Expert Systems with Applications 42 (2015) 332–340

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Integration computing and collective intelligence

Marcin Maleszka, Ngoc Thanh Nguyen *

Institute of Informatics, Wroclaw University of Technology, Poland

ARTICLE INFO

Article history:
Available online 8 August 2014

Keywords:
Collective intelligence
Integration computing
Integration functions
Hierarchical structures

ABSTRACT

Techniques for processing knowledge in collectives are more and more needed because of rapidly increasing number of autonomous sources of knowledge in the world. Collective intelligence, among others, deals with creating the knowledge of a collective which is consistent and complete. This means that it should contain all elements not belonging to the knowledge of collective members, but can be inferred on the basis of knowledge of them. For this process the methodologies for knowledge integration seem to be very useful. In this paper the authors present a framework for integrating knowledge of a collective which shows that knowledge of a collective should not be a normal sum of knowledge of its members. The model for knowledge integration using complex hierarchical structures has been also presented and analyzed.

1. Introduction

Computational models of human intelligence often focus on the individual as the only area of analysis (Pentland, 2006). In recent years the trend to consider whole groups of individuals has become more and more visible (Google N-Gram Viewer, 2013). One of the examples of such approach is multi-agent technology, which enables making decisions based on a set of autonomous sources from agents' knowledge bases (Bosse, Jonker, Schut, & Treur, 2006; Castelfranchi, 1998; Hoen & Bohte, 2003). Another example is Web Intelligence, where the knowledge is originated from different knowledge bases with different structures and most often inconsistent referring to a common subject (Fischer, Giaccardi, & Eden, 2005; Gan & Zhu, 2007; Zettsu & Kiyoki, 2006).

Collective intelligence, the more general term than that one considered in this paper, is in turn often defined as an intelligence that emerges from the collaboration and competition of many individuals (Levy, 1997; Russell, 1995). This intelligence may appear to have a mind of its own, independent of the individuals it consists of. In opinion of many authors, collective intelligence appears in a wide variety of forms of the collective knowledge state, which arises in the results of the consensus decision-making processes. Three main features of collective intelligence may be distinguished:

- Elements of the collective are autonomous and intelligent.
- Their knowledge may be inconsistent.
- The collective members are commonly task-oriented.

One of the phenomena of collective knowledge is that it is often larger than the sum of individual's knowledge. For example, if one collective member knows that "$x > y$" and another that "$y > z$", then together they also know that "$x > z$". One of possible processes, by means of which knowledge bases of individuals become the knowledge of the collective, is integration. This process has the following possible aspects:

- Several objects are merged to give a new element representing them. The initial objects are then discarded and only the new representant is used.
- Several objects create a "union" acting as a whole. In cases like data warehouse federations, the initial objects may even remain unchanged and the "union" is only a new, common perspective for the final user.
- Several objects are connected with each other. This may be done as part of the previous two aspects, as well as an independent process. It is often named "mapping" or "matching" in literature, depending on the final result.

The first two aspects are most important and most popular. They are also most useful for purposes of collective intelligence.

In practice very often knowledge on the same subject is gathered from different sources, for example in Internet. If someone wants to solve a problem he/she often asks not only one expert

---

* Corresponding author.
E-mail address: ngoc-thanh.nguyen@pwr.edu.pl (N.T. Nguyen).

http://dx.doi.org/10.1016/j.eswa.2014.07.036
0957-4174/

but several. It is normal thus in such situation the knowledge originating from different sources can be inconsistent and for making a use of it, some integration technique should be applied.

In general, a collective is understood as a set of intelligent units like experts, agent systems, or simply a set individuals which are autonomous in making decisions. Each of collective members has its own knowledge, but when collective members are asked for giving their opinions, comments, solutions for some problem, the knowledge of whole collective (i.e. the collective knowledge) may not be the normal sum of members knowledge. This phenomenon is the motivation of our work. We would like to know if the collective knowledge is larger or smaller than the sum of collective members knowledge.

To our best knowledge, in the word-wide literature there have been performed some experiments with the participation of people which showed that in general if the collective knowledge is larger, but no mathematical model has been proposed. And this is the main objective of our paper. The framework we propose in this work contains only a generic model for the additional value of knowledge of a collective. We do not show, however, what is the influence of the number of collective members on this additional value. This should be the subject of the future work.

Integration may be used as a tool to determine the knowledge of the whole collective. When the knowledge state of all individuals is integrated then some new element or a union of elements will be created. This new integrated element will represent the knowledge of the whole collective and may contain information exceeding the simple sum of individual states. The integration process itself may be conducted using multiple methods to achieve different solutions – each of them may contain different additional knowledge. In this paper we describe the integration function and the *Aug* function that we use to calculate this new knowledge. We also provide some specific results for knowledge trees, where different situations related to integration may occur.

