Assessing the Quality of a Consensus Determined Using a Multi-level Approach

Adrianna Kozierkiewicz-Hetmańska
Faculty of Computer Science and Management
Wroclaw University of Science and Technology
Wybrzeze Wyspianskiego 27
50-370, Wroclaw, Poland
Email: adrianna.kozierkiewicz@pwr.edu.pl

Marcin Pietranik
Faculty of Computer Science and Management
Wroclaw University of Science and Technology
Wybrzeze Wyspianskiego 27
50-370, Wroclaw, Poland
Email: marcin.pietranik@pwr.edu.pl

***Abstract*—The following paper investigates a multilevel approach to data integration using the widely accepted Consensus Theory. We focus on an issue related to an initial classification of raw input data into groups that can be integrated in parallel. A final consensus is a result of the integration of obtained partial outcomes. Our main research concerns an application of Fleiss' kappa value, which in the literature is a well known measure that describes how consonant the data in a selected set are. In other words - for a given set of values, the higher the value of this measure, the higher its inner consistency. Therefore, we have attempted to answer the question whether or not the initial data should be divided into coherent groups or into highly divergent subsets, that better represent the whole input. We present a theoretical background, broad description of a series of experiments that we have performed and their statistical analysis.**

***Index Terms*—consensus theory, knowledge management, knowledge integration**

I. Introduction

In recent years researchers focused on issues related to processing very large sets of heterogenous data. Covered areas of interest concern tasks such as data alignment, knowledge extraction, data integration, etc. Rapidly growing number of data sources (spreading from social networks, through distributed databases, up to federation of data warehouses) do not simplify these tasks. It increases their complexity even more, due to the necessity of initial data cleaning and asserting their consistency before the actual task is performed.

Dividing an initial set of input data into smaller, more manageable chunks, that can be processed independently, is one of the most common approaches. It has been widely accepted and adopted in a variety of both practical ([1]) and theoretical research. The eventual outcome is a result of an integration of the partial outcomes. These tasks may be related to a wide array of different applications. A particular type of this approach is a task where an integration itself of all of the input data is the expected result. If this is the case, then we talk about a multi-level data integration.

This entails two problems. The first one is determining number of steps into which the whole process should be decomposed in order to get reliable results in acceptable time. Obviously the whole procedure may consist of more than two steps. The second problem (on which we focus in the following paper) is a method of dividing the initial set of data into aforementioned smaller chunks (hereafter referred to as classes). In other words, what is a criterion that one could use to split the input data into a finite set of classes in order to acquire their best possible integration?

In order to simplify the notation we will assume that the input data comes from a finite (and possibly small) group of experts that give their opinions on some topic. These opinions are expressed using binary vectors and obviously the expected result is also a binary vector. The trivial approach would incorporate the consensus theory to perform the expected integration in one single step, but as aforementioned earlier, we would like to investigate a multi-level approach to the consensus theory. The main contribution of this paper is instigating how different approaches to grouping experts' opinions influence the quality of the final outcome. One of the foundations on which we base our work is Fleiss' kappa which is a measure of inner consistency of a set of data. It can be used to answer the question how unanimous experts from one group are. As a consequence, should we group experts into classes that are highly consistent or quite

978-1-5090-5795-5/17/$31.00 © 2017 IEEE

the contrary?

The article is structured as follows. Section II is an overview of the work that have been done in the field. Section III contains basic notions used throughout the whole paper. Section IV covers the general idea of the main approach on which we have based our work which is then further described in details in Section V. Section VI is a description of experimental verification that have performed. Section VII summarises the paper and sheds some light on our upcoming research.

II. Related Works

The multilevel consensus determination method is a quite new idea and so far it has not been widely investigated. There are some papers like [9], [7] or [8] devoted to one- and two-level consensuses and the problems related to its' determination. Authors have developed a formal framework that can be easily used to designate the consensus in one- and two- steps for assumed macro- and microstructures. Next, for some criteria the analysis of obtained consensuses has been made. The researches demonstrated that both one-level and two-level approaches in comparison to the optimal solution give acceptable outcomes- the one- and two-level algorithms give results respectively worse by 1% and 5%.

The multi-level idea has been applied in the ontology integration task [5]. The analytical analysis pointed out that for the presented algorithm the one- and multi-level integration processes give the same final ontology. However, the multi-level integration allows to decrease the time of data processing even by 20% in comparison to the one-level approach. This property is especially important in the emerging context of Big Data.

