Overcoming the Condorcet's Border in Collective Intelligence Systems

Vladislav Protasov, Zinaida Potapova
Center of Computing for Physics and Technology
Moscow, Russia
protvlad@gmail.com, zinaida.potapova@gmail.com

Eugene Melnikov
Center of Computing for Physics and Technology
Moscow, Russia
apinae1@gmail.com

***Abstract* – The paper presents a new approach of solving intellectual problems by means of collective intelligence. The essence of the approach is combination of two principles: Condorcet's principle of the jury and evolutional coordination, based on reciprocal evaluation among intellectual agents. This method exploits both generative and evaluative abilities of the agents and allows to eliminate so called Condorcet's border, which means that for obtaining correct solution every expert must make correct decision with probability greater than 0.5. The paper also observes the conditions that guarantee correct solution obtainment.**

Keywords* – *collective intelligence; genetic algorithm; coordination; generation; evaluation.

I. INTRODUCTION

In October 2010 in the "Science" magazine (USA) employees of the Center for Collective Intelligence of the Massachusetts Institute of Technology headed by Pr. Thomas V. Malone published the first English language paper that proved the effect of excess of the collective Intelligence Quotient value over both group average and maximum individual IQ values [1]. The factor was identified that determines the successfulness of solution in intelligent groups. It was called the C-factor (collective factor). The factor analysis approved the existence of the only significant component for all collective activities. This component is the C-factor.

The analysis of these experiments has provided two conclusions: first of all, the collective component, that defines group intelligence potential, exists, secondly, it can be evaluated objectively.

There are also a huge number of well-known crowdsourcing application success stories [2], based on the application of Condorcet's jury theorem [3]. The intellect gain effect during expert group work takes place due to several reasons. In the first case (Malone's) it is defined mainly by the "legislative" component. It means that at the first step some solutions are generated and during further discussion they are combined by the collective intelligence of the group into a collective solution. In the case of crowdsourcing the process often consists of only the first step – ideas generation, and if every expert has a probability of correct solution $G_P > 0.5$, then at a great number of experts the probability of the collective solution correctness tends to 1. If for every expert $G_P < 0.5$, then the probability tends to 0, which is a big problem in application of crowdsourcing.

Most papers upon collective intelligence refer the Condorcet's principle [4]. This paper provides an approach of eliminating the Condorcet's constraint of 0.5. First of all, the method will be introduced and explained. After that there will be a computer experiments report. In the end goes the definition of the condition that must hold in order to obtain correct result.

In this paper, basic principles that are related to the algorithm are discussed in Section II, while Section III describes the method itself. Section III deals with theoretical and experimental substantiation of the method. Last section provides the conclusion.

II. QUICK VIEW UPON THE BASIC PRINCIPLES

The Condorcet's jury theorem assumes that a group of individuals wishes to reach a decision by a majority vote. One of the two outcomes of the vote is correct, and each voter has a probability p of voting for the correct decision. The theorem contends that if $p > 0.5$ then with increasing the number of voters the probability of obtaining the correct solution tends to 1; in other hand, if $p < 0.5$ then with increasing the number of voters it tends to 0.

This means, for a jury group making a decision between two alternatives, if their $p > 0.5$ then a large group will make correct decision with higher probability than a small one, but if their $p < 0.5$ then one expert will make correct decision much more probably than a group.

Another approach that lies on the foundation of this paper is Genetic Algorithms (GA) [5]. Its basis is rather simple. Each solution is assumed as a biologic individual that has a genotype (sequence of bits that encodes the set of its characteristics or so called phenotype). The algorithm starts with generation of initial set of individuals (solutions). The correct solution search passes like a biological cycle of population. It consists of three stages that repeat iteratively:

1) *Mutation* of random individuals: each individual subjected for mutation changes random bits of its genotype. Usually only few bits are subjected for change – near 10%.

2) *Natural selection*: each individual has a value of the *fitness function*. Basing on this values some "weak" individuals are eliminated and population decreases, usually on 50%. The selection algorithm may vary.

3) *Crossover*: all individuals divide into pairs and every pair give birth to two children – individuals that incorporate the genotypes of their parents.

