The Emerging "Big Dimensionality"

Zhai, Yiteng; Ong, Yew-Soon; Tsang, Ivor W.

2014

Zhai, Y., Ong, Y.-S., & Tsang, I. W. (2014). The Emerging "Big Dimensionality". IEEE Computational Intelligence Magazine, 9(3), 14-26.

<https://hdl.handle.net/10356/81714>

<https://doi.org/10.1109/MCI.2014.2326099>

© 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: [<http://dx.doi.org/10.1109/MCI.2014.2326099>].

Downloaded on 26 Apr 2023 19:28:15 SGT

The Emerging “Big Dimensionality”

Yiteng Zhai, Yew-Soon Ong, *Senior Member, IEEE*, and Ivor W. Tsang

Abstract—The world continues to generate quintillion bytes of data daily, leading to the pressing needs for new efforts in dealing with the grand challenges brought by Big Data. Today, there is a growing consensus among the computational intelligence communities that data *volume* presents an immediate challenge pertaining to the scalability issue. However, when addressing *volume* in Big Data analytics, researchers in the data analytics community have largely taken a one-sided study of *volume*, which is the “Big Instance Size” factor of the data. The flip side of *volume* which is the dimensionality factor of Big Data, on the other hand, has received much lesser attention. This article thus represents an attempt to fill in this gap and places special focus on this relatively under-explored topic of “Big Dimensionality”, wherein the explosion of *features (variables)* brings about new challenges to computational intelligence. We begin with an analysis on the origins of Big Dimensionality. The evolution of feature dimensionality in the last two decades is then studied using popular data repositories considered in the data analytics and computational intelligence research communities. Subsequently, the state-of-the-art feature selection schemes reported in the field of computational intelligence are reviewed to reveal the inadequacies of existing approaches in keeping pace with the emerging phenomenon of Big Dimensionality. Last but not least, the “curse and blessing of Big Dimensionality” are delineated and deliberated.

Index Terms—Big dimensionality, big data, high dimensionality, feature selection, survey

I. INTRODUCTION

AS we embark on the new era of Big Data, many industrial leaders today are earnestly seeking for new ways to enhance and empower consumer experiences, increase productivity and sales, through making sense of the data that is now becoming ubiquitous. Grasping the fact that a majority of the data generated in the world have been produced within the last two years while we continue to create quintillion bytes daily [1]–[3], there is a real pressing need for credible research into large-scale data analytics. This has led to the rising number of researchers that devote much time and efforts in dealing with the challenges brought about by Big Data. In recent years, the core challenges of Big Data have been widely established and can be summarized under the popular 5Vs in Table I [4]–[7].

From a survey of the literature, there is a growing consensus among data scientists that each “V” brings about unique challenges to the overall task considered in Big Data analytics. For instance, **volume** presents the immediate challenges pertaining to the scalability issue of Big Data. Also, this is what directly

TABLE I
5Vs FOR BIG DATA ANALYTICS.

Volume	Refers to the massive amounts of data that have been generated across a wide range of sources. In the context of data analytics, volume can be regarded as the product of instance size and dimensionality of the data.
Velocity	Refers to the rate when considering the acquisition and update of data. The fast-moving data thus yields an imperative need for frequent decision making using reliable computational intelligence (CI) technology.
Variety	Refers to the presence of multiple data types (or different feature representations) that take roots in various sources and range from text, image, audio, video, social media, to web logs, and so on.
Veracity	Considers the diverse quality and security levels of the data. Unhealthy sparsity, missing attributes and incomplete data are some of those considered under this category of interest.
Value	Refers to the benefits that can be gained from analysis on Big Data and the value of insights that can be extracted from Big Data. Very often, industrial players and researchers regard this as the core motivation and driver behind the study of Big Data analytics.

comes to our mind when we refer to the term “Big”. However, it is worth highlighting that, researchers in the data analytics community have largely taken a one-sided study of **volume** [4]–[6], which refers to the “Big instance size” factor of the data; the corresponding factor of “Big Dimensionality”, on the other hand, has received much lesser attention in the context of Big Data analytics [8]. To date, some studies on *high dimension small sample size* problems have been reported, such as random projection, Naïve Bayes, and others [9]–[12]. Theoretical efforts on Big Data with millions of dimensions have, however, remained relatively under-explored. In contrast to previous studies, in this article, we attempt to fill in the gap by putting focus on this under-explored topic of Big Data analytics – “Big Dimensionality”, wherein the explosion of features brings about new challenges to computational intelligence. In what follows, we begin with a peek through various stages of *feature life cycle*, namely feature description, feature selection and feature evaluation, as depicted in Figure 1. For each stage, we present a brief overview that focuses on the flip side of **volume** in Big Data – by putting spotlight on the emerging phenomenon of Big Dimensionality and the challenges ahead.

A. When Features Are Born

There are many ways to solve problems, and as many ways to describe them. Thus, the means of generating features will be markedly different due to the different representations that are of interests. Depending on one’s experience, knowledge and understanding of the domain, a **variety** of feature description methods and representations can be introduced. For example, when working with images, depending on one taking a bird’s eye view, a worm’s eye view or an eagle’s eye view,

Y. Zhai and Y.-S. Ong are with the Centre for Computational Intelligence (C²i), School of Computer Engineering (SCE), Nanyang Technological University (NTU), Singapore, 639798.

E-mail: {yzhai1, asysong}@ntu.edu.sg

I. W. Tsang is with the Centre for Quantum Computation & Intelligent Systems, University of Technology, Sydney, Australia, NSW 2007.

E-mail: ivor.tsang@gmail.com

global or local features can be derived and represented with a multi-view. This makes image feature description complex in structure, while abundance in **volume** and **variety**. Even for the evening news that we are endowed with daily, researchers and engineers working in the background have to face with similar challenges of operating with the different languages in presentation. In the field of natural language processing (NLP), practitioners have to work with multiple feature types such as words, bigram/trigram templates, part-of-speech tagging templates, etc., simultaneously, in order to arrive at comprehensive representations that produce reliable predictors [13]. This myriad of feature types and their mixture are becoming a norm in many of today's real world applications [14] and together with the rapid advancements in computing and information technologies, they are major contributors of feature explosion and continue to fuel Big Dimensionality.

B. Features Are Alike, Features Are Different

Everything has two sides, hence it is natural for some features to contribute alike, while some features differently. Since it is often the case that not all features carry equal weights on the prediction models and the application domains of interest, *feature selection* is the process of retrieving a subset of relevant features from the original feature space, for the purpose of building robust, accurate and fast learning models. Serving as the enabler for fast and cost-effective predictors, while keeping checks on the requirements in measurement and storage, feature selection is often regarded as one of the most important tasks in Big Data research. The emerging phenomenon of Big Dimensionality however calls for fresh feature selection strategies that are capable of coping with the explosion of features. Particularly, one has to deal with the explosive combinatorial effects of features or the “*curse of Big Dimensionality*”, while seeking to identify a good feature subset that is of high **value** from the original “Big” set comprising potentially irrelevant, redundant, noisy and missing features (uncertain). Due to the high importance of feature selection and many significant challenges that this stage is imbued with, it has remained and will remain to be one of the core research interests in the fields of computational intelligence and Big Data research¹.

