ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

Byte The Bullet: Learning on Real-World Computing Architectures

Alessandro Ghio, Luca Oneto

DITEN - University of Genoa
Via Opera Pia 11A, I-16145, Genoa, Italy

Abstract.
Fast, effective, and reliable models: these are the desiderata of every theorist and practitioner. Machine Learning (ML) algorithms, proposed in the last decades, proved to be effective and reliable in solving complex real-world problems, but they are usually designed without taking into account the underlying computing architecture. On the contrary, the effort of contemplating the exploited computing device is often motivated by application-specific and real-world requirements, such as the need to accelerate the learning process with dedicated/distributed hardware, or to foster energy-sparing requirements of applications based on mobile standalone devices. The ESANN 2014 *Byte The Bullet: Learning on Real-World Computing Architectures* special session has pooled a compilation of the most recent proposals in this area, by encouraging submissions related to the development and the application of fast, effective, reliable techniques, which consider possibilities, potentialities and constraints of real-world computing architectures as basic cornerstones and motivations.

1 Introduction

World War II was almost reaching his apex when the term “information explosion” was first contemplated by the Oxford English Dictionary in 1941 [1]: the long story of modern Data Science started several decades before words like Big Data, Data Mining, Machine Learning or, even, Personal Computer became sexy. Generally speaking, people always collected data in different forms, leading to an often unnecessary overwhelming of stored information: with the advent of the first embryonic systems for automated data collection, this phenomenon became endemic, opening the door to a large number of studies concerning the explosion of the amount of available data [2, 3, 4, 5, 6, 7].

Indeed, automation of data collection was the main cornerstone, which scarred the beginning of a new era in the 60s: computers, tapes and disks allowed to store data at a pace which was becoming faster than human capacity of interpreting them [8, 9]. This issue has been amplifying for the last years, due to last decades advances in Information and Communication Technologies (ICT). Some examples of such key enabling technologies are: the diffusion of internet and social networks; the ever massive exploitation of low-cost distributed sensors (like the ones based on Radio Frequency IDentification – RFID – technology), also targeted towards home automation and domotics; the widespread use of smartphones [10, 11] and of apps, running on them (e.g. Foursquare or Facebook) [12, 13, 14, 15]; the improvements on database technologies for dealing with structured and unstructured data (e.g. SQL, NoSQL like Hadoop, NewSQL [16, 17]). Three open problems have been arising, as negative by-product consequences, i.e. issues related to:

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

(i) privacy

(ii) storage costs

(iii) the exploitation of the huge volume of collected data.

Concerning the first point, data gathering and analysis often snoop into people's private life (e.g. [18, 19, 20]): the 10,000% increase in Orwell's "1984" book sells after last June's National Security Agency leaks in the United States [21] shows how severely people perceive issues related to their own privacy. As a matter of fact, well-known Carnegie Mellon University Professor Tom Mitchell stated, in a *Science* paper [22, 23], that "The potential benefits of mining data are various; examples include reducing traffic congestion and pollution, limiting the spread of disease, and better using public resources such as parks, buses, and ambulance services. But risks to privacy from aggregating these data are on a scale that humans have never before faced." Privacy issues related to data collection and exploitation would deserve a special session by themselves: as these topics are out of scope with reference to *Byte The Bullet*, we can briefly trivialize that, as far as valuable services are offered to people (i.e. the final users), they will be more prone to accept kind of privacy "violations".

Concerning the second issue above, it is straightforward noting that gathering huge amounts of data leads to high costs, mainly in terms of Total Cost of Ownership (TCO), i.e. the purchase cost-per-TeraByte plus the direct and indirect operating costs (e.g. air conditioning, energy supply, etc.). Though we are discussing an issue different from privacy, the answer is somewhat similar to the one above: costs can be counterbalanced by transforming data into valuable services and offers for the final users, being them people, managers, IT specialists, or whoever.

