IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ADVANCED DATA ANALYTICS FOR LARGE-SCALE COMPLEX DATA ENVIRONMENTS

Date of current version March 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2895440

EDITORIAL

IEEE ACCESS SPECIAL SECTION EDITORIAL: ADVANCED DATA ANALYTICS FOR LARGE-SCALE COMPLEX DATA ENVIRONMENTS

Big Data is defined as an emerging paradigm that includes complex and large-scale information beyond the processing capability of conventional tools. Traditional data analytics methods have been commonly used for many applications, such as text classification, image recognition, and video tracking. For analysis purposes, these data often need to be represented as vectors. However, many other types of data objects in real-world applications contain rich feature vectors and structure information, such as chemical compounds in bio-pharmacy, brain regions in brain networks and users in social networks. Unfortunately, vector representations are very simple features that do not inherently contain the object's structure information. In reality, objects may have complicated characteristics depending on how the objects are assessed and characterized. Data may also reside in heterogeneous domains, such as traditional tabular-based data, sequential patterns, social networks, time series information, and semi-structured data. As a result, novel data analytics methods are needed to discover meaningful knowledge in advanced applications from objects with large-scale complex characteristics. In total, there were 104 submissions to this IEEE ACCESS Special Section, with 39 accepted after rigorous peer review.

The work by Han *et al.* "An effective approach for rock mass discontinuity extraction based on terrestrial LiDAR scanning 3D point clouds," presented an effective approach for rock mass discontinuity extraction based on the terrestrial LiDAR scanning 3-D point clouds. First, a Quadtree-Octree index method that retained the adjacency relationship of rock point clouds was proposed to organize point cloud data in a high-efficiency manner. Second, the normal vector (NV) calculation was directly conducted by the local triangulated irregular networks together with eight neighborhood areas. Third, a double-clustering strategy was developed based on the Quadtree-Octree index and NV calculation discussed in step one and two. The first clustering was conducted in the point cloud matrix at each station, whereas the second clustering was among multiple stations. Fourth, an extended random sample consensus algorithm was designed with two separate checks (distance and angle) to detect whether the compliance with certain constraints occurred between rock point clouds and fitting planes. The proposed method was evaluated by a real field data set in China and a public data set from Rockbench. The feasibility of the proposed method was verified by these two data sets, which indicated a promising perspective for the field engineering survey.

GNSS multi-frequency multi-system carrier phase differential positioning has become the main technology used in high-precision positioning. Until recently, the fault detection and exclusion (FDE) methods for multi-frequency multi-system carrier phase differential positioning mostly focus on processing of errors in the carrier phase domain, which cannot exclude all the faults, causing a faulted baseline resolution, e.g., a fault that occurs in resolution process. Besides, the multi-fault of multi-frequency in the carrier phase domain cannot be identified due to the multi-frequency carrier phase observation errors' high correlation. Liang *et al.* in "GNSS multi-frequency multi-system highly robust differential positioning based on an autonomous fault detection and exclusion method," presented a method of autonomous FDE based on multi-frequency multi-system carrier phase differential positioning. It focused on the procession of errors in the position domain, and detected and excluded the faults in different frequency baseline resolutions caused not only by measurement fault as in the traditional method, but also the resolution fault, which can enhance the robustness and accuracy of the differential positioning system. The experimental results showed that the method can effectively detect and exclude the failure of different frequency baseline resolutions and then the accurate multi-frequency multi-system positioning results can subsequently be effectively fused. The proposed method improved the accuracy and robustness of the differential positioning system.

There are a lot of mixed pixels in the remotely sensed imagery, which can seriously limit the utility of classification. Sub-pixel mapping (SPM) is a promising technique to solve

this problem. It can generate a fine resolution land cover map from coarse resolution fractional images, by predicting the spatial locations of different land cover classes at sub-pixel scale. However, the accuracy and detail are always limited. When the scale factor is large among sub-pixels per pixel, the data volumes are amplified and the sub-pixel distribution becomes complex. The traditional methods are carried out only by the fractions of land cover and the spatial dependence theory, which cannot satisfy the requirement of the SPM. To avoid this flaw, a new SPM method based on maximum a posteriori (MAP) model with subpixel/pixel spatial attraction theory aimed at the largescale factor was proposed by Wu *et al.* ''Sub-pixel mapping based on map model and spatial attraction theory for remotely sensed image''. First, MAP was proposed to improve the resolution of the fractional images and to obtain the initial sub-pixel locations; after that, the pixel swapping algorithm was used to optimize and produce the final SPM result.