C. Assessing A Good Feature Subset

Assessment is particularly essential to guide the search towards high **value** features in Big Data. Every search algorithm, except for uniform random search, introduces some kind of bias into its search. Different performance metrics used in the search for evaluations exhibit unique biases [17]. It is these biases that lead the search towards particular subset of features that differs from the others. In the last decade, multivariate performance measures have been regularly introduced for the reliable *evaluations* of features, leading to highly complex

¹Both the *IEEE International Conference on Data Mining (ICDM) 2013* and *IEEE World Congress on Computational Intelligence (WCCI) 2014*, organize workshops/special sessions that focus on the the issues of high dimensionality [15], [16], while seeking for notable feature selection strategies that are capable of addressing the “*curse of Big dimensionality*” and uncovering the potential “*blessing of Big dimensionality*”.

Fig. 1. Feature Life Cycle.

criteria for assessing predictive models [18]. Further, to *verify the authenticity* of the identified features for the application domains of interest, the availability of specialized human experts that are equipped with appropriate domain knowledge are essential. In this regard, a key technology that is helpful to human experts in data analytics is *visualization*. The presentation of data in different visual forms such as graphs, diagrams, charts, maps and other specialized means, can lead to easier and faster capturing of critical information as well as enhancing human understandings. Thus the technology that pushes the field forward would rely on the visualization of features, wherein a proper presentation can help in isolating the **values** of the features and subsequently figuring out the potential directions for further developments². However, traditional feature verification and visualization approaches are likely to become obsolete in the face of Big Dimensionality.

Contributions: From our survey, it is noted there has been a lack of studies that focus explicitly on analyzing the emerging trends of Big Dimensionality in the era of Big Data. The objective of this paper is specifically set out to fulfill such a role. We begin by concentrating on the influences of *advancing technology* and the arising *myriads of feature descriptors* on the origin of Big Dimensionality. An analysis on the evolution of feature dimensions is then conducted based on popular data repositories used in the data analytics and computational intelligence research communities. Subsequently, a review on the state-of-the-art feature selection methods in the field of computational intelligence reveals the insufficiencies of existing approaches in keeping pace with the explosion of dimensionality. Based on the analyses, the “*curse and blessing of Big Dimensionality*” are delineated. It is worth noting that such a study would be informative to the data analytics and computational intelligence communities since it underlines the emerging trend of Big Dimensionality and deliberates on the curse and blessing of such developments. Further, it is hoped that the acknowledgement on Big Dimensionality will serve to promote the need for greater research efforts in the subject and assist in identifying new important research directions.

²In the past, various new feature descriptors are inspired by feedbacks obtained from the verified models and visualized results [19], [20].

Fig. 2. A depiction on the evolving Pixel size of images produced by a series of advancing cell phone embedded cameras, beginning from the Year 2001 to present.

II. THE ORIGIN OF BIG DIMENSIONALITY

In this section, we focus on the key factors that accounts for the origin of Big Dimensionality. We begin with a study on the origin of features, by focusing on how they come about, how they are represented and then reveal the core bases for the upsurge in feature dimensions over the recent years as a result of the *advancements in technology* and the *myriad of feature descriptors* that have emerged. Here, we showcase the domain of image and video learning, since it is a popular domain of computer science as motivated by the rising popularity of the Internet and mobile devices. Subsequently, we present some new insights into the evolution of feature size (dimension) through analyzing three widely used data repositories of the data analytics and computational intelligence research communities.

A. Advancements in Technology

Today, the advancement in computing and information technologies is happening at a rate that is far beyond our expectations. In the cell phone manufacturing industry for example, the technology of the embedded camera is progressing by many factors each year as fuelled by the innovations in diverse avenues; ranging from the flashlight, processor, sensor size, photosensitive element to the operating system and technics (e.g., optical image stabilizer³, PureView⁴, etc.). The significant developments in the area of smart devices and image processing tools are empowering consumers with the capacity to generate extremely high resolution photos and video captures at anytime and anywhere, with great ease.

In the “*Cell Phone Activities 2012*” annual report [21], the portion of cell phone owners that uses their phone to take pictures was reported to have reached an astronomical rate of 82%, which is incidentally also the highest among all activities⁵ made on the phone in 2012. This is a rise of 6%

³http://www.usa.canon.com/cusa/consumer/standard_display/Lens_Advantage_IS

⁴<http://i.nokia.com/blob/view/-/2000652/data/4/-/Download-pureview-920.pdf>

⁵These activities include picture taking, sending/receiving text messages, accessing the internet, sending/receiving email, video recording, apps download, searching for health or medical information online and online banking.

from 2010 on such activities. With the growing popularity of online social networking services and free social media applications, including Facebook⁶ and Flickr⁷, photo-taking activities over the cell phone are expected to continue expanding at an accelerating rate.

When working with images, pixel is the basic cornerstone of data dimension considered by most feature descriptors (also known as feature generator or feature detector) and processing algorithms. Figure 2 summarizes the growth in the feature dimensions with respect to the pixel size (y-axis) and the resolution of the embedded-camera in cell phone, from the Year 2001 (*i.e.*, this is the year for the birth of camera phone) to present (x-axis). From the figure, there is a clear distinct indication of an exponentially increase in the feature dimensions (pixels) of camera phone generated images over the last decade. With respect to the NOKIA Lumia 1020 EOS, for instance, it is worth noting that presently we can easily enjoy up to an extremely high resolution of 41-megapixels on the pictures taken, which is 400 times more than the 0.11-megapixels image produced by the SHARP J-SH04 almost a decade ago. With much evidence that the rise in the dimensionality of data representation is set to continue, the emergence of Big Dimensionality is expected to further intensify the challenges in Big Data.

B. Myriad of Feature Descriptors

The advancements in computing devices and information technology have also led to new creations of features and representations, some of which can be highly detailed and sophisticated. In this Internet and mobile device dominated era, our way of life has been greatly influenced by the wide variety of media, services and applications that are now readily available online. It has been reported that multimedia content including text, image, 3D graphics, audio and video accounts for over 60% of traffic in the Internet [22].

Today, the escalation of users that enjoy spending their leisure time watching videos, browsing photos and sharing

⁶<https://www.facebook.com>

⁷<https://www.flickr.com>

Fig. 3. An illustration on the myriads of features descriptors for video content. A: The key frames of an online video; B: The different feature descriptors of image; C: Text features; D: Acoustic features; E: Motion features.

them online continue to drive the research efforts on data-centric media computing platforms that are capable of performing large-scale automatic analysis, understanding, summary, collection, organization, query and searching of multimedia content. Among the wide range of multimedia, online video has been found to account for more than half of the Internet traffic. Recently, the *YouTube Statistic* reported that 100 hours of video clips are constantly uploaded within every minute [23].

In the last decade, the pace in technological innovations of media formats and descriptors of video has risen significantly and there is much evidence that this trend will continue to rise. Video format, in particular, has evolved with increasing definition, resolution and content, advancing from 1080i, 720p, 1080p to 4K in a short period of only 5 years. In order to facilitate high accuracy learning of sophisticated video content, a myriad of feature descriptors have been introduced. In what

follows, we discussed some of the core feature descriptors of online video content that contribute to the explosion of data dimensionality.