These stages run iteratively until the *convergence condition* holds. It may be an obtainment of some fitness value or end of the population progress (maximal fitness doesn't increase). The individual with the highest fitness is selected as the final solution.

As an example of GA application let's assume a chemical experiment. We need to adjust the volumes of reagents to provide maximal heating of the mixture. The set of volume values represents an individual and is encoded into bits of its genotype. Fitness of each individual (solution) is evaluated by chemical modeling module. Since we have all the data presented in GA model, we can run the algorithm to obtain a rational solution.

III. Method of Evolutionary Solutions Coordination Introduced

As Malone's experiments proved, in experts' work, based on legislating and voting procedure, the leader effect and conflicts, hidden and evident, considerably worsen effectiveness of the group.

This paper proposes the synthesis of the two approaches that is called by the authors Method of Evolutionary Solutions Coordination (MEC), which has advantages of two methods and considerably compensate their weaknesses.

The method incorporates genetic algorithms (GA) [6], Condorcet's theorem, Tychurin's metasystem transitions theory [7] and collective intelligence systems theory, that is partially presented in this paper.

The method is defined as follows. MEC is an approach of organization of collective work upon a project with predefined objectives and rules of interaction, based on classic GA principles. Experts work is usually organized by the means of a computer network. These are the rules of organization of intellectual agents work and interaction:

1) Objectives of the project are defined
2) Experts group and their interaction method are selected
3) Frame (slots structure) of the project is created
4) The first solutions are found, they may be incomplete
5) Solutions are exchanged between the experts
6) Exit condition is checked; if it's fulfilled, algorithm halts
7) New solutions are created from the old ones through crossover
8) Some new solutions mutate
9) Go to the point 5

According to interaction rules the collective work instructions are developed considering features of certain problem, communication environment, abilities and qualification of the intellectual agents. The scheme of MEC is illustrated on Fig. 1.