Ma *et al.* in ''Robust image feature matching via progressive sparse spatial consensus,'' proposed an efficient algorithm, termed as progressive sparse spatial consensus, for mismatch removal from a set of putative feature correspondences involving a large number of outliers. The goal was to estimate the underlying spatial consensus between the feature correspondences and then remove mismatches accordingly. This is formulated as a maximum likelihood estimation problem and solved by an iterative expectation-maximization algorithm. To handle the large number of outliers, the authors introduced a progressive framework, which used matching results on a small putative set with high inlier ratio to guide the matching on a large putative set. The spatial consensus was modeled by a non-parametric thin-plate spline kernel; this enables the method to handle image pairs with both rigid and non-rigid motions. Moreover, they also introduced a sparse approximation to accelerate the optimization, which can largely reduce the computational complexity without degenerating the accuracy. The quantitative results on various experimental data demonstrated that the method can achieve better matching accuracy and can generate more good matches compared to several state-of-the-art methods.

Hyperspectral remote sensing images are typical high-dimensional data with a lot of redundant information, which impacts the classification accuracy. Feature extraction is an effective method to reduce the redundancy of hyperspectral image (HSI) and improve the classification performance. However, most feature extraction methods just consider a single structure information of HSI that will lose some valuable information. To address this drawback, Gan *et al.* in ''Feature extraction based multi-structure manifold embedding for hyperspectral remote sensing image classification,'' proposed an unsupervised feature extraction method termed multi-structure manifold embedding (MSME) for HSI classification. First, MSME utilized sparse representation to obtain the sparse coefficients of HSI data. Then, it constructed a sparse graph and a sparse hypergraph with the sparse coefficients. The authors used the sparse graph, the sparse hypergraph, and the local linear property to represent different intrinsic structures of HSI. Finally, they constructed a feature learning method with these structures to achieve an optimal projection matrix for feature extraction. MSME made full use of the complementarity of different structures to reveal the intrinsic properties of HSI and improve the discriminating power of features for classification. Experiments on the Salinas and PaviaU data sets showed that the proposed MSME algorithm achieves the best classification results than other state-of-the-art methods.

Recognizing a person of interest in cameras of different viewpoints is known as the task of person re-identification. It has been a challenging job considering the variation in human pose, the changing illumination conditions and the lack of paired samples. Previous matching techniques in the person re-identification field mainly focused on Mahalanobis-like metric learning functions. Taking advantage of the sparse representation and collaborative representation, Guo *et al.* in ''Person re-identification by weighted integration of sparse and collaborative representation,'' proposed a new approach that elaborately exploits both the globality and locality of images. First, it explored multi-feature extraction with different spatial levels. The extracted features were then projected to a common subspace which handled dimension reduction. Second, it learned a single dictionary for each level that was invariant with the changing of viewpoints. Third, it adopted a weighted fusion approach that combined the dictionary learning-based sparse representation with collaborative representation. Experiments on two benchmark re-identification data sets (VIPeR and GRID) justified the advances of our integration algorithm by comparing with several state-of-the-art methods.

The resulting big data creates opportunities to develop diverse sets of context-aware services and systems, ensuring smart city services are optimized to the dynamic city environment. Critical resources in these smart cities will be more rapidly deployed to regions in need, and those regions predicted to have an imminent or prospective need. In the work by Kotevska *et al.* ''Dynamic network model for smart city data-loss resilience case study: city-to-city network for crime analytics,'' they presented a dynamic network model for improving service resilience to data loss. The network model identified statistically significant shared temporal trends across multivariate spatiotemporal data streams and utilized these trends to improve data prediction performance in the case of data loss. Dynamics also allowed the system to respond to changes in the data streams such as the loss or addition of new information flows.

In online sequential applications, a machine learning model needs to have a self-updating ability to handle the situation, which the training set is changing. Conventional incremental extreme learning machine (ELM) and online sequential ELM are usually achieved in two approaches: directly updating the output weight and recursively computing the left pseudo inverse of the hidden layer output matrix. Jin *et al.* in ''Incremental and decremental extreme learning

machine based on generalized inverse,'' developed a novel solution for incremental and decremental ELM (DELM), via recursively updating and downdating the generalized inverse of the hidden layer output matrix. By preserving the global optimality and best generalization performance, the approach implemented node incremental ELM (N-IELM) and sampled incremental ELM (S-IELM) in a universal form, and overcame the problem of self-starting and numerical instability in the conventional online sequential ELM. This work also proposed sample DELM (S-DELM), which was the first decremental version of ELM. The experiments on regression and classification problems with real-world data sets demonstrated the feasibility and effectiveness of the proposed algorithms with encouraging performances.

