論 文 の 内 容 の 要 旨

論文題目　Subgradient-based Online and Stochastic Learning with Biases
（劣勾配ベースのバイアス考慮型オンライン学習および確率的学習）

氏　　名　大岩　秀和

Machine learning is a framework to make machines learn from data. By taking advantage of useful information in observed data, machine learning enables informative knowledge to be extracted from the data. This extracted knowledge can be used to analyze and predict unobserved data with great accuracy. A large variety of massive-scale data has recently become available for analysis in many fields along with the recent progress of gathering, storing, and processing data in machines. The increase of the data size contributes to the improvement of the predictive performance both theoretically and empirically; therefore, the demand to deal with massive data has been increasing in an efficient way. However, developing novel machine learning algorithms for massive data analysis is a challenging task due to many practical difficulties, such as achieving a short computational time with a small memory size to derive a predictive model. To solve these problems, subgradient-based online and stochastic learning frameworks have been emerging to systematically utilize massive data. These frameworks process a bunch of data one by one and iteratively update predictive models through the subgradient information with respect to the currently received datum. This incremental update procedure realizes an efficient massive data analysis with a simpler implementation and faster computation with a smaller memory space than other machine learning frameworks.

In this thesis, we find that several biases make the original performance measures of online and stochastic learning algorithms obsolete. We find out two biases in this thesis: a truncation bias and a human cognitive bias. These biases originate from involvements of data and humans, which are crucial for applying these incremental learning algorithms to practical applications. We first show how these

biases affect the original measures by using toy examples and subjective experiments. These results indicate that the standard performance measures become inappropriate to achieve the original objectives of online and stochastic learning due to the existences of these biases. In the result, state-of-the-art incremental learning algorithms are critically deteriorated even though these algorithms are guaranteed to derive the optimal solution in the conventional performance measures. To solve this difficulty, in this thesis, we reformulate objectives in online and stochastic learning by integrating the effects of these biases and develop algorithms for new framework while maintaining advantages of incremental learning procedures. We theoretically prove the effectiveness of our proposed algorithms under new objectives. In addition, we experimentally show the superiority of our algorithms by using several datasets.