Figure 3A, for example, shows three key frames of an online video taken from the online social media service provider, YouTube⁸. Each of the frame can be processed separately in the form of an image, as illustrated in Figure 3B. In image processing, from basic pixel features, researchers have embarked on the development of complex descriptions as an important step for further analysis. GIST or simply *Spatial envelope*, for instance, was introduced as a holistic descriptor that captures the core objects in the pictures [24], see Figure 3B-I. *Histogram of oriented gradients* (HOG), on the other hand, generates image features based on the gradient information of small cells that have been segregated in the image [25],

⁸<http://www.youtube.com/>

see Figure 3B-II. As HOG exhibits a silhouette of the original image, it is widely used for object detection of static imagery. Of equal importance is the *Scale-invariant feature transform* (SIFT) descriptor, which has been designed for image mapping, conducts a point-matching between different views of the same scenes [26]. As SIFT is invariant to translations, rotations and scalings in the image domain and robust to moderate perspective transformations and illumination variations, it is popular in the computer vision community, see Figure 3B-III. Subsequent extensions of SIFT include the *Speeded up robust features* (SURF) [27] and the *Gradient location and orientation histogram* (GLOH) [28] descriptors, which were designed for gains in speed and prediction accuracy, respectively. Other popular image descriptors include the *Quick shift* [29] and the *Difference of gaussians cornerness measure* [30], etc., whose feature representations are depicted in Figure 3B-IV and Figure 3B-V, respectively.

In addition to the image features, Sub-figures 3C, 3D and 3E depict the other typical forms of descriptors considered in online video learning [31]–[34], which include the motion information, audio information, (*i.e.*, acoustic feature families including *Mel-frequency cepstral coefficients*) and text information (*i.e.*, derived from the scripts and subtitles of video that were inserted by human), respectively.

With the ongoing surge in demand for enhanced user experiences and services over the Internet and mobile platforms, the pressure for highly accurate and fast processing of multimedia content involving myriads of feature descriptors can only continue to grow. As discussed, the rapid advancements in digital sensors have given birth to video images that can contain up to 4K resolution. With the myriads of feature descriptors that are available for representing video contents (*i.e.*, image, motion, acoustic and text), many millions of features or dimensions could easily transpire. Such a trend is non-isolated and can already be observed across many branches of applications. In life science research, for example, the search for a compact gene subset comprising tens of relevant biomarkers (genes) from the original thousands in microarray data is deemed as crucial to biologists before moving on to *in vitro* study [35], [36]. The rapid advancements in biotechnologies and biodevices, nevertheless, has given researchers the option of using Single-Nucleotide Polymorphism (SNP) (see Figure 4) as a new form of feature descriptor that defines the behaviors of genes. Note that this represents a quantum leap from the original thousands of features (genes) to millions of features (SNPs) that we now have to deal with. Similarly for the domain of natural language processing, the feature space is now made up of not only words but phrases or templates that appear in documents, tweet streams and webpages, which also extends to many millions of dimensions.

C. The Evolution of Feature Dimensions in Data-Centric Research

In this subsection, we present an analysis on the developments of feature dimensions by focusing on three popular computational intelligence and data analytics repositories, namely, *UC Irvine Machine Learning Repository* (UCI) [37],

TABLE II
CHARACTERISTICS OF THE DATASETS COLLECTED FROM THREE POPULAR COMPUTATIONAL INTELLIGENCE AND DATA ANALYTICS REPOSITORIES.
THE DATASETS ARE GROUPED ACCORDING TO THE DOMAIN OF APPLICATION.

(a) <i>UC Irvine Machine Learning Repository</i>			
Application Domain	Data Name	Dimension	Year
Acoustics	ISOLET	617	1994
	Chess	36	1989
game	Connect-4	42	1995
	Letter	16	1991
Image	Corel	89	1999
	Soybean	35	1987
Life Science	Molecular	58	1990
	Mammals	72	1992
	SPECTF	44	2001
	P53	5,409	2010
Multi-view	Internet AD	1,558	1998
Physics	Spectrometer	102	1988
	H ₂ O Treat. Plant	38	1993
Sociology	Insurance	86	2000
Text	Bag of words	100,000	2008
	URL	3,231,961	2009
Time-series	PEMS-SF	138,672	2011
	Gas Sensor	1,950,000	2013
Video	YouTube MVG	1,000,000	2013
(b) <i>UCI KDD Archive</i>			
Application Domain	Data Name	Dimension	Year
Demography	Internet Usage	22	1997
Life Science	E. coli	4,289	2001
Marketing	KDD1998	481	1999
Geology	Forest	54	1998
Text	Microsoft.com	294	1998
	NSF Abstracts	30,779	2003
Times-series	Control Chart	600	1999
(c) <i>LIBSVM Database</i>			
Application Domain	Data Name	Dimension	Year
Image	USPS	256	1994
	Gisette	5,000	2003
Life Science	Leukemia	7,129	1999
	Colon-cancer	2,000	1999
	Breast-cancer	7,129	2001
	News20	62,061	1995
Text	Real-sim	20,958	1998
	Sector	55,197	1998
	Rcv1	47,236	2004
	News20.binary	1,355,191	2005
	Webspam	16,609,143	2006
	SIAM	30,438	2007
	Log1p	4,272,227	2009
Education	KDD2010	29,890,095	2010

UCI KDD Archive (UCI KDD) [38] and *LIBSVM Database* (LIBSVM) [39], that have transpired in the last two decades. Further, we consider a collection of high dimensional datasets from the three repositories, whose detailed characteristics including the *year of creation*, *dimension size*, *data name* and *application domain*, are tabulated in Table II.

UCI KDD was originally introduced for use in large-scale data analytics research. This suggests why the datasets available in this archival are higher in dimensions relative to *UCI*. However, as the computational intelligence and data mining communities converge towards Big Data research, there is no longer a need to maintain the *UCI KDD* separately and was since merged with *UCI* from July, 2009. The *UCI* and *LIBSVM* repositories, cover a wide spectrum of real world datasets with various domains, ranging from game

Fig. 4. An illustration of different biomarker datasets that have been introduced in life science research from the Year 1999 to present. Note that a quantum leap from the use of genes as features (hundred thousands of dimensions) to the choice of SNP as feature descriptors (millions of dimensions) for the identification of relevant biomarkers across a range of BioInformatics or Medical Informatics related datasets can be observed at around Year 2004.

(Chess), image (Corel), life science (Leukemia), physics (Spectrometer), text (Webspam), time-series (Gas Sensor) to video (YouTube MVG) and others.

These data repositories are now becoming *de facto* benchmarks for conducting data analytics studies in many areas of computational intelligence, artificial intelligence, machine intelligence, data mining, soft computing, meta-heuristics and others. The *UCI* for instance, is among one of the top hundred most cited archives⁹ in all of computer science related publications, and continues to attract vast interests even today.

To gain understanding on the evolution of feature size (dimension) in data analytics research, we chart the dimensions of the datasets that has been used in the last two decades with respect to the year it was introduced. From Figure 5, an exponential increase in dimensionality can be observed across all three popular repositories considered in the early 2000s. For instance, the *News20.binary* is noted to have grown from ten of thousands of features (62,061) in 1995 (*News20*) to one with more than a million in dimension (1,355,191) in just a decade. This is a dramatic rise of more than 20 times. Thus, a simple forecast of the upward trend would reveal that, a feature dimensionality of up to 40 billion features is likely to arise by the year 2020.

III. THE CHALLENGES AND BLESSING OF BIG DIMENSIONALITY

The immense growth of feature dimensionality in data analytics has exposed the inadequacies of many computational intelligence methodologies that exist to date. Hence there is an urgent need for the conception of new paradigms and methodologies that can cope with the emerging phenomenon of Big Dimensionality. Correspondingly, how to solicit the key features to concisely represent the data and the prediction model well, while facilitating fast prediction and reduced storage, are among the important tasks of Big Data analytics. To this

end, we begin with a review on the state-of-the-art feature selection methods in this section. Subsequently, we proceed to reveal the core emerging challenges of feature selection when facing Big Dimensionality. Lastly, the “*blessing of Big dimensionality*” is also discussed.