Chen *et al.* in ''Multi-temporal depth motion maps-based local binary patterns for 3-D human action recognition,'' presented a local spatio-temporal descriptor for action recognition from depth video sequences, which was capable of distinguishing similar actions as well as coping with different speeds of actions. This descriptor was based on three processing stages. In the first stage, the shape and motion cues were captured from a weighted depth sequence by temporally overlapped depth segments, leading to three improved depth motion maps (DMMs) compared with the previously introduced DMMs. In the second stage, the improved DMMs were partitioned into dense patches, from which the local binary patterns histogram features were extracted to characterize local rotation invariant texture information. In the final stage, a Fisher kernel were used for generating a compact feature representation, which was then combined with a kernel-based extreme learning machine classifier. The developed solution was applied to five public domain data sets and was extensively evaluated. The results obtained demonstrated the effectiveness of this solution compared with the existing approaches.

Visual navigation for mobile robots has emerged in recent years. Among the various methods, topological navigation using visual information provides a scalable map representation for large-scale mapping and navigation. In the work by Ma *et al.* ''Robust topological navigation via convolutional neural network feature and sharpness measure,'' they improved topological mapping to make it more efficient and robust, by using a convolutional neural network (CNN) feature as the holistic image representation.

The security of streaming data should be ensured in the current complex data era in order to provide a trusted and secure network environment. To authenticate the scalable video coding (SVC) streams by fully utilizing its decoding relationship without reducing its scalability, Ma *et al.* in ''Authentication of scalable video coding streams based on topological sort on decoding dependency graph,'' established an acyclic and directed decoding dependence graph (DDG) on the logical units of SVC steams. By applying the topological sort on DDG, they obtained the hash appendence mode for different layers of the streams (i.e., spatial and temporal layers). The authentication approach can achieve much less computation cost and much lower overhead, while it can preserve higher verification rates and better quality of video as compared with other state-of-the-art methods.

Wu *et al.* ''Heterogeneous manifold ranking for image retrieval,'' proposed a heterogeneous MR (HMR) model, in which a couple of graphs built on the click and visual feature sets were fused to simultaneously encode the image ranking results. Specifically, HMR model applied different solutions to fuse the heterogeneous feature sets in terms of whether the relevance feedback mechanism is available or not. In addition, authors developed a click refinement technique to address the noisiness and sparseness problems inherent in the click-through data. It pruned the inaccurate clicks from the click-through data using a neighbor voting strategy, and then enriched the pruned data with novel yet accurate clicks based on a novel collaborative filtering (CF) approach. Extensive experiments on the tasks of click refinement and image retrieval demonstrated the superior performance of the proposed algorithms over several representative methods, especially when the click-through data was highly noisy and sparse.

In ''Bilateral two-dimensional neighborhood preserving discriminant embedding for face recognition,'' Liang *et al.* proposed a novel bilateral 2-D neighborhood preserving discriminant embedding for supervised linear dimensionality reduction for face recognition. It directly extracted discriminative face features from images based on graph embedding and Fisher's criterion. The proposed method was a manifold learning algorithm based on graph embedding criterion, which can effectively discover the underlying nonlinear face data structure. Both within-neighboring and between-neighboring information were taken into account to seek an optimal projection matrix by minimizing the intra-class scatter and maximizing the inter-class scatter based on Fisher's criterion. The performance of the proposed method was evaluated and compared with other face recognition schemes on the Yale, PICS, AR, and LFW databases. The experiment results demonstrated the effectiveness and superiority of the proposed method as compared with the state-of-the-art dimensionality reduction algorithms.

Current studies focus on indirect factors in social recommender systems such as the similarity between users, but multiple direct interactions are seldom considered, such as mentions, reposts and comments. The work by Li *et al.* ''Social recommendation with multiple influence from direct user interactions,'' addressed direct connections between users in social recommender systems and analyzed direct interactions to investigate the connection strength between users. Authors proposed a recommendation method with social influence, which made full use of information among users in social networks and introduced the mechanisms of macroscopic and microscopic influences. Real-world microblog data were applied to verify our model, and the results showed that the proposed recommendation method outperforms other state-of-the-art recommendation algorithms.

Zinc is an important trace element, and it can be used in combination with proteins to play an important biological function. In the work by Li *et al.* ''Integrative method based on linear regression for the prediction of zinc-binding sites in proteins," three types of prediction tools based on sequence were studied for the prediction of zinc-binding sites in proteins, and a novel integrated predictor termed meta-zincPrediction was presented. Multiple linear regressions were used in the proposed approach to integrate the results of the three prediction tools, and the parameters were estimated by the least square method until the optimal model was constructed.

