields. Unfortunately, most of the existing fusion methods practically target to either Cyber-Physical, Cyber-Social or Social-Physical dimensions. Meanwhile, they are application specific and lack a unified, resilient, reliable and efficient fusion method for CPSS. Also the above existing fusion methodologies primarily focus on individual environment. Hence, various application domains of CPSS are relatively isolated and unable to realize the effective inter operations. To address the above challenging problems, we propose a series of tensor-based fusion method for CPSS data.

5.1. Tensor based unified fusion (TUF) model

Although researchers have done a lot of work in this area, these efforts mainly aimed at single modal low order data, which are not suitable for data fusion in CPSS. Inspired by the work [64], Kuang et al. [146] proposed a tensor based unified fusion (TUF) model. Fig. 6 shows the process of the tensor based unified fusion model. Firstly, the objects in CPSS are represented as tensors, for example, the video data can be represented as a 4th-order tensor $T_{video} \in R^{I_h\times I_c\times I_f\times I_w}$, the social network can be represented as a matrix $T_{social-network} \in R^{I_x\times I_y}$, the XML document can be represented as $T_{XML} \in R^{I_i\times I_{id}\times I_x\times I_y}$, the GPS data can be represent as a 3th-order tensor $T_{GPS} \in R^{I_{ec}\times I_{en}\times I_{er}}$. These tensors are heterogeneous due to the diversity of order and dimensionality. In order to fuse heterogeneous data, the TUF model defined two tensor operation, Tensor Extension Operator (TEO) as Eq. 8, unified data ten-

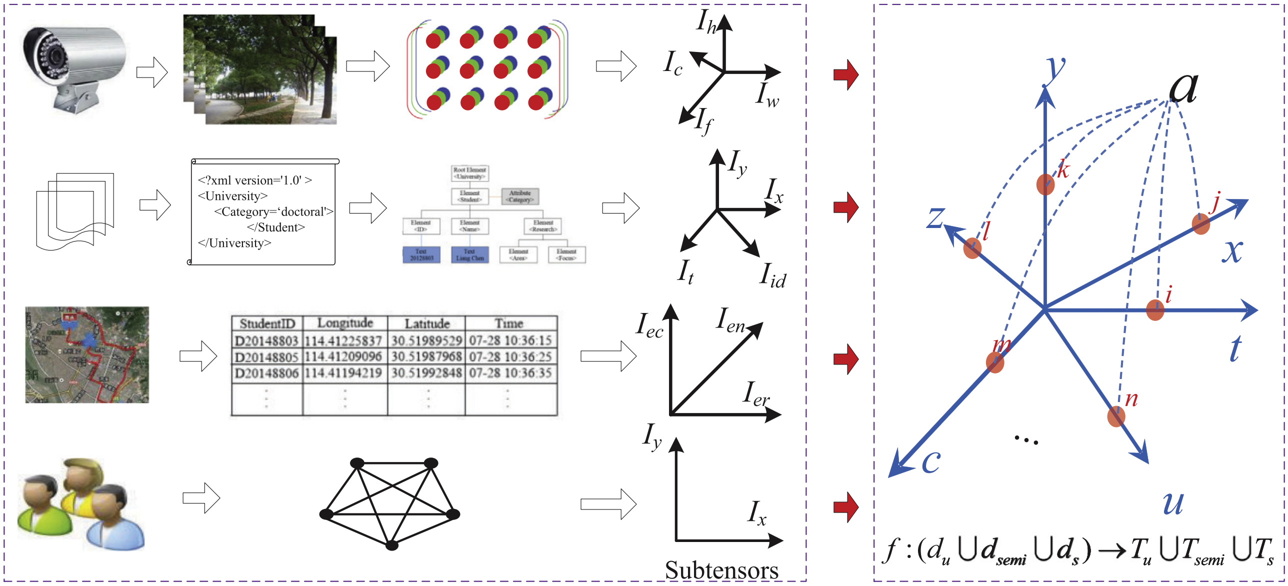


Fig. 6. The tensor-based unified fusion model. The heterogeneous data from cyber space, physical space and social space firstly are represented as low order local-space-tensors, which are fused to a higher order unified tensor using Tensor Extension Operator (TEO).

sorization (UDT) operation as Eq. 9.

$$f\ :\ A\vec{\times}B \to C, C \in R^{I_t\times I_s\times I_u\times I_1\times I_2}, \tag{8}$$

where $A \in R^{I_t\times I_s\times I_u\times I_1}$ and $B \in R^{I_t\times I_s\times I_u\times I_2}$. The operator's advantage is that the identical orders are merged and the diversity are kept:

$$f\ :\ (d_u \cup d_{semi} \cup d_s) \to \underbrace{T_u \cup T_{semi} \cup T_s}_{T}, \tag{9}$$

where d_u denotes unstructured data, d_{semi} denotes semi-structured data, d_s denotes structured data.

The heterogeneous data from cyber space, physical space and social space firstly were represented as object tensors, Eq. 8 is used to fuse the object tensors related to cyber space, physical space and social space, then local-space-tensors (LST) or subtensors are got. LST has fused the single space information. To fuse the CPSS data in a comprehensive way, the Tensor Extension Operator (TEO) is done on the LSTs, in the end we get a higher order unified tensor $T \in R^{I_u\times I_c\times I_t\times I_x\times\cdots\times I_y\times I_z}$ which consists of all information of CPSS.

Fig. 7 illustrates the unified fusion model from another point of view. In particular, the unified fusion model consists of two layer, the first is overall layer, which is a three-order tensor $I_t\times I_s\times I_u$ containing *TI, SP* and *U*, where *TI, SP* and *U* refer to the fixed order time, space and user. The second is inner layer, which can be embedded to the overall layer. The data from tri-space are fused as three sub-tensors, then the sub-tensors are embedded in the overall layer, the unified tensor is got lastly.

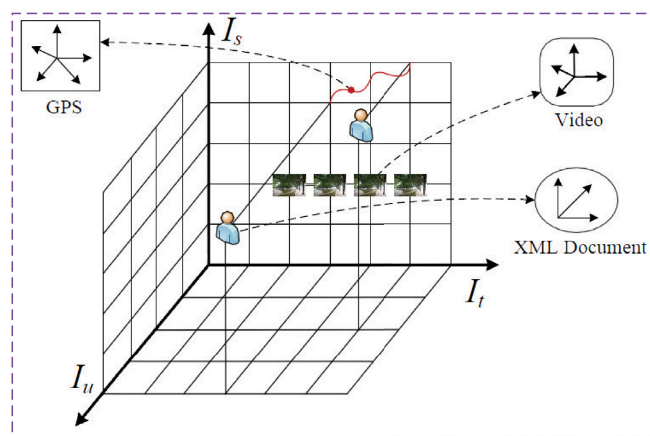


Fig. 7. The unified fusion model from [146], the unified fusion model consists of two layers, the first is overall layer, which is a three-order tensor $I_t\times I_s\times I_u$ containing *TI, SP* and *U* , the second is inner layer, which is embedded to the overall layer.

5.2. M^2T^2 *Model*

We develop the tensor based unified fusion (TUF) model and add a probability constraint on unified fusion tensor, in nature, cooperating multivariant Markov Chain (MC) with tensors. In particular, we extend the transition matrix of MC, and propose the multivariate multi-step transition tensor M^2T^2 model to fuse data arising from CPSS, including mobile sensor data, point of interesting (POI) in city, social network information. The M^2T^2 model consists of three steps, spatio data fusion, spatio-temporal fusion and human-spatio-temporal fusion.

Spatio fusion The spatio transition model is established according to the POI sequence which is obtained from GPS raw data. The corresponding transition matrix is a row stochastic matrix (i.e. the sum of matrix elements in each row equal to one). Fig. 8 shows the process of generating transition matrix which describes a user's mobility behavior modeling. The element of the matrix denotes the probability named transition probability of user's moving from one POI to another.

According to the discrete random process theory, there is a stationary distribution corresponding to the transition matrix, in other word, given any initial distribution vector, after infinite times and transition matrix multiplication, and finally converge to a vector called the steady-state distribution. According to matrix theory, the steady-state distribution vector is the eigenvector corresponding to the largest eigenvalue of transition matrix. This Markov model only fuses the space information. In order to fuse temporal information, the spatio-temporal transition tensor is proposed as follows.

