Math.Comput.Sci. (2013) 7:379–386
DOI 10.1007/s11786-013-0169-2

Foreword to the Special Focus on Mathematics, Data and Knowledge

Xiaoyu Chen · Dongming Wang · Xiangliang Zhang

Published online: 7 December 2013

Abstract There is a growing interest in applying mathematical theories and methods from topology, computational geometry, differential equations, fluid dynamics, quantum statistics, etc. to describe and to analyze scientific regularities of diverse, massive, complex, nonlinear, and fast changing data accumulated continuously around the world and in discovering and revealing valid, insightful, and valuable knowledge that data imply. With increasingly solid mathematical foundations, various methods and techniques have been studied and developed for data mining, modeling, and processing, and knowledge representation, organization, and verification; different systems and mechanisms have been designed to perform data-intensive tasks in many application fields for classification, predication, recommendation, ranking, filtering, etc. This special focus of *Mathematics in Computer Science* is organized to stimulate original research on the interaction of mathematics with data and knowledge, in particular the exploration of new mathematical theories and methodologies for data modeling and analysis and knowledge discovery and management, the study of mathematical models of big data and complex knowledge, and the development of novel solutions and strategies to enhance the performance of existing systems and mechanisms for data and knowledge processing. The present foreword provides a short review of some key ideas and techniques on how mathematics interacts with data and knowledge, together with a few selected research directions and problems and a brief introduction to the four papers published in the focus.

X. Chen (✉)
State Key Laboratory of Software Development Environment, School of Computer Science and Engineering,
Beihang University, Beijing 100191, China
e-mail: franknewchen@gmail.com

D. Wang
Laboratoire d'Informatique de Paris 6, Université Pierre et Marie Curie–CNRS, 4 place Jussieu,
BP 169, 75252 Paris cedex 05, France
e-mail: Dongming.Wang@lip6.fr

X. Zhang
Division of Computer, Electrical and Mathematical Sciences and Engineering (CEMSE),
King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia
e-mail: Xiangliang.Zhang@kaust.edu.sa

1 Interaction of Mathematics with Data and Knowledge

To study the behaviors and properties of social and natural phenomena, for example, in economy, psychology, politics, physics, biology, geography, and astronomy, scientists have kept on collecting values of qualitative or quantitative variables in terms of time, space, structure, relationship, etc. around the world. The values appearing in the form of numbers and characters, i.e., data, are obtained through observation and measurement and may be recorded nowadays in digital media. Recent advances in computing and database techniques allow one to collect a vast amount of data from many sources, such as sensor networks, network flows, web pages, cloud computing system logs, financial applications, biological microarrays and astronomic observation databases. Through data checking, processing, computation and analysis, scientific regularities, predictions, and meaningful information and patterns with respect to phenomena can be revealed and summarized as knowledge in the form of statements, procedures, models, and so on. In the processes of data mining and knowledge discovering, mathematics plays an important role in formally describing the characteristics of phenomena, making inferences and revisions, and seeking out appropriate models to explain complex systems and to study the effects of different components. In order to fit experimental measurements, various forms of mathematical models have been developed; they include but are not limited to dynamical systems, stochastic processes, differential equations, and game theoretic models. These and other types of models can overlap, with each given model involving a variety of abstract structures. Applications of different mathematical methods on knowledge discovery from data fall mainly into the following categories.

1.1 Probability and Statistics

Probability theory and statistical concepts, such as regression and Bayesian inference, play a remarkable role in knowledge discovery. Given a dataset, regression is a statistical technique used to find the best-fitting relationship between a target (dependent) variable and its predictors (independent variables) [15]. The fitting function can be linear or nonlinear with respect to a set of parameters. Basis functions, e.g., polynomial functions in different degrees and Gaussian functions with different means, variances, or kernels are usually adopted. Logistic regression with a categorical target variable is often used to predict the class label of objects. It can also be considered as an approach for binomial regression. It is worth to mention that regression may suffer an over-fitting problem, e.g., when a complex function is used to approximate a simple function underlying a small dataset. An over-fitted function model has poor predictive power and cannot be generalized. Regularization is a popular solution to the over-fitting problem. Adding a regularization term to the original function limits the search of the optimal results within a promising region. Bayesian inference tackles the over-fitting problem by introducing prior distributions over regression parameters. The posterior probability is maximized to find the optimal parameters of regression functions.