The kernel minimum square error classification (KMSEC) algorithm has been widely used in classification problems. It shows good performance on image data but has the following drawbacks: not sparse in the solutions, and sensitive to noises. The latter drawback will result in a decrease in the recognition performance. To this end, Liu *et al.* in ''An improved kernel minimum square error classification algorithm based on L2, 1-norm regularization,'' proposed an improved (IKMSEC) by using the L 2, 1-norm regularization, which can obtain a sparse representation of nonlinear features to guarantee an efficient classification performance. The comprehensive experiments showed the promising results in face recognition and image.

Co-authorship is one of the key relations in citation recommendation, but it is usually regarded as a binary relation in current graph-based models. This binary modeling of co-authorship is likely to result in information loss, such as the loss of strong or weak relationships between specific research topics. To address this issue, Guo *et al.* in ''Exploiting fine-grained co-authorship for personalized citation recommendation,'' presented a fine-grained method for co-authorship modeling that incorporates the co-author network structure and the topics of their published articles. Then, they designed a three-layered graph-based recommendation model that integrates fine-grained co-authorship as well as author-paper, paper-citation, and paper-keyword relations. The model effectively generated query-oriented recommendations using a simple random walk algorithm. Extensive experiments conducted on a subset of the anthology network data set for performance evaluation demonstrated that the method outperforms other models in terms of both Recall and NDCG.

In real application scenarios, the input images are often noisy, blurry, or suffer from other unknown degradations. Traditional face super-resolution techniques treat image noise at the pixel level without considering the underlying image structures. In order to rectify this shortcoming, Lu *et al.* in ''Robust face super-resolution via locality-constrained low-rank representation,'' proposed a unified framework for representation-based face super-resolution by introducing a locality-constrained low-rank representation (LLR) scheme to reveal the intrinsic structures of input images. The low-rank representation part of LLR clustered an input image into the most accurate subspace from a global dictionary of atoms, while the locality constraint enabled recovery of local manifold structures from local patches. In addition, low-rank, sparsity, locality, accuracy, and robustness of the representation coefficients were exploited in LLR via regularization. Experiments on the FEI, CMU face database, and real surveillance scenario showed that LLR outperformed the state-of-the-art face super-resolution algorithms (e.g., convolutional neural network-based deep learning) both objectively and subjectively.

Depth images play an important role in 3-D applications. However, due to the limitation of depth acquisition equipment, the acquired depth images are usually in limited resolution. In this work by Zhong *et al.* ''Spatially adaptive tensor total variation-tikhonov model for depth image super resolution,'' a spatially adaptive tensor total variation-Tikhonov model was proposed to solve this problem. The tensor total variation regularization was adopted to maintain sharp edges that reflect latent discontinuities in the real world, while the Tikhonov regularization ensured that depth changed smoothly inside objects. Furthermore, a fused edge map was proposed to indicate edge regions and balance both regularization terms. In edge regions, tensor total variation regularization was predominant, thus edge blurring artifacts were suppressed. In non-edge regions, Tikhonov regularization played a more important role to suppress staircasing artifacts. Specifically, texture edges were removed in the fused edge map, and texture copying artifacts were avoided. Experimental results demonstrated the effectiveness and superiority of the proposed framework.

Cui *et al.* in ''DMFA-SR: deeper membership and friendship awareness for social recommendation,'' proposed a new, deeper membership and friendship awareness for social recommendation. They first calculated the deeper membership similarity between users utilizing the improved Jaccard similarity coefficient, and the deeper friendship similarity between users using the proposed two-hop random walk algorithm. The deeper membership similarity and the deeper friendship similarity were combined in a unified way to form a comprehensive deeper social relation similarity. It adopted the matrix factorization method incorporating the deeper membership and the deeper friendship between users as a regularization term for social recommendation, and the corresponding comprehensive deeper social relationship similarity was regarded as the regularization parameter. Experiments on two real-world datasets demonstrated the superiority of the proposed recommendation model.

As an emerging voice of the customer (VOC) containing feedback, such as opinions and expectations about products, social media data have the potential use for product improvement and new product development. In the work by Ko *et al.* ''Identifying product opportunities using social media mining: application of topic modeling and chance discovery theory,'' they suggest an approach to identify product opportunities from customer reviews in social media. This approach contributes to the systematic ideation process for product opportunity analysis based on large-scale and real-time VOC.

Recently, gate-level information flow tracking (GLIFT) has been proposed to monitor the flow of information in secure hardware design by associating data objects with sensitivity labels and tracking the flow of labeled data. In the work by Tai *et al.* ''Towards quantified data analysis of information flow tracking for secure system design,'' they propose a precise multi-bit GLIFT method to perform simultaneous multi-bit flow tracking for understanding exactly which bits are affecting a data object at the same time. The proposed method provides more detailed insights into simultaneous information flow behaviors and thus allows proof of quantitative information flow data properties.