Spatio-Temporal fusion In the real world, the users mobility behavior is related with the time, for instance, the users usually are at lunch-

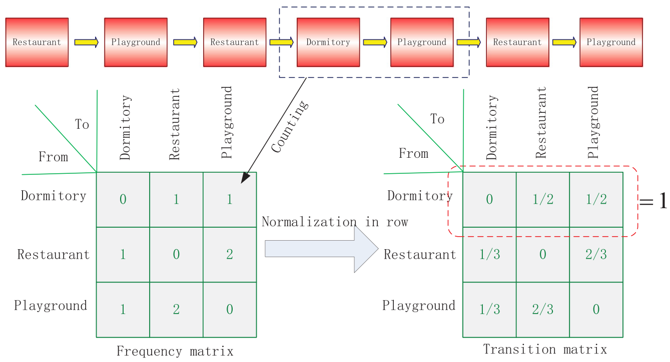


Fig. 8. The space transition tensor from [70], which describes a user's mobility behavior modeling. The element of the matrix denotes the probability named transition probability of user's moving from one POI to another.

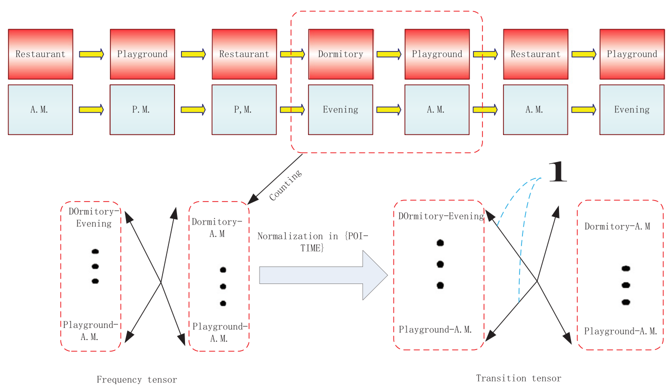


Fig. 9. The time-space transition tensor from [70]. The elements of the tensor denote the probability named transition probability of user's moving from one POI sometime to another POI at another time.

room 12 o'clock noon, but at playground 5 o'clock afternoon. So the time is a vital factor which influences the users mobility pattern. In order to fuse the time information, firstly we discretize a day into I time slices, then a new state space $S = \{\{T_1, P_1\}, \{T_i, P_J\}, \cdots \{T_I, P_J\}\}$ is built, the state space consists of POI information and time information. Based on the new state space, by using the Eq. 10, we construct the spatio-temporal transition tensor denoted as T_{ST} which is a 4th-order tensor. The 4th-order tensor merges the time and space information.

Fig. 9 shows the process of generating transfer tensor which describes a user's mobility behavior model including time and space. The transition tensor is described as $T_{ST} \in R^{I\times J\times I\times J}$ and the elements of the tensor denote the probability named transition probability of user's moving from one POI sometime to another POI at another time.

$$T_{t_1 l_1 t_2 l_2} = \frac{count((t_1, l_1) \to (t_2, l_2))}{count((t_1, l_1))}. \tag{10}$$

Human-Spatio-Temporal fusion In order to fuse social information, we firstly quantify the social attributes, a concept of mutual influences is defined as Eq. 11. If two users' trajectories are more similar, their mutual influences are greater, in a nutshell, their correlation coefficient is larger.

$$\rho_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}. \tag{11}$$

Assuming N users, by calculating the correlation coefficient for each pair of users, the correlation coefficient matrix $\Gamma \in R^{N\times N}$ can be obtained. Its elements ρ_{ij} denote the mutual influence of the $user_i$ and $user_j$.

$$\Gamma = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{bmatrix} \tag{12}$$

Normalizing the correlation coefficient matrix Γ in row, such we obtain the influence matrix (IM) Λ as follows,

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{n1} & \lambda_{n2} & \cdots & \lambda_{nn} \end{bmatrix} \tag{13}$$

All N users' influences are fused in the time-space transition tensor with the tensor operation as Eq. 14, Fig. 10 shows the process.

$$A = \Lambda \odot T_{ST}, \tag{14}$$

where $\odot$ is Kronecker product. The Human-spatio-temporal transition tensor fuses the cyber space, physical space, social space in a unified form.

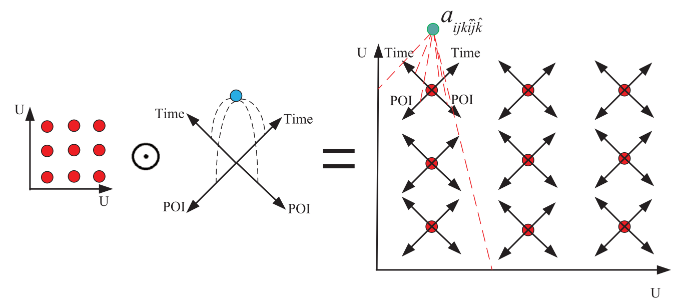


Fig. 10. The process of generating the Human-spatio-temporal transition tensor from [70]. The time-space transition tensor fuses the social network information through the tensor operation (Kronecker product), in the end, the Human-Spatio-Temporal transition tensor is got, which is a 6th-order tensor consists of the temporal information, spatio information and social information.

5.3. CPST^2 *Model*

In order to cope with the multi-space service problem, in this work [71], the Cyber-Physical-Social transition tensor ($CPST^2$) model was proposed to fuse the CPSS data in a unified form. In nature, the $CPST^2$ model is a Multi-Markov-Chain form, every Markov Chain responds to a space, there are some potential interactions among these chains. According to the Random Walk theory, the potential interactions construct a random walk among multi spaces, we study the likelihood that the object will arrive at any particular multivariate state in various spaces.

As Fig. 11 shown, the CPSS data fusion consists of two steps, firstly, cyber data and physical data construct two Markov Chains, based on the Random Walk theory, a two space transform frequency tensor $T_{cp} \in R^{J\times K\times M\times N\times O\times P\times Q}$ is got, which fuses time information, weather information, temperature information, traffic flow information from cyber world and physical world. Secondly, in order to fuse the social information, we construct a social network based on the humans' trajectory, using Pearson Correlation Coefficient (PCC) to get an influence matrix (IM) $\Gamma \in R^{I\times L}$ as described in the Section 5.2. After that, using the Kronecker Product and Tensor Reshape Operation [147], namely Reshape($\Gamma \odot T_{cp}$), we get the multi space transform tensor. The multi space transform tensor is described as $T \in R^{I\times J\times K\times L\times M\times N\times O\times P\times Q}$. Then three normalization operations responding to three space are completed, three transition tensors $T1$, $T2$, $T3$ are got. The $T1$, $T2$ and $T3$ are inde-

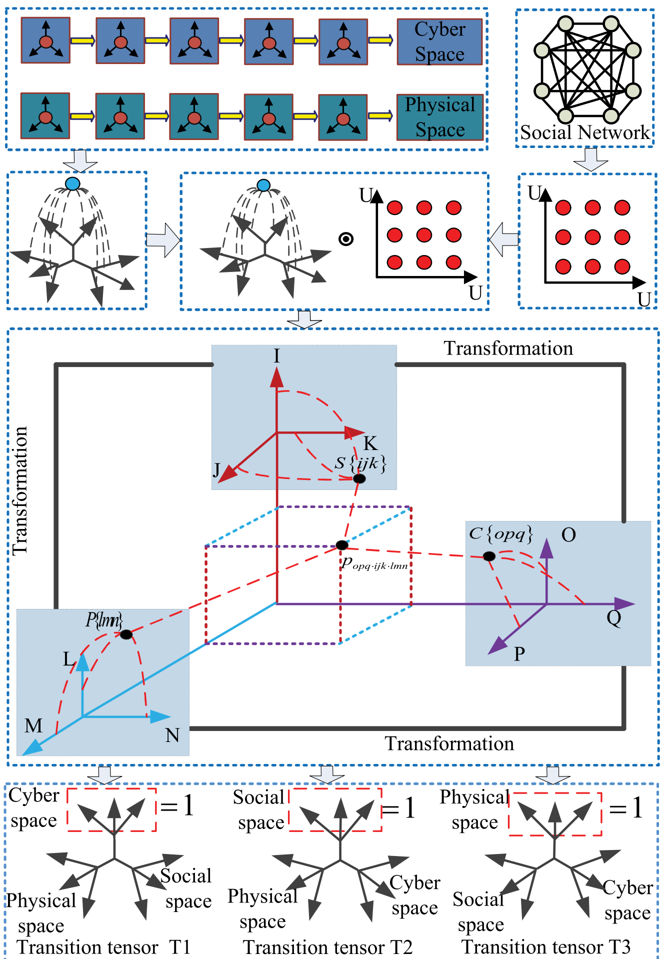


Fig. 11. The process of generating the multi-space transition tensor T from [71]. The CPSS data fusion consists of three steps, the first is constructing the frequency tensor responding to cyber space and physical space, the second is fusing social information, getting the unified tensor, the third is normalizing the unified tensor responding to multi space. In the end, the multi space transition tensors are obtained.

pendent of time and multi-column stochastic. The elements of $T1$ denote the likelihood which the state $S\{i, j, k\}$ will occur under the condition of state $P\{l, m, n\}$ and $C\{o, p, q\}$. The elements of $T2$ and $T3$ are similar to $T1$, but the state spaces change. In order to complete the multi space service, a new tensor decomposition (TD) named Intersect Tensor Power Method (ITPM) was proposed to solve three largest eigentensors related to $T1$, $T2$ and $T3$, like Markov Chain (MC) or PageRank, the three largest eigentensors are three limiting probability distributions responding to three spaces, we deem the limiting probability distributions of objects as predictive values to provide service for multi spaces respectively.

