



HOKKAIDO UNIVERSITY

Title	A study on machine learning for personalized prediction of human perception toward visual stimuli [an abstract of dissertation and a summary of dissertation review]
Author(s)	諸戸, 祐哉
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第16015号
Issue Date	2024-03-25
Doc URL	https://hdl.handle.net/2115/92390
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	doctoral thesis
File Information	Yuya_Moroto_abstract.pdf, 論文内容の要旨



学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称 博士（情報科学） 氏名 諸戸 祐哉

学 位 論 文 題 名

A study on machine learning for personalized prediction of human perception toward visual stimuli
(視覚刺激に対する人間の知覚を個別予測するための機械学習に関する研究)

This thesis summarizes studies on the construction of machine learning models specific to personalized prediction of human perception toward visual stimuli.

Machine learning has attracted significant attention in assisting humans due to its high potential and continues to respond to expectations in various fields. Specifically, after the development of deep learning technologies such as convolutional neural networks and recurrent neural networks, machine learning models can solve more complex tasks by learning a large amount of data. Recent studies on machine learning have progressed to the foundation models such as contrastive language-image pre-training and generative pre-trained transformer, and researchers have focused on the way to construct models that can effectively learn big data and conduct several tasks in a single model. Namely, they aim to develop the generalized model. Although this direction may be one advancement of machine learning, another important direction is the development of machine learning models that can be tuned for each person from the perspective of human assistance. For instance, user satisfaction in video-sharing services can be improved by personalizing the multimedia content recommender system. Therefore, the personalization of machine learning can be an effective direction of advancement.

The person-specific information is needed as a clue for training machine learning models to suit each person. One of the person-specific information is the biological information obtained from humans. Here, to introduce such information into the machine learning models, human perception should be mediated as in the actual human information processing. However, it is difficult to directly implement them to existing models in various tasks such as content recommendation and information retrieval since machine learning just recognizes the patterns in the inputs and outputs and may ignore human perception. Hence, studies on predicting human perception have been conducted to indirectly personalize machine learning. Concretely, previous studies have predicted emotion and attention as human perception from brain activity and gaze data as the data representing biological information (hereafter, biological data). In these studies, although machine learning models have been used as prediction models, these models do not necessarily consider the properties specific to biological data since their architectures are designed not specifically for biological data. In contrast to the general data in the fields of computer vision and natural language processing, biological data are difficult to handle due to their unique properties such as individual differences. Therefore, there are great demands for rethinking the machine learning models suitable for biological data.

This thesis focuses on three perspectives related to the inherent properties of biological data. The first perspective is the data volume obtained from each person. Biological data varies widely among individuals, and data obtained from various persons are difficult to handle in a uniform manner. Hence,

the machine learning models need to be trained from the limited amount of data for reflecting on individual differences. The next perspective is the relationship between stimuli and their human response. Humans constantly receive a variety of stimuli and perceive them in their daily lives, and biological data reflect on such stimuli. To effectively predict human perception, not only biological data but also the contents of stimuli should be considered. Finally, the third perspective is mutual complementation through the collaborative use of several types of biological data. Advancements in sensor technologies enable the easy and simultaneous acquisition of various types of biological data. Each type of biological data represents a different aspect of the human response, and the human perception can be more precisely predicted by collaboratively using them than one of them alone.

The purpose of this thesis is to construct machine learning models that can predict personalized human perception by incorporating the above perspectives. This thesis targets the human perception toward visual stimuli since several studies show that visual information is the most important to humans. Concretely, this thesis mainly tackles three themes to construct the machine learning models incorporating the above perspectives, respectively. First, to address the problem of the data volume, I focus on the similarities of biological data between persons. In the case of predicting human attention toward visual stimuli such as images, I propose a new method for detecting the individuals with biological data patterns similar to those of the target person. Moreover, I construct the machine learning model using the data obtained from similar persons for predicting the perception of the target person. Secondly, for analyzing the relationship between visual stimuli and biological data, I focus on the construction of the uniform representation of visual contents and gaze data including the region watched by the person. Finally, I newly propose the feature integration methods for treating several types of biological information since biological data are pre-processed for calculating features suitable for each type of data before inputting machine learning models, generally. Then, when calculating the features of gaze data, I adopt the representation based on the second perspective for considering both visual contents and biological data. In this way, I newly proposed machine learning models suitable for biological data and indicate the effectiveness of focusing on the above inherent perspective.

This thesis consists of six chapters. Chapter 1 describes the research background and the proposition of this thesis. Chapter 2 describes the related works and their problems to be solved in this thesis. Chapter 3 presents methods for few-shot personalized saliency prediction, which is the task predicting regions in images gazed at by individuals. Chapters 4 and 5 focus on human emotions as perceptions. Chapter 4 presents the methods for classifying images into emotional categories using gaze data. Chapter 5 presents the methods for multi-modal human emotion recognition based on various types of biological information. Finally, Chapter 6 concludes this thesis and clarifies the future directions.

In summary, this thesis presents several machine learning methods for personalized prediction of human perception toward visual stimuli. For constructing the machine learning models specific to personalized prediction of human perception, the proposed methods incorporate the similarities of biological data between persons and mutual complementation between different types of biological information. Furthermore, I confirm the effectiveness of the proposed methods through empirical experimentation on datasets derived from personally acquired raw data and openly available datasets.