

# Collective Intelligent Information and Database Systems

Ngoc-Thanh Nguyen<sup>a</sup> Manuel Núñez<sup>b</sup> and Bogdan Trawiński<sup>a</sup>

<sup>a</sup> Faculty of Computer Science and Management

Wroclaw University of Science and Technology, Poland

E-mail: {Ngoc-Thanh.Nguyen,Bogdan.Trawinski}@pwr.edu.pl

<sup>b</sup> Faculty of Computer Science

Complutense University of Madrid, Spain

E-mail: manuelnu@ucm.es

Collective intelligence is most often understood as group intelligence which arises on the basis of intelligences of the group members. On the one hand, Newell [19] defined an intelligent collective as a social system, which is capable to act, even approximately, as a single rational agent. On the other hand, Lévy [11] understood collective intelligence as an intelligence that emerges from the collaboration and competition of many individuals; an intelligence that seemingly has a mind of its own. These definitions refer to cognitive systems. From the computational and artificial intelligence point of view, we can think about a collective as a set of autonomous units working on some common task, for example, a multi-agent system. We can say that the collective is intelligent if it can make a use of the intelligences of its members for solving some problem, for example, a decision making problem. Thus Computational Collective Intelligence (CCI) should provide computational methods which are based on the collective intelligence aspects and which use the computational techniques for solving these problems. Regarding knowledge engineering, computational collective intelligence provides methods and techniques for determining the knowledge of a collective as a whole on the basis of collective members' knowledge. The need for processing collective knowledge is quickly increasing because of the very fast development of Internet, social networks and distributed databases. It is obvious that knowledge originating from autonomous sources for the same subject is very often inconsistent. Therefore, the aspects of inconsistency processing and integration computing are very important.

*Knowledge engineering* [9] plays a relevant role in CCI since it is necessary to use its techniques in order to represent individual information. In fact, there are studies advocating that scientific knowledge is essentially collective knowledge [23]. In this line, recent studies state that it is very important that different individuals provide orthogonal, highly unrelated, and possible contradictory knowledge to the collectivity. In other words, “the higher the inconsistency, the better the quality of collective knowledge” [20]. Another relevant interaction with knowledge engineering is ontology matching and integration [27], a fundamental activity in order to collect and process information in CCI. Since it is well known that this process is both time and resource-consuming, it is necessary to provide advanced architectures and platforms to reduce these costs [15] and good algorithms and frameworks to efficiently integrate ontologies [16,21].

Although *machine learning* and *data mining* are two independent fields of work, there are frequent interactions between them [18]. Interestingly enough CCI uses machine learning and data mining in its solutions but there are also numerous applications of CCI to improve machine learning and data mining processes. A first issue concerns the representation of data in machine learning, given the fact that efficient algorithms strongly rely on a good structure of data. Actually, the research community organizes contests to compare existing approaches and to identify future challenges [7] and, as a result, new highly efficient methods, beating previous learning methods, are developed to improve representation learning [30]. Given the huge amount of data generated by current information systems, it

is a must to use good machine learning algorithms to discover knowledge in these vast datasets. In order to reduce the amount of *useful* data, new techniques take advantage of the improvement in performance provided by parallel algorithms to process distributed data [31].

Learning from data streams is an area of increasing importance [14] and new approaches are needed to perform data stream mining and, again due to the huge amount of processed data, classify the received data, according to similarity, with good performance [22, 17]. Data stream mining uses prediction models to deal with historical data. However, these models lose accuracy if they do not frequently update *old* data with current data, so that it is necessary to deal with *concept drift* and adapt to it [5]. In particular, in order to cope with this problem when classifying streamed data, new methods must take into account temporal dependencies and splitting criteria based on misclassifications [33,24].

Another area of interest, highly related to the issue of processing huge amounts of unstructured data, is *text processing and information retrieval*. This line of research is currently very active because of the ubiquity of social networks, an obvious area of interest for CCI. Information extraction in tweets is specially challenging because traditional natural language algorithms cannot be used to process them. In particular, *named entities recognition* (NER) is difficult to perform because, due to their short nature, single tweets do not provide enough information and novel clustering techniques must be used [13]. Interestingly enough, similar problems with NER appear in other disciplines, most notably in Medicine where active learning methods are successfully used [3]. In fact, clustering methods are widely used to detect communities in social networks, where members involved in similar social objects are grouped [32], although some researchers advocate that the existence of random factors needs to be taken into account so that statistical *fuzzy* approaches are more suitable [12].

*Intelligent information systems* are very useful in areas where vast amounts of heterogenous, and usually unstructured information, must be processed. These systems are gaining popularity in Medicine, in particular when used to evaluate the quality of health care systems [1] and monitor the progress of patients [29]. In the latter case, it is necessary to take into account the special nature of vital signs so that the best performance to predict patient conditions can be obtained when using a fuzzy model.