Open problem (iii), listed above, thus clearly emerges as the core issue to deal with. As a matter of fact, famous cognitive scientist Noam Chomsky recently stated that: "Our problems are not lack of data, but understanding them" [24]. The main answer to such needs consists in approaches of learning from data, which have been clearly playing a central role since the 60s. At that time, the first analytics proposals blossomed: they mostly based on classic statistical tools, allowing to transform data into basic, though fundamental, business information [9]. Indeed, statistics have been contributing for a long time to derive business value from data, but they were not successful in answering complex questions of today's competitive world: conventional statistics reported measures of a phenomenon in a reactive setting; however, modern economy needs a more proactive approach, based on deriving precious high level patterns from data rather than limiting to reporting. Nevertheless, statistics represent the solid basis of disciplines, e.g. like Machine Learning (ML) and Data Mining (DM), which have been enabling learning from messy and large sets of data for the last decades, thus allowing to transform the gathering and collection of amounts of data from issues into potential energy. Learning from data, in this sense, allows converting potential energy into kinetic energy. This concept has been well depicted by European Consumer Commissioner Meglena Kuneva, who stated in 2009 [25] that "Data are the new oil": they are of limited use by themselves, but enclose an inestimable value if properly handled.

Choosing and/or designing the most suitable approach for learning is not a trivial task, as it depends on several and heterogeneous factors. The business value of the in-

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

formation derived from data is often underestimated by data scientists, but represents a key driving force. It mostly depends on the different stakeholders, final users, or interested individuals: personal applications, e.g. related to healthcare and wellbeing, allow to implement effective systems to improve home automation, safety and (feeling of) security of the user [26]; Business-to-Consumer (B2C) applications target personalized services for the final user (e.g. dedicated offers in retail [27], personalized medicine [28], fraud detection in credit cards [29]); Business-to-Business (B2B) applications aim at deriving high business value through, for example, pre-sales projects (e.g. customer segmentation [30] and retention [31]), condition based maintenance [32], supply chain optimization [33], etc. Moreover, other key factors are represented by, for example:

- the quantity of exploitable data: collecting, storing and/or using huge amounts of patterns could be impossible in some applications, e.g. due to (relatively) limited memory or computational power, or due to unaffordable costs related to data gathering;
- the quality of available data. Learning from data obviously requires a proper data management, in order to avoid "garbage-in, garbage-out" effects [34] as well as to reduce costs by sparing storage;
- the type of device where learning from data is performed, or where learned models are implemented and run (e.g. smartphones, sensors, PCs, multi-core architectures, cloud-computing servers shared by different applications, etc.). Note that model creation is often performed off-line on (eventually multi-core) PCs or servers, then it can be computationally weighed down in order to foster lighter runtime executions of learned models on smaller devices (e.g. [35]), if applicable. It is also worth highlighting that the choice of the destination device is a key decision in business: the way derived information is presented is often considered at least as important as its accuracy, in some applications [36];
- possible constraints on response times. In fact, the correctness of an output of a model is usually measured with reference to the so-called TAVA paradigm: that is, the extracted information must be Timing, Accurate, Valuable, and Actionable. In other words, this means that a correct response, outputted in an unfeasible time, is substantially comparable to an inaccurate one, as a delay can affect the (business) value and the actionability of the consequent decisions undertaken. Personal and B2C applications are usually run in real-time or on-demand, and require that a response is given to the user in a negligible or, at most, limited time: personalized marketing offers in retail represent a straightforward example. On the contrary, most of the B2B applications are run in batch mode and are characterized by more relaxed timing constraints (e.g. customer segmentation for marketing scopes).

These factors are pillars of a single framework for learning. For example, personal applications for human activity recognition usually exploit sparse models to allow runtime classification of activities on smartphones [10] or resource-limited devices [37]. On the contrary, B2B applications, like customer segmentation in retail, usually rely on

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

high-end servers[1] and billions of data, while time-scales are dilated. Cross-solutions (i.e. large data applications on resource-limited devices) are becoming popular as well: in this case, matrixes of hundreds of low-cost systems are used in B2C and B2B implementations, which use large amounts of data, to speed-up learning while restraining costs.

Consequently, depending on the setting defined for a particular problem, the most suitable approach for learning from data should be chosen or designed. The ensemble of the main characteristics and constrains of the targeted application can be generally defined as its *budget*. Clearly, a potentially infinite number of configurations for the budget can be derived, depending on the factors in play and on the values they can assume. Nevertheless, two macro-clusters of applications can be identified, which well depict most of real-world problems and offer a comprehensive taxonomy for this special session's scopes [38]:

- whenever the most important and/or the strictest budget constraints focus on the exploitation of a limited number of data (e.g. due to the computational limitations of the used device, because of the impossibility of gathering further samples, etc.), we refer to *small-scale problems*;
- whenever data abound and the most important and/or the strictest budget constraints are linked to execution/response time and to the business value of the generated output, we refer to *large-scale problems*.