Bayes' theorem has a successful application to the classification problem. The degree of belongingness of an object X to a category C is predicted by a conditional probability $P(C|X)$, which can be written as $P(C|X) = P(X|C)P(C)/P(X)$. $P(C)$ and $P(X)$ are the prior probabilities of C and X, and $P(X|C)$ is the probability of observing X, given that X belongs to category C. The object X is usually described by a set of variables (attributes), denoted by $\{A_1, A_2, \ldots, A_n\}$. Since $P(C)$ can be easily obtained from a given dataset, the main focus is to calculate the joint probability, e.g., $P(A_1, A_2, \ldots, A_n|C)$.

There are two main approaches for the calculation of joint probability $P(A_1, A_2, \ldots, A_n|C)$: one assumes the independency among all the variables A_i $(1 \le i \le n)$, while the other assumes conditional independency among variables. The first approach is the so-called Naive Bayes, which calculates $P(A_1, A_2, \ldots, A_n|C)$ as the product $\prod_i P(A_i|C)$ of conditional probabilities of all the variables. The second approach, Bayesian Network, characterizes the dependency among variables by a directed acyclic graph. $P(A_1, A_2, \ldots, A_n|C)$ is calculated as the product $\prod_i P(A_i|\mathtt{parent}(A_i), C)$ of the conditional probabilities of those variables which depend on their parent variables. Bayesian network can be generalized to various graphic models, which facilitate the inference for many data application problems [6].

1.2 Matrix Decomposition and Dimensionality Reduction

Matrix Decomposition is a powerful tool for revealing latent concepts and eliminating irrelevant variables, and eventually studying data in a new space, which is more meaningful than the original space. Singular Value Decomposition (SVD) is one of the most popular methods for matrix decomposition. It is also known as Latent Semantic Indexing (LSI) in text mining. It is a matrix factorization method which decomposes a matrix into (right and left) singular values and singular vectors. SVD can be used to reduce noise from a given dataset, by eliminating small singular values and reconstructing the original matrix. The right and left singular vectors can be used to learn a new representation for both documents and terms. In the case of text mining, documents are to be represented by the new set of singular values (latent topics) instead of the terms.

Principal Component Analysis (PCA) is another popular method based on matrix decomposition. It aims at finding a set of orthogonal (uncorrelated) Principal Components (PCs) from the original variable space. PCs are linear combinations of the original variables and are actually eigenvectors of the gamma matrix of X, defined as $1/N(X-\bar{X})^T(X-\bar{X})$, where $\bar{X}$ is a vector containing the mean of each variable. For each PC, its importance is measured by its corresponding eigenvalue, which indicates the variance it captures. The most important PC captures the largest variance in the data, the second most important PC captures the second largest variance in the data, and so on. PCs with low importance are eliminated. The remaining PCs represent X in a space of dimension lower than the original one and capture the underlying pattern with little loss.

One major difference between SVD and PCA lies in the obtained new presentation. PCA subtracts the variable mean and generates each PC as a linear combination of all the original variables. The PCs, used to span a new and lower-dimensional space, are not sparse; it is thus not easy for users to interpret them. Sparse PCA (SPCA) was therefore studied to find PCs with a small number of nonzero elements [30]. SPCA was formulated with a semidefinite relaxation and then solved by Semidefinite Programming in [14]. It has been studied with LASSO in [29].

Independent Component Analysis (ICA) is very similar to PCA, though it poses more strict assumptions. Instead of finding orthogonal latent variables, ICA finds non-Gaussian and statistically independent hidden variables. Statistical independence is an assumption much stronger than un-correlation. In text mining, the Independent Components (ICs) represent latent topics in the dataset, where the independence assumption means that topics are statistically independent. The goal of ICA is to minimize the statistical dependence between the basis vectors, while the goal of PCA is to minimize the re-projection error from compressed data.

In manifold learning, data lie on a nonlinear manifold embedded in high-dimensional space, e.g., for swiss roll and fish bowl. PCA, a linear dimensionality reduction method, often fails for such datasets. Two research papers published in Science 2000 [23,27] tackled the nonlinear dimensionality reduction problem. The proposed algorithms, named ISOMAP [27] and LLE [23], are the most popular ones for obtaining a global low-dimensional representation of the data. Since Euclidian neighborhood holds only locally, ISOMAP defines the geodesic distance between two points as the sum of the local Euclidian distances along the shortest path from one point to the other on the manifold. LLE (Locally Linear Embedding) preserves local geometry of the data based on the assumption that the linear function weights representing a data object by a combination of its neighbors in the original space will hold in the low-dimensional space. Both of the algorithms find the low-dimension data expression by eigenvalue and eigenvector calculation, which implicitly minimizes the reconstruction loss.