6. The proposed framework design

6.1. The fusion framework design

Through CPSS data fusion, we can further mine the value of data and enhance the function of information analysis. The multi-source information provide cross validation for the target, so as to reduce information errors and omissions, prevent decision-making errors. The combination of cyber space, physical space and social space will also inspire some novel applications in various domains. These novel CPSS applications require novel framework designs.

The CPSS data fusion framework design can be divided into three stages and seven versions, Fig. 12 shows the CPSS data fusion framework (CDFF) development.

The initial stage. In fact, researchers have added fifth levels of "user optimization" on the correction model of JDL, which endowed the JDL architecture with human cognitive function. This can be seen as the embryonic form of CPSS fusion framework. After then, some works are done about the CPSS data fusion framework design, among them, the more typical works are from [4,148,149], which construct the initial stage. The Ver1.0 [4] proposed a system level design framework, which figures out the automatic design decisions for CPSS under single user environment. This design methodology's highlight is design of CPSS flow, the framework of Ver1.0 firstly proposed the data flow, physical flow and human flow. The Ver2.0 [148] is an extended model of Ver1.0. The Ver2.0 proposed social flow, which takes advantage of a directed graph to denote the social interactions among multiple users in stead of human flow in single user scenario of Ver1.0. Each state in the upper Petri net indicates control mark in social flow, which associates with a state of cyber actor or physical object in the bottom Petri net. Also, each state in the bottom Petri net corresponds to a control mark in data flow or physical flow. And the framework aims to fulfill synthesis, allocation, optimization and integration for the generic CPSS organization and architecture in terms of total cost, as well as performance, energy consumption, etc. The Ver3.0 [149] comprehensively considered security problem, energy consumption and user satisfaction. To ensure the elastic security service for CPSS, the Ver3.0 design a generic security model, focusing on confidentiality protection service for the messages of CPSS.

The evolutional stage. Compared with the initial stage, the evolutional stage introduces tensor element for CPSS data fusion framework design. Concretely, Ver4.0 [146] firstly proposed tensor to represent the CPSS data, the framework consists of five modules, Data Collection Module (DCM), Data Tensorization Module (DTM), Data Dimensionality Reduction Module (DDRM), Data Analysis Module (DAM), Data Service Module (DSM).

Based on the Ver4.0, the Ver5.0 [70] add a probability constraint on tensor based data representation and fusion. To be specific, Ver5.0 combines tensor and Multivariate Markov Chain (MMC) model to fusion the cyber, physical, social data. This provides a strict mathematical explanation for CPSS data fusion used in prediction systems. As an extension of Ver5.0, the Ver6.0 [71] proposed a multi space service framework, which combines multi MMC model with tensors. Specifically, we construct multi transform tensors responding to multi space. the Ver5.0 and Ver6.0 frameworks all consist of four modules, Inter-Sensing Module (ISM), Data Fusion Module (DFM), Rules Module (RM) and Application Module (AM). The architecture of Ver5.0 and Ver6.0 is shown as Fig. 13, whose highlight is the rule layer, which endow the framework with intelligence. The rules in our framework are cubic rules, which essentially is Eigentensors as described in Section 2.

The future stage. From Ver1.0 to Ver6.0, although some works about CPSS data fusion have been done from various aspects, there is a lack of a unified framework which comprehensively considers inter-space sensing, heterogeneous data representation and fusion, system security and privacy, energy consumption. Therefore, learning from the advantages of the above version, we propose a tensor based unified fusion framework for CPSS as Fig. 14.

The framework consists of 7 layers. The first layer is the data source which consists of three spaces, namely cyber space, physical space and social space. The second layer is perception layer, which includes not only physical/hard sensors but also soft sensors (humans and social networks). The third layer is data representation layer. Inspired by the work Ver1.0 and Ver2.0, we develop the idea of data flow, introduce tensors to represent CPSS data and propose physical space tensor flow, social space tensor flow, cyber space tensor flow. The fourth layer is fusion layer. The layer uses the tensor based unified fusion (TUF) model from [70,71,146] (Ver4.0, Ver5.0, Ver6.0) as described in Section 5 to fuse

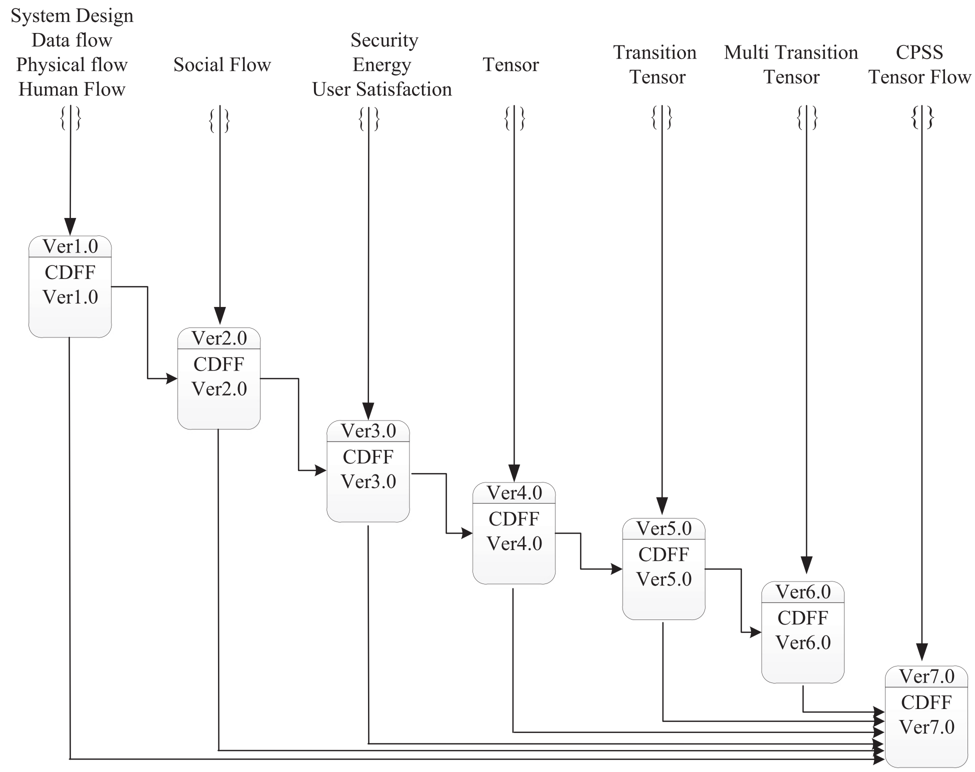


Fig. 12. The development of the CPSS data fusion framework. The CPSS data fusion framework design can be divided into three stages and seven versions. The initial stage is composed of Ver1.0 [4], Ver2.0 [148] and Ver3.0 [149]. The evolutional stage consists of Ver4.0 [146], Ver5.0 [70] and Ver6.0 [71]. The future stage is Ver7.0 which is firstly proposed in this paper.

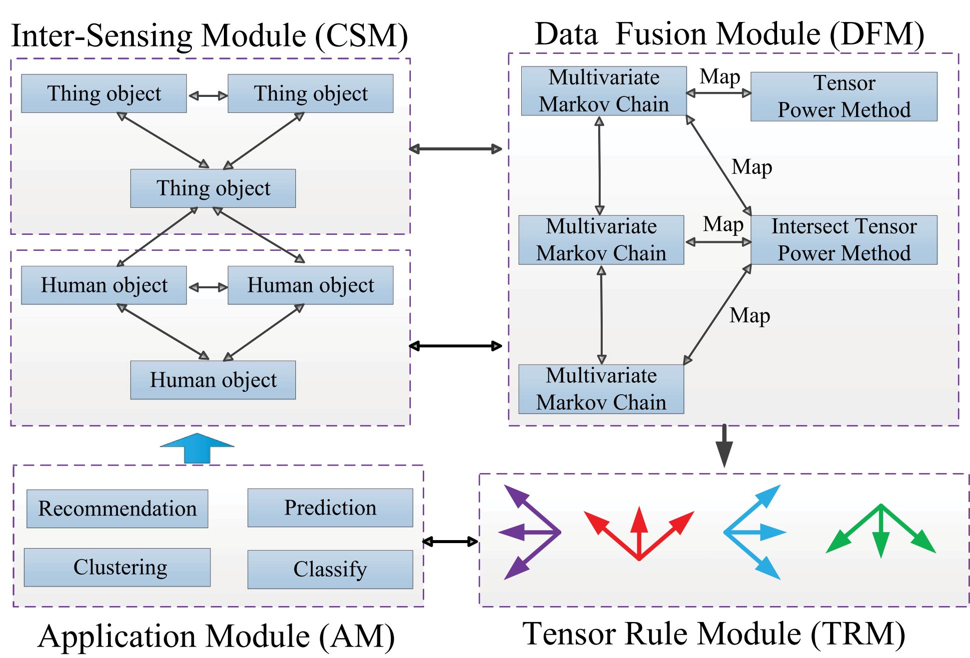


Fig. 13. The architecture of Ver5.0 and Ver6.0. Which consists of four modules, Inter-Sensing Module (ISM), Data Fusion Module (DFM), Rules Module (RM) and Application Module (AM).