Spatial data clustering has played an important role in knowledge discovery in spatial databases. However, due to the increasing volume and diversity of data, conventional spatial clustering methods are inefficient even on moderately large data sets, and usually fail to discover clusters with diverse shapes and densities. In the work by Wang *et al.* ''From partition-based clustering to density-based clustering: fast find clusters with diverse shapes and densities in spatial databases'', they propose a two-phase clustering method named KMDD (clustering by combining K-means with density and distance-based method) to fast find clusters with diverse shapes and densities in spatial databases.

Power system faults are significant problems in power transmission and distribution. Methods based on relay protection actions and electrical component actions have been put forward in recent years. However, they have deficiencies in dealing with a power system fault. In the work by Zhang *et al.* ''Data-based line trip fault prediction in power systems using LSTM networks and SVM'', a method for data-based line trip fault prediction in power systems using long short-term memory (LSTM) networks and support vector machine (SVM) is proposed. The temporal features of multisourced data are captured with LSTM networks, which perform well in extracting the features of time series for a long-time span. The strong learning and mining ability of LSTM networks is suitable for a large quantity of time series in power transmission and distribution. A discussion of practical applications is presented to demonstrate the feasibility in real scenarios.

Active learning is a promising way to reduce the labeling cost with a limited training samples initially, and then iteratively select the most valuable samples from a large number of unlabeled data for labeling in order to construct a powerful classifier. The goal of active learning is to make sure the labeled data set has no redundancy as much as possible. Uncertainty and diversity are two important criteria for active learning. Currently, a promising way to do this is developed, by combining uncertainty and diversity for active learning. However, many of these methods are designed based on the binary class or uncertainty followed by diversity strategy. It is hard to select the most valuable samples for multiple classes with binary setting with diversity and uncertainty simultaneously. Wang *et al.* in ''Multi-class active learning by integrating uncertainty and diversity,'' integrated uncertainty and diversity into one formula by multi-class settings. Uncertainty is measured by the margin minimum while diversity is measured by the maximum mean discrepancy, which is a popular way to measure the distribution between the two data sets. By minimizing the upper bound for the true risk of the integrating formula, they find the samples share not only uncertainty but also diversity with each other.

Ke *et al.* in ''Adaptive change detection with significance test,'' proposed a significance test-based change detection method that can automatically discriminate between changed and unchanged pixels in the difference image. The method adaptively considers the local contextual information, which is contained in the neighborhoods of each pixel, to derive the decision threshold. A significance test algorithm based on maximum a posteriori estimate is constructed; then, a weight is imposed to each pixel in the block to increase the change detection accuracy. The distribution of the difference image satisfying Laplace model also leads to good precision. For the experimental component, two types of images were tested.

Luo *et al.* in ''Manifold regularized distribution adaptation for classification of remote sensing images,'' performed unsupervised domain adaptation for the classification of remote sensing data by learning a shared subspace. Maximum mean discrepancy (MMD) is applied to each class, making the approach able to minimize the domain shift on a per-class basis. Furthermore, manifold regularization is employed to constrain the data manifold of both the source and target data to be preserved in the subspace. The manifold regularization in conjunction with the per-class MMD strategy is called manifold regularized distribution adaptation (MRDA) algorithm. Since the class mean of target data should be estimated by the predicted labels, spatial information and overall mean coincidence (OMC) method are integrated to improve the prediction accuracy, resulting in the Spa_OMC_MRDA approach. Experimental results on both multispectral and hyperspectral remote sensing data indicate the good performance of the proposed approach.

Accurately predicting future service traffic would be of great help for load balancing and resource allocation, which play a key role in guaranteeing the quality of service (QoS) in cloud computing. With the rapid development of data centers, the large-scale network traffic prediction requires more suitable methods to deal with the complex properties (e.g., high-dimension, long-range dependence, non-linearity, and so on). However, due to the limitations of traditional methods (e.g., strong theoretical assumptions and simple implementation), few research works could predict the large-scale network traffic efficiently and accurately. More importantly, most of the studies took only the temporal features but not the services' communications into consideration, which may weaken the QoS of applications in the data center. To this end, Cao *et al.* in ''Interactive temporal recurrent convolution network for traffic prediction in data centers,'' applied the gated recurrent unit (GRU) model and the interactive temporal recurrent convolution network (ITRCN) to single-service traffic prediction and interactive network traffic prediction, respectively. Especially, ITRCN takes the communications

between services as a whole and directly predicts the interactive traffic in large-scale network. Within the ITRCN model, the convolution neural network (CNN) part learns network traffic as images to capture the network-wide services' correlations, and the GRU part learns the temporal features to help the interactive network traffic prediction.

