An Investigation into the Relationship between Curse of Dimensionality and Dunning-Kruger Effect

Mehmet Cem Catalbas

Electronic Technology, Vocational High School Department, OSTIM Technical University Ankara/Turkey;
mcemcatalbas@gmail.com; https://orcid.org/0000-0002-9291-1180

Abstract

This study addresses a novel perspective for analyzing the source of confidence in human behavior. The concept of confidence was examined via the relationship between two phenomena in the area of machine learning and psychology, namely the Dunning-Kruger effect and the curse of dimensionality. A relationship was established between these two phenomena which were investigated in the light of neuroscience. This study claims that confidence is highly related with the total time it takes to reach specific information and this relationship is inversely proportional. Image gender classification algorithm was used to analyze this relationship for this study and the curves which were obtained as a result of this analysis was compared with the curve of Dunning-Kruger effect and curse of dimensionality. This relationship has been explained by the knowledge of human's problem-solving ability and mathematical models of memory.

Keywords: Neuroscience, Artificial Intelligence, Dunning-Kruger, Memory Model, Cognitive Science

Boyutun Laneti ve Dunning-Kruger Etkisi Arasındaki İlişkinin İncelenmesi

Öz

Bu çalışma insan davranışındaki güvenin kaynağının incelenmesi ile ilgili yenilikçi bir bakış açısı içermektedir. Güven kavramı, makine öğrenmesinin ve psikoloji biliminin iki fenomeni olan boyutun laneti ve Dunning-Kruger etkisine göre incelenmiştir. Bu iki fenomen arasındaki ilişki sinir bilimi ışığında incelenmiştir. Bu çalışmaya göre bireyin güveni o konudaki özel bilgiye erişme süresi ile alakalıdır ve bu ilişki ters orantılıdır. Cinsiyet imgelerinin sınıflandırılması sürecinde bu ilişki analiz edilmiştir ve bu süreç sonucunda elde edilen eğriler Dunning-Kruger etkisi ve boyutun laneti kavramı ile karşılaştırılmıştır. Bu ilişki insanoğlunun problem çözme yeteneği ve hafızanın matematiksel modellerine ait bilgiler kullanılarak açıklanmıştır.

Anahtar Kelimeler: Sinir Bilimi, Yapay Zeka, Dunning-Kruger, Hafıza Modeli, Bilişsel Bilim

1. Introduction

Throughout history, humankind has explored the causes of their behavior and brain activity which has been a black-box since the ancient Greeks. Neuroscience, which is known as the science of the brain, has opened a way to systematically analyze human behavior [1]. Neuroscience is a branch of science that examines the relationship of the nervous system with behavior and learning. Research into human behavior will never end. As neuroscience can develop, measure and observe living beings, its popularity increases day by day. Additionally, neuroscience is an interdisciplinary science which is related to areas such as, mathematics, linguistics, engineering and computer science [2,3]. Neuroscience provides an understanding of various diseases in the brain including Parkinson's disease and brain tumors [4]. Nowadays, research on the brain has reached a new era thanks to brain imaging techniques such as Magnetic resonance imaging (MRI), especially functional MRI [5,6]. Neuroscience also investigates how the brain processes new information and stores it which is highly important for the field of

education and there are several studies conducted in this area to increase the learning ability of a student in the classroom [7-9]. Another area of interest in neuroscience investigates the reasons behind the decision one makes regarding social life. This trending topic opens a new perspective in the area of advertising as it helps to analyses the reasons customers prefer certain products. Customers' product preferences are analyzed in many different ways by neuroscientists to make products more attractive [10]. Another area which is investigated in the scope of neuroscience is the vision ability of human beings.

Neuroscientists try to understand the stages of how a real object is transformed into a representation in the brain which is a process that also leads to the understanding and development of the areas of computer vision and artificial intelligence [11,12]. According to research in this area, the human brain performs edge detection and determines the shape at first contact after which the colors and other information available are processed. In this study, the relationship between psychology and machine learning was investigated and the source of confidence was analyzed according to this relationship. Firstly, the Dunning-Kruger effect, a phenomenon in psychology, was explained and then another term in the area of machine learning which is known as the curse of dimensionality was explained briefly.

