Domain Adaptation for Application to Smart Meters of Electricity Usage in Each Household

Hisashi Oshima*, Tomoyuki Higuchi*, and Tsuyoshi Ishizone[+]

***Graduate School of Science and Engineering, Chuo University, Japan**

[+]Graduate School of Advanced Mathematical Sciences, Meiji University, Japan

The measurement of electricity consumption by each household is becoming increasingly automated with the installation of smart meters, and the use of this data is being considered in a variety of ways. In Japan and the U.K., data is collected and stored every 30 minutes from the standpoint of data storage efficiency. One promising use of the data is to analogize the presence or absence of residents in a home. If this can be achieved, the cost of re-delivery to absent residences in home delivery, which has increased dramatically in recent years due to the Corona disaster, can be significantly reduced. Therefore, there has been considerable research in recent years on determining whether a homeowner is present or absent from his or her electricity smart meter.

From a machine learning perspective, the study of algorithms for determining presence/absence is relatively easy to formulate in the framework of supervised learning. The problem is that supervised data are rarely available. Experimentally, it is easy to obtain supervised data by installing carbon dioxide monitoring device or sensors that detect people's movements. However, because people's behavior is so varied, machine learning models based on a small number of examples of experimental data are almost useless. As for the possibility of obtaining more data, it is quite difficult to obtain cooperation for the installation of devices in homes to monitor presence and absence due to privacy and security concerns. Also, while it would be possible to solve this problem by having security companies and power companies share data, it is not feasible from a business standpoint to begin with.

Therefore, to summarize the problem from a machine learning perspective, a presence/absence discriminator must be constructed based on a small number of supervised data sets and a large amount of unsupervised data sets. A more troublesome problem arises here. The problem is that the distribution of the explanatory variable vectors differs between the supervised and unsupervised data sets, a so-called covariate shift. This cannot be avoided even in the analysis of electricity usage data.

Techniques for dealing with this type of problem are referred to in recent machine learning as domain adaptation, and domain adaptation can also be considered a form of

transfer learning. Although the definition of domain adaptation is very broad, in this study we define it simply. The domain with supervised data is usually called the source domain and the domain with unsupervised data is usually called the target domain. In domain adaptation, a domain discriminator that discriminates between source and target from the explanatory variable vector and a discriminator that discriminates between presence and absence from the source's explanatory variable vector are learned simultaneously. However, feature representation learning, which consists of explanatory variable vectors, is learned in an adversarial manner so that the differences between domains are small. In other words, the feature extractor is learned in such a way that it tricks the domain discriminator to reduce the differences between domains.

Although the application of this methodology is very broad, in this presentation we focus on the problem of presence/absence discrimination from electricity smart meters. This methodology allows us to assign pseudo-labels to unlabeled, unsupervised data from a vector of explanatory variables. Even a small increase in the accuracy of the labels can have a tremendous economic impact. In this presentation, we consider a methodology that can handle cases where the distribution differs not only between the source and target, but also between other factors for both. Due to the limited presentation time, this presentation will introduce our aims and outline our methodology. We also present the results of our application to publicly available data. Please note that our methodology is based on the real big data of a company.