Big Data Analytics is one of the most popular and widespread topics nowadays, so focussing on large-scale problems is straightforward; however, the going-smart tendency of the last decade, relying on stand-alone, portable and personal devices, has revamped the interest on the small-scale setting as well [39]. Moreover, as we will also discuss further-on in this paper and in the special session, cross-solutions, where part of the (or the whole) computational effort in large-scale applications is demanded to resource-limited devices, are assuming an ever increasing importance with reference to approaches for learning from data.

Independently of having to deal with a large- or small-scale problem, budget constraints always matter: in other words, data scientists have always to "bite the bullet" to extract the maximum value from bytes. This motivated us to organize the ESANN 2014 Special Session *Learning on Real-World Computing Architectures*, that we called, by exploiting a paraphrase, *Byte The Bullet*. This special session encouraged submissions related to the development and the application of fast, effective, reliable techniques for learning from data, which consider possibilities, potentialities and constraints of real-world computing architectures as basic cornerstones and motivations, i.e. with a clear focus on the contextualization on the actual available budget, as depicted above. In the following, we will give an overview of these research topics with a particular reference to Machine Learning (ML) techniques in Section 2. Section 3 will be finally devoted to the introduction of the works, accepted for the *Byte The Bullet* Special Session.

[1]For example, we report the minimum requirements for SAP HANA in–memory platform: it requires 128 GB RAM, eight 50 GB disks, and two 8–cores Intel CPUs. Further details can be retrieved at `http://www.saphana.com/docs/DOC-2382`.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence
and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

2 A Machine Learning Point of View on Small– & Large–Scale Learning

In this section, we briefly depict a general framework for learning from data, based on Machine Learning (ML) approaches, to which we will refer with respect to small-scale and large-scale problems.

As previously highlighted, the amount of collected data is growing faster than human capability to analyze them; we may also state that, now, the volume of data is growing faster than several devices capability to analyze them. Data storage is expensive, but deriving value from data could even become more expensive because of the necessary computing power. Conventional ML approaches demand computing resources (being them time, memory, computing power, or a combination/trade-off of them) that grow faster than the volume of data, leading computing resources to become the strictest bottleneck to implement efficient real-world applications [38, 40].

Bottou [40] introduces an interesting metaphor for what he defines the "computerized society": he identifies two types of computers in our society, i.e. *makers* and *thinkers*. Makers do business, generate revenue, and produce data proportionally to their activities; thinkers, on the contrary, analyze data to create revenue by finding competitive advantages. As the population of computers grows, the ratio between thinkers and makers must be bounded; however, data grow with the number of makers, while the number of thinkers does not grow as fast as data. In other words, the computing resources, available for learning purposes, do not grow faster than the volume of data, otherwise costs would exceed the business value of learning from data; however, conventional ML techniques conversely demand resources that grow faster than the volume of data.

Acting on the learning process is the key: depending on the targeted objectives and type of problem, one or more stages of the learning process can be interested. In particular, learning from data, according to the ML framework, usually consists of: the *training phase*, where models are created according to a learning procedure and the best possible one is chosen in the so-called *model selection* step [41]; the *feed forward phase*, where the trained model is implemented on the target device and deployed for runtime usage. Depending on the application and on the algorithm exploited for training purposes, the machine, where training is performed, and the final device for the feed forward phase could not coincide: in these cases, both computing requirements for training and implementation purposes should be contemplated, by properly tuning the effort between the two phases and the two exploited systems.

2.1 Small-Scale Problems

When dealing with small-scale problems, the main budget constraint consists in the amount of data, exploitable to derive, describe and/or use a model. This can be due to several factors: the costs related to data gathering and storing and the amount of available computing and storage resources are only some examples. Tackling small-scale problems thus involves dealing with the capability of realizing effective models while, at the same time, allowing to fit limited computational constraints: in other

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

words, working on small-scale applications is substantially equivalent to properly tailoring models and data to the target and resource-limited device (e.g. smartphones, sensors).

Such objective can be reached by acting on some or all the phases of the learning process. For example, the training phase can be directly performed on the destination device [42]; alternatively, models can be adapted in an *a–posteriori* fashion, before the feed forward phase is finally deployed, so to contemplate all the requirements and restrictions [39]. A novel research frontier consists, instead, in restructuring the whole learning procedure so to make it *device–friendly* from its basis: this represents a challenging objective, as it involves abating the boundaries among the domains of ML, statistics, signal processing, etc. However, recent results, such as the two presented in this special session [43, 44], are encouraging: they show how dealing with the whole learning process enables, respectively, reducing the number of features and building thrifty effective models for classification purposes, while sparing battery charge and avoiding device overheating.