The common part of these two terms was discussed in the light of computational neuroscience. Although these two phenomena occur at different times and in different areas, their graphical representations are quite similar. The area of computational neuroscience creates a bridge between these two phenomena and explains briefly the reason for similarity in terms of graphical representation. Additionally, this study gives information in a different perspective to understand the main reason for confidence in human behavior. According to this study, the main reason of self-confidence is not only related with the general knowledge of humans but also related with the amount of time it takes to reach a piece of information in a specific subject.

2. Dunning-Kruger Effect

The Dunning-Kruger effect, a theory received a Nobel Prize in 2000, was first proposed by Justin Kruger and David Dunning in 1999. In general, this theory examines the relationship between self-confidence and achievements with human behaviors [13, 14]. It is said that people who suffer from the Dunning-Kruger effect have high self-confidence in a subject despite having inadequate knowledge in that specific subject whereas people who do have knowledge don't declare this fact. The basic illustration of the Dunning-Kruger effect is shown in Figure 1.
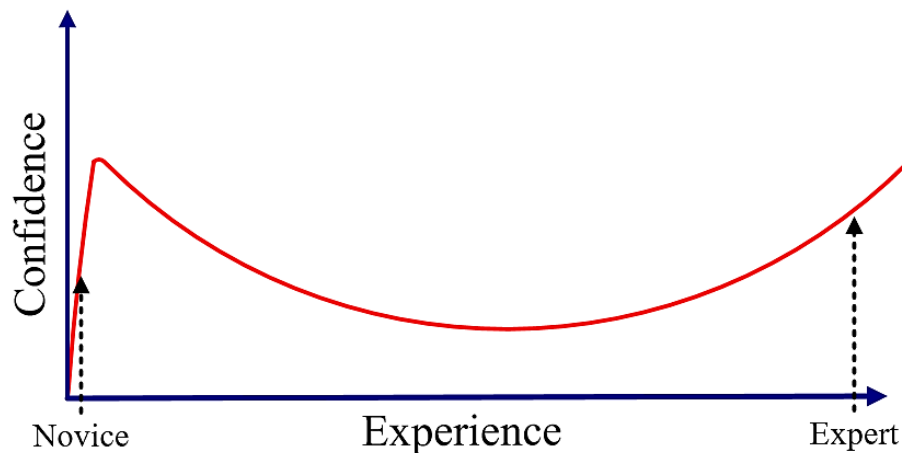

Figure 1 Dunning-Kruger confidence curve

In this study, the relationship between these two phenomena in the fields of machine learning and human behavior was investigated. The relationship between the amount of information and self-confidence is similar in various ways to the curse of dimensionality, an important phenomenon in machine learning. According to Bellman's curse of dimensionality, if there is a larger data set than there should be("required") to solve a problem, this will decrease the performance of the solution to the problem.

3. Bellman's Curse of Dimensionality

The term of the curse of dimensionality was first introduced by Belman in 1961 [15,16]. According to this term, number of feature must be an optimal size to represent data in the storage area and increase the number of the feature do not always give the satisfying result on classification, clustering or regression. In general, the cost of the algorithm increases exponentially with the dimension of the information. This cost makes the calculation impossible or too difficult for large space dimensions.

The problems with increasing dimensions (d) in the dataset are as follows:

•Sparsity on the input becomes more obvious and the redundant part in the dataset increases rapidly.
•The decreasing performance of the algorithm dramatically depends on the increase of dimensions.
•The similarity measurement or classification process of the algorithms becomes redundant due to input and test data becoming the same as each other. The purpose of using a method such as a classifier or a similarity measurement becomes useless.
•Even though the computation and data storage ability of computers have evolved significantly since the term curse of dimensionality was first discussed, computation time and area of usage in the storage are still critical issues in the area of machine learning.