The rest of this paper is organized as follows: Section 2 contains a short survey of collective intelligence research in terms of integration; Section 3 contains the notion of integration functions, including a function to determine the additional knowledge that may be gained by the process; Section 4 contains a short description of multi-level integration process; Section 5 provides multiple examples in which this research may be used in complex tree integration; and Section 6 contains some concluding remarks and avenues for future research.

2. Related works

The concept of collective intelligence may be tracked back to works in psychology area, where it was made distinct from the more broad concept of collective behavior (Weschsler, 1971). It was postulated that while individuals pool resources for task achievement, it requires the collective intelligence aspect for cross-fertilization of ideas. Thus collective intelligence was found more innovative, if not more effective, than the sum of tasks completed by individuals working separately.

This aspect of intelligence was first adopted in the field of Computer Science in works on artificial life and robotics. Early works utilized the collective aspect of animal behavior and intelligence, aiming to apply them to automatic solutions, be it agent (Ferber, 1999) or robot (Mataric, 1993; Millonas, 1992). In each case the analysis of most simple behaviors lead to observing more complex emergent ones.

This leads directly to the current definition of collective intelligence, which may have been best expressed in terms of utility functions (Tumer & Wolpert, 1999): general **world utility** is a function of the state of all agents across all time, a more specific **private utility** function is a function of only one agents state at a single moment; "*the aim of collective intelligence designer is to maximize world utility through the proper selection of private utility functions*". This definition lacks only a single important point about collective intelligence – the fact that knowledge used by world utility function may be much larger than sum of knowledge from all private utilities.

Integration is the process by which the summary knowledge of the collective may be determined. In its merging aspect it may be tracked back to works such as Margush and McMorris (1981), where mathematical solutions were necessary to find an "average" of multiple experimental results. It was mostly developed by evolutionary biologist to solve the problem of finding a real phylogenetic tree based on inconsistent data (Adams, 1986; Day, 1985). The general solution was called the median tree. It is obtained by finding the element that minimizes the sum of distances to all inputs of integration. The name is based on the fact that in one dimensional euclidean space the element minimizing the sum of distances is the median of the inputs. To solve this problem (proven to be NP-hard for trees (Amir & Keselman, 1994)) a variety of algorithms were proposed, including cluster approach (Day, 1985), triads and nestings (Adams, 1986). The median procedure was the basis for development of multiple tools in knowledge management, including consensus theory (Nguyen, 2008; Nguyen, 2002) and knowledge integration methods (Konieczny, 2000; Lin & Mendelzon, 1999). Large part of research described in this paper is based on those papers and builds on their basis. In particular, the research presented in this paper is most similar to that published in Margush and McMorris (1981), which defined the median tree as similar to our postulate $O1$.

Modern research on integration occurs in multiple areas and applications. These range from continued work on phylogenetic trees (Jansson, Shen, & Sung, 2013) and schema matching (Peukert, Eberius, & Rahm, 2012), to ontology alignment (Bock & Hettenhausen, 2012) and data warehouse federations (Kern, Stolarczyk, & Nguyen, 2012).

The continued research in the first area is due to the fact that existing algorithms are not computationally efficient and the final solution to the problem would have to deal with integration of massive structures (trees with millions of vertices). Thus current work focuses on finding solutions faster than the old algorithms, possibly in polynomial or linear time (Jansson et al., 2013).

A lot of work is also directed towards applying basic data integration techniques in real world application. A large part of this research lacks a theoretical basis, but is focused on the specific application it is intended for. These range from merging logs (Claes & Poels, 2014) to merging documents or their schemas (Peukert et al., 2012). The work presented in this paper intends to create this theoretical basis for the applications research, as well as provide the multi-level integration tool to improve computation time for existing methods.

The schema matching research area is active due to the fact that the schemas become larger in real-world applications. The algorithms may be later used as part of integration process for data and knowledge. Most existing systems are semi-automatic and work by creating mapping suggestions that are later corrected by the user. This requires human effort both before and after the matching is conducted, first in order to teach the system and then to improve the results. Current research, among others, aims to improve this process by including adaptation in subsequent uses (Peukert et al., 2012).

The area of ontology alignment is closely related to schema matching in its underlying principles, but focuses in larger part on knowledge relationships than on basic data. Currently multiple issues had been solved and good alignment and integration algorithms exist, but there are still challenges to be overcome