

Generalization Aspect of Neural Networks on Upgrading Assimilation Structure into Accommodating Scheme

Waleed, A.J. Rasheed and Hamza A. Ali

Department of Software Engineering, Faculty of Computer Science,
Al-Isra Private University, P.O. Box 22, Code 11622, Amman, Jordan

Abstract: Problem statement: Generalization feature enhancement of neural networks, especially feed forward structural model has limited progress. The major reason behind such limitation is attributed to the principal definition and the inability to interpret it into convenient structure. Traditional schemes, unfortunately have regarded generalization as an innate outcome of the simple association, referred to by Pavlov and had been modeled by piaget as the basis of assimilating conduct. **Approach:** A new generalization approach based on the addition of a supportive layer to the traditional neural network scheme (atomic scheme) was presented. This approach extended the signal propagation of the whole net in order to generate the output in two modes, one deals with the required output of trained patterns with predefined settings, while the other tolerates output generation dynamically with tuning capability for any newly applied input. **Results:** Experiments and analysis showed that the new approach is not only simpler and easier, but also is very effective as the proportions promoting the generalization ability of neural networks have reached over 90% for some cases. **Conclusion:** Expanding neuron as the generalization essential construction denoted the accommodating capabilities involving all the innate structures in conjugation with Intelligence abilities and with the needs of further advanced learning phases. Cogent results were attained in comparison with that of the traditional schemes.

Key words: Neural network modeling, generalization, psychology development

INTRODUCTION

Generalization ability of Neural Networks (NNs) is considered as the most important performance criterion^[1]. So many researchers of this domain have been making intensive efforts to promote neural network generalization ability. Learning method based on combinations of weak classifiers is reported by Chuanyi and Sheng^[2]. Weak classifiers such as linear classifiers (perceptrons) which can do a little better than making random guesses, then combined through a majority vote, resulted into good generalization performance and a fast training time.

Several methods have been studied such as fuzzification of input vector^[3], regularization^[4], result-feedback^[5], early stopping^[6], neural network ensembles^[7]. Although these methods can improve the generalization ability of NNs to some extent, however, the problem of NNs' generalization is generally still not solved or not completely solved. This can be attributed to the fact that the principle behavior of artificial neural networks is of instance-based learning. A neural

network should learn a relation from limited data and properly respond to unseen input^[8], therefore, it is impossible for NNs to solve all the problems by learning from limited examples and hence developing new methods for improving NNs' generalization ability is highly needed.

To improve NNs' generalization ability, Ishibuchi and Nii^[3] used fuzzification of input vector to avoid over fitting. Recently, a new algorithm^[5] to improve the learning performance of neural network through results-feedback, called FBBP algorithm, presented by Wu and Wang, can improve NNs' generalization ability too. This FBBP-based algorithm is an inner-and-outer layer learning method in which weight value renewing plays the dominating role with the assistance of input renewing. It minimizes the error function of neural network through the dual functioning of weight value and input vector value tuning, where tuning of the input vector is similar to fuzz the input vector. This idea brings new inspiration and people had previously devoted large amounts of time to tuning weights of NNs for improving the NNs' performance (including

Corresponding Author: Hamza A.Ali, Faculty of Computer Science and Information Technology, Al- Isra Private University, P.O. Box 22, Code 11622, Amman, Jordan

the generalization ability), but lacked new ideas. Feng *et al.*^[9] suggested an approach that appropriately shrinks or magnifies input vector, thereby makes the generalization ability of NNs improved. This algorithm is called “Shrinking-Magnifying Approach” (SMA) that finds the appropriate Shrinking-Magnifying Factor (SMF) and obtains a new neural network having better generalization ability. Ganchev *et al.*^[10] tackled generalized locally recurrent probabilistic neural networks GLRPNN, for text independent speaker verification. It is contrasted with that of Locally Recurrent PNNs, Diagonal Recurrent Neural Networks, Infinite Impulse Response and Finite Impulse Response MLP-based structures, as well as with Gaussian Mixture Models-based classifier.