A. Review of State-of-the-art Methods

Over the decades, a vast variety of feature selection methods have been proposed for handling different learning scenarios. Based on their learning paradigms, one can divide them into three categories – Filter methods, Wrapper methods and Embedded methods. Further, as feature correlation has been widely established as an important component of both feature and feature group selection, a subsection is also devoted to the topic.

1) *Filter Methods*: One simple way to select features is based on the ranking of features subject to some criteria. *Gain Ratio* [40] and *Chi-Square* [41] are among some of the earliest simple yet popular filter methods ever to be introduced. They rank features differently based on the criterion of information entropy or chi-square statistics. Such approaches however pay little attention on the feature interactions (*i.e.*, only the predictive information of the individual feature is considered, which is often referred to as *feature-level score*). Correspondingly, on applications such as microarray data analysis [42], [43] where genes possess much intrinsic linkages, such approaches often suffer from the presence of suboptimality (local optima).

To address such problems, instead of the simple *feature-level score*, direct measurements on the contributions of feature subsets have been considered. One such effort of *subset-level score* methods is the general graph-based feature selection framework under a *trace ratio* criterion [44]. As graph depicts the relationships among data in a natural and effective way, weighted undirected graphs can be readily used to incorporate the within-class and between-class information. Another notable correlation based filter is *minimum Redundancy Maximum Relevance* (mRMR) [45], which selects the correlated

⁹<http://archive.ics.uci.edu/ml/about.html>

features that contribute most to the class labels such that they are mutually far apart from each other. This is achieved by maximizing the dependency between the joint distribution of the selected features and the output labels. In summary, methods that consider the interactions among features have received very interesting and notable results on small dimensional data, but generally succumb under applications involving millions of features due to computations on feature interactions. One such example is the aforementioned YouTube MVG dataset found in the UCI repository. In general, many empirical studies have shown that filter methods do not scale well above tens of thousands of features.

2) *Wrapper and Embedded Methods*: When dealing with Big Dimensionality, scalability is a primary concern. Here we review some of the notable wrapper and embedded methods that have been shown to handle large feature size, *i.e.*, tens of thousands features, elegantly. Among them, *Support Vector Machine* (SVM) has been established to operate well on high dimensional classification problems [46]. Naturally, it is expected for numerous schemes using SVM for feature selection [47], [48] to showcase respectable prediction performance. For example, as a widely-studied wrapper approach in many different domains, the *SVM based on recursive feature elimination* (SVM-RFE) [36] operates by searching through all features individually, and eliminates the less important features iteratively. To solve high dimensional data efficiently, ℓ_1 -norm regularizer can be embedded into the SVM [49]–[51]. More recently, the *feature generating machine* [52] and *group discovering machine* [53] have been introduced with notable results on datasets comprising millions of dimensions. These frameworks operate with the *cutting plane algorithm* and *multiple kernel learning* to make coping with millions of dimensions feasible.

3) *Feature Correlation*: From our survey, feature correlation plays an essential role in filter, wrapper and embedded methods, as demonstrated in the context of biomarker networks [54] and natural language processing [13]. In particular, feature correlation plays a crucial part of *feature group selection* in computational intelligence. The *graph-guided fused lasso* is among the early approaches of feature group selection approaches and operates by identifying feature groups based on the graph-structure define over the features [55]. *Octagonal shrinkage and clustering algorithm for regression*, on the other hand, incorporate the ℓ_∞ -penalty so as to reduce similar feature pairs [56]. Further, Zhong *et al.* introduced an efficient projection step to accelerate the process of feature grouping [57]. With a non-convex optimization formulation [58], some amount of bias alleviation can be also achieved.

B. Emerging Challenges

From our analysis of the real-world datasets in popular repositories, there is little doubt that Big Dimensionality is setting upon us. The figure of dimensions in research studies on Big Data currently hovers around the million range (10^6). In Table II, it is worth noting that 7 out of 11 datasets that appeared in the last 8 years have dimensionality in the region of millions.

In this subsection, we begin with a discussion on some inadequacies of current computational intelligence methodologies, as they were not designed to cope with Big Dimensionality. Consequently, we highlight the imperative need for fresh studies on computational intelligence and feature selection paradigms that are proficient in dealing with the explosion of dimensionality and detail some of the core challenges that lies ahead.

1) *Millions of Dimensions and Beyond*: The field of “Big Data” was coined to place attention on the need for new ways in making sense of the unprecedented scale of data that are today becoming ubiquitous. In the same spirit, Big Dimensionality refers to the unprecedented number of features that is scaling to levels which now render existing state-of-the-art computational intelligence approaches inadequate. There is thus a pressing need for new approaches that can cope with this explosion of dimensionality.

In Big Dimensionality, scalability poses as the key challenge to many existing state-of-the-art methods. For instance, we cite the biomarker feature selection problem in life science as the illustrating example. The search for a compact subset of relevant biomarkers [59] from *single-nucleotide polymorphism* (SNP) is known to be critical to biologists in defining the behaviors of genes and their relevance to the disease of interests [35], [36], [60]. In the case of the *psoriasis* SNP dataset, which composes of only 0.5 million features, it took the state-of-the-art SVM-RFE and mRMR biomarker selectors more than a day of computational effort to crunch the data.

To gain further insights into the current state of feature selection research, we conducted an analysis on the dimensionality of datasets that have been used in the studies of computational intelligence, which is summarized in Figure 6. Particularly, the focus has been placed on the three flagship transactions and the magazine of the computational intelligence society, namely, the *IEEE Transactions on Evolutionary Computation* (TEC), *IEEE Transactions on Fuzzy Systems* (TFS) and *IEEE Transactions on Neural Networks and Learning Systems* (TNN¹⁰) and *IEEE Computational Intelligence Magazine* (CIM). By contrast, it is evident that as the dimensionality of the datasets continues to rise exponentially in time (see Figure 5), the complexity of the feature selection tasks being addressed began to overwhelm the algorithms proposed (see Figure 6) to date. Particularly, the dimensionality of the algorithms under-studied (as summarized in Figure 6) is significantly lagging behind those being produced (see Figure 5). From Figure 6, the statistical results further show that a majority (79.3%) of the studies analyzed remain confined to datasets with features that are less than 10,000 in dimensions. Notably, only 5.2% of the studies reported considered real world datasets with features that in the range of millions [13], [111], [114]. In summary, it is becoming clear that the explosion of dimensionality is pushing the capability limits of current computational intelligence algorithms.

2) *Handling Trillion Correlations*: With millions of features in hand, existing computational intelligence approaches

¹⁰Before 2012, this transactions was named as “*IEEE Transactions on Neural Networks*”.

Fig. 5. An illustration of the evolving feature size (dimension) in data analytics and computational intelligence datasets as introduced in the last two decades.

Fig. 6. Trends of the dataset dimension used in publications that appeared in the flagship journals (TNN, TFS and TEC) and magazine (CIM) of the computational intelligence society from Year 2010 to Year 2013 [13], [61]–[117].

that require the calculations of pairwise correlations in their algorithmic designs will have to cope with computations in the range of trillions¹¹. For example, a dataset with millions of features (10^6) would translate to trillions of pairwise correlations (10^{12}) that need to be computed. However, existing approaches that require the calculations of pairwise correlations in their algorithmic designs (e.g., filters) cannot cope with such datasets elegantly and often scored miserably, since computing

¹¹The number pairwise correlation computations is a squared of the dimensionality.

at such scale can be intractable. Note that this poses a grand challenge that has never been explicitly addressed in the field of computational intelligence and data mining research.

