

Doppelganger Effect

The doppelganger effect is a phenomenon in which people perceive an image, sound, or idea as being repeated or duplicated. This effect has been studied extensively in various fields, including psychology, neuroscience, and computer science. It can occur in a variety of contexts, from visual illusions to social interactions. Tiddleman[1] found that people are more likely to perceive two faces as duplicates when they share similar features such as eye shape and nose size. Leibovich and Henik[2] found that people perceive auditory stimuli as duplicates when the stimuli share similar acoustic features, suggesting that the doppelganger effect is not limited to the visual domain. Riddoch[3] investigated the doppelganger effect in patients with prosopagnosia, a condition in which people have difficulty recognizing faces. They found that people with prosopagnosia are still susceptible to the doppelganger effect, suggesting that the effect is not dependent on conscious face recognition. This paper will discuss the influence of the doppelganger effect and the methods which can avoid or mitigate in the practice and development of machine learning models for health and medical science.

For computer science, doppelganger effects can occur in machine learning models when two or more features are highly correlated or contain duplicated information. And if it applied in the biomedical field, it may lead to incorrect diagnoses or treatments, resulting in harm to the patients. For example, if we want to design a machine learning model to predict the risk of respiratory disease based on the dataset that includes information on age, heart rate and Blood oxygen concentration. If age and heart rate are highly correlated, the model may incorrectly attribute the increased risk of respiratory disease to heart rate when in fact it is age that is driving the association. This can lead to incorrect diagnoses and treatments.

Doppelganger effects also can occur in many different types of data. In the medical

imaging, the doppelganger effect can occur when different patients have similar anatomical features that can lead to misdiagnosis or misinterpretation of the images. For example, in brain imaging studies, different patients with similar brain structure or lesions may be mistakenly identified as the same individual. This can lead to errors in diagnosis or treatment planning. Also, in the gene sequencing studies, the doppelganger effect can occur when different individuals have similar genetic profiles. For example, two individuals may have similar genetic variations that are associated with a particular disease, but only one of them actually has the disease. This can lead to incorrect conclusions about the genetic basis of the disease.

Doppelganger effects can be explained in many ways. Our perception of reality is not always accurate, and our brain can be easily fooled by visual or auditory illusions. When we see or hear something that resembles another image or sound, our brain can quickly create a connection between the two and create the impression of a duplicate. At the same time, our memory can also contribute to the doppelganger effect. When we encounter something that reminds us of a past experience, our brain can retrieve that memory and create a link between the two experiences. This can lead to the perception of a duplicate, even if the two experiences are not exactly the same. From a quantitative angle, the doppelganger effect can be explained by the statistical distribution of genetic variation across the population. In other words, the probability of finding two individuals with similar genetic variations is not negligible, and this probability increases as the number of individuals increases.

In order to avoid doppelganger effects in biomedical machine learning models, there are several approaches that can be taken. First of all, preprocessing and feature selection. Before training a machine learning model, preprocess the data and select features carefully. Preprocessing steps such as normalization, standardization, and outlier detection can help to reduce the impact of highly correlated or duplicated data. Feature selection methods such as principal component analysis (PCA) or regularization can help to identify and remove redundant features from the dataset. Second, the Cross-

validation. By splitting the dataset into training and testing sets, cross-validation helps to identify overfitting and ensures that the model is generalizing well to new data. Third, Regularization. By adding a penalty term to the objective function, regularization encourages the model to prioritize simpler solutions that generalize better to new data. Fourth, increase the diversity of data. Ensure that the dataset used to train the machine learning model is diverse and representative of the population of interest. It can be achieved by collecting data from multiple sources, including data from different populations, institutions, and time periods. Last but not least, the feature engineering. In addition to preprocessing and feature selection, feature engineering involves creating new features that capture important information from the original features while reducing the correlation between them.

To address these challenges, it is also important to maintain transparency and accountability in the development and deployment of machine learning models for health and medical science. it includes documenting the sources and quality of the data used to train the model, ensuring that the model is interpretable and understandable to domain experts, and regularly monitoring and auditing the performance of the model in real-world settings.

In conclusion, doppelganger effects are not unique to biomedical data, but can pose significant challenges in machine learning models used for health and medical science. Careful preprocessing, feature selection, cross-validation, regularization, diversity of data, feature engineering, and transparency and accountability can all help to mitigate these effects and ensure that machine learning models in biomedical data are robust and reliable. It is important for researchers, practitioners, and policymakers to be aware of these challenges and to work together to develop and deploy machine learning models that improve health outcomes while minimizing the risk of harm.

References

- [1] Tiddeman B, Burt M, Perrett D. Prototyping and transforming facial textures for perception research[J]. IEEE computer graphics and applications, 2001, 21(5): 42-50.
- [2] Leibovich Tali, Henik Avishai. Magnitude processing in non-symbolic stimuli[J]. Frontiers in Psychology, 2013
- [3] M. Jane Riddoch and Robert A. Johnston. Are faces special? A case of pure prosopagnosia[J]. Cognitive Neuropsychology, 2008