Editorial on the special issue “Hybrid and ensemble techniques in soft computing: recent advances and emerging trends”

Przemysław Kazienko¹ · Edwin Lughofer² · Bogdan Trawinski¹

Published online: 28 October 2015
© Springer-Verlag Berlin Heidelberg 2015

The application of hybrid and ensemble methodologies in the field of soft computing (SC) and machine learning (ML) has become more visible and attractive. The relevance of these methodologies is motivated by their power of being able to express knowledge contained in data sets in multiple ways, benefiting each of the other, i.e., exploiting their diversity, thus increasing the performance of sole base models in terms of model accuracy and generalization capability by intelligent combination strategies, especially while dealing with high-dimensional complex regression and classification problems. Another main reason for their popularity is the high complementary of its components. The integration of the basic technologies into hybrid machine learning solutions facilitates more intelligent search, enhanced optimization, reasoning and hybridization methods that match various domain knowledge with empirical data to solve advanced and complex problems.

Both ensemble models and hybrid methods make use of the information fusion concept but in a slightly different way. In case of ensemble models, multiple but homogeneous, weak (base) models are combined, typically within boosting (Collins et al. 2002) and bagging approaches (Breiman and Predictors 1996), or in a more general form at the level of their individual outputs, using various fusion and combination methods (Kuncheva 2004), which can be grouped

into fixed (e.g., majority voting), and trained combiners (e.g., decision templates) (Sannen et al. 2010), exploiting model diversity on the one hand, and exploring data variation, as, e.g., caused by noise, on the other hand (Brazdil et al. 2009).

Hybrid methods, in turn, combine completely different, heterogeneous soft computing and/or machine learning approaches, seeking for homogenous solutions (Castillo and Melin 2009; Wozniak 2014), e.g., neural networks combined with evolutionary strategies for multi-objective approximation Sher (2012) or difference equation problems, or genetic fuzzy systems (Cordon et al. 2001; Fazzolari et al. 2013) for providing a reasonable interpretability/accuracy tradeoff within optimization cycles (Casillas et al. 2003). Usually, they are applied for complex optimization problems within the field of data-driven model-based design which cannot be solved with classical analytical or standard machine learning techniques (Mitchell 1997).

Both, ensemble learning and hybrid approaches, may considerably improve quality of reasoning and boost performance as well as robustness of the entire solutions (Brazdil et al. 2009; Polikar 2006). For that reason, ensemble and hybrid methods have found applications in numerous real-world problems ranging from person recognition, through medical diagnosis, bioinformatics, recommender systems and text/music classification to financial forecasting (Okun 2009).

When crawling through the isi web of knowledge database established through Thomson Reuters, we could find 37,928 publications during the last 15 years (2000–2015) dealing with hybrid or ensemble techniques when restricting the search to “Articles” and the field of “Computer Science” resp. 47,643 publications over the same time span dealing with hybrid or ensemble models, largely overlapping to the aforementioned ones. This underlines the wide acceptance of

✉ Edwin Lughofer
edwin.lughofer@jku.at

Przemysław Kazienko
kazienko@pwr.wroc.pl

Bogdan Trawinski
bogdan.trawinski@pwr.wroc.pl

¹ Wrocław University of Technology, Wrocław, Poland

² Johannes Kepler University of Linz, Linz, Austria

Fig. 1 Number of articles published dealing with hybrid AND ensemble models/techniques

and demands to such techniques in the research community as well as industry.

When performing a search over a combination of these two, only 427 publications have been found whose development over the last 15 years can be seen in the picture below. This clearly shows an increasing trend of publications in that particular field, basically due to an increased importance level within various data mining and machine learning tasks (Fig. 1).

