Technical University Munich

Data/Science/Society
Winter Term 2020/2021

The influence of transparent social behavior modelling

based on the example of the social credit system in China

Moritz Schüler
03721689
March, 2021

1 Social behavior modelling

Social behavior modelling or social modelling is a quite old field of study which evolved from the curiosity about what effects our actions. Traditional behavioral theories claim that our actions are learned by reinforcement (reward or punishment) for or by reproduction (doing the same things again and again) of an action [1]. Social modelling expands these theories in the way that it not only focuses on reinforcements but also considers the internal processes of an individual [2]. Explicitly, it shows that learning can happen solely by observation of a behavior and its consequences, also called vicarious reinforcement [2, 3].

The results of the research mainly changed the view on how humans learn things, despite being more general as it also explained why we exhibit certain behaviors. It showed that our actions are influenced by the actions of our role models (parents, friends, idols), as we strive to behave the same way and imitate them. Further study revealed that similarities of personal characteristics with the role model, i.e. age or gender (attribute similarity) play a significant role in influencing the actions of a person, meaning higher similarity leads to a higher probability of imitating the behavior [4, 5].

As these theories help understand the human behavior and also improve teaching by using better learning approaches, there is a wide range of applications these days. Based on the idea and with the help of big data and artificial intelligence the application range is broadened even further by not only applying the theories but additionally using previous behavior to predict future actions. Often the theories are used to understand taken actions (psychological context or criminology) [6, 7], as well as influence or rate behavior [8–10]. This suggests that a responsible handling of such data is very important and should be monitored, however currently there are no measurements in place.

Two applications in media are [8, 9] and [10]. The first one analyzes educating people with the help of television shows about daily social life, i.e. soap operas. It shows that there are usually three types of role models involved (a positive, a negative and an undecided) with which the viewer should emphasize. By showing the actions and reactions of these types in a social context, the viewer should learn to behave in the way of the positive role model, when in a similar situation in his own life [8, 9]. The second and more recent example is about social media, i.e. Facebook. While most people think the usage of Facebook is for free, actually the users are paying with their behavioral data for the services. This information is subsequently used

to show targeted ads or content which reinforces specific behavioral parts of ourselves. This leads us to advertise several products or lifestyles ourselves or to buy certain products. Overall one can say, that the advertising companies pay Facebook and other social media providers to change our behavior, buy their products and indirectly finance these social media companies thereby [10].

There are many more positive [11,12] as well as negative [13,14] examples for the use of behavioral data. Here it is important to note that most of them have one drawback in common, which is transparency.

It is generally not a bad thing to collect this data and make use of it, but it should be transparent to the user, such that he or she can decide to give consent or not. Nevertheless these systems are somewhat accepted by the society and do bring some benefits for both parties (user and analyst). In the following one very controversial use case, the social credit system (SCS) of China, will be analyzed in detail. The main idea is to expand a scoring to all aspects of life, but in contradiction to most systems, the SCS claims to be transparent as well as starting voluntary [15].

2 The Chinese social credit score

The Chinese social credit system was first introduced in 2014 by the State Council with "the aim to promote 'integrity in government affairs', 'commercial sincerity', 'social integrity' and 'judicial public trust'" [12]. But what does this exactly mean for the Chinese society and how will the government implement the system?

The idea behind this system of the Chinese State is to improve itself and its citizens in the way they are interacting as a society and tries to enforce this by building a guiding system. It should help them to decide with whom to make business but also how to behave and which friends to have [12,15,16]. Clearly from a western perspective this sounds undesirable as it resembles a surveillance state, violating the personal privacy. However one must respect the cultural difference and therefore be objective.

For the participants of the system - volunteers until 2020, nowadays every Chinese citizen [15] - a good score comes with benefits, like renting an apartment or a car without needing a deposit, whereas a bad score enforces some restrictions, like being banned for booking flights [14,15]. The Chinese premier Li Keqiang even said back in 2018: "Those who lose credibility will find it hard to make a tiny step in society," [17]. Regarding the functioning

of the system, the Chinese government instructed several large companies, which already operate in the market of big data, like Sesame Credit or Alibaba, to use their data and construct a trustworthiness score out of it [12,17]. Most of the data sources are traditional (financial, criminal and governmental records), however China is also trying to use video surveillance for real time assessment [17], enabling for example to punish people for obeying street lights.

For me, two questions arise: Is educating the people the real reason for building the system and how transparent can a system really be based on black box style machine learning?

2.1 Reasons

Looking closer at the start of the development of the system and the problems China was facing at that time, one can find interesting insights that are more convincing than making up for a better world. China, in contrast to many western countries, was loosing billions of dollars in the financial sector due to giving loans to people that couldn't pay them back. The difference to the West is that loans aren't given out depending on a credit score, which tells how likely it is that the person will pay back the money [15]. And there are several more implicit systems in every society, working well in their purpose, therefore concluding that China is attempting to combine those to one huge central system in hope for getting all benefits [15]. But is China also mitigating the risk and problems of these systems like opacity?