Event detection and evaluation in social networks provide an effective solution, which has become the critical basis for many real applications, such as crisis management and decision making. However, the existing methods ignore the difference of the evolution fluctuations of nodes. In order to further improve the accuracy of event detection, Wang *et al.* in ''An event detection method for social networks based on evolution fluctuations of nodes,'' proposed an event detection method for social networks based on node evolution fluctuations (NodeED). It contains a node similarity index algorithm (SimJudge) and a microevolution fluctuation detection algorithm (MicroFluc). The main work is as follows: 1) based on particle swarm optimization algorithm, SimJudge is proposed to apply the values of different similarity indexes to quantify the evolution fluctuations of nodes, and the optimal similarity index is determined for each node and 2) microFluc is proposed to integrate the evolution fluctuations of different nodes and quantify the impacts of events in the evolutions of social networks.

Transfer learning and ensemble learning are the new trends for solving the problem that training data and test data have different distributions. The work by Liu *et al.* ''Ensemble transfer learning algorithm,'' designed an ensemble transfer learning framework to improve the classification accuracy when the training data are insufficient. First, a weighted resampling method for transfer learning is proposed, which is named TrResampling. In each iteration, the data with heavy weights in the source domain are resampled, and the TrAdaBoost algorithm is used to adjust the weights of the source data and target data. Second, three classic machine learning algorithms, namely, naive Bayes, decision tree, and SVM, are used as the base learners of TrResampling, where the base learner with the best performance is chosen for transfer learning.

Hyperspectral image (HSI) is usually corrupted by various types of noise, including Gaussian noise, impulse noise, stripes, deadlines, and so on. Recently, sparse and low-rank matrix decomposition (SLRMD) has demonstrated to be an effective tool in HSI denoising. However, the matrix-based SLRMD technique cannot fully take advantage of spatial and spectral information in a 3-D HSI data. In the work by Huang *et al.* ''Hyperspectral image denoising with group sparse and low-rank tensor decomposition,'' a novel group sparse and low-rank tensor decomposition (GSLRTD) method is proposed to remove different kinds of noise in HSI, while still preserving spectral and spatial characteristics. Since a clean 3-D HSI data can be regarded as a 3-D tensor, the proposed GSLRTD method formulates a HSI recovery problem into a sparse and low-rank tensor decomposition framework. Specifically, the HSI is first divided into a set of overlapping 3-D tensor cubes, which are then clustered into groups by K-means algorithm. Then, each group contains similar tensor cubes, which can be constructed as a new tensor by unfolding these similar tensors into a set of matrices and stacking them. Finally, the SLRTD model is introduced to generate noisefree estimation for each group tensor. By aggregating all reconstructed group tensors, we can reconstruct a denoised HSI.

Wireless community networks (WCNs) are emerging as an alternative to provide wireless offloading before the deployment of 5G network. However, it is not clear about its development prospects, e.g., how to understand the WCN technology affects the adoption of users, and the competition in a communication market coexisting with different wireless technologies, i.e., LTE in unlicensed spectrum. To this end, Yang *et al.* in ''Evolution and competition, a game theoretical analysis of heterogeneous wireless networks in unlicensed spectrum,'' envision three distinct development stages of WCN: evolution, regulation, and competition. The users in commercial WCN can be divided into two parts: insiders and outsiders, according to whether they are contributing connectivity into the community.

In the system of face recognition, the traditional method of data dimension reduction method is used to rearrange the face image vectors, resulting in the structural characteristics of the data itself being lost and the recognition accuracy not being high. Mei *et al.* in ''A discriminant subspace learning based face recognition method,'' develop a data dimension reduction method based on tensor-multilinear discriminant subspace projection. The algorithm directly describes the face with tensor and projects the tensor data into the vector discriminant subspace through a new projection mode tensor to vector projection (TVP). This method finds a set of orthogonal projection vector sets to maximize the dispersion between the data classes and minimize the intra-class dispersion in the discriminant subspace. Then, the high-dimensional tensor data is mapped to low-level vector data by TVP. These vector features after dimension reduction will be the most representative feature data in the whole face data under the appropriate constraint condition. Finally, these feature data are classified by the K-nearest neighbor classifier.

Many target detectors commonly utilize a single a priori target spectral signature as an input. However, the detection results are greatly affected by the quality of the a priori target spectral signature because the spectral variability phenomenon is universal and anisotropic in hyperspectral image data. Wang *et al.* in ''A sparse representation method for a priori target signature optimization in hyperspectral target detection,'' propose a sparse representation-based method to generate an optimized target spectrum from limited target training samples, which is able to alleviate the impact of spectral variability on hyperspectral target detection. When lacking comprehensive knowledge about the target object of interest, an optimized representative target spectrum should be expected to be reconstructed by the hyperspectral data themselves in a sparse representation manner following the

characteristics of the data structure, and then be generated by a set of selected candidate pixels that contain the target signal with a varying status.