The curve of the curse of dimensionality is shown in Figure 2 and the classifier performance is represented via Equation 1.

$$ClassifierPerformance = \frac{Correct\,ClassificationPercentage(\%)}{ComputationTime} \tag{1}$$
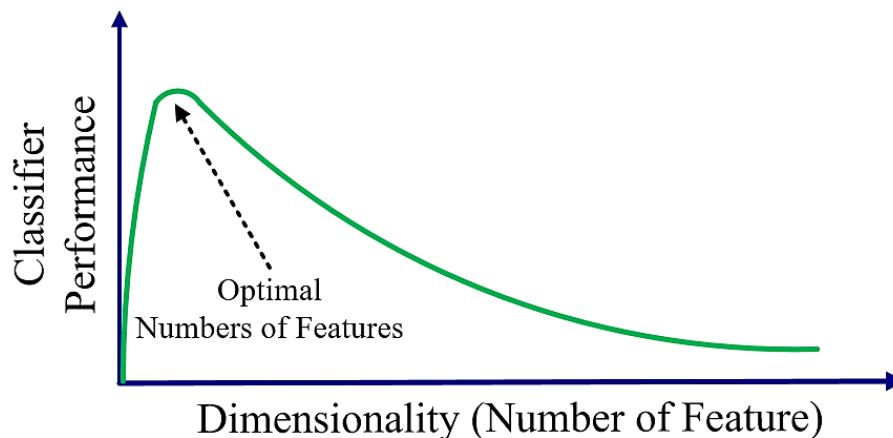

Figure 2 Bellman's curse of dimensionality

As shown in Figure 2, there is a number about an optimal number of feature for classifier performance and this peak point gives information about the optimal representation of features in a dataset with this classifier algorithm.

4. Human Problem Solving and Cognitive Science

According to recent studies in the area of neuroscience, the main reason for the decrease in cognitive ability is the size of information received over time. In a study by Ramscar et al., the processing speed of the brain was found to decrease over time with the size of the stored information and the cognitive model they proposed was tested with linguistic models for the simulation of the human brain [17]. The present study consolidated the relationship between the human brain and machine learning via classification algorithms. As noted at the beginning of this study, processing speed not only affects learning ability but also manipulates human behaviour such as confidence. Processing new information must have a computation cost. This cost is highly related to the size and location of information.

Old studies have suggested that the brains of adult individuals have become stable and do not develop over time. However, it has been observed with current studies in neuroscience that this situation is contrary to what is believed and the neural clusters of human memory are dynamically reconstructed especially at lower level clusters. In the basic studies in the field of human memory, it is foreseen that the human brain consists of three basic parts and these are as follow: after image, primary and secondary memory [18, 19]. Thanks to the developing cognitive and neuroscience today, human memory is predicted to consist of four categories: 1. sensory buffer memory (SBM), 2. short-term memory (STM), 3. long-term memory (LTM), and 4. T action buffer memory (ABM). These categories are classified according to data saving period and the type and size of the stored data [20]. For example, LTM is defined as a category of memory in which knowledge and skills are kept permanently and the storage ability of LTM is about limitless due to large number of neurons (at 100×109 level) and much more potential synaptic connections (at 108,432 level). The basics of human knowledge and long-term memory can be represented by an Object-Attribute-Relationship (OAR) model based on the synaptic structure of human memory [21]. The model of OAR is representing dynamic and hierarchical information clusters which are stored in the memory and this model provides to develop a logical model of cognitive knowledge bases. The mathematical model of OAR model which is stored at LTM is given in Equation 2 [18].