1.3 Mathematical Programming (Optimization) and Learning from Data

Knowledge and information discovery from data is often realized by learning techniques, e.g., learning hyperplanes that can separate objects in different categories (classification), learning data distribution from which a dataset is generated (model-based clustering), and learning fitting functions from a set of objects (regression). Many learning problems are eventually formulated as convex optimization ones, e.g., the problem of maximizing the log likelihood or minimizing the sum of squared errors (least square regression). Such optimization problems may be solved by using techniques of mathematical programming, such as quadratic, linear, second-order cone, semidefinite, or semi-infinite programming [4]. In general, convex problems are theoretically tractable. There are algorithms which can optimize convex objective functions subject to convex constraints, ensuring that every local minimum is always a global minimum. The accuracy, speed, and robustness of knowledge discovery are determined essentially by the employed mathematical programs.

The interaction between state-of-the-art learning algorithms and mathematical programming can be found in several instances. In Support Vector Machines (SVMs), the optimal hyperplane for separating different objects is defined as the one resulting in the maximal margin. Finding the maximal margin is then formulated as a quadratic programming problem with the help of Lagrange multipliers and duality [8]. In neural networks, the determination of network parameters within a maximum likelihood framework consists in solving a nonlinear optimization problem using, e.g., the backpropagation framework and derivatives evaluated by means of Jacobian and Hessian matrices [6]. The regression problem aims at finding a function that fits best to the given dataset, where the function parameters are found by quadratic or linear programming [7]. Based on the assumption that each given dataset is generated from a mixture of Gaussian distributions, model-based clustering finds the parameters of the mixture by maximizing the likelihood function. One identified Gaussian component corresponds to one cluster. Semidefinite programming is also commonly used, e.g., in dimensionality reduction for recognizing a dataset lying in a low-dimensional space [24].

1.4 Fuzzy Theory and Knowledge Discovery from Web Data

Fuzzy techniques have been widely utilized for knowledge discovery from web data, including web pages, images, audio, video, links among pages (provided as graphs), usage logs, user profiles, and user queries [22]. Besides their nature of huge volume, heterogeneity, and high dimension, data on web may be structured, unstructured, or semistructured and may be mixed with imprecise or incomplete information. The flexibility of fuzzy techniques allows one to handle the complexity of web data both in theory and in practice.

In particular, fuzzy sets and fuzzy logic have been developed to deal with uncertainty in the belongingness of data objects [26]. An object that cannot be certainly recognized to belong to any known groups may be assigned partially to all groups with different probabilities. Fuzzy C-means (FCM) is a direct and successful application of fuzzy sets to clustering analysis, one of the most important problems in data mining. It was proposed by Dunn in 1973 [17] and improved by Bezdek in 1981 [5]. FCM and the k-means clustering algorithm are analogous, as they both use the cluster mean as a centroid of the cluster and minimize the sum of distances between each object and its associated centroid. They differ from each other on the object assignment. In k-means, each object is assigned to exactly one cluster, which is called *hard assignment*. In FCM, however, one object is allowed to belong to more than one cluster, which is called *soft assignment*. The degree of belongingness of one object in one cluster is valued in [0, 1] and determined by its relative closeness to this cluster centroid, compared with other cluster centroids.

Rough sets are applicable to conduct clustering analysis, to induce rules, and to select useful information from incomplete or imprecise data, e.g., web data [22]. Knowledge discovery from web data supports information retrieval. The theory of fuzzy sets, rough sets, and fuzzy logic has been applied to web documentation and user clustering [18], user profile modeling, inference of web usage patterns (e.g., determining what URLs tend to be requested together) and retrieval of multimedia (including both text and images) [3]. It is also considered as a promising framework for enabling Semantic Web, an extension of the current web with well-defined information.

1.5 Advanced Mathematical Techniques

In dealing with probability distributions underlying statistical models, information geometry makes use of concepts and techniques (such as Riemannian manifold and metric, dual affine connection, and divergence) from differential geometry and provides a new perspective and a new analytic tool for studying the structure of data and knowledge models [1]. It has been applied successfully to solve various problems arising from the fields of statistical inference, time series and linear systems, multiterminal information theory, quantum estimation theory, statistical mechanics, nonlinear prediction, neural computing, mathematical programming, and portfolio [2]. Its usefulness for the elucidation of information systems, intelligent systems, control systems, physical systems, mathematical systems, etc. has also been demonstrated.

