# "Wide learning" AI technology enables highly precise learning even from imbalanced data sets
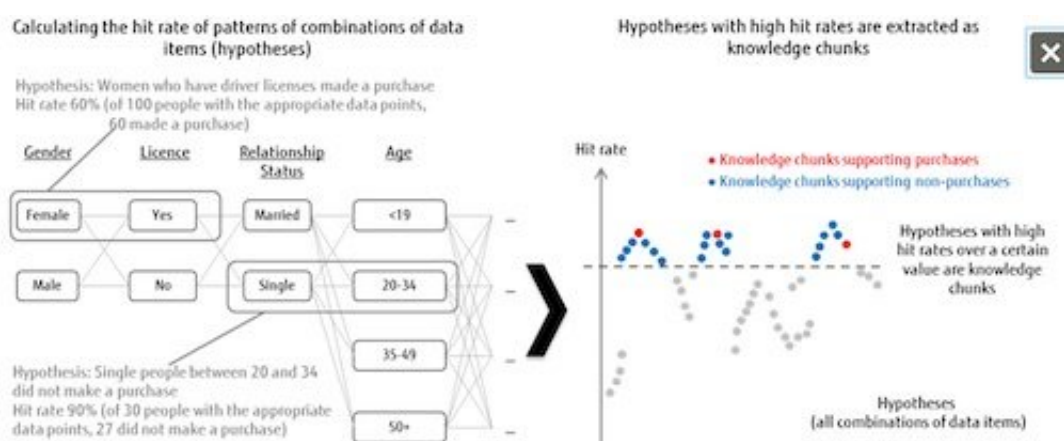
September 20 2018



Figure 1: Hypothesis listing and knowledge chunk extraction. Credit: Fujitsu

Fujitsu Laboratories Ltd. today announced the development of "Wide Learning," a machine learning technology capable of accurate judgements even when operators cannot obtain the volume of data necessary for training. AI is now often used to leverage data in a variety of fields, but the accuracy of AI may be impacted in cases where the volume of data to be analyzed is small or imbalanced. Fujitsu's Wide Learning technology enables judgements to be reached more accurately than was previously possible, and learning is achieved uniformly, no matter which hypothesis is examined, even when the data is imbalanced. It achieves this by first extracting hypotheses with a high degree of

importance, having made a large set of hypotheses formed by all of the combinations of data items, and then by controlling for the degree of impact of each respective hypothesis based on the overlapping relationships of the hypotheses. Moreover, because the hypotheses are recorded as logical expressions, humans can also understand the reasoning behind a judgement. Fujitsu's new Wide Learning technology allows for the use of AI even in areas such as healthcare and marketing, where the data needed to make judgements is scarce, supporting operations and promoting the automation of work processes using AI.

In recent years, AI technology has begun to be used in a variety of fields, including healthcare, marketing, and finance. Expectations are rising for the use of AI decision-making in support of operations and automating tasks in these areas. One challenge that remains to realizing the potential of these technologies, however, is that the data may be imbalanced. Specifically, depending on the industry it can be difficult to obtain sufficient data for training AI on the targets on which it is to make judgements. This, in effect, leaves many of these technologies unable to produce results with sufficient accuracy for practical use. Furthermore, a major reason why AI deployment lacks progress is that even when an AI provides sufficiently accurate recognition or classification performance, experts and even the developers themselves often cannot explain why the AI produced a certain answer, and if they cannot fulfill their responsibility to explain the results to the front lines of industry then AI cannot be deployed.

AI technologies based on deep learning conventionally make highly accurate judgements by being trained on large volumes of data, including ample target data to be judged. In real world scenarios, however, there are many cases in which the data is insufficient, with extremely little target data. In these cases, when faced with unknown data, it becomes difficult for AI technology to deliver highly accurate judgements. Moreover, the machine learning model for existing AI based on deep

learning is a black box model that cannot explain the reasons behind the judgements the AI makes, creating a problem with transparency. As such, moving forward it will be necessary to develop new AI technology that realizes highly accurate judgements from imbalanced data, and that is also transparent in order to solve various issues in society.

Bearing these challenges in mind, Fujitsu Laboratories has now developed Wide Learning, a machine learning technology capable of making highly accurate judgements even in cases where the data is imbalanced. The features of Wide Learning technology include the following two points.