3) *Real Time Data Analytics*: With the prevalence of social media networks and portable devices, the demands for sophisticated portable device applications (e.g., video/image concept detection, interest scene detection, spam detection, sentiment analysis, etc.) in handling big **volumes** of multimedia content is rising. In such applications, real-time performance is of utmost importance to users, since no one is willing to spend

Fig. 7. The evolution (rise) of feature dimensionality in correlation matrices.

any time waiting nowadays. In other words, achieving real-time analysis and prediction on these Big Dimensionality is a new challenge of computational intelligence on portable platforms.

4) *Visualization That Matters*: To verify and authenticate the identified features, one classical technology that has always been helpful to human experts in data analytics is none other than “visualization”. Visual analytics in particular, has been defined in [118] as “*the science of analytical reasoning facilitated by interactive visual interfaces*”. A presentation of the diabetes dataset given in Figure 7(a) showcases a simple visualization of the correlations between feature pairs using a 2D correlation matrix. With only 8 features in the diabetes dataset, the 2D correlation matrix can be embedded with correlation coefficients (using pie-charts) and feature labels (using red fonts) information that would aid in enhancing human verifications and understandings of the data. On the lung cancer dataset, which has 7 times features than the diabetes dataset, a panoramic view of the 2D correlation matrix as given in Figure 7(b) remains legible, although the correlation coefficients and feature labels can no longer be included with ease. In the case of the psoriasis¹² dataset which comprises 529,651 SNPs (as features or dimensions), a visualization of the 2D correlation matrix involving 280.5 billion grids can hardly make any sense to a human user. Thus, as the feature size continues to grow and evolve towards the phenomenon of Big Dimensionality, fresh visualization technologies and tools that can equip decision makers with the flexibility to combine creativity and domain knowledge for the identification of features that contain valuable commercial **value**, from the bulk of big dimensional features, would be absolutely essential.

C. Blessing of Big Dimensionality

Besides the challenges (curse), in this section we review the potential benefits that are attributed by the presence of Big Dimensionality, which is less widely noted than the former [119].

¹²<http://www.sph.umich.edu/csg/abecasis/casp>

The results of our experimental study on correlation frequency involving the News20 (62,601 dimensions) and News20.binary (1,355,191 dimensions) corpora are summarized in Figure 8. The statistics obtained show that 99.88% and 99.39% of the feature pairs in News20 and News20.binary, respectively, have correlation coefficients that are lower than 0.1. This implies that a majority of the feature pairs are either uncorrelated or the correlated feature pairs are extremely sparse. Moreover, Figure 8 displays a downward trend in the number of correlated feature pairs, with increasing correlation threshold. Further, it can also be observed from the figure that the correlation frequencies of the feature pairs in News20.binary are noted to be generally lower than the pairs found in News20. This indication of features becoming more sparsely correlated as the dimension scales up clearly showcases a potential blessing of Big Dimensionality that one could leverage upon, since a majority of the uncorrelated feature pairs do not contribute to the correlation matrix.

IV. CONCLUDING REMARKS

In this article, the notion of “Big Dimensionality” has been introduced. In a similar spirit to “Big Data”, the term Big Dimensionality has been coined to put attention on the need for new ways in coping with the unprecedented number of features (dimensions) that are scaling to levels that now renders existing computational intelligence approaches inadequate. Our survey has revealed the lack of studies on the evolution of data dimensionality in the era of Big Data. In particular, our analysis on three popular data repositories has uncovered an exponential increase in the dimensionality of many datasets that have been produced since early 2000s. In life science research, for instance, a quantum leap from the original thousands of genes (features) to millions of Single-Nucleotide Polymorphism in a short period of time has been observed. And there is much evidence that the upward trend of Big Dimensionality will only continue to follow, as influenced by the rapid advancements in computing and information technologies and the arising myriads of feature descriptors, where a forecast of 40 billion features in dimensions is to be

Fig. 8. Correlation frequencies of feature pairs in the News20 (62,601 dimensions) and News20.binary (1,355,191 dimensions) corpora. Each bar in the figure denotes the percentage of feature pairs (the y-axis) that satisfies the corresponding correlation thresholds (the x-axis).

expected by year 2020. Based on our detailed analyses, it has been found that the progress of feature selection methods in the field of computational intelligence are falling very much behind the rapidly rising pace of data dimensionality or Big Dimensionality. Last but not least, the core challenges of feature selection (*curse of Big Dimensionality*) and the potential benefits of dimensionality explosion in feature selection (*blessings of Big Dimensionality*) have been presented and discussed.

Through this article, it is hoped that this acknowledgement on Big Dimensionality would serve to highlight the need for renewed research efforts in the era of Big Data. Hence, it is worth noting that an interesting line of thought is to use a “divide and conquer” approach to handle this explosion of Big Data and Dimensionality. Specifically, subset of features can be identified from disparate set (subset) of data where memes that have been perceived as building blocks of structured knowledge can be derived via *Deep Learning* [120], *Memetic Computation* [121] or otherwise, for improved scalability, applicability and reusability¹³. These atomized units of memes, metamemes, or memplex can then be expressed in hierarchical nested relationships or conceptual entities for higher-order learning [123], thus forming societies of the mind for more effective problem solving.

ACKNOWLEDGMENT

This work is partially supported under the A*STAR-TSRP funding, the Singapore Institute of Manufacturing Technology - Nanyang Technological University (SIMTech-NTU) Joint Laboratory and Collaborative Research Programme on Complex Systems, and the Centre for Computational Intelligence (C²i) at NTU.

REFERENCES

- [1] “IBM Big Data Success Stories,” IBM, USA, Tech. Rep., October 2011.

¹³From the statistics generated by Thomson Reuters that targets on world-class emerging research trends, memetic computation has been recently singled out as among the top 10 research fronts in the category of “Mathematics, Computer Science and Engineering” [122].

- [2] IBM Big Data Platform - Bringing Big Data to The Enterprise. IBM. [Online]. Available: <http://www.ibm.com/software/data/bigdata/>
- [3] P. C. Zikopoulos, D. deRoos, K. Parasuraman, T. Deutsch, D. Corrigan, and J. Giles, *Harness the Power of Big Data The IBM Big Data Platform*, R. B. Melnyk, Ed. McGraw-Hill, 2012.
- [4] D. E. O’Leary, “Artificial Intelligence and Big Data,” *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 96–99, 2013.
- [5] M. Ferguson, “Architecting A Big Data Platform for Analytics,” *A Whitepaper Prepared for IBM*, 2012.
- [6] J. Gantz and D. Reinsel, *Extracting Value from Chaos*, Std., June 2011. [Online]. Available: <http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf>
- [7] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, “Big Data: The Next Frontier for Innovation, Competition and Productivity,” McKinsey Global Institute, Tech. Rep., May 2011.
- [8] P. Kröger, “Going Big in Data Dimensionality - Challenges and Solutions in Mining High Dimensional Data,” in *Symposium on Scalable Analytics*, Garching bei München, November 2012.
- [9] A. Garg, S. Har-Peled, and D. Roth, “On Generalization Bounds, Projection Profile, and Margin Distribution,” in *ICML*, Sydney, Australia, 2002, pp. 171–178.
- [10] R. J. Durrant and A. Kabán, “Sharp Generalization Error Bounds for Randomly-projected Classifiers,” in *ICML*, Atlanta, USA, 2013, pp. 693–701.
- [11] P. J. Bickel and E. Levina, “Some Theory for Fishers Linear Discriminant Function, Naive Bayes, and Some Alternatives When There Are Many More Variables Than Observations,” *Bernoulli*, vol. 10, no. 6, pp. 989–1010, December 2004.
- [12] E. Candes and T. Tao, “The Dantzig Selector: Statistical Estimation When p Is Much Larger than n ,” *The Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, December 2007.
- [13] Q. Mao and I. W. Tsang, “Efficient Multi-Template Learning for Structured Prediction,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 2, pp. 248–261, 2013.
- [14] P. Gehler and S. Nowozin, “On Feature Combination for Multiclass Object Classification,” in *ICCV*, Kyoto, Japan, 2009.
- [15] (December) The 1st International Workshop on High Dimensional Data Mining (HDM). ICDM 2013. Dallas, Texas, USA. [Online]. Available: <http://www.cs.bham.ac.uk/~axk/HDM.htm>
- [16] Special Session for WCCI 2014 “EC Generalisation in High-dimensional Input Spaces”. WCCI 2014. [Online]. Available: <http://ieee-cis.blogspot.sg/2013/11/call-for-papers-special-session-for.html>
- [17] J. Huang and C. X. Ling, “Constructing New and Better Evaluation Measures for Machine Learning,” in *IJCAI*, Hyderabad, India, 2007.
- [18] Q. Mao and I. W. Tsang, “A Feature Selection Method for Multivariate Performance Measures,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2051–2063, 2013.
- [19] H. Leather, E. Bonilla, and M. O’Boyle, “Automatic Feature Generation for Machine Learning Based Optimizing Compilation,” in *CGO*, Seattle, WA, USA, March 2009, pp. 81–91.
- [20] Y. Sun and B. Bhanu, “Image Retrieval with Feature Selection and