There are well-developed methods from topology which can deal with geometric information and properties in a way insensitive to the choice of metrics and independent of the chosen coordinate system. Such topological methods are naturally appropriate for quantitative and qualitative analysis of data structure. Data are often provided in the form of a point cloud exiting in a geometric space (e.g., Euclidean space) equipped with distance functions. By using point-cloud data sampled from geometric objects, witness complexes can be constructed to tackle the problem of efficiently computing topological invariants of the objects in a robust manner [25]. To answer the questions of how to infer high-dimensional structure from low-dimensional representations and how to assemble discrete points into global structure, one may employ topological methods to replace sets of data points by families of simplicial complexes, indexed with proximity parameters, to analyze the complexes using the theory of persistent homology, and to encode the persistent homology of each dataset in the form of a parameterized version of a Betti number (called a *barcode*) [19]. In the process of reasoning about the nature of clustering methods, the notion of functoriality plays a key role [10]. To recover geometric and topological features from a point cloud approximating a topological type, possibly with outliers or background noise, distance functions may be used successfully to address many crucial issues [11].

The rapid development of data analysis and processing techniques has benefited considerably from the rich supply of powerful and sophisticated mathematical methods and tools, ranging from probability theory and statistics to dynamical systems, and from linear algebra and combinatorics to differential geometry. On the other hand, the study of regression, classification, clustering, and many other issues related to data analysis and processing has furnished a large variety of theoretical and practical problems that have stimulated and will continue stimulating the development of pure, applied, and computational mathematics. Computational statistics, information geometry, symbolic-numeric computation, evolutionary computing, constraint and equation solving, and automated reasoning and verification are just a few representative examples of fast emerging areas of modern interdisciplinary science, where mathematics meets data and knowledge.

2 Research Directions and Challenging Problems

Data and knowledge are now everywhere with us. They are influential on many aspects of our work and daily life. Characterized by their volume, velocity, variety, and veracity, their underling evolving regularities have become more and more complex than ever. It is interesting to see whether dynamical systems with massive data and complex knowledge have common features or obey some general laws as in physics and biology. As a new domain of science, data and knowledge need be studied systematically by exploring their features and inherent laws. Such features and laws have to be formulated and analyzed by using and developing mathematical languages and methods, in order to establish a solid mathematical foundation for the science of data and knowledge. On the foundation, new methods, mechanisms, strategies, etc. may then be developed to manage systems of data and knowledge and to study their properties and behaviors. Foundational research in data and knowledge science is expected to have a significant impact on future development of mathematics, opening new research directions and leading to revolutionary theories and thoughts. In fact, data, information, and knowledge can be described mathematically as three hierarchical spaces S_1, S_2, and S_3, respectively, in which objects, functions, and relations may be defined. The space S_i can be constructed from S_{i-1} by introducing relations among objects in S_{i-1} for $i = 2, 3$. A thorough study of the data space S_1, the information space S_2, and the knowledge space S_3 will consolidate the science of data, information, and knowledge.

In order to explore knowledge from big data, mathematical models have to be adapted to meet at least four expectations: (1) high efficiency in computing time and memory usage; (2) fast convergence to approximate solutions (rather than exact solutions with high computational cost); (3) high performance when feeding unknown data objects;

(4) high scalability to large-scale datasets. One has to face the following difficulties and challenges when developing mathematical models, methods, and theories.