In [14] and [13] authors have considered the two-stage approach to the consensus. For this goal, clustering methods are used to classify a large collective into smaller chunks. In [13] *k-means* algorithm has been used. The representatives of new, smaller collectives have beed determined in the first stage of the consensus choice. Next, each representative has been assigned a weight value depending on the number of members in the corresponding collective. The experimental results have demonstrated that the weighted approach is helpful in reducing the difference between the two-stage and the single-stage consensus choice in determining the knowledge of a large collective in reference to the non-weighted approach.

In [13] authors have presented an improvement method of the two-stage consensus choice in determining the knowledge of a large collective by taking into a consideration the problem of a susceptibility to consensus. The experimental evaluation have confirmed that the improvement method is useful in minimising the difference between single-stage and two-stage consensus choice approaches in terms of the quality of the final results.

As it was mentioned, the idea of a multi-level approach to consensus determination is quite new and missing in the literature. However, the paralleling of data processing is not so uncommon topic and is also known as the MapReduce. It has been developed from the data analysis model of the information retrieval field and is a cloud-based technology. The general idea of MapReduce is based on breaking data processing into small units of work that can be run in parallel across several nodes in the cluster. Until now, several MapReduce architectures have been developed for handling big collections of data. Hadoop is a one of the most popular open-source implementation of MapReduce for an analysis of large datasets [17] which uses a distributed user-level filesystem to manage the storage of resources across the cluster [15].

Researches have demonstrated the effectiveness of MapReduce across a wide range of domains, including large-scale machine learning problems, clustering problems, extracting data to produce reports of popular queries, extracting properties of web pages for new experiments and products, processing of satellite image data, language model processing for statistical machine Translation or large-scale graph computations [2]. However, there are still many opened problems and challenges ([16], [4]) like: meeting the acceptable processing speed, understanding data, addressing data quality, displaying meaningful results, dealing with outliers, etc.

III. Basic Notions

Let U be a finite, nonempty set of a universe of objects which could be considered as the potential elements of a knowledge referring to a certain world. The symbol 2^U denotes a powerset of U, that is the set of all subsets of U. By $\Pi_b(U)$ we denote the set of all b-element subset (with repetitions) of the set U for $b \in N$. Therefore $\Pi(U) = \bigcup_{b \in N} \Pi_b(U)$ is the set of all nonempty subsets with repetitions of the universe U. Each X which belongs to $\Pi(U)$ is called a knowledge profile (a profile) and could be interpreted as the set of experts' opinions and each $x \in X$ as the opinion of a certain collective member.

Definition 1: The macrostructure of the set U is a distance function denoted by δ with a signature $\delta : U \times U \to [0,1]$ which satisfies the following conditions ([9]):

1) $\forall_{v,u \in U}, \delta(v,u) = 0 \Leftrightarrow v = u$
2) $\forall_{v,u \in U}, \delta(v,u) = \delta(u,v)$

Definition 2: For an assumed distance space (U, δ), the consensus choice problem requires establishing the consensus choice function. By a consensus choice function in space (U, δ) we mean a function:

$$C : \Pi(U) \to 2^U \tag{1}$$

By $C(X)$ we denote the representation of $X \in \Pi(U)$ and for each $c \in C(X)$ we denote a consensus of a profile X. Therefore, each element of $C(X)$ represents a consistent knowledge state of an assumed collective. We can interpret it as the final decision determined based on experts' opinions.

In [11], [10] and [9] authors present 10 postulates for consensus choice functions: *reliability, unanimity, simplification, quasi-unanimity, consistency, Condorcet consistency, general consistency, proportion, 1-optimality* and *2-optimality*. The last two, namely *1-optimality* and *2-optimality*, play the important role in solving the consensus choice problem. The former postulate requires the consensus to be as near as possible to elements of the profile and could be recognised as the best representative of the profile. The latter allows to determine the most 'fair' consensus. In order to formally define them, let us assume the following notions:

- $\delta^1(x, X) = \sum_{y \in X} \delta(x, y)$
- $\delta^2(x, X) = \sum_{y \in X} (\delta(x, y))^2$

Definition 3: For a profile $X \in \Pi(U)$ a consensus choice function C satisfies the postulate of:

- *1-optimality* iff $(x \in C(X) \Rightarrow (\delta^1(x, X) = \min_{y \in U} \delta^1(y, X))$,
- *2-optimality* iff $(x \in C(X) \Rightarrow (\delta^2(x, X) = \min_{y \in U} \delta^2(y, X))$.