2.2 Transparency

Firstly, it has to be explained why transparency is that important. In Germany for example, there exists the so called Schufa score, which tells how credit worthy you are. Using this for giving out loans sounds intuitive, but it is also already widely in use for proving the credibility when intending to rent or buy an apartment. As can be seen, the score has a huge, not always positive influence on many citizens lifes without showing frankly how the score is obtained or even what one's score is unless you pay for it [14].

The SCS has an even bigger impact on your perspectives and possibilities, which is why knowing your score and the actions that lead to a negative scoring is even more crucial in order to keep away from the blacklist. To analyze the transparency of the SCS, two notions of transparency will be introduced. The first one is about knowing how the score itself is calculated, for example having access to the source code and data. The other aspect of

transparency is knowing the score and what affects it. For simple systems this differentiation might not be needed, however, when using machine learning models there is a big difference in types of transparency. Additionally, for people who don't have access or aren't educated to understand code the first notion would result in an opaque system.

Compared to existing technologies which fulfill neither the first nor the second notion of transparency, the SCS achieves transparency in the second sense. Official documents and speeches mention what is considered as a good and bad actions and what are the up- and downsides of the score. But how useful is this? Does it solve the problem of rating systems?

To answer this have a look at another instance of a system, which is transparent in the way that everyone can look up the guidelines or the good and bad actions: the law. Nearly everyone has access to read and study it. Nevertheless it is such a complex system, that only few people understand it and there exists a job focused especially on understanding it. It gets even more complex because lawyers themselves specialize in one area of the law, therefore this does not seem to be a very useful and user-friendly kind of transparency. On the other hand, if a system would be too simple it could be easily abused. Nevertheless the law is working because there is a very common understanding of what is allowed. But is this the same for the SCS?

In [15] it is mentioned that Chinese people were questioned on the street whether they know the system and the terrifying result was: they don't! They are not aware of the system, that they are monitored, that there exists a score rating their trustworthiness and especially the linked consequences. Otherwise the SCS could be seen as a system which only formalizes norms, which are common knowledge. But how to prove this is the case? This requires at least transparency in the first sense as well as a white box machine learning model. In the current application a machine learning model would try to learn the norms from data, which means that it is very biased towards the training data. Also, how would someone encode norms in data? Additionally doesn't this system violate the definition of norms, because norms naturally arise out of a society and are a set of unspoken rules that nearly everyone agrees to [18].

This reveals that some kind of mechanism to regulate the system is needed, because even if a white box model is used, how could the society tell which norms they want to obey and would not be overruled by made up rules of the government to keep everyone in line with their thinking [15]. A similarity could be drawn to the unnoticed way of how social media uses

advertisement to change the opinion of people to their liking as presented in [10].

The solution could be a slow implementation process which starts with less sensitive domains and expands step by step - just like the SCS does - but adding two additional important properties [19]. One being to instruct the people better, so that everybody knows about the system's existence. The other one being able to influence the rating collectively. A very interesting approach to do the latter is described by Kevin Kelly in [20], called "watching the watchers", "where the watchers and the watched [...] transparently track each other" [15]. Whilst transparency being a crucial part, the next chapter will more generally elaborate on the chances and dangers the SCS creates.

2.3 Chances and Dangers

One very often discussed problem every society faces is the wage gap. The poor are kept down and the gap between the rich and the poor gets bigger and bigger. But does the SCS allow to fight this problem?

Nowadays, education is key to getting a good job. While good grades or money often determine the school, the SCS changes this to only allow children from parents with good scores [15]. It seems like a step in the right direction, but as the score also depends on the financial situation and the social environment [15] this might even strengthen the wage gap because of a kind of inherent negative bias towards poor people [12].

Another danger stemming from the opacity might be corruption. It is already a major problem of China, which it is trying to combat, however adding another semi-transparent system creates a new space for corruption [21]. As long as some data is fed to the system by people, these people might be paid to look over exhibited bad behavior.

In contrast, a good score opens up several opportunities like getting a loan with better conditions to become self-employed. Another chance is that such a system introduces objectivity. Regarding the law, each violation is discussed on a per case basis, leading to different sentences even though the crime was very similar. With an objective system, that only uses the data to assess the severity of the violation, this is clearly an improvement.

One major drawback of this and several similar systems is, however, that they don't fight the root cause of problems but instead punish people for having a bad score or behaving in a bad manner. This leads to two problems, one being the reinforcement of a bad score. If you don't have access to the benefits anymore like renting a car without a deposit, you maybe

can't attend a job interview only because you are immobile. How far such reinforcement policies can go can be seen in a popular Black Mirror episode about a rating system [22] where the main actor ends up being trapped with a bad score. How closely related this might be to the real SCS can be shown by this quote from the deputy director of the NDRC, Lian Weiliang: "When we discuss restoring credit, it's mainly aimed at minor or regular offences. Those who have committed serious offences or violations will not be taken off the blacklist [...] their untrustworthy record will be kept for a long time, according to law" [17]. This implies that once someone has a bad rating, it impedes to integrate back to a normal social life again. On top of the reinforcement, the system only relies on punishing people for bad behavior instead of finding ways to keep people from bad actions in the first place.

