

Chapter 1

Advances in Feature Selection for Data and Pattern Recognition: An Introduction

Urszula Stańczyk, Beata Zielosko and Lakhmi C. Jain

Abstract Technological progress of the ever evolving world is connected with the need of developing methods for extracting knowledge from available data, distinguishing variables that are relevant from irrelevant, and reduction of dimensionality by selection of the most informative and important descriptors. As a result, the field of feature selection for data and pattern recognition is studied with such unceasing intensity by researchers, that it is not possible to present all facets of their investigations. The aim of this chapter is to provide a brief overview of some recent advances in the domain, presented as chapters included in this monograph.

Keywords Feature selection · Pattern recognition · Data mining

1.1 Introduction

The only constant element of the world that surrounds us is its change. Stars die and new are born. Planes take off and land. New ideas sprout up, grow, and bear fruit, their seeds starting new generations. We observe the comings and goings, births and deaths, neglect and development, as we gather experience and collect memories, moments in time that demand to be noticed and remembered.

Human brains, despite their amazing capacities, the source of all inventions, are no longer sufficient as we cannot (at least not yet) grant anyone a right to take a

U. Stańczyk (✉)

Institute of Informatics, Silesian University of Technology, Akademicka 16,
44-100 Gliwice, Poland

e-mail: urszula.stanczyk@polsl.pl

B. Zielosko

Institute of Computer Science, University of Silesia in Katowice, Będzińska 39,
41-200 Sosnowiec, Poland

e-mail: beata.zielosko@us.edu.pl

L.C. Jain

University of Canberra, Canberra, Australia

e-mail: jainlakhmi@gmail.com

© Springer International Publishing AG 2018

U. Stańczyk et al. (eds.), *Advances in Feature Selection for Data and Pattern*

Recognition, Intelligent Systems Reference Library 138,

https://doi.org/10.1007/978-3-319-67588-6_1

direct peek at what is stored inside. And one of irresistible human drives is to share with others what we ourselves notice, experience, and feel. Mankind has invented language as means of communication, writing to pass on our thoughts to progeny and descendants, technologies and devices to help us in our daily routines, to seek answers to universal questions, and solve problems. Both human operators and machines require some instructions to perform expected tasks. Instructions need to be put in understandable terms, described with sufficient detail, yet general enough to be adapted to new situations.

A mother tells her child: “*Do not talk to strangers*”, “*take a yellow bus to school*”, “*when it rains you need an umbrella*”, “*if you want to be somebody, you need to study hard*”. In a factory the alarm bells ring when a sensor detects that the conveyor belt stops moving. In a car a reserve lights up red or orange when the gasoline level in a tank falls below a certain level. In a control room of a space centre the shuttle crew will not hear the announcement of count down unless all systems are declared as “go”. These instructions and situations correspond to recognition of images, detection of motion, classification, distinguishing causes and effects, construction of associations, lists of conditions to be satisfied before some action can be taken.

Information about environment, considered factors and conditions are stored in some memory elements or banks, retrieved when needed and applied in situations at hand. In this era of rapid development of IT technologies we can observe unprecedented increase of collected data, with thousands and thousands of features and instances. As a result, on one side we have more and more data, on the other side, we still construct and look for appropriate methods of processing which allow us to point out which data is essential, and which useless or irrelevant, as we need access to some established means of finding what is sought and in order to do that we must be able to correctly describe it, characterise it, distinguish from other elements [15].

Fortunately, advances in many areas of science, developments in theories and practical solutions come flooding, offering new perspectives, applications, and procedures. The constant growth of available ways to treat any concept, paths to tread, forces selection as an inseparable part of any processing.

During the last few years feature selection domain has been extensively studied by many researchers in machine learning, data mining [8], statistics, pattern recognition, and other fields [11]. It has numerous applications, for example, decision support systems, customer relationship management, genomic microarray analysis, image retrieval, image and motion detection, and text categorisation [33]. It is widely acknowledged that a universal feature selection method, applicable and effective in all circumstances, does not exist, and different algorithms are appropriate for different tasks and characteristics of data. Thus for any given application area a suitable method (or algorithm) should be sought.

The main aim of feature selection is the removal of features that are not informative, i.e., irrelevant or redundant in order to reduce dimensionality, discover knowledge, and explore stored data [1]. The selection can be achieved by ranking of variables according to some criterion or by retrieving a minimum subset of features that satisfy some level of classification accuracy. The evaluation of feature selection

technique or algorithm can be measured by the number of selected features, performance of learning model, and computation time [19].