For the same distance space it is possible to determine many consensuses. To evaluate their reliability we use a consensus' quality that is defined as follows ([10]):

Definition 4: Let $X \in \Pi(U)$ and $x \in C(X)$. By the quality of a consensus x in a profile X we call the following value:

$$Q^i(x, X) = 1 - \frac{\delta^i(x, X)}{card(X)} \tag{2}$$

where: $i \in \{1, 2\}$.

If the consensus x in the profile X satisfies the *1-optimality* criterion, we calculate $Q^1(x, X)$. Otherwise we use $Q^2(x, X)$.

IV. General Idea Of One- And Two-level Consensus Determination Method

In this work we assume that we ask many experts for their opinion or we process data from various sources. Thus, we have to deal with a knowledge of a collective, and based on it we have to make the final decision. It has been proved in [10] that the Consensus Theory (mentioned in the previous Section) can be useful in determining a consistent knowledge of a collective. It can be formally defined as follows: *for given n experts opinions $EO_1, EO_2, ..., EO_n$ one should determine a final decision $EO*$ which is the best representation of given input opinions.*

Any kind of a database or, in general, a source of knowledge can be treated as an expert's opinion. Therefore, the experts opinions should be stored as the real number, binary vectors or as more complex structures as graphs or ontologies etc. In a real situation designating the final decision based on experts opinions is not an easy task. This problem is even more complicated if experts opinions are geographically distant and on the level of raw data their transfer entails major delays. Moreover their content may be sensitive or confidential and for that reason the efficient and simple method for determining the final representation is a crucial requirement. A division of a task of determining a decision based on experts opinions into smaller subtasks and their parallelisation can solve this problem. Unfortunately, it entails the consecutive difficulty concerning the initial classification of data sources into groups and the integration the partial results into the final outcome.

The general idea of the two-level approach to the consensus choice requires to conduct an initial division of the sequence of n experts' opinions into k classes. For each class, the included opinions are integrated using an ordinary integration algorithm. The final consensus is obtained as a consensus of partial consensuses. Obviously, the procedure can be divided into even more stages which outputs serve as inputs of subsequent steps. The general idea of mentioned procedure is presented in Figure 1. The formal definition of one- and two-level consensus choice and is given below:

Definition 5: For a given distance space (U, δ) a one-level consensus is defined as:

$$c_1^*(X) = \{x \in U : \delta^i(x, X) = \min_{y \in U} \delta^i(y, X)\} \tag{3}$$

where: $i \in \{1, 2\}$.

Based on the above definition, the multi-level approach can be defined as:

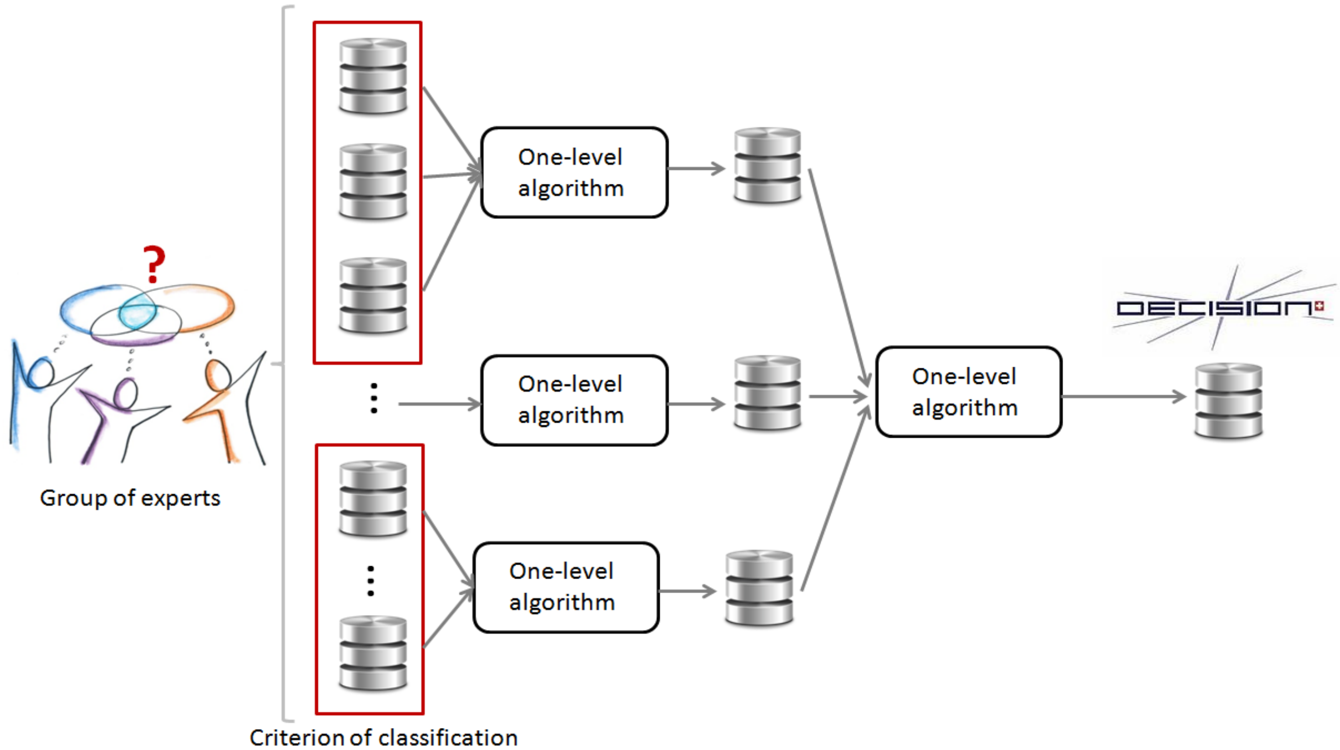

Fig. 1. General idea of the multilevel integration

Definition 6: Let us assume that for a given space (U, δ) a profile $X \in \Pi(U)$ is consisted of the consensuses obtained during m-1 level, where $m \geq 2$: $X = \{x^1_{m-1} \in c^1_{m-1}(X), x^2_{m-1} \in c^2_{m-1}(X), ..., x^n_{m-1} \in c^n_{m-1}(X)\}$. Let us assume that profile X is primarily divided into k classes: $X_1, X_2, ..., X_k$. A m-level consensus is determined in the following way:

1) For each profile $X_1, X_2, ..., X_k$ calculate the one-level consensus $X' = \{x^1 \in c^*_1(X_1), x^2 \in c^*_1(X_2), ..., x^k \in c^*_1(X_k)\}$.
2) For a new profile X' calculate the one-level consensus $c^*_m(X) = c^*_1(X')$.

V. The Proposed Approach To Multi-Level Consensus Determination

As it was presented in the previous section the multi-level consensus determination method required designation of the following elements:

1) The formal representation of experts' opinions.
2) Criterion of classification of experts' opinions.
3) One-level consensus determination method.

In this paper we assume that we ask n experts about their opinions for a complex problem. Experts are asked to answer for N different questions. In other words they assess the N objects and they can choose one from two possible answers i.e. *yes* or *no*, *a* or *b* etc. We try to find the best solution of our problem which can be interpreted as the consensus of the experts' answers. The profile X containing experts' opinions (knowledge of a collective) stored as n binary vectors of the length equal to N. Formally, we can define distance space as: $U = \{u_1, u_2, ...\}$ where elements of the universe are binary vectors, $\delta(w, v) = \sum_{j=1}^{N} |w^j - v^j|$ for such $w, v \in U$ that $w = (w^1, w^2, ..., w^N), v = (v^1, v^2, ..., v^N)$, $v^q, w^q \in \{0, 1\}$, $q \in \{1, ..., N\}$ and the profile $X = \{a_1, a_2, ..., a_n\} \in \Pi(U)$, where: $a_i = (a_{i1}, a_{i2}, ..., a_{iN}), i \in \{1, ..., n\}$.

In our approach we assume that the profile X is primarily divided into k classes based on Fleiss' kappa measure [3], that is calculated in the following way:

$$\kappa = \frac{P - P_e}{1 - P_e} \tag{4}$$

where:

- $P_e = \sum_{j=1}^{2} p_j^2$, $p_j = \frac{1}{Nn} \sum_{i=1}^{N} n_{ij}$,
- n_{ij}- the number of experts who assigned the *i*-th object to the *j*-th category, $j \in \{1, 2\}$,
- $P = \frac{1}{N} \sum_{i=1}^{N} P_i$, $P_i = \frac{1}{n(n-1)} \sum_{j=1}^{2} n_{ij}(n_{ij} - 1)$.

Thus, our criterion of classification is the value of Fleiss' kappa measure, denoted as κ.