The current paper proposes a modified structure based on Pavlov and Piaget theorems^[11,12] in order to enhance the generalization capability of Feed Forward neural networks. It is designed to merge both into a classical and generalization learning characteristics simultaneously in one network simulating human conduct in relation with responses of the different mental activities adopted for various levels of timing consideration in output generation. Basically this structure incorporates an extra layer attached to the output layer of a traditional network with the capability of dynamically adjustable neuronal threshold during both of training and testing phases. Besides, a convenient procedure has been adopted for training the whole network with the aid of Genetic Algorithm. The procedure involves two learning cycles; the first deals with the traditional scheme and an additional neuron expanding the last layer and the second cycle deals with the additional layer and with the attributes of the output generated from the last expanded layer of the traditional scheme besides the required output of the training data. The first cycle stands for Pavlov learning assimilating capability and the second cycle substantiates Piaget arguing through the accommodating capability. Different testing data have been used in a wide range of experiments. Adequate results of success are gained and that in turn used to approve the validity of the proposed model.

Background, pavlov and piaget generalization structural interpretation: Throughout the intensive studies of human brain, neural networks appear as one of the successful and efficient abstracting models. These models prompted enormous interest of researches in psychology and physiology besides other related supporting applied sciences and medical investigations. The concrete basis, used to establish the main concept, is envisaged to lay on Pavlov theorem of conditional

simple association^[12]. This theorem has been conjugated with Hebb’s theorem to simulate the weighting characteristics of the reticular formation of the in between cell connections of the nervous system, especially the synaptic junctions^[13]. However, there were no literal interpretation to the natural processing carried out in the brain as a system with its associated behavior and constituents.

Based on the foregoing discussion and that of the psychological fundamentals, it could be stated that Pavlov theorem is faithfully interpreted and implemented with the traditional neural network models, but unfortunately to what relates Piaget’s theorem, these networks failed to do so. It is known through the literature of the developed models, generalization is envisaged as an intuitive and as side effect of the connection schemes. While the significant deduction, as Piaget argued, generalization is an active learned process rather than being passive behavior of an association scheme. This might address the major obstacle stands behind improving the generalization capability of the traditional connection schemes where generalization enhancement had been attributed to data selection and net layering dimension scales as major trajectories of the efforts devoted for the developing purposes^[14].

MATERIALS AND METHODS

The proposed model involves dynamic response in data generation. The model consists of two distinct parts, a traditional neural network consisting of input layer, number of hidden layers and an output layer with biasing neuron, referred to hereafter as the traditional connection scheme, extended by an additional output layer with its own biasing neuron too, as shown in Fig. 1. This extra layer differs from those of the common preceding ones in the connection layout by its neuronal threshold setting mechanism and control of its variations. The weights matrix of the traditional layers is adjusted during the training phase and kept constant in the testing phase, whereas the additional layer keeps on changing its neuronal thresholds on both of the training and testing phases. Moreover, a convenient procedure has been adopted for training the whole network with the aid of Genetic Algorithm.

Traditionally, NNs are static structures after being trained. Therefore, signals propagate from the input to the output layers via the hidden layers on fixed values of connection weights and threshold values. By recalling that these values of connection weights imply the main data association, it seems difficult to vary any weight during testing phase in order to avoid any

arbitrary generation of outputs. Hence, the only permissible action to vary the attributes of the scheme denotes threshold tuning of a specified layer for new inputs, on condition that this layer should maintain the association of the training set unchanged.

Model architecture: Figure 1 shows the schematic diagram of the propose network. The first block shows the traditional network scheme (or Atomic scheme), while the second block shows the suggested additional output layer (or Supporting Layer). This layer extends signal propagation of the whole net in order to generate the output in two modes. The first mode deals with the required output of trained patterns with predefined settings, while the second mode tolerates output generation dynamically with tuning capability for any newly applied input.

In order to enable threshold tuning of the supporting layer to take place, the last layer of the atomic scheme is expanded with a new neuron called the band selector neuron. The output of this neuron is utilized as a bias to the supporting layer, Fig. 2.