Apart from the point of view of pattern recognition tasks, feature selection is also important with regard to knowledge representation [17]. It is always preferable to construct a data model which allows for simpler representation of knowledge stored in the data and better understanding of described concepts.

This book is devoted to recent advances in the field of feature selection for data and pattern recognition. There are countless ideas and also those waiting to be discovered, validated and brought to light. However, due to space restriction, we can only include a sample of research in this field. The book that we deliver to a reader consists of 14 chapters divided into four parts, described in the next section.

1.2 Chapters of the Book

Apart from this introduction, there are 14 chapters included in the book, grouped into four parts. In the following list short descriptions for all chapters are provided.

Part I Nature and Representation of Data

Chapter 2 is devoted to discretisation [10, 13]. When the entire domain of a numerical attribute is mapped into a single interval, such numerical attribute is reduced during discretisation. The problem considered in the chapter is how such reduction of data sets affects the error rate measured by the C4.5 decision tree [26] generation system using cross-validation. The experiments on 15 numerical data sets show that for a Dominant Attribute discretisation method the error rate is significantly larger for the reduced data sets. However, decision trees generated from the reduced data sets are significantly simpler than the decision trees generated from the original data sets.

Chapter 3 presents extensions of under-sampling bagging ensemble classifiers for class imbalanced data [6]. There is proposed a two phase approach, called Actively Balanced Bagging [5], which aims to improve recognition of minority and majority classes with respect to other extensions of bagging [7]. Its key idea consists in additional improving of an under-sampling bagging classifier by updating in the second phase the bootstrap samples with a limited number of examples selected according to an active learning strategy. The results of an experimental evaluation of Actively Balanced Bagging show that this approach improves predictions of the two different baseline variants of under-sampling bagging. The other experiments demonstrate the differentiated influence of four active selection strategies on the final results and the role of tuning main parameters of the ensemble.

Chapter 4 addresses recently proposed supervised machine learning algorithm which is heavily supported by the construction of an attribute-based decision graph (AbDG) structure, for representing, in a condensed way, the training set associated with a learning task [4]. Such structure has been successfully

used for the purposes of classification and imputation in both stationary and non-stationary environments [3]. The chapter provides the motivations and main technicalities involved in the process of constructing AbDGs, as well as stresses some of the strengths of this graph-based structure, such as robustness and low computational costs associated to both training and memory use.

Chapter 5 focuses on extensions of dynamic programming approach for optimisation of rules relative to length, which is important for knowledge representation [29]. “Classical” optimising dynamic programming approach allows to obtain rules with the minimum length using the idea of partitioning a decision table into subtables. Basing on the constructed directed acyclic graph, sets of rules with the minimum length can be described [21]. However, for larger data sets the size of the graph can be huge. In the proposed modification not the complete graph is constructed but its part. Only one attribute with the minimum number of values is considered, and for the rest of attributes only the most frequent value of each attribute is taken into account. The aim of the research was to find a modification of an algorithm for graph construction, which allows to obtain values of rule lengths close to optimal, but for the smaller graph than in “classical” case.

Part II Ranking and Exploration of Features

Chapter 6 describes an overview of reasons for using ranking feature selection methods and the main general classes of this kind of algorithms, with definitions of some background issues [30]. There are presented selected algorithms based on random forests and rough sets, and a newly implemented method, called Generational Feature Elimination (GFE) is introduced. This method is based on feature occurrences at given levels inside decision trees created in subsequent generations. Detailed information about its particular properties, and results of performance with comparison to other presented methods, are also included. Experiments were performed on real-life data sets as well as on an artificial benchmark data set [16].

Chapter 7 addresses ranking as a strategy used for estimating relevance or importance of available characteristic features. Depending on applied methodology, variables are assessed individually or as subsets, by some statistics referring to information theory, machine learning algorithms, or specialised procedures that execute systematic search through the feature space. The information about importance of attributes can be used in the pre-processing step of initial data preparation, to remove irrelevant or superfluous elements. It can also be employed in post-processing, for optimisation of already constructed classifiers [31]. The chapter describes research on the latter approach, involving filtering inferred decision rules while exploiting ranking positions and scores of features [32]. The optimised rule classifiers were applied in the domain of stylometric analysis of texts for the task of binary authorship attribution.