For the data, collected and stored in the described way, the one-level consensus satisfying the *1-optimality criterion* is determined based on computing the number of occurrences of the *0* and *1* in the experts' opinions. The final decision is the one that has been chosen more often. The base algorithm presented below is taken from [9]:

Algorithm 1 One-level consensus determination method

Require: $X = \{a_1, a_2, ..., a_n\}$

1: Set $j = 1$;
2: Set $f_j = 0$;
3: Set $i = 1$;
4: IF $a_{ij} = 1$ THEN $f_j + +$;
5: $i + +$;
6: IF $i \leq n$ GOTO 4 ;
7: $j + +$;
8: IF $j \leq N$ GOTO 2 ;
9: Set $j = 1$;
10: IF $f_j \geq \frac{n}{2}$ THEN $x_j^* = 1$ ELSE $x_j^* = 0$
11: $j + +$;
12: IF $j \leq N$ GOTO 10;

VI. Experimental Verification

The proposed methods of the multi-level consensus determination procedure has been implemented in a dedicated experimental environment. For the simplicity of calculations we have considered only the two-levels approach.

In our previous work [6] we have proved that if the profile consists of binary vectors then the sufficient condition to determine a reliable consensus is the odd number of the cardinality of profile. Therefore in our experiments we have chosen the number of experts n equal to 9 and the number of classes $k = 3$ in order to assert the susceptibility to a consensus. Additionally, we have assumed that each expert has expressed their binary opinion for $N = 10$ objects.

For the purpose of the experiment the experts' opinions have been chosen randomly. The experiments have been done for uniform distribution of zeros and ones and for occurrences of one of the options predominate respectively in 10%, 20%, 30% and 40%. All calculations have been repeated 100 times.

Figure 2 contains the results of the experiments where the quality of the final consensus (obtained in the second level) has been compared with the distribution of the

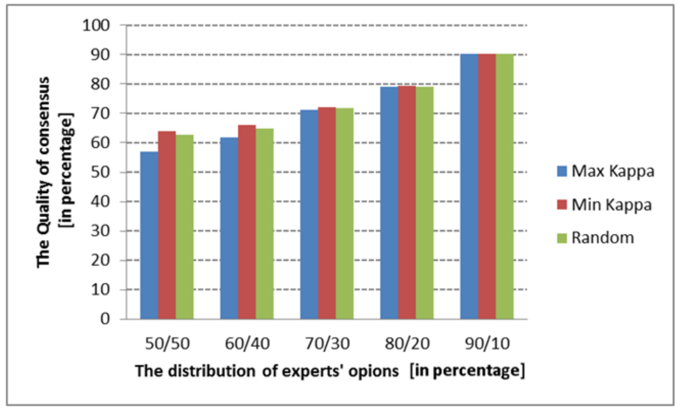

Fig. 2. The results of experiments

input experts opinions and methods of experts' grouping in the first stage of the considered approach. In our experiments we have investigated three strategies of grouping of experts' opinions: based on minimal value of Fleiss' kappa, maximal value of Fleiss' kappa and randomly chosen groups. We denote these strategies as: *MinKappa*, *MaxKappa* and *Random*.

The experiments showed that, if the distribution of input experts' opinions is highly divergent (meaning that there is no one predominant opinion) then the choice of the criterion of experts' grouping is more significant from the point of view of quality of the final consensus. On the other hand, it is also obvious, that if experts tend to agree more frequently (meaning that one of the options is a particular favourite) then the quality of the final consensus is higher.

The best quality of the final decision has been reached if in the first level of our algorithm the experts' opinion are grouped based on the minimal value of Fleiss' kappa. This hypothesis has been statistically proved on significance level $\alpha = 0.05$. Before selecting a proper test we have analysed the distribution of all obtained data by using the Lilliefors test. We have decided that sample *MinKappa* do not come from a normal distribution because we have obtained p-values less than $1 * 10^{-3}$, therefore, for the further analysis the *Kruskal-Wallis* test have been used.

The statistical difference between the method of grouping have occurred only for more divergent experts' opinions. Therefore, we have considered only two cases where the proportion of zeros to ones are $50\%/50\%$ and $60\%/40\%$. The obtained results are presented in Table I.

In both cases we have obtained p-value less than 1×10^{-3}, therefore, we can reject the null hypothesis stating that the criterion of grouping the experts' opin-

TABLE I
THE MEAN OF THE RANKS FOR PARTICULAR SAMPLE

Distribution	MinKappa	MaxKappa	Random
50%/50%	197.2050	78.81	174.485
60%/40%	187.32	100.895	163.285

ions on the first level has no influence on the quality of the final consensus. The analysis of the means of the ranks allows to decide that the best strategy for grouping input experts' opinions is the one based on the minimal value of Fleiss' kappa measure. It means that in order to improve the quality of the final consensus one should classify initial opinions into groups that are internally highly inconsistent.