2.2 Large-Scale Problems and Cross-Solutions

In large-scale problems, the computational complexity becomes the main budget constraint when envisioning the analysis of large amounts of data [38]: as a matter of fact, the analyses should be computed and completed in a limited amount of time to avoid compromising the timeliness and, consequently, the business value of the extracted information. Two roads can be driven to target such scope: (a) modify the learning algorithm (both related to training and run-time execution issues), so that it scales (sub)linearly with data; (b) reduce the amount of exploited samples. A combination of solutions (a) and (b) clearly represents the best trade-off, involving the analysis of the statistical benefits and the computational costs of contemplating an increasing volume of data.

While statistical aspects have been widely explored in literature (e.g. [45]), the computational complexity of learning procedures has seldom been taken into account [46]. One of the few trials was Valiant's learnability [47]: he excluded exponential time algorithms and relied only on polynomial time ones. However, as they are usually too slow, this theory led to few actual results. As an alternative, an analysis of approximate optimization was proposed [48, 46], where a framework, that takes into account the effects of approximation on learning procedures, has been developed.

Approximation is indeed one of the keywords in large-scale learning, altogether with parallelization (e.g. [49], that will be presented in this session). Whenever allowed by the nature of the problem and of the exploited algorithm, distributing the learning effort on different machines is key to allow limiting the computational burden related to the analysis of large data volumes. Nevertheless, costs could be remarkably affected by the exploitation of several parallel workstations: in order to deal with costs issues without giving up parallelism, a cross-fertilization with activities on learning on resource-limited devices has been allowing, in the last years, to exploit large matrixes of inexpensive devices in place of conventional computing architectures for large-scale learning (e.g. [50], accepted for this special session). The computational effort can be effectively demanded to these devices (e.g. Graphical Processing Units – GPUs); how-

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

ever, such applications are particularly challenging, as they marry the main difficulties afflicting both large-scale and small-scale learning.

3 Contributions to the ESANN 2014 Special Session on Learning on Real-World Computing Architectures

The *Byte the Bullet* special session collected research results from 4 groups, dealing with issues related to the implementation, speed-up and improvement of Machine Learning (ML) approaches, when the exploited computing architecture is taken into account. The topics, which the accepted papers deal with, cover the whole range of macro-clusters of the taxonomy, presented in the previous section: two papers cope with small-scale learning; one paper focuses on issues related to challenging cross-solutions; finally, one paper deals with large-scale learning. Each paper is briefly introduced in the following.

As discussed above, the implementation of Machine Learning (ML) algorithms on stand-alone small-scale devices allows incorporating in these systems new services and advanced functionalities without the need of resorting to remote computing systems. Nevertheless, small-scale systems suffer of issues related to their resource-limited nature, like limited battery capacity and processing power. In order to deal with such limitations, in [44], authors propose to merge two ideas that emerged in recent literature, i.e. local and bit-based hypothesis spaces, in order to build thrifty models. The authors show the effectiveness of their proposal, both in terms of accuracy and enhanced battery duration, by carrying out experiments on a smartphone in a Human Activity Recognition application [51].

When coping with small-scale learning, ML approaches can also be used to improve the efficiency of ultra-low-power sensor interfaces. The authors, in [43], show how adaptive feature extraction circuits can be assisted by hardware embedded learning to dynamically activate only most relevant features. This selection is done in a context and power cost-aware way, where also context dependence of different feature sets is explained. By presenting results obtained on a proof-of-principle Voice Activity Detector, enhanced with the proposed context- and cost-dependent voice/noise classifier, they show an average circuit power savings of 75%, with negligible accuracy loss.

Feature selection is an horizontal topic, which can be covered by analyzing it from different perspectives. It is the case of the paper [50], which deals with large-scale learning in astronomy and proposes a cross-solution, based on Graphical Processing Units (GPUs), to speed-up Nearest Neighbor (NN) models. In fact, their performance depends on the underlying metric, and in particular on the selection of a meaningful subset of informative features. In [50], the authors propose an efficient parallel implementation of incremental feature selection for NN models, which relies on the use of GPUs and shows a significant computational speed-ups over its sequential single-core competitor of up to two orders of magnitude. The applicability of the proposal is shown on the problem of the detection of distant galaxies in the universe.