Chapter 8 discusses the use of a method for attribute selection in a dispersed decision-making system. Dispersed knowledge is understood to be the knowl-

edge that is stored in the form of several decision tables. Different methods for solving the problem of classification based on dispersed knowledge are considered. In the first method, a static structure of the system is used. In more advanced techniques, a dynamic structure is applied [25]. Different types of dynamic structures are analyzed: a dynamic structure with disjoint clusters, a dynamic structure with inseparable clusters, and a dynamic structure with negotiations. A method for attribute selection, which is based on the rough set theory [24], is used in all of the described methods. The results obtained for five data sets from the UCI Repository are compared and some conclusions are drawn.

Chapter 9 contains the study of knowledge representation in rule-based knowledge bases. Feature selection [14] is discussed as a part of mining knowledge bases from a knowledge engineer's and from a domain expert's perspective. The former point of view is usually aimed at completeness analysis, consistency of the knowledge base and detection of redundancy and unusual rules, while in the latter case rules are explored with regard to their optimization, improved interpretation and a way to improve the quality of knowledge recorded in the rules. In this sense, exploration of rules, in order to select the most important knowledge, is based in a great extent on the analysis of similarities across the rules and their clusters. Building the representatives for created clusters of rules bases on the analysis of the premises of rules and then selection of the best descriptive ones [22]. Thus this approach can be treated as a feature selection process.

Part III Image, Shape, Motion, and Audio Detection and Recognition

Chapter 10 explores recent advances in brain imaging technology, coupled with large-scale brain research projects, such as the BRAIN initiative in the U.S. and the European Human Brain Project, as they allow to capture brain activity in unprecedented detail. In principle, the observed data is expected to substantially shape the knowledge about brain activity, which includes the development of new biomarkers of brain disorders. However, due to the high dimensionality selection of relevant features is one of the most important analytic tasks [18]. In the chapter, the feature selection is considered from the point of view of classification tasks related to functional magnetic resonance imaging (fMRI) data [20]. Furthermore, an empirical comparison of conventional LASSO-based feature selection is presented along with a novel feature selection approach designed for fMRI data based on a simple genetic algorithm.

Chapter 11 introduces the notion of classes of shapes that have descriptive proximity to each other in planar digital 2D image object shape detection [23]. A finite planar shape is a planar region with a boundary and a nonempty interior. The research is focused on the triangulation of image object shapes [2], resulting in maximal nerve complexes from which shape contours and shape interiors can be detected and described. A maximal nerve complex is a collection of filled triangles that have a vertex in common. The basic approach is

to decompose any planar region containing an image object shape into these triangles in such a way that they cover either part or all of a shape. After that, an unknown shape can be compared with a known shape by comparing the measurable areas covering both known and unknown shapes. Each known shape with a known triangulation belongs to a class of shapes that is used to classify unknown triangulated shapes.

Chapter 12 presents an experimental study of several methods for real motion and motion intent classification (rest/upper/lower limbs motion, and rest/left/right hand motion). Firstly, EEG recordings segmentation and feature extraction are presented [35]. Then, 5 classifiers (Naïve Bayes, Decision Trees, Random Forest, Nearest-Neighbors, Rough Set classifier) are trained and tested using examples from an open database. Feature subsets are selected for consecutive classification experiments, reducing the number of required EEG electrodes [34]. Methods comparison and obtained results are given, and a study of features feeding the classifiers is provided. Differences among participating subjects and accuracies for real and imaginary motion are discussed. Chapter 13 is an extension of the work presented where the problem of classifying audio signals using a supervised tolerance class learning algorithm (TCL) based on tolerance near sets was first proposed [27]. In the tolerance near set method (TNS) [37], tolerance classes are directly induced from the data set using a tolerance level and a distance function. The TNS method lends itself to applications where features are real-valued such as image data, audio and video signal data. Extensive experimentation with different audio-video data sets was performed to provide insights into the strengths and weaknesses of the TCL algorithm compared to granular (fuzzy and rough) and classical machine learning algorithms.

Part IV Decision Support Systems

Chapter 14 overviews an application area of recommendations for customer loyalty improvement, which has become a very popular and important topic area in today's business decision problems. Major machine learning techniques used to develop knowledge-based recommender system, such as decision reducts, classification, clustering, action rules [28], are described. Next, visualization techniques [12] used for the implemented interactive decision support system are presented. The experimental results on the customer dataset illustrate the correlation between classification features and the decision feature called the promoter score and how these help to understand changes in customer sentiment.

