critical thinking of Machine Learning applicated in data visualization

FIT5141 Assignment 1

Siyang Feng

28246993

**Abstract**

This article talks about what the machine learning uses in data visualization domain. First of all, history of AI is introduced and then we introduce what is data visualization and machine learning. After that, some application of machine learning is talked about. It includes technology in biology and medicine and other aspects. Finally, the advantages and disadvantages of this technology Is considered because it will influence economy and human wellbeing.

Catalogue

[Introduction 4](#_Toc491247070)

[History of Artificial Intelligence 4](#_Toc491247071)

[Introduction to Data Visualization 4](#_Toc491247072)

[Introduction to Machine Learning 5](#_Toc491247073)

[Applications in Data Visualization Area 6](#_Toc491247074)

[Research of Gene using Machine Learning 6](#_Toc491247075)

[Deep Learning in Disease Analyzation 7](#_Toc491247076)

[Medical Image Improvement 8](#_Toc491247077)

[Other Applications 10](#_Toc491247078)

[Critical Thinking in Machine Learning 12](#_Toc491247079)

[Conclusion 13](#_Toc491247080)

[Reference 14](#_Toc491247081)

# Introduction

## **History of** **Artificial Intelligence**

Artificial intelligence (AI) seems like a new research theory which is wildly used in deferent regions nowadays. The aim of this theory is through doing some simulation or other technologies to release human’s brain to more creative work, such as how to design and construct a more intelligence system [5]. Actually, the concept of AI has already put forward in 1950’s. in AI field, there is a famous question: “When can we say that a system constructed by human designer is intelligence” [4]. Alan M. Turing gives a solution with the help of imitation game, Turing Test, which is the first time to figure out the concept of AI. Turing mentioned that if a machine can communicate with human and cannot be figured out, this machine is considered with intelligence [10]. Unfortunately, because of the limitation of hardware and the large computation of AI algorithm, AI developed so slow that a lot of scientists gave up this field.

About half century later, from 1993, scientist regarded IA more cautiously and with the development of hardware and running time reducing, more complex algorithms were considered to be used in AI. AI becomes a hotspot again in research. Deep Blue, a chess-playing super computer developed by IBM, won its first game on 10th February 1996. In May 1997, it defeated Garry Kasparov and became the first computer system which defeated a champion who reigns the world under standard chess tournament controls [14]. History is always similar. With the development of IA, AlphaGo, a board game computer program, was developed by Google DeepMind in London in October 2015. It defeated Lee Sedol in a five-game match in March 2016 and in 2017, it beat Ke Jie who is the No. 1 ranked player at that time in the three-game match. AlphaGo combined Monte Carlo tree search algorithm and machine learning, especially artificial neural network which is one of deep learning method.

Machine learning, a part of important methods in IA research, is wildly used in industry development. This article will talk about the technology of machine learning used in data visualization region.

## **Introduction to Data Visualization**

Data visualization is the region of study to use the visual representation of data and information. This method makes people more easily to communicate with data. Actually, analyzing with numerical data seems like a huge challenge for human in multidimensional data in the period of big data. However, data visualization could show more features directly.

Normally, most data visualization nowadays continues to use graphical primitively. It is easy to understand to immigrate numerical data into a two-dimensional graph and data statistics always use this method. However, with the development of data analyzing algorithm, data visualization is wildly used in almost every region, such as biology and medicine, weather and climate, aerospace and so on. There are more and more data represented by images or videos. Thus, there is another trend that the data be visualized is also visual data. Usually, a project with this kind of data will create a new visual representation improving the two-dimensional graph into visual spatial variable [8]. This new trend will save a lot of time for people avoid imaging plane into a spatial variable. It sometimes needs a great deal of experience or a strong space imagination. It may limit creation in some words.

## **Introduction to Machine Learning**

Machine learning as a branch of AI, is still rapidly developing. In another word, machine learning is a method to achieve artificial intelligence through helping machine to learn some experience.

In recent years, there has been an explosion development of machine learning especially in deep learning models and it leads to groundbreaking results in a lot of regions such as natural language processing [13], biology, and computer vision [6]. The main theory of machine learning is to design and analyze algorithms to lead machine learning automatically and react with the different scenario to achieve intelligence.