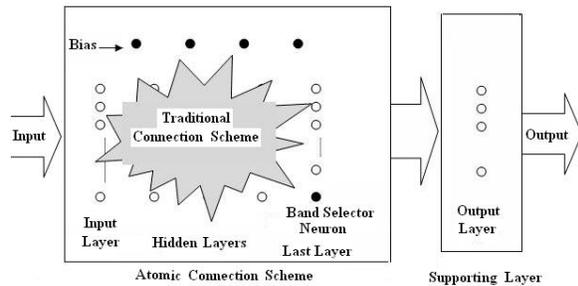


Fig. 1: Model structure

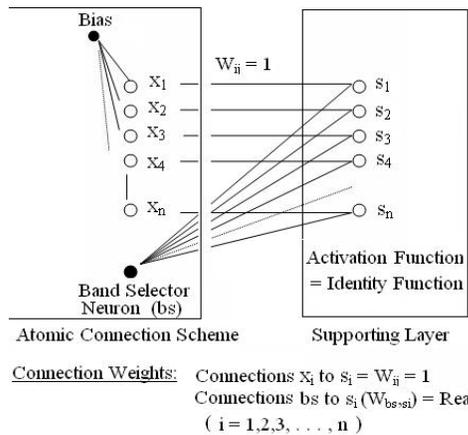


Fig. 2: Supporting layer connection characteristics

Hence, unlike the rest layers of the atomic scheme, wherein a bias input is adjusted and kept fixed afterwards, the supporting layer tunes its neuronal threshold values in accordance to the output of the band selector neuron continually. In fact, this output is made to be regulated as a function of the input of the whole model.

However, the major association attributes of the supporting layer denotes the weight values of the connections needed to link the band selector neuron to its neurons and they are committed to the second cycle of model training. This cycle, definitely, will be commenced when the first cycle terminates and obtains the needed association in similar manner to that of the classical phase of training in traditional nets. The only difference here is that an extra output value is added to each pattern of the training set, as an additional argument representing band selector output. It must be noted here that neurons of the last layer in the atomic scheme are connected to their counterparts of the supporting layer with unity weight and in one-to-one configuration.

Model training: Due to the requirements of the given specifications, genetic algorithm is adopted to determine the overall connection scheme of the presented model. Although there are no anomalous restrictions to apply dedicated activation function or limit bounds to the input and output levels, it is found more applicable to use identity mode of activation function to the supporting layer. This function offers efficient error compensation when output drifts are detected on the preceding layer of the atomic scheme and thus it tends to recall the required output at the supporting layer responses throughout the training.

A pre-organization is needed to facilitate the training; patterns of the training set emergently are ought to be divided into two groups. The first group ideally involves the most primitive pattern associations, while the second group involves the patterns that are supposed to support the generalization capability. These patterns, in general, are extended by an extra argument in their related outputs. The value of this argument is given a null estimation (zero) to all patterns of the first group and a random number to the patterns of the second group, as shown in Fig. 3.

Furthermore, the training can be characterized by two stages as follows:

- Figure 3b is used as the training data for the first stage. The concerned structure of the neural network denotes the atomic structure involving the band selector expanding the last layer

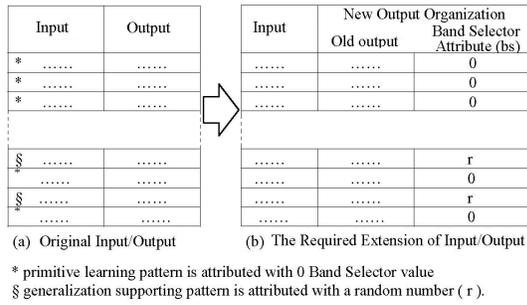


Fig 3: Training set organization

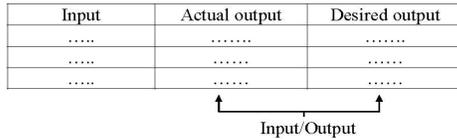


Fig. 4: The training patterns for the second stage

- The last layer of the atomic scheme is connected to the supporting layer by one-to-one connections and the bias of each neuron at this layer is derived from the band selector neuron. Therefore, the training here is conducted in order to determine the weights of the bias connections only

Figure 4 shows the input/output pattern for the second training phase including the original input. The association of the training here considers the actual output of last neurons of the traditional net as inputs that are to be laid to generate the desired output itself including the effect of the band selector neuron.

RESULTS

Probably the major problem which researchers confront in the course of testing any proposed neural network structure is the standardization issue of the compared schemes. Structural constituents of layering organization, neuronal compositions of each layer and the data of the underlying applications used are the main parameters addressed into this context. Anyway results could not be judged perfectly certain without any doubts. That is because of the absence of identical simulation programming coding, data representation and training algorithms. However, it is intended in this work to standardize the comparison parameters between the traditional nets and the presented structure as much as possible. Specifying same constituents with different examples and utilizing common data, which have been provided on Proben1 set [10], denotes all the possible trends that have been implemented to conduct the experimentation task. In this task, genetic algorithm is used as the training tool.