CPSS tensor flows. The fifth layer is data processing layer. In this layer, some tensor decomposition (TD) algorithms are used for data mining or Knowledge Discovery in Database (KDD). Until now, our group has proposed some TD algorithms, tensor power method (TPM) [70], intersect tensor power method (ITPM) [71], high order orthogonal tensor singular value decomposition (HO-OTSVD). The sixth layer is the rule layer.

Our group has proposed a new concept about Eigentensor described in Section 2. In our proposed data fusion framework, we deem Eigentensors as cubic rules or tensor rules. Compared with traditional rule, tensor rule is a multi-model rule, it is not only a logical rule, but also gives the rule value. For example, a traditional rule is described as follows, if A occurs, then B occurs, $A \Rightarrow B$. A tensor rule $T_{rule} \in R^{human \times time \times location}$ de-

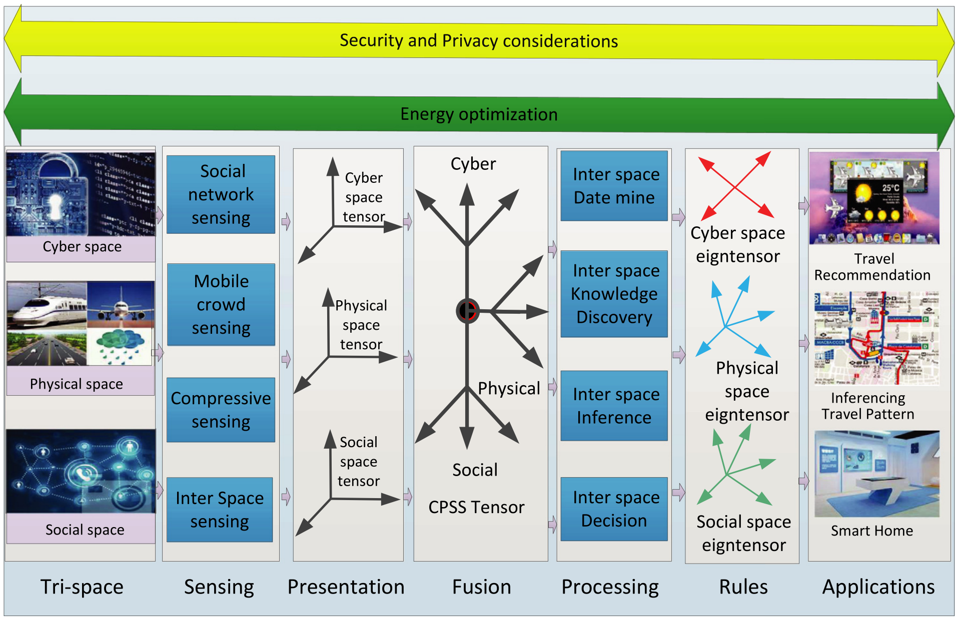


Fig. 14. The proposed tensor based Cyber-Physical-Social big data analysis framework design for CPSS. The framework consists of 7 layers, the first layer is the data source, the second layer is the perception layer, the third layer is the data representation layer, the fourth layer is the fusion layer, the fifth layer is the data processing layer, the sixth layer is the rule layer, the last layer is the application layer.

notes the cubical User-Spatio-Temporal probability distribution, which will be described in detail in Section 6.2. The tensor rules endow the framework with high intelligence. The last layer is the application layer, based on the rules from the sixth layer, this layer provides proactive and personalized services for humans.

6.2. Case study

A Case study of the proposed fusion method and framework are illustrated in this section as Fig. 15. The space data, time data and GPS trajectory data from CPSS are collected by various sensors, for example, i)physical sensors; ii)human with GPS devices, which always are called social sensors and the sensing paradigm always is named mobile crowd sensing, then represented as local tensors, namely cyber space tensors, physical space tensors, social space tensors. Among them, the data from social network are represented as a influence matrix (IM), namely Λ, the space data and time data are fused as a 4th-order tensor T_{ST}, then we used unified fusion (TUF) model to fuse the local tensors, namely, $A = \Lambda \odot T_{ST}$, then a 6th-order transition tensor $A \in R^{Uer \times Time \times Location \times Uer \times Time \times Location}$ is constructed, these orders respectively represent the user (U), location (*Location*), time slice (*Time*). Based on the 6th-order transition tensor, the Tensor Power Method (TPM) algorithm is used to solve eigentensor E as Fig. 16, the E represents the people travel rules in the spatio space, a tensor rule about the humans distribution in time and space can be described as $T_{rule} \in R^{human \times time \times location}$, which has three models, the first is users, the second is time, the third is locations, an element of $T_{rule} \in R^{human \times time \times location}$ denotes the probability of some human being at some POI in some time. This eigentensor reflects human's spatio-temporal probability distribution. In other word, the eigentensor represents the user mobility rules in spatio-temporal space.

According to the cubical User-Spatio-Temporal probability distribution ($Pr[user, time, location]$), we can predict the user mobility pattern. As shown in Fig. 17, we find that the prediction accuracy of tensor-based unified fusion model (Human-Time-Space model) is highest, followed by the Time-Space model and the Space-only model, the Time-only model is lowest. The results show that our tensor-based unified fusion model improves prediction accuracy significantly, this is because we use tensor to fuse the multi-aspects data from CPSS, which consider the comprehensive context factors and multi-mode associations of Cyber-Physical-Social space.

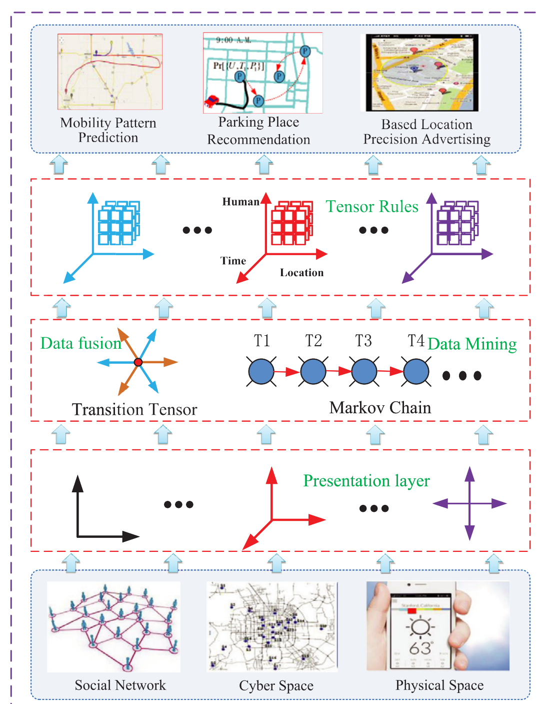


Fig. 15. A case study of the proposed fusion method and framework. The data from CPSS are represented as local tensors, then these tensors are fused with the proposed methods as a 6th-order transition tensor, which is combined with multivariant Markov Chain to complete the Knowledge Discovery (KDD). Then we get a cubical User-Spatio-Temporal probability distribution. According to the tensor rule, the human mobility pattern is predicted.

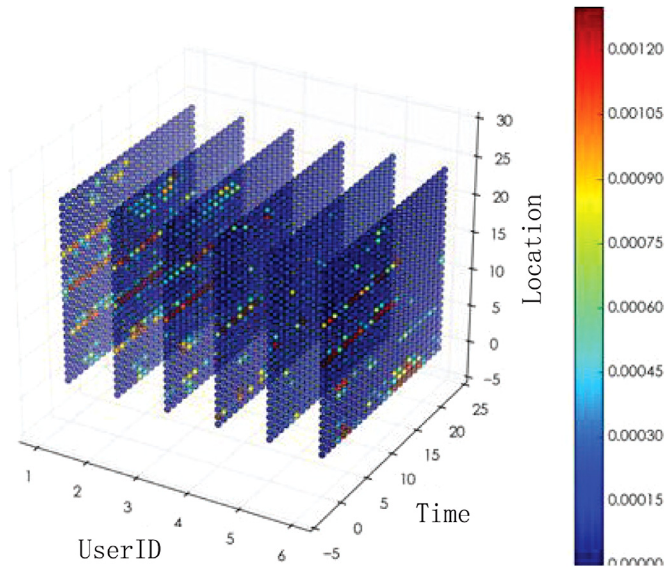


Fig. 16. The eigentensor E, which is a tensor rule $T_{rule} \in R^{human \times time \times location}$, the tensor rule has three models, the first is users, the second is time, the third is location, an element of $T_{rule} \in R^{human \times time \times location}$ denotes the probability of some human being at some POI in some time [70].