$$OAR \triangleq (O, A, R) \tag{2}$$

where O is a finite set of objects identified by unique symbolic names, A is a finite set of attributes for characterizing each object, and R is a set of relations between objects and attributes. The illustration of OAR model is shown in Figure 3.
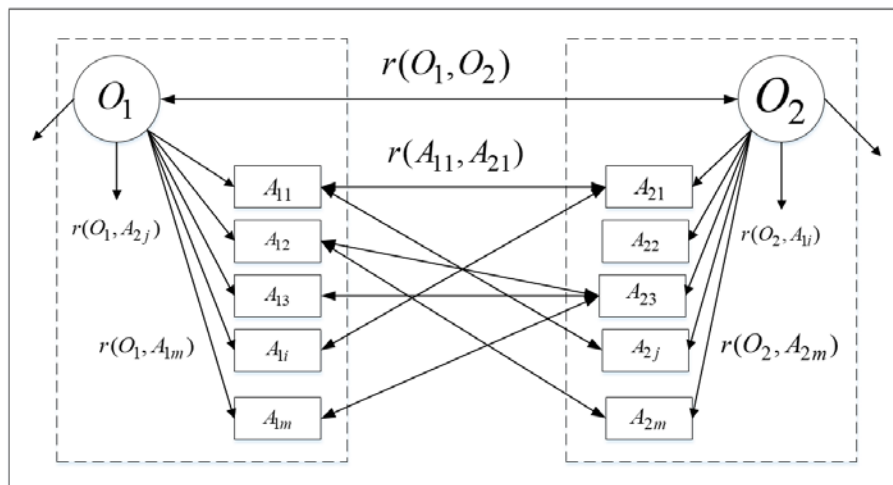

Figure 3 The OAR model of human memory architecture

As shown in the OAR model, the memory architecture of human is consist of the enormous number of networks. These memory structures are similar to the neuron structure, which is the basic building block of the process a, dynamically solving problems. Because neuron structures are also a network structure that contains many interactive and complex relationships with each other. These neuron structures are at the basis of artificial neural networks that form the basis of today's machine learning.

In this study, radial basis function (RBFN) neural networks were used to model the classification process of the human brain [22]. RBF neural networks are generally used to create a nonlinear function between inputs and outputs and the topology of RBF is a form of three-layered feed-forward neural networks. The first layer is defined as the input layer, the hidden layer consists of structures containing a large number of nonlinear RBF activation functions and the third layer is defined as the output layer. The

preliminary structure that separates the RBF neural network from traditional multilayer neural networks is the structure of the activation function and Gaussian transfer function is used as an activation function in RBFN structure [23]. The illustration of Gaussian transfer function is given in Figure 4. The value of radius (σ) is determines the behavior of the model to be created on RBFN. The larger value of selection on sigma causes underfitting, while the low sigma value leads to overfitting on the stage of function approximation or classification.
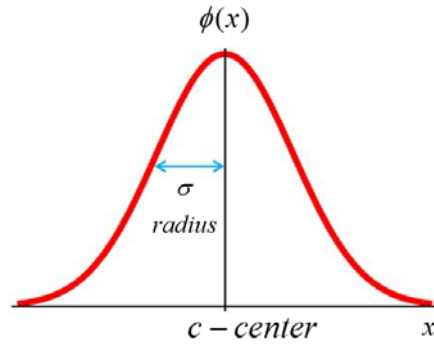

Figure 4 Gaussian Activation Function
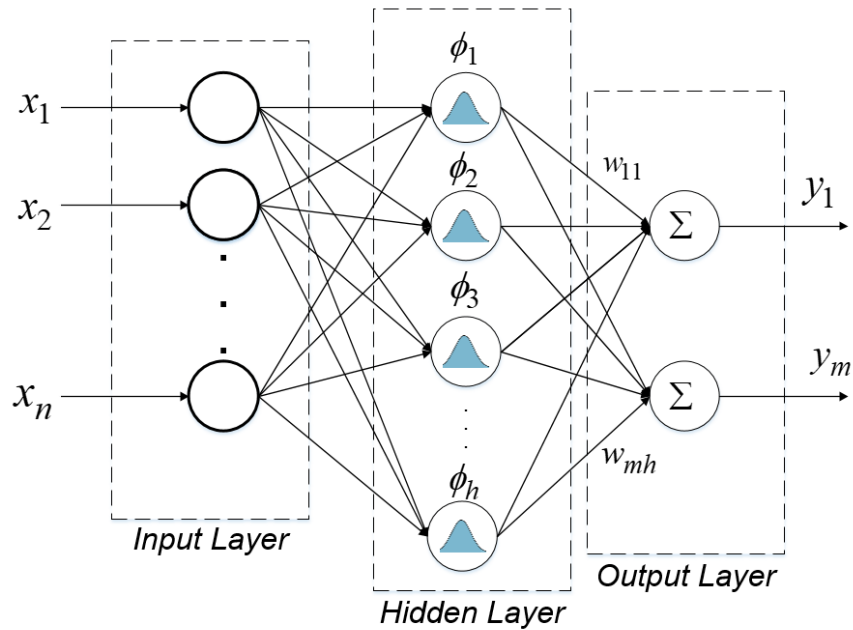
The structure of RBFN is given in Figure 5.