Peng *et al.* in ''Negative ε dragging technique for pattern classification,''propose the negative ε dragging technique for robust classification of noisy and contaminated data. Different from the naïve ε dragging technique, the negative ε dragging technique argues that robust results can be obtained by properly reducing the class margin of conventional least squares regression when performing classification on noisy data. The underlying rationale of the negative ε dragging technique assumes that setting a relatively small class margin for the training procedure of least squares regression leads to desirable generalization capability, which, therefore, contributes considerably to boosting the classification performance for the data corrupted with noise.

Road detection with high precision from very high resolution remote sensing imagery is very important in a huge variety of applications. However, most existing approaches do not automatically extract the road with a smooth appearance and accurate boundaries. Shi *et al.* in ''Road detection from remote sensing images by generative adversarial networks,'' proposed a novel end-to-end generative adversarial network to construct a convolutional network based on adversarial training that could discriminate between segmentation maps coming either from the ground truth or generated by the segmentation model. The proposed method could improve the segmentation result by finding and correcting the difference between ground truth and result output by the segmentation model.

The active learning method involves searching for the most informative unmarked samples by query function, submitting them to the expert function for marking, then using the samples to train the classification model in order to improve the accuracy of the model and use the newly acquired knowledge to inquire into the next round, with the aim of getting the highest accuracy of classification using minimal training samples. Yang *et al.* in ''Active learning for visual image classification method based on transfer learning,'' details the various principles of active learning and develops a method that combines active learning with transfer learning.

Wang *et al.* in ''Mixed similarity diffusion for recommendation on bipartite networks,'' propose a mixed similarity diffusion model to integrate both explicit feedback and implicit feedback. First, cosine similarity between users is calculated by explicit feedback, and they integrate it with resource allocation index calculated by implicit feedback. They further improve the performance of the mixed similarity diffusion model by considering the degrees of users and items at the same time in diffusion processes. Some sophisticated experiments are designed to evaluate our proposed method on three real-world data sets. Experimental results indicate that recommendations given by the mixed similarity diffusion perform better on both the accuracy and the diversity than that of most state-of-the-art algorithms.

In the era of big data, reduced models capable of reducing big data graphs to estimate personalized PageRank are limited. Personalized PageRank is a page rank calculation where random jumps are only allowed to a subset of start nodes. Pirouz and Zhan in ''Toward efficient hub-less real time personalized PageRank'' propose a novel, fast, accurate and less resource-intensive algorithm to the personalized PageRank problem. FAST Personalized PageRank is utilized to find the target node set. Using the mentioned target set, the algorithm gives an estimation of the closeness of any pair of nodes in the graph. As the time taken by the estimation of personalized PageRank is directly proportional to the network size, in this article a node reduction method is used to prune the graph. In this pruning model, most popular nodes also known as hubs are found using personalized page vector. To decrease the entropy and reduce the number of alternate paths to the target nodes, popular nodes are identified and flagged. The flagged nodes are then given a lower priority in the computation. This way the redundant path will be ignored in the computation process. After pruning the graph, estimation results achieve an improved time complexity.

Here, we would like to thank all the authors who submitted their high-quality research activities to this Special Section. We are also grateful to the reviewers for their valuable comments to improve the quality of the articles. Moreover, we appreciate the support from both Prof. Michael Pecht, former Editor-in-Chief of IEEE Access, and Bora M. Onat, Managing Editor of IEEE Access. We also appreciate the help of Margery Meyer, Rebecca Hytowitz, Anna Melhorn, Kristen Cignavitch, Ian Barbour, and Kim Shumard throughout the publication process. Finally, we hope that the readers will enjoy reading the articles of this Special Section in IEEE Access.

JIA WU
Macquarie University, Sydney, NSW, Australia

SHIRUI PAN
University of Technology Sydney, Ultimo, NSW, Australia

JUNJUN JIANG
National Institute of Informatics, Tokyo, Japan

ZHIHUA CAI
China University of Geosciences (Wuhan), Wuhan, China

BO DU
Wuhan University, Wuhan, China

YINGJIE TIAN
Chinese Academy of Sciences, Beijing, China

SHUAIQIANG WANG
The University of Manchester, Manchester, U.K.