*Intelligent database systems* play a role in many different areas. In particular, they can be used in Computer Science to improve existing methodologies. This is the case of using databases to help in the replication of experiments in Software Engineering [2,6]. This is an area where, again, statistic information is frequently used to construct robust methods [10]. This is specially relevant if datasets are non-normal.

The next kind of systems that we consider within the topics covered in this special issue are *decision support systems*. They are particular relevant in the context of CCI if we consider them in the scope of decision making and knowledge engineering, most notably, in the context of decisional DNA [25], a structure suited to obtain knowledge in decision making processes, and virtual engineering objects [26].

We finish this brief overview of the field with the application of *computer vision techniques* to video surveillance and object detection. This is a line of work where huge amounts of data must be adequately processed and analyzed. Since data are observed and/or collected from distributed locations and intelligent decisions must emerge, in particular when an imminent danger is recognized, this field is in the scope of CCI. The number of cameras installed in public areas is steadily increasing with the consequent increase of the amount of images to be processed. One of the current concerns is the detection of abandoned objects because of the threat that they can represent. Therefore, new robust and efficient algorithms are being developed, taking into account varying circumstances such as lighting changes [28]. Similarly, the detection of pedestrians, as well as the objects that they carry, is an active line of work [4,8] where probabilistic approaches are widely used to decide whether the pedestrian is carrying potentially dangerous artifacts.

The aim of this special issue is to present to the research community a comprehensive collection of articles including the most relevant and recent achievements in the broad field of Collective Intelligent Information and Database Systems. We have been able to cover most methodological, theoretical and practical aspects of Collective Intelligence, and its relation to databases, understood as the form of intelligence that emerges from the collaboration and competition of many individuals (artificial and/or natural). This special issue includes, in particular, extended and revised versions of papers selected from the 2016 edition of the ACIIDS conference and the 2015 edition of the ICCI conference. In addition, we called for high qual-

ity, up-to-date contributions in the broad field of Collective Intelligent Information and Database Systems.

The topics of interest for the special issue are those considered in the intersection between the ICCCI and ACIIDS Conference series. The papers in this special issue are distributed according to the following specific categories:

- Knowledge engineering and semantic web (papers 1–9).
- Text processing and information retrieval (papers 10 and 11).
- Machine learning and data mining (papers 12–25).
- Social networks and recommender systems (paper 26).
- Agent and multi-agent systems (paper 27).
- Intelligent information systems (papers 28–30).
- Database systems and software engineering (papers 31 and 32).
- Decision support and control systems (papers 33–37).
- Computer vision techniques (papers 38 and 39).

After a careful reviewing process, we have selected 39 papers to conform this special issue. All submitted papers, including the extended versions of conference papers, were peer-reviewed and selected on the basis of quality and relevance to the special issue.

We would like to thank the authors of the submitted papers for their interest in the special issue and the high quality of their contributions. They are the most important piece to conform a relevant and interesting scientific work. We would also like to thank the members of the Guest Editorial Board, and their subreviewers, because their careful work and dedication have been fundamental for the success of this special issue. The list of members can be found at <https://sites.google.com/site/sejifs2016/guest-editorial-board>. Finally, we would like to thank Van Du Nguyen (Wroclaw University of Technology) for his help with the web site.