VII. Conclusions And Future Works

This paper covers the topic concerning initial classification of data into subgroups in the context of the multilevel approach to their integration. We have incorporated and used a widely known Fleiss' kappa measure in order to investigate the impact that the internal consistency of these subgroups has on the quality of the final integration.

The performed experiments and the statistical analysis of their outcomes clearly showed that the considered grouping should be based on the lowest value of Fleiss' kappa measure. Assuming that the data represents some opinions of experts, in order to determine the best representation of all of their ideas one should put them into groups with other experts with highly different opinions. Then the ordinary multilevel integration should be performed. Such approach is consistent with an intuition stating that each of such groups should be the best representation of the whole population.

In our upcoming publications we plan to address data representations more expressive than binary vector, such as tuples of natural number, value ranges etc. We would also like to perform more comprehensive experiments involving surveys taken from real people.

References

[1] Dean J., Ghemawat S.: MapReduce: simplified data processing on large clusters. Communications of the ACM, 51(1), 107-113. (2008)

[2] Jeffrey D. and Sanjay G , MapReduce:Simplified Data Processing on Large Clusters, COMMUN. ACM, VOL. 51, pp 107-113, (2008)

[3] Fleiss, J. L.: Measuring nominal scale agreement among many raters, Psychological Bulletin, Vol. 76, No. 5 pp. 378–382, (1971)

[4] Grolinger K., Hayes M., Higashino W. A., L'Heureux A., Allison D.: Challenges for MapReduce in Big Data, Electrical and Computer Engineering Publications. Paper 44, (2014)

[5] Kozierkiewicz-Hetmańska A., Pietranik M., The knowledge increase estimation framework for ontology integration on the concept level, Journal of Intelligent and Fuzzy Systems, pp. 1-12, (2016)

[6] Kozierkiewicz-Hetmańska A.: Analysis of susceptibility to the consensus for a few representations of collective knowledge, International Journal of Software Engineering and Knowledge Engineering. 2014, vol. 24, no 5, s. 759-775 (2014)

[7] Kozierkiewicz-Hetmanska A.: Comparison of one-level and two-level consensuses satisfying the 2-optimality criterion, Computational Collective Intelligence. Technologies and Applications Volume 7653 of the series Lecture Notes in Computer Science pp 1-10 (2012)

[8] Kozierkiewicz-Hetmanska A. , Nguyen N.T.: A Comparison Analysis of Consensus Determining Using One and Two-level Methods, Volume 243: Advances in Knowledge-Based and Intelligent Information and Engineering Systems, pp. 159 - 168, (2012)

[9] Nguyen N. T: Consensus Choice Methods and their Application to Solving Conflicts in Distributed Systems, Wroclaw University of Technology Press, (2002)

[10] Nguyen N. T.: Advanced methods for inconsistent knowledge management, Springer Science and Business Media, (2008)

[11] Nguyen N. T.: Processing inconsistency of knowledge in determining knowledge of a collective, Cybernetics and Systems 40(8) 670-688,(2009)

[12] Nguyen N. T.: Inconsistency of knowledge and collective intelligence, Cybernetics and Systems 39(6) 542-562 (2009)

[13] Nguyen V. D., Nguyen N.T: A Two-Stage Consensus-Based Approach for Determining Collective Knowledge, Advanced Computational Methods for Knowledge Engineering, Springer International Publishing, pp. 301-310 (2015)

[14] Nguyen V. D., Nguyen N.T, Hwang D.: An Improvement of the Two-Stage Consensus-Based Approach for Determining the Knowledge of a Collective. Computational Collective Intelligence vol 9875 of the series Lecture Notes in Computer Science, pp. 108-118 (2016)

[15] Maitrey S., Jha C.K.: MapReduce: Simplified Data Analysis of Big Data, Procedia Computer Science 57 , pp. 563 – 571 (2015)

[16] Siddaraju Dr., Sowmya C. L., Rashmi K., Rahul M.: Efficient Analysis of Big Data Using Map Reduce Framework, International Journal of Recent Development in Engineering and Technology Volume 2, Issue 6 (2014)

[17] Sugha P., Gunavathi R.: A Survey Paper on Map Reduce in Big Data, International Journal of Science and Research, Volume 5 Issue 9 (2016)