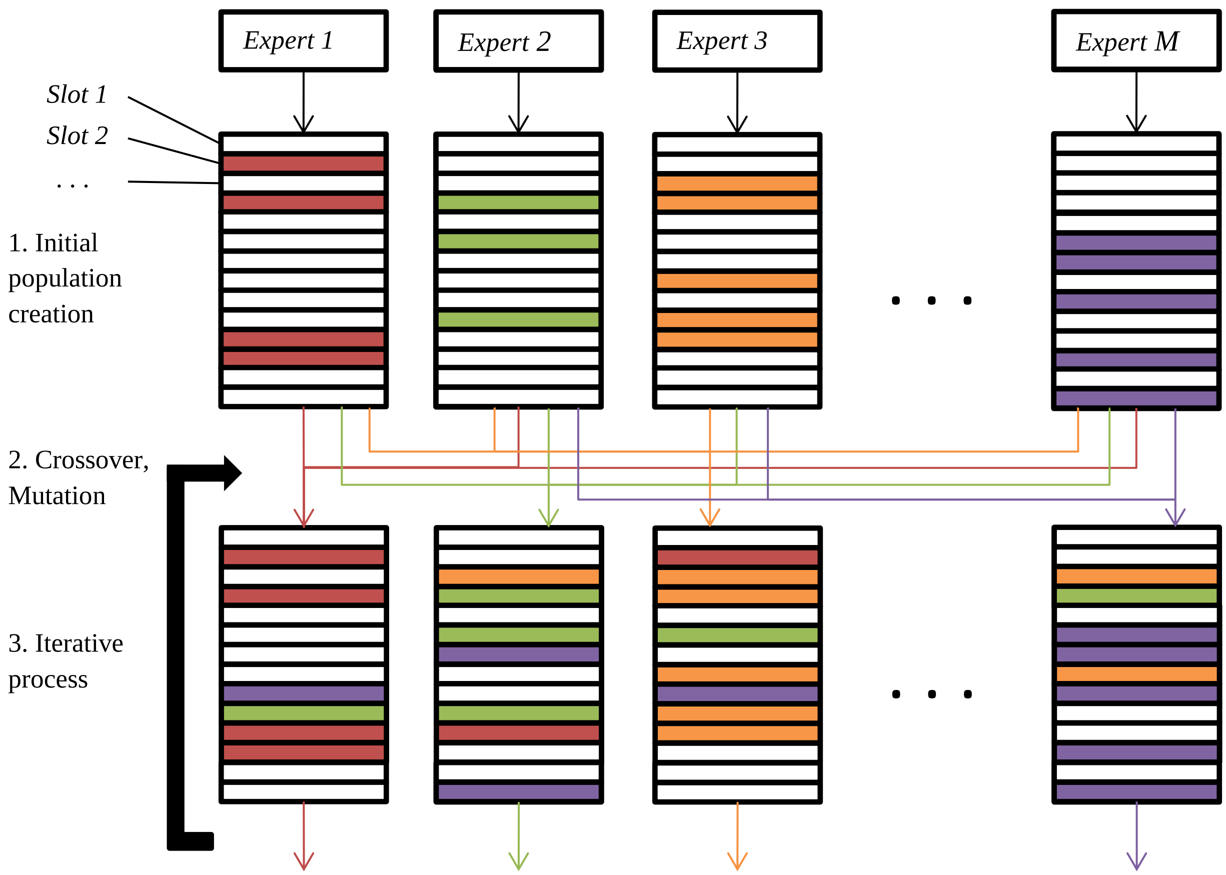

Figure 1. MEC scheme

 ISBN: 978-1-61208-269-1

On the zero iteration experts fill slots of the project according to their knowledge. In Figure 1 these slots are marked with dark rectangles. On the iterations of coordination, each expert checks others' variants and picks some of those that he considers to be the best and fills blank cells of his variant with them. After performing several iterations, the process converges to a population of equally filled solutions. For accelerating the process, it can be adjusted with the Condorcet's rule – a slot is considered as filled, if more than half of experts have made the same decision on it.

Experiments on solving complicated intellectual problems from different creative areas by expert groups have demonstrated the MEC efficiency. With this method high level checkmate problems were successfully solved by collective intelligence, when the group members could not solve the problems individually. A witnesses group effectively restored an identikit, IQ measurement [8] upon verbal Eysenck's tests [9] demonstrated an extremely high intellectual level of the group. There are many papers in Russian that refer to these results [10][11][12][13][14], but they are unknown for English-speaking audience. This paper partially makes up this lack.

Most papers upon negotiation of agents group offer approaches to find some trade-off in preferences of several individuals [15][16]. They use predefined utility functions of the agents to obtain Pareto-optimal solutions. Supposed methods are considered to demonstrate high quality of negotiating in multi-party cases. There are also some papers [17] that describe complex approaches – concluding contracts are defined by utilizing several strategies of negotiation. All these theories aim to obtain a solution in case of multiple preferences of different parties. This signifies that each party has its own assumption upon correct solution. Such circumstances appear not frequently enough on complicated multidimensional problems. In real life situations not every agent can provide a value for each component of the solution. MEC deals with such situations throughout utilization of agents' ability of estimating other agents' solutions. In most cases even if an expert can't provide his own solution of a problem, at least he can express his opinion upon an existing solution.

IV. MODELING THE CALCULATION PROCESS

The main way, considered to explore the proposed method in this paper, is computer modeling. The project, subjected for solving by a group of virtual experts, is meant to consist of K slots that must be filled with correct answers after several stages of coordination according to the vote majority.

Let us draw up a computer model of a slot filling process with applying it to all slots of the project successively. The decision-making process is divided into several stages. The first stage is individual decisions creation, further follow stages of iterative coordination of the solutions by a group of M experts. Let us suppose that on the stage of individual decisions creation every expert gives a correct solution with probability $0 < \mathrm{G_P} < 1$, an incorrect solution with probability $0 < \mathrm{G_N} < 1$ and with probability $\mathrm{G_V} = 1 - \mathrm{G_P} - \mathrm{G_N}$ he gives no solution. On coordination stages an expert, who didn't make a decision, choses among other experts' solutions – correct one with probability $\mathrm{E_P}$, incorrect one with probability $\mathrm{E_N}$ – and choses no solution with probability $\mathrm{E_V} = 1 - \mathrm{E_P} - \mathrm{E_N}$.