## References

- [1] A. Aktas, S. Cebi, and I. Temiz. A new evaluation model for service quality of health care systems based on AHP and information axiom. *Journal of Intelligent & Fuzzy Systems*, 28(3):1009–1021, 2015.
- [2] J. C. Carver, N. Juristo, M. T. Baldassarre, and S. Vegas. Replications of software engineering experiments. *Empirical Software Engineering*, 19(2):267–276, 2014.
- [3] Y. Chen, T. A. Lasko, Q. Mei, J. C. Denny, and H. Xu. A study of active learning methods for named entity recognition in clinical text. *Journal of Biomedical Informatics*, 58:11–18, 2015.
- [4] D. Damen and D. Hogg. Detecting carried object from sequences of walking pedestrians. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(6):1056–1067, 2012.
- [5] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, and A. Bouchachia. A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4):article 44, 2014.
- [6] O. S. Gómez, N. Juristo, and S. Vegas. Understanding replication of experiments in software engineering: A classification. *Information & Software Technology*, 56:1033–1048, 2014.
- [7] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. C. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R.-T. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio. Challenges in representation learning: A report on three machine learning contests. *Neural Networks*, 64:59–63, 2015.
- [8] V.-D. Hoang, L. M. Ha, and K.-H. Jo. Hybrid cascade boosting machine using variant scale blocks based HOG features for pedestrian detection. *Neurocomputing*, 135(7):357–366, 2014.
- [9] S. L. Kendal and M. Creen. *An introduction to knowledge engineering*. Springer, 2007.
- [10] B. A. Kitchenham, L. Madeyski, D. Budgen, J. Keung, P. Brereton, S. Charters, S. Gibbs, and A. Pohthong. Robust statistical methods for empirical software engineering. *Empirical Software Engineering*, in press, 2016.
- [11] P. Lévy. *Collective Intelligence*. Plenum/Harper Collins, 1994.
- [12] H. J. Li. The comparison of significance of fuzzy community partition across optimization methods. *Journal of Intelligent & Fuzzy Systems*, 29(6):2707–2715, 2015.
- [13] X. Liu and M. Zhou. Two-stage NER for tweets with clustering. *Information Processing & Management*, 49(1):264–273, 2013.
- [14] N. Lopes and B. Ribeiro. *Machine Learning for Adaptive Many-Core Machines - A Practical Approach*, volume 7 of *Studies in Big Data*. Springer, 2015.
- [15] Y. Lv, Y. Ni, H. Zhou, and L. Chen. Multi-level ontology integration model for business collaboration. *The International Journal of Advanced Manufacturing Technology*, 84(1):445–451, 2016.
- [16] M. Maleszka and N. T. Nguyen. A method for complex hierarchical data integration. *Cybernetics and Systems*, 42(5):358–378, 2011.
- [17] D. Mena Torres and J. S. Aguilar-Ruiz. A similarity based approach for data stream classification. *Expert Systems with Applications*, 41(9):4224–4234, 2014.
- [18] R. S. Michalski, I. Bratko, and M. Kubat, editors. *Machine Learning and Data Mining: Methods and Applications*. Wiley, 1998.
- [19] A. Newell. *Unified theories of cognition*. Harvard University Press, 1990.
- [20] V. D. Nguyen, N. T. Nguyen, and H. B. Truong. A preliminary analysis of the influence of the inconsistency degree on the quality of collective knowledge. *Cybernetics and Systems*,

- 47(1-2):69–87, 2016.
- [21] M. Pietranik and N. T. Nguyen. A multi-attribute based framework for ontology aligning. *Neurocomputing*, 146:276–290, 2014.
- [22] S. Pramod and O. P. Vyas. Data stream mining: A review on windowing approach. *Global Journal of Computer Science and Technology Software & Data Engineering*, 12(11):26–30, 2012.
- [23] J. Ridder. Epistemic dependence and collective scientific knowledge. *Synthese*, 191(1):37–53, 2014.
- [24] L. Rutkowski, M. Jaworski, L. Pietruczuk, and P. Duda. A new method for data stream mining based on the misclassification error. *IEEE Transactions on Neural Networks and Learning Systems*, 26(5):1048–1059, 2015.
- [25] C. Sanín, C. Toro, Z. Haoxi, E. Sánchez, E. Szczerbicki, E. Carrasco, W. Peng, and L. Mancilla-Amaya. Decisional DNA: A multi-technology shareable knowledge structure for decisional experience. *Neurocomputing*, 88:42–53, 2012.
- [26] S. I. Shafiq, C. Sanín, C. Toro, and E. Szczerbicki. Virtual engineering object (VEO): Toward experience-based design and manufacturing for industry 4.0. *Cybernetics and Systems*, 46(1-2):35–50, 2015.
- [27] P. Shvaiko and J. Euzenat. Ontology matching: State of the art and future challenges. *IEEE Transactions on Knowledge & Data Engineering*, 25(1):158–176, 2013.
- [28] Y. L. Tian, R. Feris, H. Liu, A. Humpapur, and M.-T. Sun. Robust detection of abandoned and removed objects in complex surveillance videos. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(5):565–576, 2011.
- [29] C. Vallejos de Schatz, F. Schneider, P. Abatti, and J. Nievola. Dynamic fuzzy-neural based tool for monitoring and predicting patients conditions using selected vital signs. *Journal of Intelligent & Fuzzy Systems*, 28(6):2579–2590, 2015.
- [30] Y. Yang and Q. M. J. Wu. Multilayer extreme learning machine with subnetwork nodes for representation learning. *IEEE Transactions on Cybernetics*, in press, 2016.
- [31] A. A. Yildirim, C. Özdoğan, and D. Watson. Parallel data reduction techniques for big datasets. In W.-C. Hu and N. Kaabouch, editors, *Big Data Management, Technologies, and Applications*, pages 72–93. IGI Global, 2014.
- [32] Z. Zhao, S. Feng, Q. Wang, J. Z. Huang, G. J. Williams, and J. Fan. Topic oriented community detection through social objects and link analysis in social networks. *Knowledge-Based Systems*, 26:164–173, 2012.
- [33] I. Zliobaite, A. Bifet, J. Read, B. Pfahringer, and G. Holmes. Evaluation methods and decision theory for classification of streaming data with temporal dependence. *Machine Learning*, 98(3):455–482, 2015.