Figure 1 Classification of Machine Learning

Figure 1 indicates the classification of machine learning. There are four main parts. Supervised learning is now widespread used in lots of regions. People mark the data with features and use these marked data to train model. Then, the trained model is used to predict. Statistic is very important in this part. Comparing with supervised learning, unsupervised learning uses unmarked data to train prediction model. It is more difficult than supervised learning. Clustering algorithm is famous in this part. Semi-supervised learning is a method which is between the two methods mentioned before. The last one, reinforcement learning, react through environment observation.

Figure 2 Relationship in AI

Figure 2 indicates the relationship in AI and machine learning. AI is the knowledge based technology. Under the AI, there is machine learning. Machine learning includes logistic regression and linear regression and other methodologies. Next level is representation learning. Autoencoder is a very important part of representation learning. The final level and the most famous level is deep learning. It creates unprecedented success in commercial application. Convolution neural network (CNN) deep neural network (DNN) and other neural networks are belonging to deep learning. Because there is deep node layer and more complex algorithm.

Machine learning is widespread in almost all regions in this society. This article only talks about machine learning in data visualization region nowadays. People are more confidence by using machine learning to solve real problems and it does work well.

# Applications in Data Visualization Area

## **Research of Gene using Machine Learning**

There is a hotspot to use machine learning in biology especially in visualization field. In biomedical region, it is extremely important for a doctor and investigator to understand the predictions of a model. Gene as one of the important parts of biology research, investigator always uses the testing data to image its structure to get the visual representaton in past years. This method can often be influenced by one’s emotion or previous theory.

In gene prediction region, there are different neural network models and deferent visualization strategies. Often, neural network models are used to improve visualization strategy. Combining different neural models with different visualization strategy will get the difference performance which will help people to analyze gene and make predictions [2]. One article talks about three deep learning architectures to classify the sequences which are recurrent neural network (RNN), convolution neural network (CNN) and convolutional-recurrent neural network (CNN-RNN). And it also mentioned three popular visualization strategies: Temporal Output Scores, Saliency Maps and Class Optimization [7]. they combined the different machine learning algorithm with different visualization method to compare the efficiency of getting features.

Same as DNA, machine learning is also used to analyze RNA expression. It is used to recognize complex patterns and make an intelligent decision. Moreover, it is also used for improving the outcome of target prediction and RNA gene detection [16] and then put the result as a visualized result which will be convenience for future research.

Gene prediction is still an important part of computational biology study but actually, it is also a big challenge for world nowadays that no one can figure out what the real look gene is. The parts we predict are only the features and machine learning is exactly used in feature prediction. It is used to identify the regions of genomic DNA through encoding genes. This technology will help people know more about gene and find gene defect. In the future, we believe that machine learning will be an important method to predict real gene structure and people may change gene of human to improve their feature and repair the gene defect. It is known that there are a lot of viruses are RNA virus. These kinds of viruses are hard to control because single chain RNA is easy to change. Prediction of RNA will make it more efficient to control RNA virus. However, if there is someone to solve problems of DNA and RNA. There must be some people to produce more problems in this region. Machine learning improves the prediction of gene and it means people must pay their attention in the protection of gene at the same time.

## **Deep Learning in Disease Analyzation**

Gene is only one part of biology and medicine region using machine learning, such as support vector machine (SVM) [3] and data visualization. Disease analyzation in cell level also needs machine learning in data visualization. A website uploaded thousands of 3D stem cell images which shows no two cells are identity even if one is cloned by another. This kind of 3D cell images are all predicted through deep learning algorithm and the cell lines changed by gene-edition tool CRISPR [9]. These images are hopeful to help people better understand some diseases and cancer prediction makes a great improvement in this part.

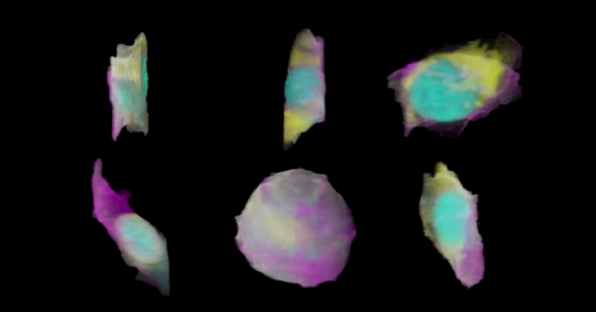


Figure 3 Three-dimensional Views of Human Stem Cells

Figure 3 shows the predicted stem cells in the website. Disease analyzing through machine learning is popular in the past five years. People already put machine learning method into cancer research and make a good progress. In 2016, BIDMC and medical college in Harvard develop an AI system which is used to detect breast cancer and the accuracy rate is up to 92% [3]. Although there are a lot of achievement in disease detection, the cost is so very high that normal people cannot regard it as a part of normal physical examination. It is still a long trip to popularize this technology.