All the probabilities are determined with consideration that correctness of solutions can be estimated. It is necessary for building the mathematical model, but in further application it isn't essential – real life problems are not usually provided with fitness function and each solution is estimated by subjective opinion of an expert. And for MEC it isn't matter if the estimation process is formulized or defined by certain expert's opinion.

Moreover, the assumption that correctness of solution is a binary function (correct/incorrect) is also considered only for durable mathematical substantiation. It can be discarded f.e. with application of fuzzy logic (every solution is correct and incorrect with some degrees). All further deductions can be completed with probabilistic functions of correctness. We assume that taking them into consideration is not essential and theoretical substantiation on strict correctness can be propagated to fuzzy functions.

The algorithm of the decision-making process that lies upon the computer model can be easily described on cellular automata language. Let us suppose that experts fill the automata cells basing on the following rules.

Initial state of the cellular automata is determined by $\mathrm{G_P}$ and $\mathrm{G_N}$ parameters, it is represented by the vector $\mathrm{B_i}, \mathrm{i} = 1, 2, \ldots \mathrm{M}$.

$$\mathrm{B_i} = \begin{cases} 1, & 0 < \xi \le \mathrm{G_P} \\ 2, & \mathrm{G_P} < \xi \le \mathrm{G_P} + \mathrm{G_N}, \\ 0, & \xi > \mathrm{G_P} + \mathrm{G_N} \end{cases}$$

where ξ is a random number in the interval (0, 1). 1 means that an expert has made a correct decision, 2 means an incorrect decision, 0 – the expert has not made a decision. Mean of the number of ones in the automata equals $\mathrm{M} \cdot \mathrm{G_P}$, of the number of twos - $\mathrm{M} \cdot \mathrm{G_N}$, of the number of zeros - $\mathrm{M} \cdot \mathrm{G_V}$.

The subsequent state C_i of the cellular automata, that imitates solutions coordination process, is filled according to the following rules.

1. If $B_i > 0$, then $C_i = B_i$ (experts have filled their sells and now are waiting for other experts to fill up the cells).
2. If $B_i = 0$, then then the *i*-th expert uniformly choses another cell *j* such that $B_i > 0$ and randomly updates his ξ in the interval (0, 1). Depending on value of ξ he chooses one of three options:
   - 2.1. If $0 < \xi \le E_P$, then if $B_j = 1$, then on the second state of the expert $C_i = 1$, if $S_j = 2$, then the expert ignores wrong decision and saves 0 to the C_i cell.
   - 2.2. If $E_P < \xi \le E_P + E_N$, then if $B_j = 1$, then the expert makes wrong decision – he ignores the correct 1 and saves 0 to the C_i. If $B_j = 2$, he saves 2 to the C_i by mistake.

 ISBN: 978-1-61208-269-1

2.3. If $\xi > E_P + E_N$, then the expert cannot estimate the B_j cell solution and saves 0 to the C_i.

3. C array is saved to B. If it still has zeros, then proceed from the point 2, otherwise – move to the point 4.
4. At this stage the group's solution is defined by the votes majority. If ones dominate, then it's considered, that the group has made a correct decision. If twos dominate, then the decision is wrong.

Basing on this algorithm, a computer program was developed, that allowed to find G, the probability of making correct decision for one expert in the end of convergent iterative process, according to predefined parameters of the model: $\mathrm{G_P}$, $\mathrm{G_N}$, $\mathrm{E_P}$, $\mathrm{E_N}$, when he use his expert ability of choosing between others' solutions together with the ability of ideas generation. To find P – the probability of finding correct solution by votes majority in the group of M experts – let us use famous formula, that follows from the Condorcet's Theorem [3]:

$$\mathrm{P} = \sum_{\mathrm{i=0}}^{\frac{\mathrm{M-1}}{2}} \mathrm{C_M^i G^{M-i} (1-G)^i} \qquad (1)$$

One of the research objectives was to explore the dependence of the group intellectual potential upon the model parameters. As the intellectual potential, the function $\mathrm{IP} = \frac{1000}{\mathrm{M}}$ was considered. It is inversely proportional to the number of experts that have found correct solution with probability $1 - \varepsilon$, where ε is a predefined small value. In our experiments $\varepsilon = 0.001$.

