

Design, implementation and validation of AI-inspired information systems

Alfredo Cuzzocrea

Published online: 1 November 2013
© Springer Science+Business Media New York 2013

While there is an emerging and always-growing interest for novel paradigms appeared recently (e.g., *social networks*, *Cloud computing*, *NoSQL databases*, *Big Data*, and so forth), *Artificial Intelligence* (AI) always plays a critical role in *next-generation Information Systems*. Indeed, as technology and paradigms pervade our life, there is a challenging need for smarter and more sophisticated Information Systems, for instance using innovative methodologies like *crowdsourcing*. As a consequence, it is natural to foresee the advancement of a novel class of Information Systems, which we call as *AI-Inspired Information Systems*. Basically, these are Information Systems which incorporate in their critical layers (i.e., *design, implementation, validation*) AI methodologies, yet extending their roots to classical foundations, with, indeed, exciting innovations.

From this main evidence, there emerges a great interest in methodologies that can improve the variegated aspects of Information Systems life-cycle, with special emphasis on the *modeling, representation, querying, retrieval, reasoning, mining* and *end-user phases*. On the other hand, AI-inspired approaches tend to be computationally expensive, so that complexity analysis of proposed solutions must be rigorously evaluated and reasonable execution bounds should be derived. As a consequence, it naturally follows that designing, implementing and validating AI-Inspired Information Systems not only conveys in the need for elegant and scalable approaches, but also these approaches need to be theoretically-sound and exposing bounded complexities.

Along this line, this special issue on “*Design, Implementation and Validation of AI-Inspired Information Systems*” of *Journal of Intelligent Information Systems* focuses on latest research results and open research challenges on the problem of designing, implementing and validating Information Systems, based on AI paradigms, according to principles and guidelines provided above. With the aim of adequately fulfilling both theoretical and practical issues of the investigated research context, this special issue contains five papers, which have gone through two rigorous review rounds before being accepted for the final inclu-

A. Cuzzocrea (✉)
ICAR-CNR & University of Calabria, Via P. Bucci, 41C, 87036 Rende, Cosenza, Italy
e-mail: cuzzocrea@si.deis.unical.it
URL: <http://si.deis.unical.it/~cuzzocrea>

sion. Some of the contributions of this special issue have been invited for submission as best papers from the session “*Advanced Knowledge-based Systems*” of the *15th International Conference on Knowledge-based and Intelligent Information & Engineering Systems* (KES 2011), held in Kaiserslautern, Germany, during September 12–14, 2011, led by the Guest Editor. In the following, we provide a summary of the papers contained in this special issue.

The first paper, titled “*Consistent Structuring of Inconsistent Knowledge*”, by Mark Burgin and Kees (C.N.J.) de Vey Mestdagh, focuses the attention on *inconsistency handling*, which is recognized as one of the central problems in many areas of AI. In literature, there are different approaches to deal with contradictions and other types of inconsistency. Following this main trend, authors develop an approach based on *logical varieties and prevarieties*, which are complex structures constructed from *logical calculi*. Being locally isomorphic to a logical calculus, *globally logical varieties* form a logical structure, which allows representation of inconsistent knowledge in a consistent way and provides much more flexibility and efficacy for AI than standard logical methods. Problems of logical variety immersion into a logical calculus are studied as well. Such immersions extend the local structure of a logical calculus to the global structure of a logical variety. The obtained results demonstrate when it is possible to use standard logical tools, such as logical calculi, and when it is necessary to go beyond this traditional technique. Finally, a particular logical variety, the so-called *Logic of Reasonable Inferences*, applied to the design of legal knowledge based systems is described, as a further confirmation of the reliability of the proposed model in real-life information systems.

The second paper, titled “*Design of Distributed Database Systems: An Iterative Genetic Algorithm*”, by Sukkyu Song, provides cost models and solutions to two annoying problems of *distributed database system design*, namely *operation allocation* and *data allocation*. Operation allocation refers to *query execution plans* determining which sub-queries should be allocated to which sites in a computer network, so that query processing costs are minimized. Data allocation concerns with allocating relations to sites so that the performance of the target distributed databases is improved. In the paper, author develops innovative techniques for operation allocation and data allocation problem, using three objective functions: (i) *total time minimization*, (ii) *response time minimization*, (iii) *combination of total time and response time minimization*. Author rigorously formulates these problems, and provides *analytical cost models* for each so-defined objective function. Finally, since the problem is known to be *NP-hard*, author proposes a *heuristic solution based on Genetic Algorithm* (GA), whose experimental assessment with comparison to the baseline exhaustive enumeration approach shows a clear superiority.