## **Medical Image Improvement**

Image processing as a part of special data visualization method. It is totally differenced with traditional. As mentioned in previous sentence, it uses image processing technology to identify through marking features or changing two-dimensional pictures or videos into spatial data. The original data is visualized and the processed data is also visualized which is easier to figure out useful information. With increasing of data which is recorded by images and videos, this kind of technology is wildly used in our life. This section, medical images as an example to talk about.

AR technology, one of the data visualization method, is considered to use in trauma and orthopedic surgery to support surgeons in challenging tasks. Investigators and doctors could use AR to deeply understand the relationships between implants, anatomy and their tools [11]. However, it becomes more and more difficult for traditional AR technology to recognize and differentiate structure in the overlaid images quickly. With the population increment of machine learning. This year, people consider adding machine learning technology into AR to identify anatomy and relevant object in X-ray and Kinect and create a special alpha map which combine Kinect and X-ray images [12].

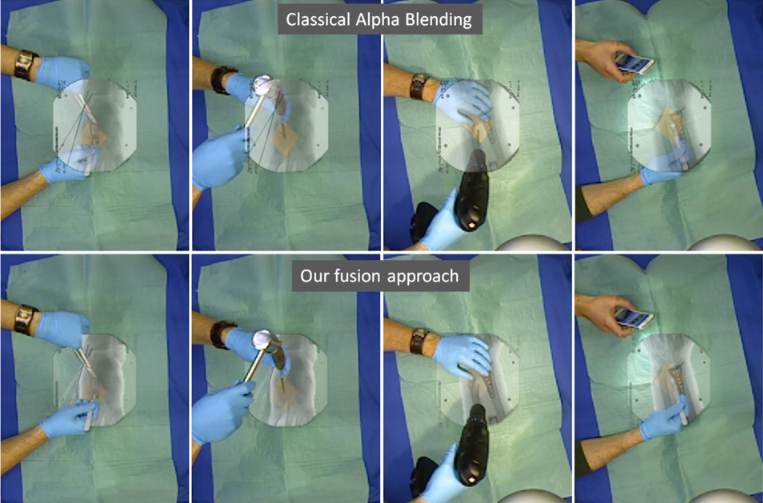


Figure 4 Comparing between Classical Alpha Map and Learning-based Map

Figure 4 indicates the difference between classical alpha map surgical and machine learning based map. From the figure, Learning-based AR improves the understanding and depth perception of the surgical scene.

Actually, surgical scene in only one part of medical image research. Clinical decision which almost depends on the medical image, such as multi-slice CT, tomosynthesis, electrical impedance tomography, will be easy and persuasive by improving the algorithms with machine learning. For this kind of issue, the international workshop on Machine learning in Medical imaging 2013 has established.

This kind of technology gives a better perspective for doctor to do clinic decision and surgical scene. It does not change the traditional technology but only improving which is well acceptable for most of doctors and patients. It just visible some features which is hard to discover for human eyes through machine learning. Moreover, combining machine learning and medical visualization could reduce the request of doctor’s experience. Machine learning technology used here is a kind of experience collected technology which can collect thousands of doctors’ experience. It just likes a professor to find out some features. However, the most disadvantage is also the advantage of this technology. Experience-based technology will be sensitive to normal features but sometimes it will neglect some rare features. Thus, if the features change, it will be hard to recognize and it is lack of creativity. In this part, an experienced doctor is still very important.

## **Other Applications**

There will talk about computer vision and it also called machine review. This technology uses sensor to get original image and translate it into data image which is easier for machine to understand. This technology is a famous part of data visualization which is widespread used with machine learning. It is a complex science which combines a lot of subjects such as imaging processing, neurobiology and so on. Machine review uses visual sensor to simulate people’s eyes and use computer to simulate brain. Then algorithm and program are used to simulate recognize of things. This technology will replace people to finish some works with visualization. The previous technology, medical images improvement is actually a portion of computer vision. The difference is that the previous technology only uses machine learning in improving the visualization details. But traditional computer vision pays their more focus on images recognition.

Computer vision as an important domain in both data visualization and machine learning. It is widespread used in robot vision, industrial manufacture, aerospace, map plotting and other domains.