- Relevance Feedback,” in *ICIP*, Hong Kong, China, 2010, pp. 3209–3212.
- [21] M. Duggan and L. Rainie, “Cell Phone Activities 2012,” Washington, D.C., Tech. Rep., November 2012.
- [22] W. Roush, “TR10: Peering into Video’s Future,” *MIT Technology Review* March 12, 2007, March. [Online]. Available: <http://www.technologyreview.com/Infotech/18284/>
- [23] (2013) YouTube Statistic. [Online]. Available: <http://www.youtube.com/yt/press/statistics.html>
- [24] A. Oliva and A. Torralba, “Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope,” *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.
- [25] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” in *CVPR*, San Diego, CA, USA, 2005, pp. 886–893.
- [26] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Key-points,” *International Journal of Computer Visions*, vol. 60, no. 2, pp. 91–110, November 2004.
- [27] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, “Speeded-Up Robust Features (SURF),” *Journal Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [28] K. Mikolajczyk and C. Schmid, “A Performance Evaluation of Local Descriptors,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [29] A. Vedaldi and S. Soatto, “Quick Shift and Kernel Methods for Mode Seeking,” in *ECCV*, Marseille, France, 2008, pp. 705–718.
- [30] J. L. Crowley and A. C. Parker, “A Representation for Shape Based on Peaks and Ridges in the Difference of Low Pass Transform,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 2, pp. 156–170, March 1984.
- [31] I. Laptev, M. Marszałek, C. Schmid, and B. Rozenfeld, “Learning Realistic Human Actions from Movies,” in *CVPR*, Anchorage, AK, USA, 2008.
- [32] G. Toderici, H. Aradhye, M. Pasca, L. Sbaiz, and J. Yagnik, “Finding Meaning on YouTube: Tag Recommendation and Category Discovery,” in *CVPR*, 2010.
- [33] H. Wang, A. Kläser, C. Schmid, and C.-L. Liu, “Action Recognition by Dense Trajectories,” in *CVPR*, Colorado Springs, USA, June 2011, pp. 3169–3176.
- [34] Z. Xu, Y. Yang, I. W. Tsang, N. Sebe, and A. Hauptmann, “Feature Weighting via Optimal Thresholding for Video Analysis,” in *ICCV*, Sydney, Australia, December 2013, pp. 3440–3447.
- [35] R. Díaz-Uriarte and S. A. de Andrés, “Gene Selection and Classification of Microarray Data Using Random Forest,” *BMC Bioinformatics*, vol. 7, no. 3, 2006.
- [36] I. Guyon, J. Weston, M. Stephen Barnhill, and V. Vapnik, “Gene Selection for Cancer Classification using Support Vector Machines,” *Machine Learning*, vol. 46, no. 1-3, pp. 389–422, 2002.
- [37] K. Bache and M. Lichman. (2013) UCI Machine Learning Repository. University of California, Department of Information and Computer Science. Irvine, CA, USA. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [38] S. Hettich and S. D. Bay. (1999) The UCI KDD Archive. University of California, Department of Information and Computer Science. Irvine, CA, USA. [Online]. Available: <http://kdd.ics.uci.edu>
- [39] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2(3), no. 27, 2011.
- [40] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann Publishers, Inc., 1993.
- [41] H. Liu and R. Setiono, “Chi2: Feature Selection and Discretization of Numeric Attributes,” in *International Conference on Tools with Artificial Intelligence*, Herndon, VA, USA, 1995, pp. 388–391.
- [42] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, “Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring,” *Science*, vol. 286, pp. 531–537, 1999.
- [43] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, and H. Liu, “Advancing Feature Selection Research,” Arizona State University, Tech. Rep., 2011.
- [44] F. Nie, S. Xiang, Y. Jia, C. Zhang, and S. Yan, “Trace Ratio Criterion for Feature Selection,” in *AAAI*, vol. 2, Chicago, Illinois, United States, 2008, pp. 671–676.
- [45] H. Peng, F. Long, and C. Ding, “Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [46] V. N. Vapnik, *Statistical Learning Theory*, S. Haykin, Ed. Wiley, September 1998.
- [47] A. L. Blum and P. Langley, “Selection of relevant features and examples in machine learning,” *Artificial Intelligence*, vol. 97, pp. 245–271, 1997.
- [48] I. Guyon and A. Elisseeff, “An Introduction to Variable and Feature Selection,” *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [49] J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani, “1-norm Support Vector Machines,” in *NIPS*, Vancouver, B.C., Canada, 2003.
- [50] P. Bradley and O. L. Mangasarian, “Feature Selection via Concave Minimization and Support Vector Machines,” in *ICML*. Madison, Wisconsin, United States: Morgan Kaufmann, 1998, pp. 82–90.
- [51] G. M. Fung and O. L. Mangasarian, “A Feature Selection Newton Method for Support Vector Machine Classification,” *Computational Optimization and Applications*, vol. 28, no. 2, pp. 185–202, July 2004.
- [52] M. Tan, L. Wang, and I. W. Tsang, “Learning Sparse SVM for Feature Selection on Very High Dimensional Datasets,” in *ICML*, Haifa, Israel, 2010, pp. 1047–1054.
- [53] Y. Zhai, M. Tan, I. W. Tsang, and Y.-S. Ong, “Discovering Support and Affiliated Features from Very High Dimensions,” in *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, J. Langford and J. Pineau, Eds. Edinburgh, Scotland, United Kingdom: Omnipress, July 2012, pp. 1455–1462.
- [54] J. T. Dudley and A. J. Butte, “Identification of Discriminating Biomarkers for Human Disease Using Integrative Network Biology,” *Pacific Symposium on Biocomputing*, pp. 27–38, 2009.
- [55] S. Kim and E. P. Xing, “Statistical Estimation of Correlated Genome Associations to a Quantitative Trait Network,” *PLoS Genet.*, vol. 5, no. 8, p. e1000587, 2009.
- [56] H. D. Bondell and B. J. Reich, “Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR,” *Biometrics*, vol. 64, pp. 115–123, March 2008.
- [57] L. W. Zhong and J. T. Kwok, “Efficient Sparse Modeling with Automatic Feature Grouping,” in *ICML*, Bellevue, WA, USA, 2011.
- [58] S. Yang, L. Yuan, Y.-C. Lai, X. Shen, P. Wonka, and J. Ye, “Feature Grouping and Selection Over an Undirected Graph,” in *KDD*, Beijing, China, August 2012.
- [59] S. Orrù, E. Giurelli, M. Casula, A. Loizadda, R. Murru, M. Mulargia, M. Masala, D. Cerimele, M. Zucca, N. Aste, P. Biggio, C. Carcassi, and L. Contu, “Psoriasis is associated with a SNP haplotype of the corneodesmosin gene (CDSN),” *Tissue Antigens*, vol. 60, no. 4, pp. 292–298, 2002.
- [60] A. J. Brookes, “The essence of SNPs,” *Gene*, vol. 234, no. 2, pp. 177–186, July 1999.
- [61] B.-Y. Sun, X. Zhang, J. Li, and X.-M. Mao, “Feature Fusion Using Locally Linear Embedding for Classification,” *IEEE Trans. Neural Netw.*, vol. 21, no. 1, pp. 163–168, 2010.
- [62] J. Gao, J. Zhang, and D. Tien, “Relevance Units Latent Variable Model and Nonlinear Dimensionality Reduction,” *IEEE Trans. Neural Netw.*, vol. 21, no. 1, pp. 123–135, 2010.
- [63] S. W. Lee and Z. Bien, “Representation of a Fisher Criterion Function in a Kernel Feature Space,” *IEEE Trans. Neural Netw.*, vol. 21, no. 2, pp. 333–339, 2010.
- [64] P. Pulkkinen and H. Koivisto, “A Dynamically Constrained Multiobjective Genetic Fuzzy System for Regression Problems,” *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 1, pp. 161–177, 2010.
- [65] Y. Aksu, D. J. Miller, G. Kesidis, and Q. X. Yang, “Margin-Maximizing Feature Elimination Methods for Linear and Nonlinear Kernel-Based Discriminant Functions,” *IEEE Trans. Neural Netw.*, vol. 5(21), pp. 710–717, May 2010.
- [66] Z. Zhu, S. Jia, and Z. Ji, “Towards a Memetic Feature Selection Paradigm,” *IEEE Comput. Intell. Mag.*, vol. 5, no. 2, pp. 41–53, May 2010.
- [67] L. Zhou, L. Wang, and C. Shen, “Feature Selection With Redundancy-Constrained Class Separability,” *IEEE Trans. Neural Netw.*, vol. 21, no. 5, pp. 853–858, 2010.
- [68] D. Lim, Y. Jin, Y.-S. Ong, and B. Sendhoff, “Generalizing Surrogate-assisted Evolutionary Computation,” *IEEE Trans. Evol. Comput.*, vol. 14, no. 3, pp. 329–355, May 2010.
- [69] M. J. Gacto, R. Alcalá, and F. Herrera, “Integration of an Index to Preserve the Semantic Interpretability in the Multiobjective Evolutionary Rule Selection and Tuning of Linguistic Fuzzy Systems,” *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 3, pp. 515–531, 2010.
- [70] K. Zhang and J. T. Kwok, “Clustered Nyström Method for Large Scale Manifold Learning and Dimension Reduction,” *IEEE Trans. Neural Netw.*, vol. 21, no. 10, pp. 1576–1587, 2010.