The third paper, titled “*Query-driven Approach of Contextual Ontology Module Learning using Web Snippets*”, by Nesrine Ben Mustapha, Marie-Aude Aufaure, Hajer Baazaoui Zghal, and Henda Ben Ghzala, moves the attention on the problem of *automatically building Ontology modules that cover search terms of users in ontology-based question answering on the Web*. Indeed, as authors study, existing approaches for Ontology module extraction aim at solving the problem of identifying *Ontology fragment candidates* that are relevant for the application. Nevertheless, these approaches consider the *input of predefined Ontologies* only, and neglect the underlying semantics kept in texts. In order to fulfill this gap, the paper proposes an approach for supporting *contextual Ontology module learning* that cov-

ers particular *search terms* via analyzing past user queries and searching for *Web snippets* provided by the traditional search engines. The obtained contextual modules are then used for query reformulation. In order to validate their proposed Ontology learning approach, authors provide an experimental analysis where they consider two different criteria: (i) the *semantic cotopy measure* of discovered Ontology modules and (ii) the *precision measure* of search results obtained by using the so-obtained Ontology modules for query reformulation. Experimental results expose a positive behavior.

The fourth paper, titled “*KIPTC: A Kernel Information Propagation Tag Clustering Algorithm*”, by Guandong Xu, Yu Zong, Ping Jin and Rong Pan, introduces a *novel tag clustering algorithm based on kernel information propagation, called KIPTC*. As authors correctly highlight, in *social annotation systems*, users annotate *digital data sources* by using *tags* which are freely-chosen textual descriptions. Tags are then used to index, annotate and retrieve resource as *resource additional metadata*. Poor retrieval performance remains a major challenge of *most social annotation systems* resulting from several problems of ambiguity, redundancy and less semantic nature of tags. In order to deal with these problems, authors argue that *Clustering* can represent a useful tool to improve the retrieval performance of social annotation systems, and propose the KIPTC algorithm. KIPTC makes use of the well-known *kernel density estimation method* of the *neighborhood directed graph* used in *kNN Clustering* as a start context to reveal the *prestige rank* of tags in tagging data. The random walk with restart algorithm is then employed to determine the center points of tag clusters. The main strength of the proposed approach is the capability of partitioning tags from the perspective of tag prestige rank rather than the intuitive similarity calculation itself. Experimental studies on six real world data sets demonstrate the effectiveness and superiority of the KIPTC algorithm against other state-of-the-art clustering approaches in terms of various evaluation metrics.

Finally, the fifth paper, titled “*Semi-Supervised Hybrid Clustering by Integrating Gaussian Mixture Model and Distance Metric Learning*”, by Yihao Zhang, Junhao Wen, Xibin Wang and Zhuo Jiang, considers the context of *Semi-Supervised Clustering*, which aims at aiding and biasing the *Unsupervised Clustering* by employing a small amount of supervised information. As authors recognize, the supervised information is generally given as *pairwise constraints*, which are used to either modify the objective function or to learn the distance measure. Many previous work have shown that while clustering algorithms based on *distance metrics* are significantly better than clustering algorithms based on *probability distributions* in some datasets, there exist totally opposite results in other datasets, so how to balance the two methods becomes a key problem. Based on this main evidence, authors propose a *Semi-Supervised Hybrid Clustering algorithm* providing a framework which integrates distance metric into *Gaussian mixture models* which consider not only the intrinsic geometry information but also the probability distribution information of data. The “*Kullback-Leibler Divergence*” is used as distance measure to regularize the target objective function. Experiments on several UCI and real-world datasets demonstrate the effectiveness and the efficiency of the proposed Semi-Supervised Clustering algorithm.

The Guest Editor would like to express his sincere gratitude to the Editors-In-Chief of *Journal of Intelligent Information Systems*, Prof. Larry Kerschberg and Prof. Zbigniew W. Ras, for accepting his proposal of a special issue focused on design, implementation and validation of AI-inspired information systems, and for assisting him whenever required. The Guest Editor would also like to thank all the reviewers who have worked within a tight schedule and whose detailed and constructive feedbacks to authors have contributed to substantial improvement in the quality of final papers.