Last but not least, the problem of distributed computing in large-scale learning is also faced by [49]. In this paper, the authors focus on the problem of designing efficient and fast K-Means clustering approaches, relying on online learning models: they

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

work in fully distributed, asynchronous networks without any central control, where a huge number of computational units is assumed. The authors base their analysis on the well-known Gossip Learning Framework (GoLF) [52], by proposing two variants that noticeably speed-up the convergence of the original algorithm. The authors' proposal is empirically benchmarked against several state-of-the-art distributed baseline algorithms in different computational scenarios, showing that the proposed approach is accurate, fast-converging, and robust against network failures.

References

[1] G. Press. A very short history of Big Data. `http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/`, 2013.

[2] D. J. De Solla Price and L. Weber. Science since babylon. *Physics Today*, 14:46, 1961.

[3] I. de Sola Pool. Tracking the flow of information. *Science*, 221(4611):609–613, 1983.

[4] P. J. Denning. Saving all the bits. *Americal Scientist*, 78:402–405, 1990.

[5] P. Lyman and H. Varian. How much information? `http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/`, 2003.

[6] Editorial. Community cleverness required. Nature, 2008.

[7] M. Hilbert. How to measure "how much information?". *International Journal of Communication*, 6:1042–1055, 2012.

[8] A. Gore. The digital earth: understanding our planet in the 21st century. *Australian surveyor*, 43(2):89–91, 1998.

[9] SQL Data Mining. History of Data Mining. `http://www.sqldatamining.com/index.php/data-mining-basics/history-of-data-mining`, 2012.

[10] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J.-L. Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2013.

[11] D. Lupton. Understanding the human machine. *IEEE Technology and Society Magazine*, 32(4), 2013.

[12] C. Shepard, A. Rahmati, C. Tossell, L. Zhong, and P. Kortum. Livelab: measuring wireless networks and smartphone users in the field. *ACM SIGMETRICS Performance Evaluation Review*, 38(3):15–20, 2011.

[13] P. Gilbert, B.G. Chun, L.P. Cox, and J. Jung. Vision: automated security validation of mobile apps at app markets. In *International workshop on Mobile cloud computing and services*, 2011.

[14] J. Lindqvist, J. Cranshaw, J. Wiese, J. Hong, and J. Zimmerman. I'm the mayor of my house: examining why people use foursquare-a social-driven location sharing application. In *Conference on Human Factors in Computing Systems*, 2011.

[15] D. Wang and Z. Xiang. The new landscape of travel: A comprehensive analysis of smartphone apps. In *Information and Communication Technologies in Tourism*, 2012.

[16] R. Cattell. Scalable sql and nosql data stores. *ACM SIGMOD Record*, 39(4):12–27, 2011.

[17] K. A. Doshi, T. Zhong, Z. Lu, X. Tang, T. Lou, and G. Deng. Blending sql and newsql approaches: Reference architectures for enterprise big data challenges. In *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, 2013.

[18] M. R. Wigan and R. Clarke. Big data's big unintended consequences. *IEEE Computer*, 46(6):46–53, 2013.

[19] S. Lohr. Big Data Is Opening Doors, but Maybe Too Many. `http://www.nytimes.com/2013/03/24/technology/big-data-and-a-renewed-debate-over-privacy.html?pagewanted=all&_r=1&`, 2013.

[20] P. Tucker. Has Big Data Made Anonymity Impossible? `http://www.technologyreview.com/news/514351/has-big-data-made-anonymity-impossible/`, 2013.

[21] C. Riley. Sales of Orwell's '1984' spike after NSA leak. `http://money.cnn.com/2013/06/12/news/1984-nsa-snowden/`, 2013.

[22] T. Mitchell. Mining our reality. *Science*, 326:1644–1645, 2009.

[23] E. Siegel. *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*. John Wiley & Sons, 2013.

[24] E. Rosenbaum. A brief history of big data, the Noam Chomsky way. `http://www.cnbc.com/id/101215758`, 2013.

[25] M. Kuneva. Roundtable on Online Data Collection, Targeting and Profiling – Keynote Speech. `http://europa.eu/rapid/press-release_SPEECH-09-156_en.htm`, 2009.

[26] J.-L. Reyes-Ortiz, A. Ghio, D. Anguita, X. Parra, J. Cabestany, and A. Catala. Human activity and motion disorder recognition: Towards smarter interactive cognitive environments. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2013.