- Modeling data with large volume and high velocity. The volume of available data increases rapidly. For example: thousands of observation samples are collected from sensor networks every second; millions of network traffic records are produced per minute at one router; billions of messages are posted via social media everyday; billions of readings are generated by NASA's observation satellites per day. Such big data with large volume and high velocity bring challenges about storage, management, processing, and modeling. Mathematical models for information retrieval and knowledge discovery need be developed to handle them efficiently. Moreover, dynamical big data systems tend to be stochastic, evolutionary, and possibly chaotic. So more accurate dynamical models should be introduced to approximate such systems at high order, in order to capture their complex behaviors.
- Handling data uncertainty. For problems of uncertain nature or with unclear features, collected data can only characterize them inexactly. For example: erroneous and inappropriate data are usually reported by sensors due to environmental factors, packet loss, and/or low battery power [21]; uncertain readings are often returned from tagged objects monitored by RFID (radio-frequency identification) systems for unpredictable occurrences [28]. Mathematical theories (in particular, probability theory) and methods have to be combined with heuristics, statistical information, knowledge bases, fuzzy reasoning, etc. for modeling and handling uncertainties in datasets [16]. Advanced techniques of approximation, sampling, testing, and inferencing need be developed for uncertainty handling in data mining and knowledge discovery.
- Analyzing data with connections. Table-data, managed in tables with observations by rows and fixed features by columns, have been well studied. As social media has become popular, data with connections (e.g., social graphs with nodes representing users and edges showing links among users) have become a new focus of attention. There are many interesting problems, such as social community detection (node clustering), link prediction, and social influence inference, which are worthy of extensive exploration. Sequential data (e.g., DNA sequences) are structured data of another kind, in which ordered data objects are not independent. Effective mathematical methods need be sought for discovering interesting patterns (e.g., maximum common subsequences) from data.
- Managing scientific knowledge. The explosion of digitalized knowledge leads to the increase of complexity in acquiring, representing, organizing, processing, and applying such knowledge from scientific domains. Mathematical knowledge has been used as a fundamental tool in science, technology, and engineering. How to manage mathematical knowledge in particular and complex scientific knowledge in general is a challenging question that deserves systematic investigation. Although a considerable amount of research effort has been made (see, e.g., [9,20] for the state of the art in the general case of mathematics and [12] in the special case of geometry), the question is still far from being settled.

3 Introduction to Papers in the Focus

This special focus contains four original research papers selected from 11 submissions according to the standard refereeing procedure of the journal. Of the four papers, two deal with problems in data modeling and analysis, and the other two deal with problems in knowledge organization and management. We complete the foreword by a brief introduction to these papers.

The paper by Mostafa M. Abbass and Hazem M. Bahig proposes an efficient algorithm for identifying DNA motifs. Among a set of DNA sequences, motifs are subsequences with same length and similar patterns. Finding motifs in a set of DNA sequences is important for biological studies, e.g., genetic drug target identification and human disease investigation. One effective solution, by using the so-called Voting Algorithm, was proposed in [13]. However, the voting algorithm has high complexity and thus has difficulties to discover long motifs. Abbass and Bahig target on improving the speed of the algorithm for motif identification. The gain of efficiency is mainly due to the reduction of the first input sequences.

The paper by Cristian Cruz, William Lima Leão, and David Rohde investigates the sensitivity of the number of clusters in a Gaussian mixture model to prior distributions. Gaussian mixture has been shown to be an effective model-based clustering approach. However, it is sensitive to the initial setting of prior distribution, and may produce non-global optimal results if not well initialized. Based on the observation that the posterior distribution of the number of clusters is sensitive to the prior distribution of variance, Cruz and others propose two distinct Bayesian approaches with practitioners guidance in setting prior distributions.

The paper by Elena Aladova, Eugene Plotkin, and Tatjana Plotkin studies the problem of equivalence of knowledge bases, which are certain kinds of databases, designed for storing and extracting information. Based on a purely universally-algebraic (with some elements of category theory) approach to the issue, Aladova and others consider three notions of equivalence of knowledge bases, i.e., informational equivalence, elementary equivalence, and isotypeness. In particular, they show that isotypeness implies both elementary and informational equivalences (the latter implication positioned as the main result), whereas elementary equivalence does not imply informational equivalence. Additionally, they conjecture that informational equivalence implies isotypeness.

The paper by Xiaoyu Chen and Dongming Wang addresses the problem of identification, formalization, and specification of standard geometric knowledge objects, which are essential for geometric knowledge management. A retrievable and extensible data structure is proposed to specify knowledge objects with embedded knowledge and a language adapted from the language of first-order logic is presented to formalize knowledge statements. The underlying ideas of the approach have been successfully used in geometric theorem proving, knowledge base creation, and electronic document generation.

Finally, we thank the authors and numerous referees for their support and contributions.