In this special issue, the intension was to follow this trend and make it more transparent to the soft computing community by drawing a broad picture of recent advances of hybrid and ensemble methods in soft computing and machine learning and especially a combination of both, emphasizing the usage of fuzzy systems, neural networks, and all types of evolutionary algorithms (genetic algorithms, memetic algorithms, differential evolution, particle swarm optimization, etc., to name a few), to employ them as base learners for ensembles and within particular hybridization schemes (e.g., neuro-fuzzy systems, neuro-evolution). Multi-objectivity will play a central role in all hybrid schemes, where any form of evolutionary algorithms is employed.

A specific focus is placed on intelligent fusion strategies which are going far beyond pure (weighted) majority voting, thus also including some trainability and cascability in terms of base learners combination and confidence level output strategies. In this context, model selection may play a crucial role to remove any superfluous information from the ensembles. Stability plays an important role, especially when base learners may be weak (in case of boosting approaches, for instance) or the noise level is high. Interpretability may become an important issue, e.g., in case of human-machine interaction systems, where humans may interact with the system in an enriched context, significantly going beyond monitoring purposes and providing plain feedback in the form of rewards.

A specific emphasis of this special issue is also given by a recently emerging trend in the research field of hybrid and ensemble techniques, and that is, methods and algorithms which are able to perform **online processing on data streams** (Gama 2010), supporting **step-wise adaptation of model ensembles in incremental manner** as well as **evolving components** (Lughofer 2011; Angelov et al. 2010).

When searching *isi web of knowledge* by connecting hybrid and ensemble models and techniques with phrases *incremental*, *online*, we could find only a couple of articles published so far. This underlines that this particular research direction is still under-represented in the community.

In the context of incremental, adaptive ensemble and hybrid techniques both, temporal as well as spatial adaptation capabilities, would be of interest, that is, being able to mining model ensembles and hybrid systems in a data stream mining context (temporal case) as well as in a spatial data site mining context (spatial case). The former case leads to the possibility to use the novel methods in fast online real-world applications such as sequential video analysis, online system identification in multiple sensor networks, time-series analysis and prediction, the latter to the possibility to use them in VLDBs (very large data bases), huge web mining or cloud computing environments.

Incrementality plays a key role to prevent cost-intensive re-training cycles and thus to keep modeling efforts smart. Strong dynamic aspects and drifting situations (Klinkenberg 2004; Shaker and Lughofer 2014), as, e.g., caused by new operating modes, changing system characteristics or non-stationary environmental influences (Sayed-Mouchaweh and Lughofer 2012), are challenges to be appropriately handled on demand and integrated on the fly into the ensembles and hybrid models.

Another emerging path within this research field addresses active and semi-supervised learning tasks for hybrid models or model ensembles to reduce efforts and costs for operators and machines, as reducing the number of requested target values for model updates. Methods in this direction have been indeed studied in several approaches before (Lughofer 2012), but using conventional models without applying any ensemble schemes.

Additionally, the “dynamization” of evolutionary algorithms (EAs), respectively induced hybrid approaches when being coupled with fuzzy systems (termed as genetic fuzzy systems; Cordon 2011) or with neural networks (termed as neuro-evolution Sher 2012), is still a highly challenging issue to cope with stream mining demands or quickly changing environments. This is because conventional EAs are usually pretty slow to be reasonably applied within online applications and/or non-stationary environments.

In summary, this special issue draws a round picture of the recent advances in hybrid and ensemble methods within different learning environments, supporting static, dynamic or

online processing as well as guaranteeing robustness of the solutions in case of high-dimensional data with significant noise. Under this scope, it includes several contributions dealing with real-world applications that demand hybridization and ensemble techniques to improve robustness and performance.

In sum, 56 papers have been submitted, out of which 21 papers could be finally accepted after several revision rounds based on careful reviews by various international well-known researchers in this particular field of research—we are grateful for their great support and in large parts quick responses during the whole reviewing process.