HAISHUAI WANG
Harvard University, Cambridge, MA, USA

JIA WU is currently a Lecturer with the Department of Computing, Faculty of Science and Engineering, Macquarie University, Sydney, Australia. His current research interests include data mining and machine learning. Since 2009, he has authored or co-authored over 60 refereed journal and conference papers, such as the IEEE TNNLS, the IEEE TKDE, the IEEE TCYB, ACM TKDD, *Pattern Recognition*, IJCAI, AAAI, ICDE, ICDM, SDM, CIKM, PAKDD, and IJCNN in these areas. He is a member of the IEEE. He has successfully organized a special session in IJCNN 2016 and IJCNN 2017. He and his research group received the IJCNN'17 Best Student Paper Award. He is an Associate Editor of the *ACM Transactions on Knowledge Discovery from Data* and the *Journal of Network and Computer Applications* (Elsevier).

SHIRUI PAN received the Ph.D. degree in computer science from the University of Technology Sydney, where he is currently a Research Associate with the Centre for Artificial Intelligence. He has published over 40 research papers in top-tier journals and conferences, including the IEEE TKDE, the IEEE TCYB, *Pattern Recognition*, IJCAI, ICDM, SDM, CIKM, and ICDE. His research interests include data mining and machine learning.

JUNJUN JIANG received the Ph.D. degree from the School of Computer, Wuhan University, Wuhan, China, in 2014. He has been a Project Researcher with the National Institute of Informatics, Tokyo, Japan, since 2016. He is currently an Associate Professor with the School of Computer Science, China University of Geosciences. He has authored or co-authored more than 60 scientific articles (including nine IEEE TRANSACTIONS papers) and has obtained ten Chinese patents. His research interests include image processing, computer vision, and machine learning. In 2014, he was a recipient of the IBM China Excellent Student Scholarship. He received the 2015 ACM Wuhan Doctoral Dissertation Award.

ZHIHUA CAI received the B.Sc. degree from Wuhan University, Wuhan, China, in 1986, the M.Sc. degree from the Beijing University of Technology, Beijing, China, in 1992, and the Ph.D. degree from the China University of Geosciences, Wuhan, in 2003, where he is currently a Faculty Member with the School of Computer Science. He has published over 50 research papers in journals and international conferences, such as the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE TRANSACTIONS ON CYBERNETICS, APPLIED SOFT COMPUTING, *Information Sciences*, *Knowledge-Based Systems*, and *Knowledge and Information Systems*. His main research areas include data mining, machine learning, evolutionary computation, and their applications.

BO DU received the B.S. and Ph.D. degrees in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote sensing, Wuhan University, Wuhan, China, in 2005 and 2010, respectively, where he is currently a Professor with the School of Computer. He has more than 40 research papers published in the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS), the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND APPLIED REMOTE SENSING (JSTARS), and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. Five of them are ESI highly cited papers. His major research interests include pattern recognition, hyperspectral image processing, and signal processing. He is a Senior Member of the IEEE. He received the best reviewer awards from the IEEE GRSS for his service to the IEEE JSTARS, in 2011, and the ACM rising star awards for his academic progress, in 2015. He was the Session Chair of both the International Geoscience and Remote Sensing Symposium 2016 and the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing. He also serves as a Reviewer for 20 Science Citation Index magazines, including the IEEE TGRS, TIP, JSTARS, and GRSL.

YINGJIE TIAN received the bachelor's degree in mathematics, in 1994, the master's degree in applied mathematics, in 1997, and the Ph.D. degree in management science and engineering. He is currently a Professor with the Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences. He has published four books on SVMs, one of which has been cited over 1,500 times. His research interests include support vector machines, optimization theory and applications, data mining, intelligent knowledge management, and risk management.

SHUAIQIANG WANG received the B.Sc. and Ph.D. degrees in computer science from Shandong University, China, in 2004 and 2009, respectively. In 2009, he visited Hong Kong Baptist University as an Exchange Doctoral Student. He was a Lecturer (equivalent to Assistant Professor in USA) with the Alliance Manchester Business School, The University of Manchester, U.K., in 2017, an Assistant Professor with the University of Jyväskylä, Finland, from 2014 to 2017, an Associate Professor with the Shandong University of Finance and Economics, China, from 2011 to 2014, and a Postdoctoral Research Associate with Texas State University, USA, from 2010 to 2011. He is currently a Research Scientist with the Data Science Laboratory, JD.COM. His research interests include recommender systems, information retrieval, and data mining.

HAISHUAI WANG received the Ph.D. degree in computer science from the Center of Artificial Intelligence, University of Technology Sydney. He was a Postdoctoral Associate with the Department of Computer Science, Washington University in St. Louis. He is currently a Research Fellow with the Department of Biomedical Informatics, Harvard University. He has been involved in real-world applications, including social network analysis and medical data analysis. He also has been with a number of IT companies for internships, including IBM, Hitachi, and Intel. His research interests include data mining and machine learning, specialized in the areas of time-variant graph and time-series learning and classification, and deep learning.