7. Future researches and challenges

In this section, we summarize future researches of CPSS data fusion, also some challenges are mainly analysed.

7.1. Security and privacy

Compared to traditional data analysis, CPSS data space has changed from single space to multi-space, it is bound to bring new security issues.

System security CPSS data are collected from various types of mobile phones, smart terminals, and wearable devices, which bring new threats (such as Trojan virus) to computing system [150]. Hence, how to avoid computing system suffering from the threat is a new challenge.

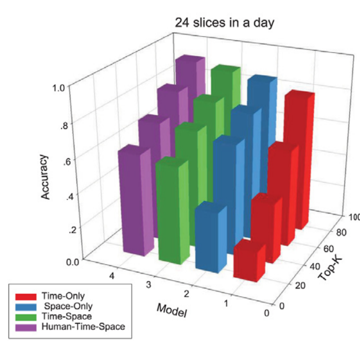
Privacy Inter-space data are out of data owners' control, therefore the access control is very difficult. Since one space is untrusted to another space, users' privacy becomes a key concern. Specifically, when data are outsourced to another space, how to securely and effectively access (e.g. query and search) and securely compute users' data become significant challenges in CPSS data fusion.

Another challenge is how to design privacy-aware architecture. To effectively preserve the privacy of a huge amount of CPSS participants, the privacy-aware fusion system is needed. One effort is the AnonySense architecture proposed in [151], which was focused on the privacy-aware applications in crowd sensing.

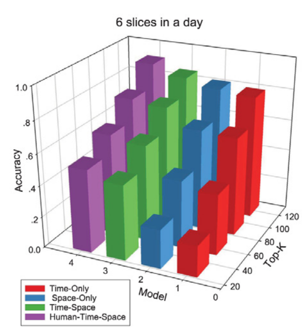
Transmission security In CPSS, more and more sensitive information are transmitted among Tri-space with Wireless Sensor Networks (WSNs) system. More and more researchers become to pay attention to transmission security. Some works have been studied for this problem [10,152], but these works mainly focus on traditional cyber attacks. It is necessary to consider authentication, key agreement, data confidentiality and data integrity protection in CPSS.

7.2. Dynamic parameters learning

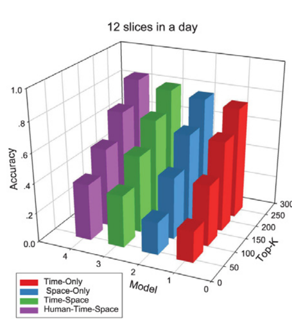
The CPSS data flows in large-scale application scenario always are dynamically evolving and need to be re-estimated on-the-fly. However, the CPSS data flows always are captured by fixed specification [4]. The



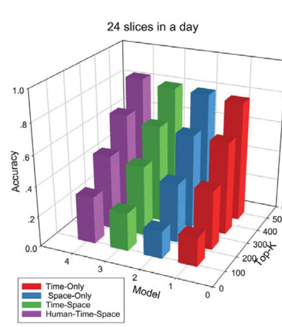
(a) Four slices in a day



(b) Six slices in a day



(c) Twelve slices in a day



(d) Twenty four slices in a day

Fig. 17. Comparison of prediction accuracy using various Top-k values and time slices in a day. The prediction accuracy of tensor-based unified fusion model (Human-Time-Space model) is highest, followed by the Time-Space model and the Space-only model, the Time-only model is lowest [70].

adaptation design of CPSS is a difficult work because environment parameters are not known in advance. The future work should aim at learning the dynamic parameters.

7.3. More intelligent paradigms

With the develop of Artificial Intelligence (AI), CPSS will step into a new smart state. How to endow CPSS with intelligence is a future work that we will focus on. The Ning et al. [11] have developed CPSS and proposed cyber-physical-social-thinking (CPST) hyperspace, this will pose new features in data fusion, it should aim at proposing new method for data fusion in CPST.

7.4. Energy

The energy consumption plays an important role in CPSS design. The development of next-generation smart energy systems required the innovative computational methodologies and models. How to design the energy-aware and data-driven CPSS has become a critical future work for the sustainable development [71,153].

8. Conclusions

CPSS data fusion involves a wide range of disciplines. In this paper, we firstly summarize the characteristics, purposes, advantages of CPSS data. A general definition of CPSS data fusion is proposed to clarify the concept of information fusion in CPSS. Then we review data representation and various data fusion methods, such as CPS data fusion and CSS data fusion, furthermore, a series of tensor-based data fusion methods are proposed for CPSS. Lastly, we survey CPSS data fusion architectures, and propose a tensor-based comprehensive CPSS data fusion framework. Also, the trends and challenges of CPSS data fusion are summarized and analyzed finally.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.inffus.2018.11.002.