Table 1: Experimentation application fields

Application	No. of inputs		No. of outputs		No. of training patterns	No. of generalization patterns	No. of testing patterns
	Binary	Real	Binary	Real			
Cancer diagnosis	-	9	2	-	350	175	174
Glass types	-	9	6	-	107	54	53
Solar flar	-	24	-	3	533	267	266
Majority functions	7	-	1	-	64	32	32
Random association	-	6	-	2	60	15	25

Experiments involve wide range of application fields, these fields are shown in Table 1. Table 1 also shows the number of data items allocated for training and testing. The provided application data is usually divided into two sets constituting 80 and 20% ratios of the universe for training and testing purposes, respectively. In the current work, the 80% sample set is further been subdivided into two groups in order to cover the requirements of the first and second training stages of the proposed network.

As an experimental example, the cancer diagnosis application data are employed for investigation of the proposed concept using neural network with different constituent models. Various numbers of hidden layers and hidden neurons were implemented resulting into various network specifications as detailed below.

NN with one hidden layer: No. of Generations = 600, Population size = 50, Mutation rate = 2%, Selection method = Rank, No. of hidden cells = 10, Hidden layer activation function = sigmoid, Output layer activation function = sigmoid, Selection rate = 80% out of the overall chromosomes.

NN with two hidden layers: No. of Hidden cells at hidden layer 1 = 10, No. of Hidden cells at hidden layer 2 = 5, Hidden layers 1 and 2 and output layer activation function = sigmoid, No. of Generations = 800, Population size = 50, Mutation rate = 5%, Selection method = Rank, Selection rate = 80% out of the overall chromosomes.

NN with three hidden layers: No. of Hidden cells at hidden layer 1 = 14, No. of Hidden cells at hidden layer 2 = 10, No. of Hidden cells at hidden layer 3 = 6, All Hidden layers 1, 2 and 3 and output layer activation function = sigmoid, No. of Generations = 1000, Population size = 50, Mutation rate = 5%, Selection method = Rank, Selection rate = 80% out of the overall chromosomes.

The main measured factor in all the three studied networks is the Mean Square Error (MSE) and the

Table 2: Cancer diagnosis using one hidden layer NN

Trial No.	Error 1	Error 2	Error 3	Error 4	Generalization iteration (Stage 2)	Error 5	Error 6	Generalization Improved
1	11.439	4.239	30.415	10.665	554	10.661	2.406	Yes
2	10.703	3.751	29.441	12.774	5	12.453	2.538	Yes
3	11.334	2.703	27.520	11.302	4	11.006	3.278	No
4	10.911	4.900	31.717	11.058	22	11.042	3.655	Yes
5	10.326	2.586	30.716	11.557	20	11.516	2.228	Yes
6	11.168	2.394	31.469	11.239	13	11.131	2.038	Yes
7	10.111	4.098	28.139	11.139	600	15.547	4.176	No
8	11.115	3.283	34.108	10.927	188	10.927	2.127	Yes
9	8.706	3.130	29.861	11.408	148	11.405	3.137	No
10	10.757	3.055	30.751	11.049	0	11.040	2.662	Yes

Table 3: Cancer diagnosis using two hidden layers NN

Trial No.	Error 1	Error 2	Error 3	Error 4	Generalization iteration	Error 5	Error 6	Generalization Improved
1	11.457	4.294	29.203	10.253	166	10.234	2.140	Yes
2	10.464	2.894	29.968	11.112	9	11.056	1.769	Yes
3	9.510	3.755	32.096	12.573	11	12.147	2.886	Yes
4	10.007	3.544	27.708	10.933	58	10.917	3.042	Yes
5	9.202	3.974	29.877	10.583	8	10.521	2.995	Yes
6	10.047	2.807	31.817	11.658	67	11.552	2.465	Yes
7	9.274	2.299	26.892	8.207	45	8.201	1.742	Yes
8	9.613	3.425	28.805	10.818	45	10.809	2.308	Yes
9	9.830	2.258	29.259	11.647	500	11.643	2.677	No
10	2.767	2.767	30.760	10.646	0	10.642	2.318	Yes