Chapter 15 presents a discussion on an alternative attempt to manage the grids that are in intelligent buildings such as central heating, heat recovery ventilation or air conditioning for energy cost minimization [36]. It includes a review and explanation of the existing methodology and smart management system. A suggested matrix-like grid that includes methods for achieving the expected minimization goals is also presented. Common techniques are limited to central management using fuzzy-logic drivers, and redefining of the model is

used to achieve the best possible solution with a surplus of extra energy. In a modified structure enhanced with a matrix-like grid different ant colony optimisation techniques [9] with an evolutionary or aggressive approach are taken into consideration.

1.3 Concluding Remarks

Feature selection methods and approaches are focused on reduction of dimensionality, removal of irrelevant data, increase of classification accuracy, and improvement of comprehensibility and interpretability of resulting solutions. However, due to the constant increase of size of stored, processed, and explored data, the problem poses a challenge to many existing feature selection methodologies with respect to efficiency and effectiveness, and causes the need for modifications and extensions of algorithms and development of new approaches.

It is not possible to present in this book all extensive efforts in the field of feature selection research, however we try to “touch” at least some of them. The aim of this chapter is to provide a brief overview of selected topics, given as chapters included in this monograph.

References

1. Abraham, A., Falcón, R., Bello, R. (eds.): *Rough Set Theory: A True Landmark in Data Analysis*, Studies in Computational Intelligence, vol. 174. Springer, Heidelberg (2009)
2. Ahmad, M., Peters, J.: Delta complexes in digital images. Approximating image object shapes, 1–18 (2017). [arXiv:1706.04549v1](https://arxiv.org/abs/1706.04549v1)
3. Bertini Jr., J.R., Nicoletti, M.C., Zhao, L.: An embedded imputation method via attribute-based decision graphs. *Expert Syst. Appl.* **57**(C), 159–177 (2016)
4. Bi, W., Kwok, J.: Multi-label classification on tree- and dag-structured hierarchies. In: Getoor, L., Scheffer, T. (eds.) *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 17–24. ACM, New York, NY, USA (2011)
5. Błaszczyszki, J., Stefanowski, J.: Actively balanced bagging for imbalanced data. In: Kryszkiewicz, M., Appice, A., Ślęzak, D., Rybiński, H., Skowron, A., Raś, Z.W. (eds.) *Foundations of Intelligent Systems - 23rd International Symposium, ISMIS 2017, Warsaw, Poland, June 26–29, 2017, Proceedings. Lecture Notes in Computer Science*, vol. 10352, pp. 271–281. Springer, Cham (2017)
6. Branco, P., Torgo, L., Ribeiro, R.P.: A survey of predictive modeling on imbalanced domains. *ACM Comput. Surv.* **49**(2), 31:1–31:50 (2016)
7. Breiman, L.: Bagging predictors. *Mach. Learn.* **24**(2), 123–140 (1996)
8. Deutsch, I., Gediga, G.: *Rough set data analysis: a road to noninvasive knowledge discovery*. Mathoδos Publishers, Bangor (2000)
9. Dorigo, M., Gambardella, L.: Ant colony system: a cooperative learning approach to the traveling salesman problem. *IEEE Trans. Evolut. Comput.* **1**(1), 53–66 (1997)
10. Fayyad, U.M., Irani, K.B.: On the handling of continuous-valued attributes in decision tree generation. *Mach. Learn.* **8**(1), 87–102 (1992)