Using predictive policing in the US showed that by observing suspects with high likelihood of committing crimes only led to overfull prisons, but having no effect on the crime rate itself [11, 23]. However using the same information to offer suspects internships helped to get people from committing a crime in the first place, leading to a safer environment and a lower crime rate as well as keeping the prisons from overflowing [11]. All in all, this technology opens up remarkable opportunities but still must be used with caution.

3 Conclusion

Overall the Chinese Social Credit System is a very interesting system with promising chances as well as possible dangers. It tackles the transparency problem of existing rating systems, but still leaves much room for improvement. However it might be a better approach to have a semi-transparent system that the citizens are aware of than having multiple completely opaque systems in the background, like in the West [15]. Both systems nevertheless show the high responsibility when developing such systems as well as the need for control mechanisms.

4 Epilogue

The course as well as writing this essay revealed the high responsibility of a data analyst and the importance to emphasize the purpose of the analysis when a new system is designed. The solution of the problem should not only be the main target, but also the causes for the problem should be approached. On top of that, even if especially for companies the economical outcome is always of importance, when building a system that influences people and their behavior, notably ethical thinking should be included as a major part of the design process. Therefore, I will try to strengthen ethics in AI and create awareness of the consequences our analysis might have.

References

[1] A. Sho, "Current diversification of behaviorism," https://www.researchgate.net/publication/333700078_Current_Diversification_of_Behaviorism, 2019.

[2] A. Bandura, "Social Learning Theory," General Learning Press, 1971.

[3] A. Bandura, "Social learning and personality development," Holt, Rinehart Winston of Canada Ltd, 1963.

[4] E. Berscheid, E. Walster, "Interpersonal Attraction," Addison-Wesley Publishing Company, 1969.

[5] A. Bandura, "Self Efficiacy: The Exercise of Control," Worth, 1997.

[6] R. Burgess, R. Akers, "A Differential Association-Reinforcement Theory of Criminal Behavior," Social Problems, 1966.

[7] P. H. Miller, "Theories of developmental psychology," Worth, 2011.

[8] A. Bandura, "Social Cognitive Theory for Personal and Social Change by Enabling Media," LEA's communication series, 2004.

[9] A. Singhal, E. M. Rogers, W. J. Brown, "Harnessing the potential of entertainment-education telenovelas", Kluwer Academic Publishers, 1993.

[10] J. Orlowski, "The social dilemma," Netflix, 2020.

[11] P. Assaro, "AI Ethics in Predictive Policing," IEEE Technology Society, 2019.

[12] D. M. Síthigh, M. Siems, "The Chinese social credit system: A model for other countries?," Department of Law, https://cadmus.eui.eu/bitstream/handle/1814/60424/LAW_2019_01.pdf, 2019.

[13] "Facebook-Cambridge Analytica data scandal," Wikipedia, https://en.wikipedia.org/wiki/Facebook-Cambridge_Analytica_data_scandal, 2021.

[14] C. Schaer, "Germany edges toward Chinese-style rating of citizens," https://www.handelsblatt.com/english/politics/big-data-vs-big-brother-germany-edges-toward-chinese-style-rating-of-citizens/23581140.html?ticket=ST-1696666-34y5bwncn5gCrb3HqACo-ap4, 2018.

[15] R. Botsman, "Big data meets Big Brother as China moves to rate its citizens," https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion, 2017.

[16] "Definition of the concept of social credit system," International Credit Monitoring Network of International Credit Standards Net, https://www.ice8000.org/cxb/16.html.

[17] A. Lee, "What is China's social credit system and why is it controversial?," https://www.scmp.com/economy/china-economy/article/3096090/what-chinas-social-credit-system-and-why-it-controversial, 2020.

[18] "Social norm," Wikipedia, https://en.wikipedia.org/wiki/Social_norm, 2019.

[19] "Fuzhou's social credit system has been established as a national demonstration," Fuzhou government, http://www.fuzhou.gov.cn/gzdt/rcyw/201908/t20190814_3014335.htm, 2019.

[20] K. Kelly, "The Inevitable," Viking, 2016.

[21] "How does corruption hinder China's development?," China Power, https://chinapower.csis.org/china-corruption-development/, 2020.

[22] J. Wright "Nosedive," Netflix: Black mirror, Season 3, episode 1, 2016.

[23] A. Liptak, "InmateCount in U.S. Dwarfs Other Nations," New York Times, https://www.nytimes.com/2008/04/23/us/23prison.html?searchResultPosition=1, 2008.