- [71] H. Huang and H. He, "Super-Resolution Method for Face Recognition Using Nonlinear Mappings on Coherent Features," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 121–130, 2011.
- [72] M. G. Epitropakis, D. K. Tasoulis, N. G. Pavlidis, V. P. Plagianakos, and M. N. Vrahatis, "Enhancing Differential Evolution Utilizing Proximity-Based Mutation Operators," *IEEE Trans. Evol. Comput.*, vol. 15, no. 1, pp. 99–119, 2011.
- [73] X. Yang, G. Zhang, J. Lu, and J. Ma, "A Kernel Fuzzy c-Means Clustering-Based Fuzzy Support Vector Machine Algorithm for Classification Problems With Outliers or Noises," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 1, pp. 105–115, 2011.
- [74] C.-H. Li, B.-C. Kuo, and C.-T. Lin, "LDA-Based Clustering Algorithm and Its Application to an Unsupervised Feature Extraction," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 1, pp. 152–163, 2011.
- [75] Z. Deng, K.-S. Choi, F.-L. Chung, and S. Wang, "Scalable TSK Fuzzy Modeling for Very Large Datasets Using Minimal-Enclosing-Ball Approximation," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 2, pp. 210–226, 2011.
- [76] R. N. Mahdi and E. C. Rouchka, "Reduced HyperBF Networks: Regularization by Explicit Complexity Reduction and Scaled Rprop-Based Training," *IEEE Trans. Neural Netw.*, vol. 22, no. 5, pp. 673–686, 2011.
- [77] J.-B. Yang and C. J. Ong, "Feature Selection Using Probabilistic Prediction of Support Vector Regression," *IEEE Trans. Neural Netw.*, vol. 22, no. 6, pp. 954–962, 2011.
- [78] H. K. Singh, A. Isaacs, and T. Ray, "A Pareto Corner Search Evolutionary Algorithm and Dimensionality Reduction in Many-Objective Optimization Problems," *IEEE Trans. Evol. Comput.*, vol. 15, no. 4, pp. 539–556, 2011.
- [79] R. Alcalá, M. J. Gacto, and F. Herrera, "A Fast and Scalable Multiobjective Genetic Fuzzy System for Linguistic Fuzzy Modeling in High-Dimensional Regression Problems," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 4, pp. 666–681, 2011.
- [80] B. Wang, S. Wang, and J. Watada, "Fuzzy-Portfolio-Selection Models With Value-at-Risk," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 4, pp. 758–769, 2011.
- [81] J. Alcalá-Fdez, R. Alcalá, and F. Herrera, "A Fuzzy Association Rule-Based Classification Model for High-Dimensional Problems With Genetic Rule Selection and Lateral Tuning," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 5, pp. 857–872, 2011.
- [82] F. Oveis, S. Oveis, A. Erfanian, and I. Patras, "Tree-Structured Feature Extraction Using Mutual Information," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 1, pp. 127–137, 2012.
- [83] Q. Hu, W. Pan, L. Zhang, D. Zhang, Y. Song, M. Guo, and D. Yu, "Feature Selection for Monotonic Classification," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 1, pp. 69–81, 2012.
- [84] Y. Huang, D. Xu, and F. Nie, "Semi-Supervised Dimension Reduction Using Trace Ratio Criterion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 3, pp. 519–526, 2012.
- [85] A. Stuhlsatz, J. Lippel, and T. Zielke, "Feature Extraction With Deep Neural Networks by a Generalized Discriminant Analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 4, pp. 596–608, 2012.
- [86] X. Li and X. Yao, "Cooperatively Coevolving Particle Swarms for Large Scale Optimization," *IEEE Trans. Evol. Comput.*, vol. 16, no. 2, pp. 210–224, 2012.
- [87] M. Antonelli, P. Ducange, and F. Marcelloni, "Genetic Training Instance Selection in Multiobjective Evolutionary Fuzzy Systems: A Coevolutionary Approach," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 2, pp. 276–290, 2012.
- [88] Y. Hou, I. Song, H.-K. Min, and C. H. Park, "Complexity-Reduced Scheme for Feature Extraction With Linear Discriminant Analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 6, pp. 1003–1009, 2012.
- [89] E. K. Burke, M. R. Hyde, and G. Kendall, "Grammatical Evolution of Local Search Heuristics," *IEEE Trans. Evol. Comput.*, vol. 16, no. 3, pp. 406–417, 2012.
- [90] T. Blackwell, "A Study of Collapse in Bare Bones Particle Swarm Optimization," *IEEE Trans. Evol. Comput.*, vol. 16, no. 3, pp. 354–372, 2012.
- [91] C.-W. Seah, I. W. Tsang, and Y.-S. Ong, "Transductive Ordinal Regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1074–1086, 2012.
- [92] Y.-C. Chen, N. R. Pal, and I.-F. Chung, "An Integrated Mechanism for Feature Selection and Fuzzy Rule Extraction for Classification," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 4, pp. 683–698, 2012.
- [93] F. Hassanzadeh, M. Collan, and M. Modarres, "A Practical Approach to R&D Portfolio Selection Using the Fuzzy Pay-Off Method," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 4, pp. 615–622, 2012.
- [94] L. W. Zhong and J. T. Kwok, "Efficient Sparse Modeling With Automatic Feature Grouping," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 9, pp. 1436–1447, 2012.
- [95] P. Padungweang, C. Lursinsap, and K. Sunat, "A Discrimination Analysis for Unsupervised Feature Selection via Optic Diffraction Principle," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 10, pp. 1587–1600, 2012.
- [96] M. Ramona, G. Richard, and B. David, "Multiclass Feature Selection With Kernel Gram-Matrix-Based Criteria," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 10, pp. 1611–1623, 2012.
- [97] K. Neshatian, M. Zhang, and P. Andreae, "A Filter Approach to Multiple Feature Construction for Symbolic Learning Classifiers Using Genetic Programming," *IEEE Trans. Evol. Comput.*, vol. 16, no. 5, pp. 645–661, 2012.
- [98] S. Xiang, F. Nie, G. Meng, C. Pan, and C. Zhang, "Discriminative Least Squares Regression for Multiclass Classification and Feature Selection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 11, pp. 1738–1754, 2012.
- [99] T. Hancock and H. Mamitsuka, "Boosted Network Classifiers for Local Feature Selection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 11, pp. 1767–1778, 2012.
- [100] T. C. Havens, J. C. Bezdek, C. Leckie, L. O. Hall, and M. Palaniswami, "Fuzzy c-Means Algorithms for Very Large Data," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 6, pp. 1130–1146, 2012.
- [101] G. Dudek, "An Artificial Immune System for Classification With Local Feature Selection," *IEEE Trans. Evol. Comput.*, vol. 16, no. 6, pp. 847–860, 2012.
- [102] J. Chen, Z. Ma, and Y. Liu, "Local Coordinates Alignment With Global Preservation for Dimensionality Reduction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 1, pp. 106–117, 2013.
- [103] D. K. Saxena, J. A. Duro, A. Tiwari, K. Deb, and Q. Zhang, "Objective Reduction in Many-Objective Optimization: Linear and Nonlinear Algorithms," *IEEE Trans. Evol. Comput.*, vol. 17, no. 1, pp. 77–99, 2013.
- [104] F. E. B. Otero, A. A. Freitas, and C. G. Johnson, "A New Sequential Covering Strategy for Inducing Classification Rules With Ant Colony Algorithms," *IEEE Trans. Evol. Comput.*, vol. 17, no. 1, pp. 64–76, 2013.
- [105] H. Xie and M. Zhang, "Parent Selection Pressure Auto-Tuning for Tournament Selection in Genetic Programming," *IEEE Trans. Evol. Comput.*, vol. 17, no. 1, pp. 1–19, 2013.
- [106] W. Ashlock and S. Datta, "Evolved Features for DNA Sequence Classification and Their Fitness Landscapes," *IEEE Trans. Evol. Comput.*, vol. 17, no. 2, pp. 185–197, 2013.
- [107] A. Kaban, "Fractional Norm Regularization: Learning With Very Few Relevant Features," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 6, pp. 953–963, 2013.
- [108] J. A. Sanz, A. Fernández, H. Bustince, and F. Herrera, "IVTURS: A Linguistic Fuzzy Rule-Based Classification System Based On a New Interval-Valued Fuzzy Reasoning Method With Tuning and Rule Selection," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 3, pp. 399–411, 2013.
- [109] U. Bhowan, M. Johnston, M. Zhang, and X. Yao, "Evolving Diverse Ensembles Using Genetic Programming for Classification With Unbalanced Data," *IEEE Trans. Evol. Comput.*, vol. 17, no. 3, pp. 368–386, 2013.
- [110] E. Lughofer and O. Buchtala, "Reliable All-Pairs Evolving Fuzzy Classifiers," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 4, pp. 625–641, 2013.
- [111] M. Tan, I. W. Tsang, and L. Wang, "Minimax Sparse Logistic Regression for Very High-Dimensional Feature Selection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24(10), pp. 1609–1622, 2013.
- [112] H. Izakian, W. Pedrycz, and I. Jamal, "Clustering Spatiotemporal Data: An Augmented Fuzzy C-Means," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 5, pp. 855–868, 2013.
- [113] A. Basak, S. Das, and K. C. Tan, "Multimodal Optimization Using a Biobjective Differential Evolution Algorithm Enhanced With Mean Distance-Based Selection," *IEEE Trans. Evol. Comput.*, vol. 17, no. 5, pp. 666–685, 2013.
- [114] G. Skolidis and G. Sanguinetti, "Semisupervised Multitask Learning With Gaussian Processes," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 12, pp. 2101–2112, 2013.
- [115] K. Subramanian, S. Suresh, and N. Sundararajan, "A Metacognitive Neuro-Fuzzy Inference System (McFIS) for Sequential Classification Problems," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 6, pp. 1080–1095, 2013.

