Assignment 4

Huanye Liu

December 3, 2019

Question 1

First of all, I want to brief on the definition of the "black box" model for prediction. The black box system was initially introduced into the software industry as software system with only inputs and outputs but internal implementations unknown to users. This approach conforms to the encapsulation principle in software engineering and becomes one paradigm for the software function testing[1]. In contrast, the RNN model we learned in class could be a metaphor of the real black box system, because we can design the internal structures by setting the configurations in the model configuration dictionary model_cfg. In addition to the internal structure of the network, the inner workings of each cell are also clear: each cell computes the sum product of the its inputs and its corresponding parameters, which are specified in the algorithm. Therefore, both the structure of the model and the algorithm are transparent to users. But we still think of such model as the "black box" mainly because of the complexity of its structure and the "meaningless" parameters linking different layers of cells.

On the one hand, the complexity of the internal structure introduces a large number of non-linear relationships between every layer's inputs and outputs, which makes the relationship between the final outputs and the initial inputs extremely hard to explain or interpret. In contrast, the relatively simple structure of theory-based model is easily recognized by human, amenable to interpretation and thus easier structure adaptation to future needs and derivation of new models. Similarly for the model parameters of the "black box", it is more difficult to find their direct correspondences in our physical world or social life compared with those of theory-based models, and therefore harder to quickly adjust them in the right direction to meet practical needs, for example, the predictions of culture phenomena in various contexts. Using the theory-based models, we can apply more specific models guided by the relevant theory to predictions in some particular contexts, and if the prediction results are not satisfactory or we want to try data from a different domain, we can further adapt the model and adjust related parameters with the aid of the relevant theory. This kind of model structure adaptations or parameter adjustments may not be easy for the ”black box” due to the lack of theory as directions of changes, and maybe can only be done by trial and error. For example, the recursive neural network model we learned in class allows us to change the model structure by setting the parameters in the dictionary model_cfg. But generally we have no better idea on how we should adjust them to improve the final results given a particular set of training and test data, for example, the rnn_bidirectional and the word_level, than just trying different options and compared the results.

On the other hand, those advantages of using the theory-based model for prediction of culture phenomena relies heavily on the quality of the relevant theory and practicality of implementing the theory in computational models. In many cases, no theory even exists to capture the features or patterns of certain culture phenomena, and here the ”black box” models can be better options. For example, some culture phenomena based on natural languages, images and most types of multimedia contents, which potentially involve complex patterns and structures, have been generally acknowledged to be ideal subjects for analysis and prediction using the ”black boxes” such as deep learning models[2][3][4]. First, we have not discovered or invented by far the perfect theories to explain the patterns and constructions of these complex culture phenomena or legisigns. Second, even if one day we have the relevant theories at hand, we could image that they are more likely to be too complex to implement by specifying those theories or rules one by one, so the theory-based models may still not apply.

Question 2

Three defining features of a genuine semiotic machines were explicitly proposed and elaborated in Noth's paper: self-control, creativity and the ability of transforming signs into actions[5]. For the first feature, Noth gave a working definition of self-control in that "a machine lacks self-control if it is completely controlled by its input", which is related to the capacity of "self-reference, autonomy in relation to its environment, self-maintenance and self-reproduction". For the second feature, Noth put the creativity in the context of "final causation" in which "a sign is not determined by a mechanical force, but by a semiotic norm or habit that allows for a certain creativity in sign production and interpretation", and he also defined the "efficient causation" as the characteristic of deterministic machines in contrast of the "final causation" which entails creativity. For the third feature, Noth said that "according to the principle of the unity of sign and action, the pragmatic dimension of sign processing is a further criterion of fully developed semiosis", and "learning from environmental experience and automatic self-correction are hence further essential...".

In my opinion, among the three features, the ability of transforming signs into actions could be the most feasible one that machines can obtain in the present day, because the two tasks it mainly involves as specified by Noth, 1)learning from the environmental experience and 2)automatic self-correction, are both within our grasp using the current AI technology. Roughly speaking, two paradigms are widely adopted in today's AI and machine learning circles: learning from data and learning by rule. The former approach could be applied to the first task of "learning from environmental experience" and the latter approach the second task of "automatic self-correction" by using, for example, the feedback paradigm in control theory and reinforcement learning. In fact, these two tasks have already been investigated and put into practices in many applications now, and maybe the most prominent example is the Mars rover[6].