References

[1] B. Guo, Z. Wang, Z. Yu, Y. Wang, N.Y. Yen, R. Huang, X. Zhou, Mobile crowd sensing and computing: the review of an emerging human-powered sensing paradigm, ACM Comput. Surv. 48 (1) (2015) 7.
[2] S. Misra, S. Goswami, C. Taneja, Multivariate data fusion-based learning of video content and service distribution for cyber physical social systems, IEEE Trans. Comput. Social Syst. 3 (1) (2016) 1–12, doi:10.1109/TCSS.2016.2561200.
[3] D. Zhang, D. Zhang, H. Xiong, L.T. Yang, V. Gauthier, Nextcell: predicting location using social interplay from cell phone traces, IEEE Trans. Comput. 64 (2) (2015) 452–463.
[4] J. Zeng, L.T. Yang, J. Ma, A system-Level modeling and design for cyber-Physical–Social systems, ACM Trans. Embedded Comput. Syst. 15 (2) (2016) 1–26.
[5] F.Y. Wang, The emergence of intelligent enterprises: from CPS to CPSS, IEEE Intell. Syst. 25 (4) (2010) 85–88.
[6] A. Sheth, P. Anantharam, C. Henson, Physical-cyber-social computing: an early 21st century approach, IEEE Intell. Syst. 28 (1) (2013) 78–82.
[7] J. Ma, L.T. Yang, B.O. Apduhan, R. Huang, L. Barolli, M. Takizawa, Towards a smart world and ubiquitous intelligence: a walkthrough from smart things to smart hyperspaces and ubickids, Int. J. Pervasive Comput. Commun. 1 (1) (2005) 53–68.
[8] X. Wang, L.T. Yang, H. Liu, M.J. Deen, A big data-as-a-service framework: state-of-the-art and perspectives, IEEE Trans. Big Data (2017), doi:10.1109/TBDATA.2017.2757942.
[9] L.T. Yang, X. Wang, X. Chen, J. Han, J. Feng, A tensor computation and optimization model for cyber-physical-social big data, IEEE Trans. Sust. Comput (2017), doi:10.1109/TSUSC.2017.2777503.
[10] J. Feng, L.T. Yang, G. Dai, W. Wang, D. Zou, A secure high-Order-Lanczos based orthogonal tensor SVD for big data reduction in cloud environment, IEEE Trans. Big Data (2018), doi:10.1109/TBDATA.2018.2803841.
[11] H. Ning, H. Liu, J. Ma, L.T. Yang, R. Huang, Cybermatics: cyber-physical-social–thinking hyperspace based science and technology, Future Gen. Comput. Syst. 56 (2015) 504–522.
[12] A. Iera, G. Morabito, L. Atzori, The social internet of things, in: Proceedings of the IEEE International Conference on Cloud Engineering, 2015. 1–1
[13] L. Atzori, A. Iera, G. Morabito, M. Nitti, The social internet of things (siot) - when social networks meet the internet of things: concept, architecture and network characterization, Comput. Netw. 56 (16) (2012) 3594–3608.
[14] L. Atzori, D. Carboni, A. Iera, Smart things in the social loop: paradigms, technologies, and potentials, Ad. Hoc. Netw. 18 (7) (2014) 121–132.
[15] Z. Liu, D.S. Yang, D. Wen, W.M. Zhang, W. Mao, Cyber-physical-social systems for command and control, IEEE Intell. Syst. 26 (4) (2011) 92–96.
[16] D. Lahat, T. Adali, C. Jutten, Multimodal data fusion: An overview of methods, challenges and prospects, in: Proceedings of the IEEE, 2015, pp. 1449–1477.
[17] P. Jiang, J. Winkley, C. Zhao, R. Munnoch, G. Min, L.T. Yang, An intelligent information forwarder for healthcare big data systems with distributed wearable sensors, IEEE Syst. J. 10 (3) (2016) 1147–1159.
[18] X. Wang, L.T. Yang, J. Feng, X. Chen, M.J. Deen, A tensor-based big service framework for enhanced living environments, IEEE Cloud Comput. Mag. 3 (6) (2016) 36–43.
[19] S. Helal, W. Mann, H. Elzabadani, J. King, Y. Kaddoura, E. Jansen, The gator tech smart house: a programmable pervasive space, Computer (Long Beach Calif) 38 (3) (2005) 50–60.
[20] D.J. Cook, A.S. Crandall, B.L. Thomas, N.C. Krishnan, Casas: a smart home in a box, Computer (Long Beach Calif) 46 (7) (2013) 62–69.
[21] J.A. Kientz, S.N. Patel, B. Jones, E. Price, E.D. Mynatt, G.D. Abowd, The georgia tech aware home., in: Extended Abstracts Proceedings of the 2008 Conference on Human Factors in Computing Systems, 2008, pp. 3675–3680.
[22] Y.W. Kao, S.M. Yuan, User-configurable semantic home automation, Comput. Standards Interfaces 34 (1) (2012) 171–188.
[23] L. Cui, F.R. Yu, Q. Yan, When big data meets software-defined networking: SDN for big data and big data for SDN, IEEE Netw. 30 (1) (2016) 58–65.
[24] A. Mercian, M.P. Mcgarry, M. Reisslein, W. Kellerer, Software-defined networking: a comprehensive survey, Proc. IEEE 103 (1) (2015) 14–76.
[25] D. Kreutz, F.M.V. Ramos, P. Esteves Verissimo, C. Esteve Rothenberg, S. Azodolmolky, S. Uhlig, Software-defined networking: a comprehensive survey, Proc. IEEE 103 (1) (2014) 10–13.
[26] L. Jun, L. Tingting, C. Gang, Y. Hua, Mining and modelling the dynamic patterns of service providers in cellular data network based on big data analysis, Wireless Commun. Over Zigbee Automotive Inclination Measur. China Commun. 10 (12) (2013) 25–36.
[27] L. Kuang, L.T. Yang, X. Wang, P. Wang, A tensor-based big data model for QoS improvement in software defined networks, IEEE Netw. 30 (1) (2016) 30–35.
[28] V. Bychkovsky, K. Chen, M. Goraczko, H. Hu, B. Hull, A. Miu, E. Shih, Y. Zhang, H. Balakrishnan, S. Madden, The cartel mobile sensor computing system, in: Proceedings of the International Conference on Embedded Networked Sensor Systems, 2006, pp. 383–384.
[29] P. Zhou, Y. Zheng, M. Li, How long to wait?: predicting bus arrival time with mobile phone based participatory sensing, in: Proceedings of the International Conference on Mobile Systems, Applications, and Services, 2014, pp. 379–392.
[30] J. Yuan, Y. Zheng, X. Xie, G. Sun, T-Drive: enhancing driving directions with taxi drivers' intelligence, IEEE Trans. Knowl. Data Eng. 25 (1) (2013) 220–232.
[31] T. Ernst, K. Uehara, K. Mitsuya, Network mobility from the internetcar perspective, in: Proceedings of the International Conference on Advanced Information Networking and Applications, 2003, pp. 19–25.
[32] K.I. Farkas, J. Heidemann, L. Iftode, Guest editors' introduction: intelligent transportation and pervasive computing, IEEE Pervasive Comput. 5 (4) (2006) 18–19.
[33] M. Morris, J. Lundell, E. Dishman, B. Needham, New perspectives on ubiquitous computing from ethnographic study of elders with cognitive decline, Lect. Notes Comput. Sci. 2864 (2003) 227–242.
[34] T. Adlam, R. Faulkner, R. Orpwood, K. Jones, J. Macijauskiene, A. Budraitiene, The installation and support of internationally distributed equipment for people with dementia., IEEE Trans. Inf. Technol. Biomed. 8 (3) (2004) 253–257.
[35] M. Alwan, S. Dalal, D. Mack, S.W. Kell, B. Turner, J. Leachtenauer, R. Felder, Impact of monitoring technology in assisted living: outcome pilot, IEEE Trans. Inf. Technol. Biomed. 10 (1) (2006) 192–198.
[36] F. Vergari, S. Bartolini, F. Spadini, A. D'Elia, A smart space application to dynamically relate medical and environmental information, in: Proceedings of Design, Automation & Test in Europe Conference & Exhibition, 2010, pp. 1542–1547.
[37] A.T. Campbell, S.B. Eisenman, N.D. Lane, E. Miluzzo, The rise of people-centric sensing, IEEE Internet Comput. 12 (4) (2008) 12–21.
[38] Y. Wang, L. Wei, A.V. Vasilakos, Q. Jin, Device-to-device based mobile social networking in proximity (msnp) on smartphones: framework, challenges and prototype, Future Gen. Comput. Syst. 74 (2017) 241–253.
[39] B. Pan, Y. Zheng, D. Wilkie, C. Shahabi, Crowd sensing of traffic anomalies based on human mobility and social media, in: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2013, pp. 344–353.
[40] D. Yao, C. Yu, H. Jin, Q. Ding, Human mobility synthesis using matrix and tensor factorizations, Inf. Fusion 23 (2015) 25–32.
[41] R.P. Mahler, Statistical multisource-multitarget information fusion, Artech House, 2007.
[42] S. Yadav, G. Shroff, E. Hassan, P. Agarwal, Business data fusion, in: Proceedings of International Conference on Information Fusion, 2015, pp. 1876–1885.
[43] M.A. Sloane, Method and apparatus for promoting products and influencing consumer purchasing decisions at the point-of-purchase, 1999. US Patent 5,918,211.
[44] R.A. Nieto, Data fusion for automated business decisions, 2009. US Patent 7,580,878.
[45] Y. Wang, W. Dai, Q. Jin, J. Ma, Bcinet: a biased contest-based crowdsourcing incentive mechanism through exploiting social networks, IEEE Trans. Systems, Man, and Cybernetics (2018).