Table 4: Cancer diagnosis using three hidden layers NN

Trial No.	Error 1	Error 2	Error 3	Error 4	Generalization iteration	Error 5	Error 6	Generalization Improved
1	11.922	3.130	32.781	13.432	7	13.432	2.838	Yes
2	10.710	4.469	32.096	12.495	7	12.160	2.184	Yes
3	10.460	5.684	31.570	12.233	3	11.974	2.866	No
4	12.761	2.616	31.904	11.440	77	11.386	3.071	Yes
5	11.426	4.275	33.069	10.990	1	10.970	2.617	Yes
6	11.113	2.938	31.372	13.294	2	13.044	2.651	Yes
7	11.866	2.886	29.944	12.381	79	12.366	2.874	Yes
8	12.264	5.590	33.291	12.561	12	12.338	2.801	Yes
9	11.062	3.941	31.585	13.005	118	12.975	1.824	Yes
10	12.452	2.255	35.717	15.481	6	15.477	0.924	Yes

generalization improvement. Hence, in Table 2-4, error 1 and 2 are the main MSE values of traditional based scheme resulted from the training and testing phases respectively. Error 3 and 4 are MSE values of the last layer (first training stage) with and without band selector neuron error, respectively. Error 5 and 6 are MSE values of the generalization based supporting scheme resulted from the final training (second training) stage and testing phase, respectively.

In addition, the genetic based training algorithm has alternately been set to a wide range of specifications. Towards the decision making of whether the structure had achieved an improvement or not, the error calculations for different stages have been traced and given along for each experiment as it is shown in (Table 2-4). A final decision is given at the last column of the concerned table. Similar outcomes are attained

along other application experiments which had been conducted in the same manner.

Generalization improvement is noticeably clear in the neural networks with two and three hidden layers as compared with one hidden layer. From Table 3 and 4, 90% generalization was reported.

DISCUSSION

In spite of having the belief to address a common definition to Generalization that is inspired by both of psychologists and neural network specialists, it seems that there is a serious contradiction in interpreting the functioning nature of such feature from the structural proposition.

Neural networks specialists have agreed on the definition of Generalization to be explained as the

ability of the network to respond to input that it has not seen before. This input may be partial or incomplete. For that, generalization takes the ability to learn and self adjust a step further. Therefore, it could be concluded that this system by itself can "hypothesize" a response. Of course, concise definition for the elaboration of what is meant by "first time seen input" denotes incomplete patterns with all of their possible modalities.

On the other side of the psychologists, it is worth to mention that such feature, i.e. Generalization, is expressed in a wider scope than that of the verbatim description of the first confined definition. Generalization, in this field of study, is classified to fall into different levels. The most primitive generalization is bringing together grouping objects on the basis of individual random feature (syncretic combinations). The more complicated level is the integrating generalization that attributes a newly formulated feature to the similar object of an underlined property of a given sample. And finally, the most complex level of generalization is that in which a distinct line is drawn between specific and generic characteristic, incorporating the object into a certain conceptual system^[15].

This scope of declaration has been thoroughly studied by Piaget to summarize convenient logical principles that is investigated in the current work. Piaget referred indirectly to the related definition of neural network Generalization term as being involved into the exploration of new object and phenomena and to the so called derived secondary reactions. This capability is supported along the progress of intelligence on different levels of association initiated by stages of assimilations and accommodation towards higher levels of more complex generalizations^[12].

However, the different reactions of generalizations or creative responses to uncomforted input are subjected to proposal of learning progresses including intelligence that is indeed an outcome of successive claims of distinguishing instinctive and learned behavior topics. In this course, it is argued that it has long been recognized, in fact, that great many instinctive behavior patterns which in their general outlines can be regarded as innate, yet depend for their detailed form and orientation upon specific learning.

In order to grasp the main guidelines sought by the above discussion, the work reformulates the psychological definitions along with the structural properties to an adequate interpretation as given:

- Neural Networks studies depend on functional definition rather than the structural approach of Generalization comprehension

- Generalization is a learnable capability and not an innate conduct
- Generalization is a higher level behavior that could be established as a structure on a primitive structure which represents a low level conduct of association. These two levels interpret the accommodating and assimilating stages of Piaget argument
- Due to the fact of considering the generalization as a learnable capability, learning pattern should involve two groups. The first stands for the simple association and the second stands for the generalization ability
- Human responses differ between stimulating outputs of simple association and outputs of generalization. The two types can be distinguished by their time responses. The first is the faster, while the second is the slower. This had overviewed the reason to be attributed to the nature of the structure itself. Although both of the activities are realized by a common structure, the slower one should consist of static features, whereas the faster should consist of dynamic features which need to settle before contributing output generation. This opinion has been utilized efficiently for the overall design of the proposed model