In aerospace, people use probabilistic autoencoders to model galaxy shape. Up till now, supervised learning has made great success in a lot of domain. However, in galaxy shape modeling, it is impossible because there is no correct output. Thus, autoencoder which belongs to unsupervised learning method is used in this part. Julia neural network model used 43444 galaxy images to train variational autoencoder (VAE) [1].



Figure 5 Spiral Galaxy image from NASA

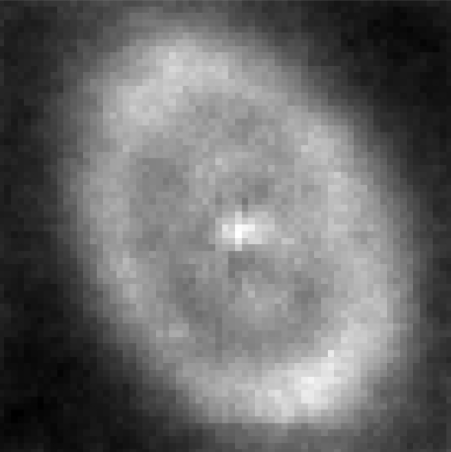


Figure 6 69x69 Pixel Image of Specific Galaxy

Figure 5 is the original image of spiral galaxy which is pictured by NASA. And figure 6 is 69x69 pixel specific galaxy image which is generated by Jon McAullife and Jeffrey Regier. The second figure is generated by the images like previous one. Then, it will be trained by VAE to create galaxy shape model. This technology helps people know wildly and deeply of galaxy and aerospace.

Similar technology can also be used in finding extreme weather with climate simulations. In previous technology, extreme weather is found through record of satellite and traditional analyzation and simulation algorithm. This kind of algorithm usually use multivariate threshold and hand-coded model which is subjective [1]. However, training a deep learning system to study a pattern classifier will be more objective.

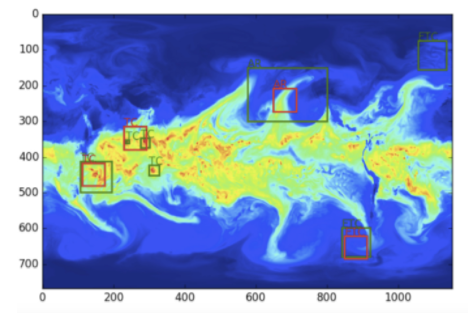


Figure 7 Result of Weather Patterns Analyzation

Figure 7 is the visual result of weather patterns analyzation which is provided by LBNL and Evan Racah. It uses semi-supervised architecture to analyze. This combines the marked features and unmarked features to achieve a predicted result. It was used to predicted extreme weather that maybe thousands of people will escape from the death of extreme weather, such as snowstorm, seismic sea wave ground sea and earthquake.

# Critical Thinking in Machine Learning

Machine learning is still a hotspot topic for almost every subject in these years and it is sure to keep hot for a long time in future.

It extends the usage area of data visualization and deepens the research. In biotechnology and genomics subject, people used different deep learning network to analyze and predict the structure of gene. It might be a new method to discover anthropogenesis and find the relationship between human and other animals. The more important one is that through gene features analyzation and prediction, people could find out more gene defects. With the in-depth research of gene, some gene defends will be renovated by artificial gene to improve into a stronger species. What’s more, human can even create a new species which might be more intelligence than intelligent machine. However, it is extremely difficult with current technology. Moreover, when the research of gene prediction is mature, the opposite technology, like artificial gene virus, may be also mature. In the future, there may be more challenge in this region. It still needs much money and a long time on this domain researching for industry effects.

Biological cytology is another region using machine learning. It always is used to observe pathological cells such as cancer cells. Cancer observation through data visualization using machine learning analysis is very hot point in this level researching and it gets a great success in recent years. These technologies would be helpful to detect cancer cell earlier. Wild using this technology could decrease the rate the death suffering cancer. With the development of this technology, monitoring normal cells to avoid cytopathy would be possible in the future.

Improving image quality in medicine is also a popular subject using machine learning. Pixels in AR and other photos are calculated and analyzed through machine learning. It performances outstanding on feature analysis and pattern recognition which works well to figure out disease and display more clearly about surgery scene. Machine learning is so famous that almost everyone has heard about this. They may not know the details but they must know its achievement in speech recognition and face recognition. It gives great confidence for everyone on recognition. Thus, doctor will be confident on surgery scene and disease images which are processed by machine learning. Machine learning is so objective and experienced that people trust it. Wildly using this technology is sure to increase the human wellbeing in the future. However, increasing experience of machine is not same as increasing doctors’ experience which means creation of people might be limited because people take more relies on the machine decision. This technology can decrease the stress of doctor but at the same time, it sometimes not good for doctors to improve their skills.