The papers can be finally sorted into five major groups of research (in the order of their appearance), appearing partially as sole batch, partially as adaptive online approaches:

- Pure ensemble modeling techniques (Papers 1–4)
- Hybrid ensembles (Papers 5 and 6)
- Hybrid neural network based approaches (Papers 7–9)
- Hybrid models for forecasting, partially with fuzzy aspects (Papers 10–13)
- Hybrid evolutionary algorithm basic (Papers 14–17) and applied (Papers 18–20)

Acknowledgments At the end, we hope that this special issue sheds light on some novel ideas of hybrid and ensemble methods. In particular, we would like to gratefully acknowledge and sincerely thank all the reviewers for their insightful comments and criticism of the manuscripts. Our thanks go also to the authors for their contributions and collaboration. Finally, we are grateful to the Editor-in-chiefs *Antonio di Nola* and *Vincenzo Loia* and the editorial office assistants of *Soft Computing* for their great support during the review and production process of this special issue. The second guest editor also acknowledges the LCM GmbH for financial support, which, as a part of K2 project, is financed using funding from the Austrian COMET-K2 programme and supported by the Austrian federal government, the federal state of Upper Austria, the Johannes Kepler University and all of the scientific partners which form part of the K2-COMET Consortium.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- Angelov P, Filev D, Kasabov N (2010) *Evolving intelligent systems—methodology and applications*. Wiley, New York
- Brazdil P, Giraud-Carrier C, Soares C, Vilalta R (2009) *Metalearning*. Springer, Berlin

- Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140
- Casillas J, Cordon O, Herrera F, Magdalena L (2003) Interpretability issues in fuzzy modeling. Springer, Berlin
- Castillo O, Melin P (2009) Hybrid soft computing models for systems modeling and control. In: Meyers RA (ed) *Encyclopedia of complexity and systems science*. Springer, New York, pp 4696–4713
- Collins M, Schapire R, Singer Y (2002) Logistic regression. Adaboost and Bregman distance. *Mach Learn* 48(1–3):253–285
- Cordon O, Herrera F, Hoffmann F, Magdalena L (2001) Genetic fuzzy systems. evolutionary tuning and learning of fuzzy knowledge bases, *Advances in Fuzzy Systems: Applications and Theory*. World Scientific
- Cordon O (2011) A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: designing interpretable genetic fuzzy systems. *Int J Approx Reason* 52(6):894–913
- Fazzolari M, Alcalá R, Nojima Y, Ishibuchi H, Herrera F (2013) A review of the application of multiobjective evolutionary fuzzy systems: current status and further directions. *IEEE Trans Fuzzy Syst* 21(1):45–65
- Gama J (2010) *Knowledge discovery from data streams*. Chapman & Hall/CRC, Boca Raton
- Klinkenberg R (2004) Learning drifting concepts: example selection vs. example weighting. *Intell Data Anal* 8(3):281–300
- Kuncheva L (2004) *Combining pattern classifiers: methods and algorithms*. Wiley, Chichester
- Lughofer E (2011) *Evolving fuzzy systems—methodologies. Advanced concepts and applications*. Springer, Berlin
- Lughofer E (2012) Hybrid active learning (HAL) for reducing the annotation efforts of operators in classification systems. *Pattern Recognit* 45(2):884–896
- Mitchell TM (1997) *Machine learning*. McGraw-Hill International Editions, Singapore
- Okun O (2009) *Applications of supervised and unsupervised ensemble methods*. Springer, Berlin
- Polikar R (2006) Ensemble based systems in decision making. *IEEE Circuits Syst Mag* 6(3):21–45
- Sannen D, Lughofer E, Van Brussel H (2010) Towards incremental classifier fusion. *Intell Data Anal* 14(1):3–30
- Sayed-Mouchaweh M, Lughofer E (2012) *Learning in non-stationary environments: methods and applications*. Springer, New York
- Shaker A, Lughofer E (2014) Self-adaptive and local strategies for a smooth treatment of drifts in data streams. *Evol Syst* 5(4):239–257
- Sher GI (2012) *Handbook of neuroevolution*. Springer, Berlin
- Wozniak M (2014) *Hybrid classifiers—methods of data. Knowledge and classifier combination*. Springer, Berlin