11. Fiesler, E., Beale, R.: *Handbook of Neural Computation*. Oxford University Press, Oxford (1997)
12. Goodwin, S., Dykes, J., Slingsby, A., Turkay, C.: Visualizing multiple variables across scale and geography. *IEEE Trans. Visual Comput. Graphics* **22**(1), 599–608 (2016)
13. Grzymała-Busse, J.W.: Data reduction: discretization of numerical attributes. In: Klösger, W., Zytkow, J.M. (eds.) *Handbook of Data Mining and Knowledge Discovery*, pp. 218–225. Oxford University Press Inc., New York (2002)
14. Guyon, I.: An introduction to variable and feature selection. *J. Mach. Learn. Res.* **3**, 1157–1182 (2003)
15. Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L. (eds.): *Feature Extraction: Foundations and Applications*. Springer, Heidelberg (2006)
16. Guyon, I., Gunn, S.R., Asa, B., Dror, G.: Result analysis of the NIPS 2003 feature selection challenge. In: *Proceedings of the 17th International Conference on Neural Information Processing Systems*, pp. 545–552 (2004)
17. Jensen, R., Shen, Q.: *Computational Intelligence and Feature Selection*. Wiley, Hoboken (2008)
18. Kharrat, A., Halima, M.B., Ayed, M.B.: MRI brain tumor classification using support vector machines and meta-heuristic method. In: *15th International Conference on Intelligent Systems Design and Applications, ISDA 2015, Marrakech, Morocco, December 14–16, 2015*, pp. 446–451. IEEE (2015)
19. Liu, H., Motoda, H.: *Computational Methods of Feature Selection*. Chapman & Hall/CRC, Boca Raton (2008)
20. Meszlényi, R., Peska, L., Gál, V., Vidnyánszky, Z., Buza, K.: Classification of fMRI data using dynamic time warping based functional connectivity analysis. In: *2016 24th European Conference on Signal Processing (EUSIPCO)*, pp. 245–249. IEEE (2016)
21. Moshkov, M., Zielosko, B.: *Combinatorial Machine Learning - A Rough Set Approach*, *Studies in Computational Intelligence*, vol. 360. Springer, Heidelberg (2011)
22. Nowak-Brzezińska, A.: Mining rule-based knowledge bases inspired by rough set theory. *Fundamenta Informaticae* **148**, 35–50 (2016)
23. Opelt, A., Pinz, A., Zisserman, A.: Learning an alphabet of shape and appearance for multi-class object detection. *Int. J. Comput. Vis.* **80**(1), 16–44 (2008)
24. Pawlak, Z.: Rough sets. *Int. J. Comput. Inf. Sci.* **11**, 341–356 (1982)
25. Przybyła-Kasperek, M., Wakulicz-Deja, A.: A dispersed decision-making system - the use of negotiations during the dynamic generation of a system's structure. *Inf. Sci.* **288**(C), 194–219 (2014)
26. Quinlan, J.R.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco (1993)
27. Ramanna, S., Singh, A.: Tolerance-based approach to audio signal classification. In: Khoury, R., Drummond, C. (eds.) *Advances in Artificial Intelligence: 29th Canadian Conference on Artificial Intelligence, Canadian AI 2016, Victoria, BC, Canada, May 31–June 3, 2016*, *Proceedings*, pp. 83–88. Springer, Cham (2016)
28. Raś, Z.W., Dardzinska, A.: From data to classification rules and actions. *Int. J. Intell. Syst.* **26**(6), 572–590 (2011)
29. Rissanen, J.: Modeling by shortest data description. *Automatica* **14**(5), 465–471 (1978)
30. Rudnicki, W.R., Wrzesień, M., Paja, W.: All relevant feature selection methods and applications. In: Stańczyk, U., Jain, L.C. (eds.) *Feature Selection for Data and Pattern Recognition*, pp. 11–28. Springer, Heidelberg (2015)
31. Stańczyk, U.: Selection of decision rules based on attribute ranking. *J. Intell. Fuzzy Syst.* **29**(2), 899–915 (2015)
32. Stańczyk, U.: Weighting and pruning of decision rules by attributes and attribute rankings. In: Czachórski, T., Gelenbe, E., Grochla, K., Lent, R. (eds.) *Proceedings of the 31st International Symposium on Computer and Information Sciences, Communications in Computer and Information Science*, vol. 659, pp. 106–114. Springer, Cracow (2016)
33. Stańczyk, U., Jain, L. (eds.): *Feature Selection for Data and Pattern Recognition*, *Studies in Computational Intelligence*, vol. 584. Springer, Heidelberg (2015)

34. Szczuko, P.: Real and imaginary motion classification based on rough set analysis of EEG signals for multimedia applications. *Multimed. Tools Appl.* (2017)
35. Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M.: Brainstorm: a user-friendly application for MEG/EEG analysis. *Intell. Neuroscience* **2011**(8), 8:1–8:13 (2011)
36. Utracki, J.: Building management system—artificial intelligence elements in ambient living driving and ant programming for energy saving—alternative approach. In: Piętka, E., Badura, P., Kawa, J., Wieclawek, W. (eds.) 5th International Conference on Information Technologies in Medicine, ITIB 2016 Kamień Śląski, Poland, June 20–22, 2016 Proceedings, vol. 2, pp. 109–120. Springer, Cham (2016)
37. Wolski, M.: Toward foundations of near sets: (pre-)sheaf theoretic approach. *Math. Comput. Sci.* **7**(1), 125–136 (2013)