Figure 5 Generic illustration of RBFN structure

The output of activation function ϕ_i is given in Equation 3 and x is equal to input value, c_i is ith center, σ_j is spread of jth transfer function, and $\|.\|$ is Euclidean norm.

$$\phi_i(\|x - c_i\|) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma_j^2}\right) \tag{3}$$

The output (y_k) of the node m at the output layer can be determined as Equation 4.

$$y_k = \sum_{j=1}^{h} w_{jk}\phi_j(x) \tag{4}$$

RBFN is a popular alternative to MLP structure due to rapid training process and a simple structure and it has the capability of universal approximation.

5. Image Classification and Curse of Dimensionality

In this study, the optimal numbers of the feature selection process were realized with an image classification problem. In this problem, gender classification was performed with principal component analysis (PCA) and Radial Basis Function (RBFN) [24,25]. The proposed technique was tested using Aberdeen facial images from Psychological Image Collection at Stirling (PICS image database) [26]. In this dataset, there were 130 input images for each gender and 70 of them were selected for training image while the rest of the images were used for the testing process. The detection of classifier performance was obtained as a result of the cross-validation of the dataset. The flowchart of the classification process is shown in Figure 6.
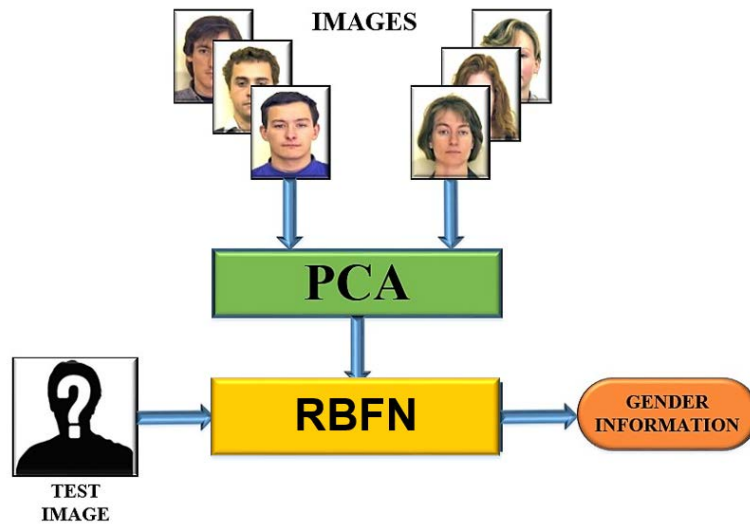