- [116] W. Dong, T. Chen, P. Tino, and X. Yao, "Scaling Up Estimation of Distribution Algorithms for Continuous Optimization," *IEEE Trans. Evol. Comput.*, vol. 17, no. 6, pp. 797–822, 2013.
- [117] H. Someya, "Striking a Mean- and Parent-Centric Balance in Real-Valued Crossover Operators," *IEEE Trans. Evol. Comput.*, vol. 17, no. 6, pp. 737–754, 2013.
- [118] J. J. Thomas and K. A. Cook, *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center, 2005.
- [119] D. L. Donoho, "High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality," in *American Mathematical Society Conf. Math Challenges of the 21st Century*, 2000.
- [120] I. Arel, D. C. Rose, and T. P. Karnowski, "Deep Machine Learning - A New Frontier in Artificial Intelligence Research," *IEEE Comput. Intell. Mag.*, vol. 5, no. 4, pp. 13–18, November 2010.
- [121] X. Chen, Y.-S. Ong, M.-H. Lim, and K. C. Tan, "A Multi-Facet Survey on Memetic Computation," *IEEE Trans. Evol. Comput.*, vol. 15, no. 5, pp. 591–607, October 2011.
- [122] C. King and D. A. Pendlebury, "Research Fronts 2013," Tech. Rep., April 2013.
- [123] R. Meuth, M.-H. Lim, Y.-S. Ong, and D. C. Wunsch, "A Proposition on Memes and Meta-Memes in Computing for Higher-Order Learning," *Memetic Comp.*, vol. 1, no. 2, pp. 85–100, 2009.