Computer experiments were performed with different values of the parameters G_P and G_N. Their variations are given in the Table I. There are also presented the evaluated values of M (top left corner of a cell) and IP (bottom right corner of a cell) for $\frac{E_N}{E_P} = 2$ and $\frac{E_N}{E_P} = 0.5$. Obviously, the cases where $G_P + G_N > 1$ are impossible. Furthermore, some small values of these parameters where $G_N > G_P$ also give no results because the number of experts M in these cases tends to infinity.

It is shown that intellectual potential of the experts differs by degrees and the legislative component brings a considerable contribution to the intellect structure. Moreover, it is the very thing that allows to overcome the Condorcet's "border".

4. Conditions of correct solution obtainment

Basing on the experiments the empirical inequality was defined that bounds up parameters of the model. When it holds, it's always possible to find M, such that guarantees obtaining correct solution with probability 0.999 for the group:

$$\mathrm{G_P} + \frac{\mathrm{E_P G_V}}{\mathrm{E_P + E_N}} > 0.5 \qquad (2)$$

For convenience, in practical applications, the constraint (2) can be deducted to:

$$\frac{\mathrm{E_P}}{\mathrm{E_N}} > \frac{1-2\mathrm{G_P}}{1-2\mathrm{G_P}} \qquad (3)$$

Here are some consequences of the inequality (3):

1. If experts have low ability of decision generation ($G_P < G_N$), then for obtaining a correct solution it's necessary for them to have high abilities of evaluation ($E_P > E_N$).
2. If experts have high ability of decision generation ($G_P > G_N$), then for obtaining a correct solution it's not necessary for the group to have high evaluation abilities ($E_P < E_N$).
3. If experts' abilities of both generation and evaluation are both low, they cannot obtain a correct solution. Moreover, if the portion of such experts in the expert group increases, then correctness of the group work result decreases.

Convergence of MEC to the correct solution and fulfillment of the inequality (3) in quantitative form were inspected also in condition of normal distribution of the model parameters. Just like in Condorcet's research, in this paper, basing on computer experiments, it's deducted that the results obtained in case of normal distribution of parameters in limits of statistical errors, concurred with results of inspection, preformed with average values of the parameters.

Let us consider a practical example of the MEC usage. One of the experiments that were made by the authors is an IQ test. The test consists of a set of questions, i.e. 50. Every question is provided with several answers only one of which is correct. There are 10 students (experts) participating in the test solving.

According to the MEC model solution template can be divided into 50 slots – one for a question. On the first iteration of MEC each of 10 students proposes answers for every question he is sure and leaves unanswered those he has doubt with. On the second and other iterations each expert receives a solution of another randomly chosen expert and rates his answers. If some answer of the expert being rated seems more correct that the one which the rater has in susceptive slot then he replaces his old slot value or blank slot with the value of that user. These iterations proceed until all solutions are similar (the students have come into conclusion). Thus, the solution that is left in the solutions population is considered as the most correct one.

On multiple experiments the method demonstrated high increase of the intellect level of real students groups.

V. Conclusion and Future Work

On the results of the research, the following conclusions were made:

- Simple and efficient iterative method of collective decision-making was proposed and explored.

TABLE I. VALUES OF THE MODEL PROPERTIES

$E_N/E_P = 0.5$									
G_P / G_N	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	129 7.75	81 12.3	55 18.2	39 25.6	29 34.5	21 4.,6	17 58.8	13 76,9	9 111.1
0.2	535 1.87	233 4.29	127 7.87	81 12.3	55 18,2	37 27,0	27 37.0	21 46.7	-
0.3	-	2147 0.465	531 1.88	233 4.29	129 7.75	81 12.3	55 18.2	-	-
0.4	-	-	-	2143 0.467	533 1.88	235 4.26	-	-	-
$E_N/E_P = 2$									
G_P / G_N	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	-	-	-	535 1.87	129 7.75	55 18.2	29 34.5	17 58.8	9 111.1
0.2	-	-	-	2145 0.466	233 4.29	81 12.3	39 25.6	21 46.7	-
0.3	-	-	-	-	535 1.87	127 7.87	55 18.2	-	-
0.4	-	-	-	-	2147 0.465	235 4.26		-	-

- The conditions were determined, that can guarantee obtainment of correct solution by a group of experts using the Method of Evolutionary Solutions Coordination.
- The experts' competence of legislation and decisions evaluation was engaged together with the ideas generation competence. It helps experts groups to operate more efficiently and eliminate the Condorcet's border.