Figure 6 Gender detection from the images

The error percentage of gender classification according to the number of features dimension is shown in Figure 7. As can be seen from the figure, the increase of the selected feature number is meaningless after the first four features after which this kind of selection only causes a decreasing performance in the algorithm.
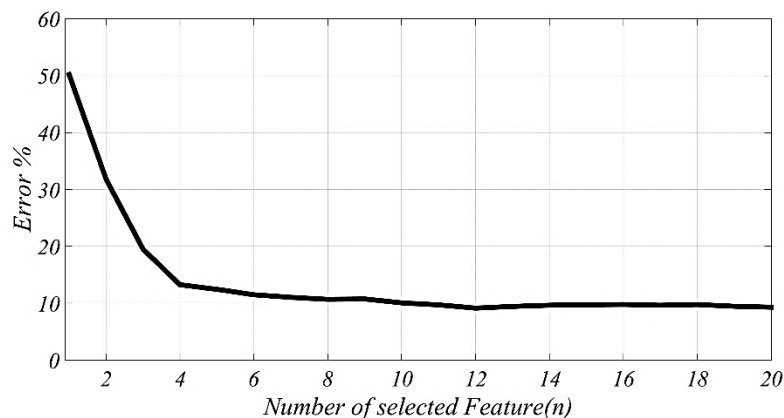

Figure 7 Percentage of error

The computation time of the algorithm versus the selected feature size is shown in Figure 8. As expected, the total computation time of the classifier increased with feature size.
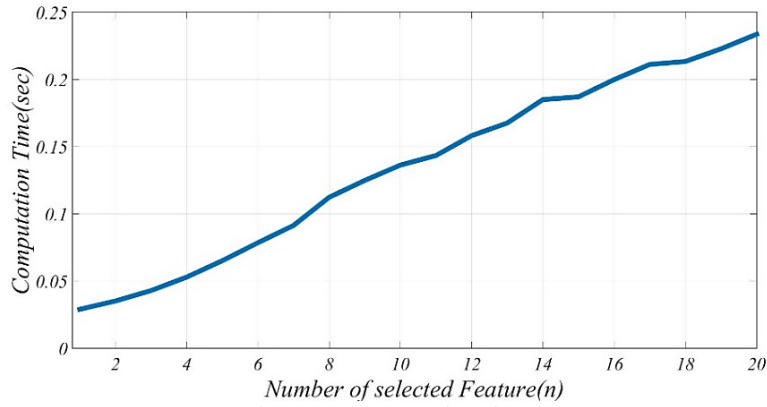
Figure 8 Computation time

The classifier performance of the gender classification process was performed according to Equation 1. As a result of this computation, a specific curve very similar to the curse of dimensionality curve was obtained as shown in Figure 9.
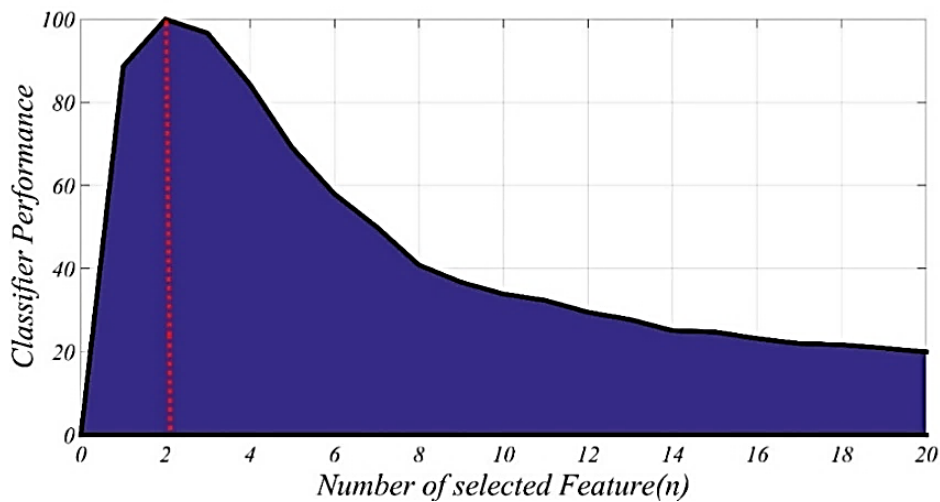