[46] Y. Zheng, H. Zhang, Y. Yu, Detecting collective anomalies from multiple spatio-temporal datasets across different domains, in: Proceedings of the Sigspatial International Conference on Advances in Geographic Information Systems, 2015, pp. 1–10.
[47] Y. Zheng, Methodologies for cross-domain data fusion: an overview, IEEE Trans. Big Data 1 (1) (2015) 16–34.
[48] L. Chen, J. Jakubowicz, D. Yang, D. Zhang, G. Pan, Fine-grained urban event detection and characterization based on tensor cofactorization, IEEE Trans. Hum. Mach. Syst. 47 (3) (2017) 380–391.
[49] Y. Wang, W. Dai, B. Zhang, J. Ma, A.V. Vasilakos, Word of mouth mobile crowdsourcing: increasing awareness of physical, cyber, and social interactions, IEEE Multimedia 24 (4) (2017) 26–37.
[50] S. Kuznetsov, E. Paulos, Participatory sensing in public spaces: activating urban surfaces with sensor probes, in: Proceedings of the ACM Conference on Designing Interactive Systems, 2010, pp. 21–30.
[51] A. Parker, S. Reddy, T. Schmid, K. Chang, G. Saurabh, M. Srivastava, M. Hansen, J. Burke, D. Estrin, M. Allman, Network system challenges in selective sharing and verification for personal, social, and urban-scale sensing applications, in: Proceedings of the 5th Workshop Hot Topic in Networks(HotNet-V), 2006.
[52] E. Paulos, T. Jenkins, Urban probes: encountering our emerging urban atmospheres, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2005, pp. 341–350.
[53] R.N. Murty, A. Gosain, M. Tierney, A. Brody, A. Fahad, J. Bers, M. Welsh, Citysense: a vision for an urban-scale wireless networking testbed, Technical Report TR-13-07 Harvard University (2007).
[54] B. Khaleghi, A. Khamis, F.O. Karray, S.N. Razavi, Multisensor data fusion: a review of the state-of-the-art, Inf. Fusion 14 (1) (2013) 28–44.
[55] I. Bloch, Information combination operators for data fusion: a comparative review with classification, IEEE Trans. Syst. Man Cybern. 26 (1) (1996) 52–67.
[56] D.L. Hall, J. Llinas, An introduction to multisensor data fusion, Proc. IEEE 85 (1) (1997) 6–23.
[57] D. Smith, S. Singh, Approaches to multisensor data fusion in target tracking: a survey, IEEE Trans. Knowl. Data Eng. 18 (12) (2006) 1696–1710.
[58] A. Ardeshir Goshtasby, S. Nikolov, Image fusion: advances in the state of the art, Inf. Fusion 8 (2) (2007) 114–118.
[59] Y. Wang, A.V. Vasilakos, Q. Jin, J. Ma, Pprank: economically selecting initial users for influence maximization in social networks, IEEE Syst. J. 11 (4) (2017) 2279–2290.
[60] I. Corona, G. Giacinto, C. Mazzariello, F. Roli, C. Sansone, Information fusion for computer security: state of the art and open issues, Inf. Fusion 10 (4) (2009) 274–284.
[61] H. Wache, T. Voegele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, S. Hübner, Ontology-based integration of information-a survey of existing approaches, in: IJCAI-01 Workshop: Ontologies and Information Sharing, 2001, pp. 108–117.
[62] G.L. Rogova, V. Nimier, Reliability in information fusion: Literature survey, in: Proceedings of the 7th International Conference on Information Fusion, 2004, pp. 1158–1165.
[63] J.T. Yao, V.V. Raghavan, Z. Wu, Web information fusion: a review of the state of the art, Inf. Fusion 9 (4) (2008) 446–449.
[64] L. Sorber, Data fusion: tensor factorizations by complex optimization, Ph.D. thesis, Faculty of Engineering, KU Leuven, Leuven, Belgium (May 2014).
[65] Y. Wang, X. Jia, Q. Jin, J. Ma, Mobile crowdsourcing: framework, challenges, and solutions, Concurrency and Computation: Practice and experience 29 (3) (2017) e3789.
[66] L. Qi, Eigenvalues of a real supersymmetric tensor, J Symb. Comput. 40 (6) (2008) 1302–1324.
[67] L.H. Lim, Singular values and eigenvalues of tensors: a variational approach, in: Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2006, pp. 129–132.
[68] X. Li, L. Yan, N. Qin, H. Ran, A novel semi-supervised short text classification algorithm based on fusion similarity, in: Proceedings of International Conference on Intelligent Computing, 2017, pp. 309–319.
[69] Q. Zhang, L.T. Yang, Z. Chen, P. Li, An improved deep computation model based on canonical polyadic decomposition, IEEE Trans. Syst. Man Cybern. Syst. PP (99) (2017) 1–10.
[70] P. Wang, L.T. Yang, Y. Peng, J. Li, X. Xie, M^2T^2:The multivariate multi-step transition tensor for user mobility pattern prediction, IEEE Trans. Network Sci. Eng., Accepted (2018).
[71] P. Wang, L.T. Yang, G. Qian, F. Lu, The cyber-physical-social transition tensor service framework, IEEE Transactions on Sustainable Computing (2018), doi:10.1109/TSUSC.2018.2799940.
[72] F.E. White, Data fusion lexicon, The Data Fusion Lexicon Subpanel of the Joint Directors of Laboratories (1991).
[73] L.A. Klein, Sensor and data fusion concepts and applications, SPIE Optical Eng. press, 1999.
[74] H. Boström, S.F. Andler, M. Brohede, R. Johansson, A. Karlsson, J.V. Laere, L. Niklasson, M. Nilsson, A. Persson, T. Ziemke, On the definition of information fusion as a field of research, Technical Report HS-IKI-TR-07-006,University of skövde (2007).
[75] I.V. Mechelen, A.K. Smilde, A generic linked-mode decomposition model for data fusion, Chemometrics and Intelligent Laboratory Systems 104 (1) (2010) 83–94.
[76] V.D. Calhoun, T. Adali, Feature-based fusion of medical imaging data, IEEE Trans. Inf. Technol. Biomedicine A Publ. IEEE Eng. Medicine Biol. Soc. 13 (5) (2009) 711–720.
[77] E. Blasch, J. Llinas, D. Lambert, P. Valin, S. Das, C. Chong, M. Kokar, E. Shahbazian, High level information fusion developments, issues, and grand challenges: Fusion 2010 panel discussion, in: Proceedings of International Conference on Information Fusion, 2011, pp. 1–8.
[78] E.P. Blasch, R. Breton, P. Valin, E. Bosse, User information fusion decision making analysis with the C-OODA model, in: Proceedings of the International Conference on Information Fusion, 2011.
[79] A. Jones, E. Subrahmanian, A. Hamins, C. Grant, Humans' Critical role in smart systems: a smart firefighting example, IEEE Internet Comput. 19 (3) (2015) 28–31.
[80] K.L. Huang, S.S. Kanhere, W. Hu, Are you contributing trustworthy data?: the case for a reputation system in participatory sensing, in: Proceedings of the International Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems, 2010, pp. 14–22.
[81] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[82] S. Vigneshwari, M. Aramudhan, Social information retrieval based on semantic annotation and hashing upon the multiple ontologies, Indian J. Sci. Technol. 8 (2) (2015) 103–107.
[83] T.D. Cao, T.H. Phan, A.D. Nguyen, An ontology based approach to data representation and information search in smart tourist guide system, in: Proceedings of International Conference on Knowledge & Systems Engineering, 2011, pp. 171–175.
[84] S. Hachem, T. Teixeira, V. Issarny, Ontologies for the internet of things, in: Proceedings of the 8th Middleware Doctoral Symposium, 2011, pp. 1–6.
[85] R. Stegers, P. Fekkes, H. Stuckenschmidt, Web semantics: Science, services and agents on the world wide web, Elsevier Science Publishers B. V., 2005.
[86] M. Signoretto, L.L. De, J.A. Suykens, A kernel-based framework to tensorial data analysis., Neural Networks the Official Journal of the International Neural Network Society 24 (8) (2011) 861–874.
[87] T.G. Kolda, B.W. Bader, J.P. Kenny, Higher-order web link analysis using multilinear algebra, in: Proceedings of the 5th IEEE International Conference on Data Mining, 2005, pp. 242–249.
[88] I. Kotsia, I. Patras, Support tucker machines, in: Proceedings of the Computer Vision and Pattern Recognition, 2011, pp. 633–640.
[89] F. Garcia, B. Mirbach, B. Ottersten, F. Grandidier, ngel Cuesta, Pixel weighted average strategy for depth sensor data fusion, in: IEEE International Conference on Image Processing, 2010, pp. 2805–2808.
[90] R.E. Kalman, A new approach to linear filtering and prediction problems, J. Basic Eng. Trans. 82 (1960) 35–45.
[91] Y. Barshalom, X.R. Li, Estimation and tracking: Principles, techniques, and software, Artech House, 1993.
[92] D. Willner, C.B. Chang, K.P. Dunn, Kalman filter algorithms for a multi-sensor system, in: Proceedings of IEEE International Conference on Decision and Control, 1976, pp. 570–574.
[93] B.P. Carlin, T.A. Louis, Bayes and Empirical Bayes Methods for Data Analysis, Chapman & Hall/CRC, 1998.
[94] G.J. Klir, B. Yuan, Fuzzy sets and fuzzy logic: Theory and applications, Prentice-Hall, Inc., 1994.
[95] P. Manjunatha, A.K. Verma, A. Srividya, Multi-sensor data fusion in cluster based wireless sensor networks using fuzzy logic method, in: Proceedings of the 3th IEEE International Conference on Industrial and Information Systems, 2009, pp. 1–6.
[96] W. Su, T.C. Bougiouklis, Data fusion algorithms in cluster-basedwireless sensor networks using fuzzy logic theory, in: Proceedings of the 11th WSEAS International Conference on Communications, 2007.
[97] K. Kapitanova, H.S. Sang, K.D. Kang, Using fuzzy logic for robust event detection in wireless sensor networks, Ad. Hoc. Netw. 10 (4) (2012) 709–722.
[98] Z. Pawlak, R. SoWinski, Rough set approach to multi-attribute decision analysis, Eur. J. Oper. Res. 72 (3) (1994) 443–459.
[99] H. Wang, Y. Chen, Sensor data fusion using rough set for mobile robots system, in: Proceedings of the IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications, 2006, pp. 1–5.
[100] S.F. Jiang, C.M. Zhang, J. Yao, Eigen-level data fusion model by integrating rough set and probabilistic neural network for structural damage detection, Adv. Struct. Eng. 14 (2) (2011) 333–350.
[101] L. Xu, Y. Chen, P. Cui, Improvement of D-S evidential theory in multisensor data fusion system, in: Proceedings of the 6th World Congress on Intelligent Control and Automation, 4, 2004, pp. 3124–3128.
[102] R.R. Yager, On the dempster-shafer framework and new combination rules, Inf. Sci. (Ny) 41 (2) (1987) 93–137.
[103] O. Basir, X. Yuan, Engine fault diagnosis based on multi-sensor information fusion using dempstercshafer evidence theory, Inf. Fusion 8 (4) (2007) 379–386.
[104] W. Wu, Y. Leung, W. Zhang, Connections between rough set theory and dempster-shafer theory of evidence, Int. J. Gen Syst 31 (4) (2002) 405–430.
[105] A. Skowron, Grzymal, J. A-Busse, From rough set theory to evidence theory, John Wiley & Sons, Inc., 1994.
[106] J.W. Guan, D.A. Bell, Evidence theory and its applications, vol. 2., Elsevier Science Inc., 1991.
[107] T.L. Chen, P.W. Que, Research on invalidation of d-s evidential theory in data fusion, J. Transducer Technol. 23 (12) (2004) 25–27.
[108] Stoyan, Dietrich, Stochastic Geometry and its Applications, Wiley, 1987.
[109] R. Mahler, Random sets: Unification and computation for information fusion - a retrospective assessment, in: Proceedings of 7th International Conference on Information Fusion, 2004, pp. 1–20.
[110] R.P.S. Mahler, Random sets in information fusion. an overview, Springer New York, 1997.