CONCLUSION

Obviously, the significant remarks concluding the outcome of these experiments reflect the logical interpretation of the psychological postulates to the developed structure of the present work. In particular, to consider traditional schemes (Feed Forward models) and their related training phases as Pavlov dependent schemes which in turn denotes the assimilating capabilities of the human. Whereas, considering the extended structure with the expanding neuron as the generalization essential construction that denotes the accommodating capabilities involving all the innate structures in conjugation with Intelligence abilities and with the needs of further advanced learning phases. Such point of view expresses the main orientation of the work in proposing an adequate structural concept to interpret the way in order to promote the simple association for higher level of mental capability.

REFERENCES

1. Hagan, M.T., H.B Demuth and M. Beale, 2002. Neural Network Design. China Machine Press, CITIC Publishing House, Beijing, ISBN: 0-534-94332-2.

2. Ji, C. and Ma, 1997. Combinations of weak classifiers. *IEEE Trans. Neural Networks*, 8: 32-42. DOI: 10.1109/72.554189
3. Ishibuchi, H. and M. Nii, 1998. Fuzzification of input vector for improving the generalization ability of neural networks. *Proceeding of the International Joint Conference on Neural Networks*, Anchorage, May 4-8, Alaska, pp: 1153-1158. DOI: 10.1109/FUZZY.1998.686281
4. Weigand, A.S., D.E. Rumelhart and B.A. Huberman, 1991. Generalization by Weight Elimination with Application to Forecasting. In: *Advances in Neural Information Processing Systems*, Lippman, R.P., J.E. Moody and D.J. Touretzky (Eds.). Morgan Kaufmann, San Mateo, CA., pp: 575-582.
5. Wu, Y. and S. Wang, 2004. A new algorithm to improve the learning performance of neural network through result-feedback. *J. Comput. Res. Develop.*, 41: 488-492. (In Chinese). DOI: 10.1109/WCICA.2004.1341928
6. Sarle, W.S., 1995. Stopped training and other remedies for over fitting. *Proceedings of the 27th Symposium on the Interface*, pp: 352-360. <http://mars.elcom.nitech.ac.jp/~anto/NEURO/inter95.ps.gz>
7. Opitz, O. and R. Maclin, 1999. Popular ensemble methods: An empirical study. *J. Artif. Intell. Res.*, 11: 169-198. <http://www.d.umn.edu/~rmaclin/publications/opitz-jair99.pdf>
8. Inohira, E., T. Uoi and H. Yokoi, 2008. Generalization capability of neural networks for generation of coordinated motion of hybrid prosthesis with a healthy arm. *Int. J. Innovat. Comput. Inform. Control*, 4: 471-484. <http://www.ijcic.org/07-044-1.pdf>
9. Feng, N., F. Wang and Y. Qiu, 2006. Novel approach for promoting the generalization ability of neural networks. *Int. J. Signal Process. Waset*, 2: 131-135. <http://www.waset.org/ijsp/v2/v2-2-20.pdf>
10. Ganchev, T., D.K. Tasoulis, M.N. Vrahatis and N. Fakotakis, 2007. Generalized Locally recurrent probabilistic neural networks for text independent speaker verification. *Proc. EuroSpeech*, 3: 1673-1676. DOI: 10.1016/j.neucom.2006.05.012
11. Engelbrecht, A., 2001. Sensitivity analysis for selective learning by feed forward neural networks. *Fund. Inform.*, 46: 219-252. <http://portal.acm.org/citation.cfm?id=1219998>
12. Piaget, J., 1979. *The Origin of Intelligence in the Child*. Routledge and Kegan Paul, Great Britain.
13. Nielson, R.H., 1989. *Neurocomputing*. Addison Wesley Publishing, USA., ISBN: 0-201-09355-3, pp: 433.
14. Robert, J., R.J. Sternberg and J.C. Kaufman, 2002. *The Evolution of Intelligence*. Published by Lawrence Erlbaum Associates, ISBN: 080583267X, pp: 390.
15. Bednya, G.Z. and S.R. Harrisb, 2008. Working sphere/engagement and the concept of task in activity theory. *Interact. Comput.*, 20: 251-255. DOI: 10.1016/j.intcom.2007.07.005