Figure 9 Classifier performance

The selection of the first two components is necessary to achieve optimal classification performance for this gender classification process. Additionally, this curve is also very similar to the curve of the Dunning-Kruger effect. In the image classification process, using the RBFN structure, it was provided to simulate the classification skills of people on computer systems [27]. The results of studies in the field of cognitive science are supportive of this similarity. Hong et al conducted analyzes on students to examine the relationships between cognitive parameters and self-confidence in the area of education [28]. In the study in which 6 different hypotheses were analyzed and the partial least squares (PLS) method was used to analyze the research hypotheses. The hypothesis which is related to our research paper is 6th hypothesis and the hypothesis question is analyses the relationship between self-confidence enhancements (SCE) relevant to learning progress (LP) is examined in the 6th hypothesis. Path graphic showing the relations obtained as a result of the analysis is given in Figure 10.
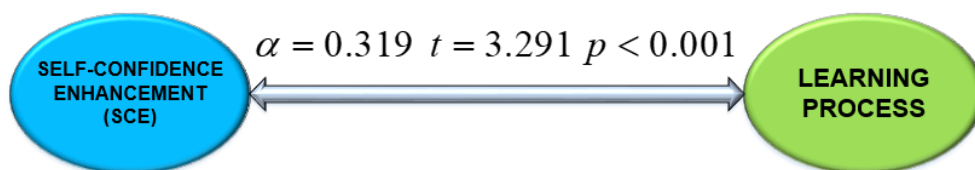

$$\alpha = 0.319 \quad t = 3.291 \quad p < 0.001$$

Figure 10 Relationship between SCE and LP

As a result of the analysis, there is a positive correlation between SCE to the learning process and this result supports the relationship between the self-confidence and performance of data access as we suggest in this article.

6. Results

In this study, the source of confidence in human behaviour was explained according to the curse of dimensionality. Image gender classification and a number of component selections were performed to obtain classifier performance. The obtained curve was found to be very similar to the curse of dimensionality. This study explained the reason for confidence being not only inadequate amounts of information about a specific subject but also the processing time information on the brain. This study provides a new perspective to the Dunning-Kruger effect in psychology and also uses neuroscience to explain the source of confidence. In this study, the storage area was defined as the human brain. This enhancement is the increase in total area and time it takes to reach specific information on the brain like a storage area on computer system. This means there must be a way to the optimal representation of data on the brain to increase confidence. The phenomena about machine learning and psychology are not as far apart as it is thought. In fact, they are highly correlated with each other and can be used to describe the origin of human behavior. Human memory structure and types are explained. The results of the studies supporting the proposed relationship were analyzed within the scope of the article. As a result of this study, processing speed and the total time it takes to reach correct information are highly correlated with confidence in human behavior and this result was explained in this multidisciplinary study utilizing information theory, machine learning and neuroscience.

Conflicts of Interest

No conflict of interest was declared by the authors.

References

[1]   S. Finger, *Origins of neuroscience: a history of explorations into brain function.* Oxford University Press, USA, 2001.

[2]   J. Ward, *The student's guide to cognitive neuroscience*. Psychology Press, 2015.

[3]   P. Dayan and F. A. Laurence," Theoretical neuroscience: computational and mathematical modeling of neural systems", *Journal of Cognitive Neuroscience*, vol.15, no.1, pp.154-155.

[4]   T. C. Napier, J. C. Corvol, A. A. Grace, J.D. Roitman, J. Rowe, J., V. Voon and A. P. Strafella," Linking neuroscience with modern concepts of impulse control disorders in Parkinson's disease", *Movement Disorders*, vol.30, pp.141-149, 2015.

[5]   M. Filippi, *fMRI techniques and protocols*. Humana press, 2016.

[6]   A. Poldrack Russell, "The role of fMRI in cognitive neuroscience: where do we stand?", *Current opinion in neurobiology*, 18223-227, 2008.

[7]   U. Goswami, "Neuroscience, education and special education", *British Journal of Special Education*, vol.31, pp.175-183, 2004.

[8]   Y. Immordino, M. Helen , A. Damasio, "We feel, therefore we learn: The relevance of affective and social neuroscience to education", *Mind, brain, and education*, vol.1, pp. 3-10, 2007.