The obtained results have become the beginning of large research, that is divided into two branches. The first is exploring applications of MEC with creation of different variations of the method and different coordination models. For example, so-called Dynamic Slots variation was developed to apply the method in machine translation. Also MEC was utilized in collaborative text creation and concepts visualization. For all these applications software products were created. The second direction of the related work is improvement of MEC. A considerable part of theoretic research is related to elitist modification of the iterative process. It means that in creation and legislation stages experts' weights differ according to their creative and evaluative skills. Several algorithms were developed for calculating these skills.

Implementation of the coordination process in collective intelligence algorithms opens new fields of application for information technologies. Not everybody is able to create great solution for certain problem, but comparison of several solutions with picking the most impressive one is a simpler task, and usage of this point gives wide opportunities. In order to apply these concepts in practice we need to discover new models of collective work organization, that, first of all, can be attractive for experts, e.g., in the Wide Web, and in the second, will provide solutions for some actual problems.

ACKNOWLEDGEMENTS

This work was supported by the Russian Foundation for Basic Research, grant #13-07-00958 "Development of the theory and experimental research of a new information technology of self-managed crowdsourcing".

REFERENCES

[1] Woolley A. W., Chabris C. F., Pentland A., Hashmi N., and Malone T. W. Evidence for a Collective Intelligence Factor in the Performance of Human Groups // Science. 2010. V. 330 pp. 686–688.

[2] P. Venetis, H. Garcia-Molina, K. Huang, and Polyzotis N. Max Algorithms in Crowdsourcing Environments // April 16–20, 2012, Lyon, France.

[3] Le Marquis de Condorcet. Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. 1785. Les Archives de la Revolution Française, Pergamon Press.

[4] Bachrach Y., Minka T., Guiver J., and Graepel T. How To Grade a Test Without Knowing the Answers | A Bayesian Graphical Model for Adaptive Crowdsourcing and Aptitude Testing // 2012 pp. 1–2.

[5] Mitchell M. An Introduction to Genetic Algorithms, MIT press //1996. pp. 2–23.

[6] Kosorukoff A. Human Based Genetic Algorithm // Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana-Campaign. January 2001.

[7] Turchin V. The phenomenon of science. A cybernetic approach to human evolution // Columbia University Press, New York, 1977.

[8] Stern W. The Psychological Methods of Intelligence Testing (G. Whipple, Trans.). // Baltimore: Warwick and York.

[9] Eysenck H. J. The Structure and Measurement of Intelligence. // University of London. Institute of Psychiatry.

[10] Гузеев В.В. Преподавание. От теории к мастерству. – М.: НИИ школьных технологий, 2009. pp. 261–264

[11] Кузнецов В.В. Технологии интернет-образования . http://www.masters.donntu.edu.ua/2006/fvti/grach/library/st2.htm

[12] Козленко А.Г. Информационная культура и/или компьютер на уроке биологии.

[13] Чернухин Ю.В. и Здоровеющев В.В. Сетевой эволюционно-генетический алгоритм принятия коллективных решений в человеко-машинных системах.

[14] Затуливетер Ю.С. Информация и эволюционное моделирование. Труды международной конференции «Идентификация и проблемы управления» SICPRO 2000, Москва. pp. 1529-1571.

[15] Faratin P., Sierra C., and Jennings N.R. Using similarity criteria to make negotiation trade-offs. Proceedings Fourth International Conference on MultiAgent Systems. IEEE Comput. Soc. //2000.

[16] Ehtamo H., Ketteunen E., and R. Hamalainen, Searching for Joint Gains in Multi-Party Negotiations. European Journal of Operational Research. //2001

[17] Klein M., Faratin P., Sayama H., and Bar-Yam Y. Negotiating Complex Contracts. Sloan School of Management 3 Cambridge Center, NE20-336. //2007