For other applications of machine learning in data visualization, most of the technologies are still in research stage. Some can establish mature model and wildly used in a lot of industries regions. And it makes a great impact toward traditional industry. It releases labors from boring and experience-based work into more challenge works and will even change the life style of people. Regarding on machine learning, every change will be possible no matter it is how unimaginable.

# Conclusion

In conclusion, machine learning gives impacts for almost all industries. It can replace people to do some works and release labor for more challenge work. With the improving of machine learning technology and widely used in industries, it will increase the efficiency of industrial manufacture. Meanwhile, it will stimulate economic growth. In this aspect, machine learning will increase human wellbeing. However, in another aspect, widely used of machine learning will replace much labor. There is a question that what those people should do without their works. Maybe in the future, this problem will be solved, but now, it is still a problem which people should worry about.

# Reference

[1] A look at deep learning for science. (2017).

[2] Alipanahi, B., Delong, A., Weirauch, M., & Frey, B. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nature Biotechnology, 33(8), 831-838. http://dx.doi.org/10.1038/nbt.3300

[3] Al-Kadi, O. (2015). A multiresolution clinical decision support system based on fractal model design for classification of histological brain tumours. Computerized Medical Imaging And Graphics, 41, 67-79. http://dx.doi.org/10.1016/j.compmedimag.2014.05.013

[4] Flasiński, M. (2016). Introduction to Artificial Intelligence (pp. pp. 3-13). Switzerland: Springer International Publishing.

[5] Huang, L. (2010). Artificial Intelligence. In 2010 The 2nd International Conference. Singapore: Computer and Automation Engineering (ICCAE). Retrieved from http://ieeexplore.ieee.org.ezproxy.lib.monash.edu.au/document/5451578/

[6] Krizhevsky, A., Sutskever, I., & Hinton, G. (2017). ImageNet classification with deep convolutional neural networks. Communications Of The ACM, 60(6), 84-90. http://dx.doi.org/10.1145/3065386

[7] Lanchantin, J., Singh, R., Wang, B., & Qi, Y. (2016). Deep Motif Dashboard: Visualizing and Understanding Genomic Sequences Using Deep Neural Networks. Pacific Symposium on Biocomputing 2017. Retrieved from https://arxiv.org/abs/1608.03644

[8] Manovich, L. (2011). What is visualisation?. Visual Studies, 26(1), 36-49. http://dx.doi.org/10.1080/1472586x.2011.548488

[9] Maxmen, A. (2017). Machine learning predicts the look of stem cells. Nature. http://dx.doi.org/10.1038/nature.2017.21769

[10] McCorduck, P. (2004). Machines Who Think. Natick: CRC Press.

[11] Navab, N., Heining, S., & Traub, J. (2010). Camera Augmented Mobile C-Arm (CAMC): Calibration, Accuracy Study, and Clinical Applications. IEEE Transactions On Medical Imaging, 29(7), 1412-1423. http://dx.doi.org/10.1109/tmi.2009.2021947

[12] Pauly, O., Diotte, B., Fallavollita, P., Weidert, S., Euler, E., & Navab, N. (2015). Machine learning-based augmented reality for improved surgical scene understanding. Computerized Medical Imaging And Graphics, 41, 55-60. http://dx.doi.org/10.1016/j.compmedimag.2014.06.007

[13] Roberts, A. (2015). The Use of Natural Language Processing to Transform Health Records Information. European Psychiatry, 30, 148. http://dx.doi.org/10.1016/s0924-9338(15)30124-3

[14] Saletan, W. (2007). The triumphant teamwork of humans and computers. Slate Magazine. Retrieved 15 August 2017, from http://www.slate.com/articles/health\_and\_science/human\_nature/2007/05/chess\_bump.html

[15] Shen, D., Wu, G., Zhang, D., Suzuki, K., Wang, F., & Yan, P. (2015). Computerized Medical Imaging and Graphics. ELSEVIER. http://dx.doi.org/0895-6111

[16] Yousef, M., & Allmer, J. (2014). miRNomics: MicroRNA Biology and Computational Analysis (pp. 257-278). Totowa, NJ: Humana Press.