[9]   B. Aberšek, *Cognitive science in education and alternative teaching strategies*. Cambridge Scholars Publishing, 2017.

[10]   B. Sung, N. J. Wilson, J. H. Yun, and E. J. LEE, "What can neuroscience offer marketing research?", *Asia Pacific Journal of Marketing and Logistics*, pp. 4-23, 2019.

[11]   S. Ullman, "Using neuroscience to develop artificial intelligence", *Science*, vol. 363, no. 6428, pp. 692-693, 2019.

[12]   G. Deco, E. T. Rolls, *Computational neuroscience of vision*. Oxford university press, 2007.

[13]   D. Dunning," The Dunning–Kruger effect: On being ignorant of one's own ignorance", *In Advances in experimental social psychology*, vol.44, pp. 247-296, 2011.

[14]   G. Pennycook, R. M. Ross, D. J. Koehler and J. A., Fugelsang," Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence", *Psychonomic Bulletin & Review*, vol.24, no.6, pp. 1774-1784, 2017.

[15]   N. Altman, M. Krzywinski, "The curse (s) of dimensionality", *Nat Methods*, vol.15, pp. 399-400, 2018.

[16]   V. Zelenyuk, "Aggregation of inputs and outputs prior to Data Envelopment Analysis under big data", *European Journal of Operational Research*, vol.282, no.1, pp. 172-187, 2020.

[17]   M. Ramscar, P. Hendrix, C. Shaoul, P. Milin, H. Baayen, "The myth of cognitive decline: Non   linear dynamics of lifelong learning", *Topics in cognitive science*, vol.6, no.1, pp.5-42, 2014.

[18]   Y. Wang, and V. Chiew, "On the cognitive process of human problem solving", *Cognitive systems research*, vol.11, no.1, pp.81-92, 2010.

[19]   H. Kuai, X. Zhang, Y. Yang, J.Chen, B. Shi, N. Zhong, "THINKING-LOOP: The Semantic Vector Driven Closed-Loop Model for Brain Computing", *IEEE Access*, vol.8, pp.4273-4288, 2020.

[20]   Y. Wang et al., "Brain-Inspired Systems: A Transdisciplinary Exploration on Cognitive Cybernetics, Humanity, and Systems Science Toward Autonomous Artificial Intelligence", *IEEE Systems, Man, and Cybernetics Magazine*, vol. 6, no. 1, pp. 6-13, Jan. 2020

[21]   R. Shivhare, A. K.Cherukuri, and J. Li, ,"Establishment of cognitive relations based on cognitive informatics", *Cognitive Computation*, vol.9, no.5, pp.721-729, 2017.

[22]   H. Faris, I. Aljarah and S. Mirjalili, *Evolving radial basis function networks using moth–flame optimizer*. In Handbook of neural computation, Academic Press, pp. 537-550, 2017.

[23]   C. Zhang, H. Wei, L. Xie, Y. Shen, and K. Zhang, "Direct interval forecasting of wind speed using radial basis function neural networks in a multi-objective optimization framework", *Neurocomputing*, vol. 205, pp.53-63, 2016.

[24]   I.T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol.374, no.2065, 20150202, 2016.

[25]   G. R. Naik, *Advances in Principal Component Analysis: Research and Development*. Springer, 2017.

[26]   I. Craw, The aberdeen set of Psychological Image Collection at Stirling (PICS) database (PICS), URL: http://pics.psych.stir.ac.uk/zips/Aberdeen.zip [Accessed: 19-July-2020].

[27]   N. Kriegeskorte,"Deep neural networks: a new framework for modeling biological vision and brain information processing", *Annual review of vision science*, vol.1, pp.417-446, 2015.

[28]   J. C. Hong, , M. Y. Hwang, , K. H. Tai and C. R. Tsai, "An exploration of students' science learning interest related to their cognitive anxiety, cognitive load, self-confidence and learning progress using inquiry-based learning with an iPad", *Research in Science Education*, vol.47, no.6, pp.1193-1212, 2017.