References

1. Amari, S.-I.: Information geometry on hierarchy of probability distributions. IEEE Trans. Inf. Theory **47**(5), 1701–1711 (2001)
2. Amari, S.-I., Nagaoka, H.: Methods of Information Geometry. American Mathematical Society, Providence, RI (2000)
3. Arotaritei, D., Mitra, S.: Web mining: a survey in the fuzzy framework. Fuzzy Sets Syst. **148**(1), 5–19 (2004)
4. Bennett, K.P., Parrado-Hernandez, E.: The interplay of optimization and machine learning research. J. Mach. Learn. Res. **7**, 1265–1281 (2006)
5. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
6. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
7. Bradley, P.S., Fayyad, U.M., Mangasarian, O.L.: Mathematical programming for data mining: formulations and challenges. INFORMS J. Comput. **11**(3), 217–238 (1999)
8. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov. **2**(2), 121–167 (1998)
9. Carette, J., Farmer, W.M.: A review of mathematical knowledge management. In: Carette, J., Dixon, L., Coen, C.S., Watt, S.M. (eds.) Intelligent Computer Mathematics, volume 5625 of Lecture Notes in Artificial Intelligence, pp. 233–246. Springer, Berlin (2009)
10. Carlsson, G.: Topology and data. Bull. Am. Math. Soc. **46**(2), 255–308 (2009)
11. Chazal, F., Cohen-Steiner, D., Mrigot, Q.: Geometric inference for probability measures. Found. Comput. Math. **11**(6), 733–751 (2011)
12. Chen, X., Wang, D.: Management of geometric knowledge in textbooks. Data Knowl. Eng. **73**, 43–57 (2012)
13. Chin, F.Y., Leung, H.C.: Voting algorithms for discovering long motifs. In: Chen, Y.-P.P., Wong, L. (eds.) Proceedings of the 3rd Asia-Pacific Bioinformatics Conference, volume 1 of Series on Advances in Bioinformatics and Computational Biology, pp. 261–271. World Scientific Publishing Company, Singapore (2005)
14. d'Aspremont, A., Ghaoui, L.E., Jordan, M.I., Lanckriet, G.R.G.: A direct formulation for sparse PCA using semidefinite programming. SIAM Rev. **49**(3), 434–448 (2007)
15. Devore, J.L., Berk, K.N.: Modern Mathematical Statistics with Applications. Springer, New York (2007)
16. Diao, Y., Li, B., Liu, A., Peng, L., Sutton, C.A., Tran, T.T.L., Zink, M.: Capturing data uncertainty in high-volume stream processing. In: Proceedings of the 4th Biennial Conference on Innovative Data Systems Research, Asilomar, California, USA, January 4–7, 2009. (Online Proceedings)
17. Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybern. **3**, 32–57 (1973)
18. Friedman, M., Last, M., Zaafrany, O., Schneider, M., Kandel, A.: A new approach for fuzzy clustering of web documents. In: Proceedings of IEEE International Conference on Fuzzy Systems, volume 1, pp. 377–381. IEEE Computational Intelligence Society (2004)

19. Ghrist, R.: Barcodes: the persistent topology of data. Bull. Am. Meteorol. Soc. **45**, 61–75 (2008)
20. Kerber, M. (ed.): Management of Mathematical Knowledge—Special issue of Mathematics in Computer Science, vol 2. Birkhäuser, Basel (2008)
21. Lian, X., Chen, L.: A generic framework for handling uncertain data with local correlations. Proc. VLDB Endow. **4**(1), 12–21 (2010)
22. Ropero, J., Leun, C., Carrasco, A., Gumez, A., Rivera, O.: Fuzzy logic applications for knowledge discovery: a survey. Int. J. Adv. Comput. Technol. **3**(6), 187–198 (2011)
23. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**(5500), 2323–2326 (2000)
24. Scheinberg, K., Peng, J., Terlaky, T., Shuurmans, D., Jordan, M., Poggio, T.: Mathematical programming in machine learning and data mining. In: 5-day Worksop Held by Banff International Research Station for Mathematical Innovation and Discovery, January 14–19 (2007)
25. Silva, V.D., Carlsson, G.: Topological estimation using witness complexes. In: Alexa, M., Gross, M., Pfister, H., Rusinkiewicz, S. (eds.) Proceedings of the First Eurographics Conference on Point-Based Graphics, pp. 157–166. Eurographics Association Aire-la-Ville (2004)
26. Simovici, D.A., Djeraba, C.: Mathematical Tools for Data Mining: Set Theory, Partial Orders, Combinatorics. Advanced Information and Knowledge Processing. Springer, London (2008)
27. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**(5500), 2319–2323 (2000)
28. Tran, T.T., Peng, L., Diao, Y., McGregor, A., Liu, A.: CLARO: modeling and processing uncertain data streams. VLDB J. **21**(5), 651–676 (2012)
29. Trendafilov, N., Jolliffe, I.T., Uddin, M.: A modified principal component technique based on the LASSO. J. Comput. Graph. Stat. **12**, 531–547 (2003)
30. Zou, H., Hastie, T., Tibshirani, R.: Sparse principal component analysis. J. Comput. Graph. Stat. **15**(2), 265–286 (2006)