[111] R.P. Mahler, Representing rules as random sets, i: statistical correlations between rules, Inf. Sci. (Ny) 88 (1–4) (1996) 47–68.
[112] D.L. Hall, M. Mcneese, J. Llinas, T. Mullen, A framework for dynamic hard/soft fusion, in: Proceedings of the International Conference on Information Fusion, 2008, pp. 1–8.
[113] D.L. Hall, M.D. Mcneese, D.B. Hellar, B.J. Panulla, W. Shumaker, A cyber infrastructure for evaluating the performance of human centered fusion, in: Proceedings of the International Conference on Information Fusion, 2009, pp. 1257–1264.
[114] M. Prentice, M. Kandefer, S.C. Shapiro, Tractor: A framework for soft information fusion, in: Proceedings of the 13th International Conference on Information Fusion, 2010, pp. 1–8.
[115] R.R. Yager, A framework for reasoning with soft information, Inf. Sci. (Ny) 180 (8) (2010) 1390–1406.
[116] J. Llinas, Application of the jdl data fusion process model to hard/soft information fusion in the condition monitoring of aircraft, in: SPIE Defense,Security Sensing Conference, 2014.
[117] J.T. Bernardo, Cognitive and functional frameworks for hard/soft fusion for the condition monitoring of aircraft, in: Proceedings of the 18th International Conference on Information Fusion (Fusion), 2015, pp. 287–294.
[118] S.C. Shapiro, D.R. Schlegel, Natural language understanding for soft information fusion, Springer International Publishing, 2016.
[119] J.T. Bernardo, A methodology for hard/soft information fusion in the condition monitoring of aircraft, SPIE Def. Secur. Sens. 8756 (2013) 701–704.
[120] M. Prentice, S.C. Shapiro, Using propositional graphs for soft information fusion, in: Proceedings of the 14th International Conference on Information Fusion, 2011, pp. 1–7.
[121] L.I. Smith, A tutorial on principal components analysis, Inf. Fusion 51 (3) (2002) 52.
[122] C. Canali, S. Casolari, R. Lancellotti, A quantitative methodology to identify relevant users in social networks, in: IEEE International Workshop on Business Applications of Social Network Analysis, 2011, pp. 1–8.
[123] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.
[124] D.D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, in: International Conference on Neural Information Processing Systems, 2000, pp. 535–541.
[125] S.E. Palmer, Hierarchical structure in perceptual representation, Cogn. Psychol. 9 (4) (1977) 441–474.
[126] A. Saha, V. Sindhwani, Learning evolving and emerging topics in social media: a dynamic nmf approach with temporal regularization, in: Proceedings of the ACM International Conference on Web Search and Data Mining, 2012, pp. 693–702.
[127] K. Matsutani, M. Kumano, M. Kimura, K. Saito, K. Ohara, H. Motoda, Combining activity-evaluation information with nmf for trust-link prediction in social media, in: IEEE International Conference on Big Data, 2015, pp. 2263–2272.
[128] Z. Yang, Q. Li, W. Liu, Y. Ma, M. Cheng, Dual graph regularized nmf model for social event detection from flickr data, World Wide Web-internet Web Inf. Syst. 20 (5) (2017) 995–1015.
[129] D. Zhang, F. Wang, L. Si, Composite hashing with multiple information sources, in: Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2011, Beijing, China, July, 2011, pp. 225–234.
[130] M.M. Bronstein, A.M. Bronstein, F. Michel, N. Paragios, Data fusion through cross–modality metric learning using similarity-sensitive hashing, in: Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, 2010, pp. 3594–3601.
[131] Y. Zhen, D.Y. Yeung, A probabilistic model for multimodal hash function learning, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012, pp. 940–948.
[132] Y. Liu, J. Shao, J. Xiao, F. Wu, Y. Zhuang, Hypergraph spectral hashing for image retrieval with heterogeneous social contexts, Neurocomputing 119 (16) (2013) 49–58.
[133] A. Slutsky, X. Hu, Y. An, Hash-based stream lda: Topic modeling in social streams, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2014, pp. 151–162.
[134] Z. Wu, M. Zou, 2014 Special issue: an incremental community detection method for social tagging systems using locality-sensitive hashing, Neural Networks Official J. Int. Neural Network Soc. 58 (10) (2014) 14–28.
[135] A.M. Shalita, B. Karrer, I. Kabiljo, A.D. Sharma, A. Presta, A. Adcock, H. Kllapi, M. Stumm, Social hash: an assignment framework for optimizing distributed systems operations on social networks, in: Proceedings of Networked Systems Design and Implementation, 2016, pp. 455–468.
[136] W. Zhang, L.I. Rui-Shan, X.Y. Gao, Poi-fusion classification based on similarity model, Periodical of Ocean University of China 9 (2014) 014.
[137] W. Xi, E.A. Fox, W. Fan, B. Zhang, Z. Chen, J. Yan, D. Zhuang, Simfusion: measuring similarity using unified relationship matrix, in: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2005, pp. 130–137.
[138] H. Hotelling, Relations between two sets of variates, Biometrika 28 (3/4) (1936) 321–377.
[139] J.R. Kettenring, Canonical analysis of several sets of variables, Biometrika 58 (3) (1971) 433–451.
[140] X. Zhu, Z. Huang, H.T. Shen, J. Cheng, C. Xu, Dimensionality reduction by mixed kernel canonical correlation analysis, Pattern Recognit. 45 (8) (2012) 3003–3016.
[141] L. Cen, W. Ser, Z.L. Yu, Speech emotion recognition using canonical correlation analysis and probabilistic neural network, in: Proceedings of International Conference on Machine Learning and Applications, 2008, pp. 859–862.
[142] H. Zhang, X. Liu, Cross-Media semantics mining based on sparse canonical correlation analysis and relevance feedback, Springer Berlin Heidelberg, 2012.
[143] A.P. Singh, G.J. Gordon, Relational learning via collective matrix factorization, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008, pp. 650–658.
[144] C. He, H. Li, X. Fei, Y. Tang, J. Zhu, A topic community-based method for friend recommendation in online social networks via joint nonnegative matrix factorization, in: Proceedings of Third International Conference on Advanced Cloud and Big Data, 2016, pp. 28–35.
[145] P. Cui, F. Wang, L.F. Sun, S.Q. Yang, A joint matrix factorization approach to unsupervised action categorization, in: Proceedings of Eighth IEEE International Conference on Data Mining, 2008, pp. 767–772.
[146] L. Kuang, F. Hao, L.T. Yang, M. Lin, A tensor-based approach for big data representation and dimensionality reduction, IEEE Trans. Emerg. Top Comput. 2 (3) (2014) 280–291.
[147] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, SIAM Rev. 51 (3) (2009) 455–500.
[148] J. Zeng, L.T. Yang, J. Ma, M. Guo, Hyperspaceflow: a system-level design methodology for smart space, IEEE Trans. Emerg. Top. Comput. 4 (4) (2016) 568–583.
[149] J. Zeng, L.T. Yang, M. Lin, Z. Shao, D. Zhu, System-level design optimization for security-critical cyber-physical-social systems, ACM Trans. Embedded Comput. Syst. 16 (2) (2017) 39.
[150] M. Kefayati, M.S. Talebi, H.R. Rabiee, B.H. Khalaj, On secure consensus information fusion over sensor networks, in: Proceedings of the IEEE/ACS International Conference on Computer Systems and Applications, 2007, pp. 108–115.
[151] C. Cornelius, A. Kapadia, D. Kotz, D. Peebles, M. Shin, N. Triandopoulos, Anonysense: privacy-aware people-centric sensing, in: Proceedings of the International Conference on Mobile Systems, Applications, and Services, 2008, pp. 211–224.
[152] S. Roy, M. Conti, S. Setia, S. Jajodia, Secure data aggregation in wireless sensor networks, IEEE Transs. Inf. Forensics Security 7 (3) (2008) 1040–1052.
[153] P. Wang, L.T. Yang, J. Li, An edge cloud-assisted cpss framework for smart city, IEEE Cloud Comput. 5 (5) (2018) 37–46.