# 1. Creates combinations of data items to extract large volumes of hypotheses
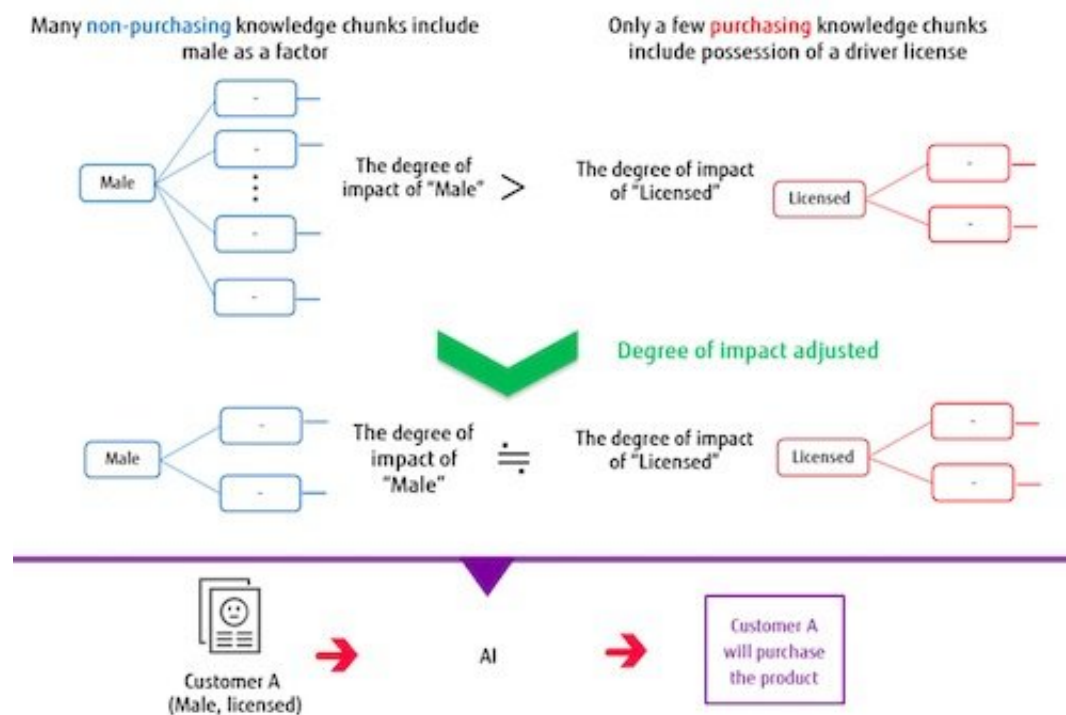


Figure 2: When making a classification model, the knowledge chunks impact adjustment. Credit: Fujitsu

This technology treats all combination patterns of data items as hypotheses, and then determines the degree of importance of each hypothesis based on the hit rate for the label category. For example, when analyzing trends in who purchases certain products, the system combines all sorts of patterns from the data items for those who did or did not make purchases (the category label), such as single women between 20-34 who have driver licenses, and then analyzes how many hits it gets in the data of those who actually made purchases when these combination patterns are taken as hypotheses. The hypotheses that achieve a hit rate above a certain level are defined as important hypotheses, called "knowledge chunks". This means that even when target data is insufficient, the system can extract all hypotheses worth looking into, which may also contribute to the discovery of previously unconsidered explanations.

## 2. Adjusts the degree of impact of knowledge chunks to build an accurate classification model

The system builds a classification model based on multiple extracted knowledge chunks and on the target label. In this process, if the items making up a knowledge chunk frequently overlap with the items making up other knowledge chunks, the system controls their degree of impact so as to reduce the weight of their influence on the classification model. In this way, the system can train a model capable of accurate classifications even when the target label or the data marked as correct is imbalanced. For example, in a case where men who did not make a purchase make up the vast majority of an item purchase dataset, if the AI is trained without controlling the degree of impact, then the knowledge chunk that includes whether or not a person has a license, independent of gender, will not have much influence on the

classification. With this newly developed method, the degree of impact of knowledge chunks including male as a factor is limited due to the overlap of this item, while the impact of the smaller number of knowledge chunks that include whether a person has a license becomes relatively larger in training, building a model that can correctly categorize both men and possession of a license.

Fujitsu Laboratories conducted a trial of this technology, applying it to data in areas such as digital marketing and healthcare. In a test using benchmark data in the marketing and healthcare areas from the UC Irvine Machine Learning Repository, this technology improved accuracy by about 10-20% compared to deep learning. It successfully reduced the probability that the system would overlook customers likely to subscribe to a service or patients with a condition by about 20-50%. In the marketing data, of the approximately 5,000 customer data entries used in the test, only about 230 were for purchasing customers, making for an imbalanced set. This technology reduced the number of potential customers excluded from sales promotions from 120, the result of deep learning analysis, to 74. Moreover, as the knowledge chunks that form the basis for this technology have a logical expression format, the ability to explain the reasoning behind a judgement is also useful in implementing this technology in society. Even when it is determined that corrections to a model are necessary, based on results from new data, it is possible to make more appropriate revisions, because users can understand the reasons for results.

Fujitsu Laboratories will continue to apply this technology to tasks that demand the reasoning behind AI judgements, such as in financial transactions and medical diagnoses, and to tasks that handle low frequency phenomena, such as fraud and equipment breakdowns, with the goal of commercializing it as a new machine learning technology supporting Fujitsu Limited's Fujitsu Human Centric AI Zinrai in fiscal 2019. Fujitsu Laboratories will also make effective use of this

technology's characteristic capability for explanation, continuing research and development into topics such as improved support for making judgements and decisions in tasks to which it is applied, and into the overall system design, including collaboration with humans.

Provided by Fujitsu

Citation: "Wide learning" AI technology enables highly precise learning even from imbalanced data sets (2018, September 20) retrieved 10 May 2024 from https://phys.org/news/2018-09-wide-ai-technology-enables-highly.html