However, I think the two other features could be two obstacles that are way harder for machines to overcome in the present day, and relatively speaking, the creativity feature may be slightly closer to our reach. The creativity is mainly related to the "final causation" as Noth proposed, which involves setting a general end or target and leaving room for the machine to decide how to proceed and interpret. The intuitive way to realize the machine creativity is to design a non-deterministic machine with some rule-based constraints specifying the "semiotic norm or habit", but the fact is that no real non-deterministic machine exists given the level of science and technology in the present day, even if the machine uses the so called non-deterministic algorithms which are essentially based on the pseudo random generator. Yes, they are only pseudo random. So in this sense, the absolute creativity is something we could not expect from machines directly, we could, however, create the illusion of a "creative" machine by including the changing external environments as inputs, adapting the configurations of machine parameters to the changing inputs, and repeating the process time and time again, each time producing a different results, which seems to be "creative" on the surface. For example, the generative neural network we learned in class made me feel it is so "creative" that it can "create" song lyrics in Taylor Swift's style. The key is the changing inputs every time we use to train the model, but both the model's internal structure and the learning algorithm are deterministic. In fact, this is the basic approach taken by the famous machine AlphaGo which beat Lee Sedol in 2016[7]. Do we not say that "AlphaGo" is creative? At least it is "creative" enough to beat the world champion!

Finally the feature of self-control may be the most difficult one to realize given the current level of human understanding about what the machine essentially is. Because based on Noth's definition, this feature is closely related to living organisms which are way more advanced than machines, and there still exists a big gap between machines and living organism in several aspects such as self-reference, self-maintenance, and even self-reproduction. Just consider Noth's working definition of self-control, how can we design a machine not completely controlled by its input given the current level of science and technology? Such kind of machine should at least need some kind of internal "metabolism" and "minds", which are not fully determined by external inputs, to generate energy and "thoughts" that would lead its behavior just like all the living organisms do. But if the machine had this ability, then it would no longer be a machine anymore: it would have crossed the big gap and become one of the living organisms.

Therefore, since machines in the present day have already obtained the ability of transforming signs into actions in many cases, and partially acquired the "creativity", at least on the surface, to complete certain tasks, we are more likely to encounter culture phenomena in all kinds of digital forms, and a considerable portion of them are pumped out every day by machines instead of "first hand" human creations. As a result, some may think of digital culture as cheap, low quality with poor taste compared with their manually crafted counterparts by human. They feel uneasy and even being harassed by those digital "bombs". In their eyes, digital culture is nothing but a synonym of fast food culture, which could only impair people's ability of deep thinking and lower their levels of appreciation of the true art because machines can replace human to think about and perceive the outside world. But others may consider digital culture contributes to a more functional, efficient and transparent society, so they enjoy, embrace and advocate the digital culture, and try to expand the digital circle as far as possible and apply it to every corner of human activities. Now commercial banks even resort to AI for creating marketing language to get across to their potential customers[8]. So do the two opposite attitudes toward digital culture tend to be more bipolar? Or could the tension be released? I think the answers depend heavily on the how we could shape the "ethos" of AI and, maybe more importantly, to what extent can we lead and control the direction of technology development in the future.

References

[1]Beizer, Boris; Black-Box Testing: Techniques for Functional Testing of Software and Systems, 1995

[2]Schmidhuber, J. (2015). "Deep Learning in Neural Networks: An Overview". Neural Networks. 61: 85–117

[3]Cireşan, Dan; Meier, Ueli; Masci, Jonathan; Schmidhuber, Jürgen (August 2012).

"Multi-column deep neural network for traffic sign classification". Neural Networks. Selected Papers from IJCNN 2011. 32: 333–338

[4]Jozefowicz, Rafal; Vinyals, Oriol; Schuster, Mike; Shazeer, Noam; Wu, Yonghui (2016). "Exploring the Limits of Language Modeling"

[5]Nöth, Winfried (2002). "Semiotic Machines." In: Cybernetics and Human Knowing, Vol. 9, Nr. 1. Essex, UK:Imprint Academic

[6]https://www.geeksforgeeks.org/how-does-nasa-use-machine-learning

[7]https://en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol

[8]Pasquarelli (2019) "Chase Commits to AI After Machines Outperform Humans in Copywriting Trials"