[27] J. Han, M. Kamber, and J. Pei. *Data mining: concepts and techniques*. Morgan kaufmann, 2006.

[28] D. Wegener, S. Rossi, F. Buffa, M. Delorenzi, and S. Ruping. Towards an environment for data mining based analysis processes in bioinformatics & personalized medicine. In *IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW)*, pages 570–577, 2011.

[29] P. K. Chan, W. Fan, A. L. Prodromidis, and S. J. Stolfo. Distributed data mining in credit card fraud detection. *IEEE Intelligent Systems and their Applications*, 14(6):67–74, 1999.

[30] T. Jiang and A. Tuzhilin. Improving personalization solutions through optimal segmentation of customer bases. *IEEE Transactions on Knowledge and Data Engineering*, 21(3):305–320, 2009.

[31] K. Ng and H. Liu. Customer retention via data mining. *Artificial Intelligence Review*, 14(6):569–590, 2000.

[32] D. Raheja, J. Llinas, R. Nagi, and C. Romanowski. Data fusion/data mining-based architecture for condition-based maintenance. *International Journal of Production Research*, 44(14):2869–2887, 2006.

[33] E. H. Frazelle. *Supply chain strategy*. McGraw-Hill Trade, 2001.

[34] L. D. Paulson. Data quality: A rising e-business concern. *IT Professional*, 2(4):10–14, 2000.

[35] E. Alba, D. Anguita, A. Ghio, and S. Ridella. Using variable neighborhood search to improve the support vector machine performance in embedded automotive applications. In *IEEE International Joint Conference on Neural Networks*, 2008.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence
and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

[36] R. Sulo, S. Eick, and R. Grossman. Davis: a tool for visualizing data quality. *Posters Compendium of InfoVis*, 2005:45–46, 2005.

[37] Y. Hanai, J. Nishimura, and T. Kuroda. Haar-like filtering for human activity recognition using 3d accelerometer. In *Signal Processing Education Workshop Digital and Signal Processing Workshop*, 2009.

[38] L. Bottou and O. Bousquet. The tradeoffs of large scale learning. In *Neural Information Processing Systems*, 2008.

[39] D. Anguita, A. Ghio, S. Pischiutta, and S. Ridella. A support vector machine with integer parameters. *Neurocomputing*, 72(1):480–489, 2008.

[40] L. Bottou. The Tradeoffs of Large–scale Learning. `http://leon.bottou.org/talks/largescale`, 2008.

[41] D. Anguita, A. Ghio, L. Oneto, and S. Ridella. In-sample and out-of-sample model selection and error estimation for support vector machines. *IEEE Transactions on Neural Networks and Learning Systems*, 23(9):1390–1406, 2012.

[42] D. Anguita, A. Boni, and S. Ridella. A digital architecture for support vector machines: theory, algorithm, and fpga implementation. *IEEE Transactions on Neural Networks*, 14(5):993–1009, 2003.

[43] S. Lauwereins, K. Badami, W. Meert, and M. Verhelst. Context– and cost-aware feature selection in ultra–low–power sensor interfaces. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2014.

[44] D. Anguita, A. Ghio, L. Oneto, and S. Ridella. Learning with few bits on small–scale devices: From regularization to energy efficiency. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2014.

[45] V. N. Vapnik. *Statistical learning theory*. Wiley–Interscience, 1998.

[46] L. Bottou and O. Bousquet. The tradeoffs of large-scale learning. *Optimization for Machine Learning*, pages 351—368, 2011.

[47] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[48] S. Shalev-Shwartz and N. Srebro. Svm optimization: inverse dependence on training set size. In *International Conference on Machine learning*, 2008.

[49] A. Berta, I. Hegedűs, and R. Ormándi. Lightning fast asynchronous distributed k-means clustering. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2014.

[50] F. Gleseke, K. Polsterer, C. E. Oancea, and C. Igel. Speedy greedy feature selection: Better redshift estimation via massive parallism. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2014.

[51] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz. Energy efficient smartphone-based activity recognition using fixed-point arithmetic. *Journal of Universal Computer Science*, 19:1295–1314, 2013.

[52] R. Ormándi, I. Hegedűs, and M. Jelasity. Asynchronous peer-to-peer data mining with stochastic gradient descent. In *Euro-Par 2011 Parallel Processing*, pages 528–540, 2011.