

# Feature Concepts as Pattern Language for Data-Federative Innovations

Yukio Ohsawa<sup>1</sup>, Sae Kondo<sup>2</sup>, and Teruaki Hayashi<sup>1</sup>

<sup>1</sup>The University of Tokyo, <sup>2</sup>Mie University  
<sup>1</sup> 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan,  
 ohsawa@sys.t.u-tokyo.ac.jp

## Abstract

To ensure that all papers in the publication have a uniform appearance, the authors must adhere to the following instructions: Feature concepts, an essential tool for data-federative innovation processes, are introduced here as a language to express the model of knowledge to be acquired from data. A feature concept can be represented by a simple feature, such as a single variable, or by a conceptual illustration of the abstract information obtained from the data. Useful feature concepts for satisfying the latent or explicit requirements in society, or the market of data, are found to have been elicited so far via creative communication among stakeholders. Here, the contribution of feature concepts to useful findings is shown with a couple of use cases, for example, explanation of change in markets and earthquakes.

## Introduction

The necessity to elicit information about the data-use contexts, that is, the situations where to use data and/or receive the services or products created based on data, has been positioned as a key scope in creating a solution for satisfying a requirement in businesses. Although participants enjoyed workplaces for innovations using/reusing data [Ohsawa et al. 2013], a marketplace for data-federative innovation) have been urged to speak out requirements and ideas for their satisfaction so that they can add or revise the DJs and store the used ideas in the background database, the missing links between data and the requirement cannot be covered. To cope with this problem, in this study, a method is introduced to illustrate the abstract image of the information to be acquired using datasets for requirement satisfaction.

## Feature concepts

A *feature concept* is an abstract image of the information or knowledge to be acquired using data linked to the method, that is, how, why, and the dataset(s), that is, what, should be used to satisfy a requirement. In the examples shown below, we discover that human creativity in data utilization has been enhanced by eliciting, using, and sharing concepts in various forms. These concepts, if the creator explicitly represents, are regarded as feature concepts. Below, we consider the feature concepts illustrated in Fig.1. For example,

unsupervised machine learning methods, such as clustering with cutting noise events (e.g. [Fränti and Yang 2018]) are algorithms in which a hidden cluster is restored from data including scattered noise signals. Thus, *embedded clusters*, as the desired information to be acquired from the data, can be interpreted as a feature concept, as shown in Fig.1. In addition, the decision tree [Quinlan 86] realizes a *tree* as a feature. Feature concepts, if represented explicitly via the communication of participants in the data market, play the role of bridging social requirements and features in datasets, as illustrated in Fig. 2.

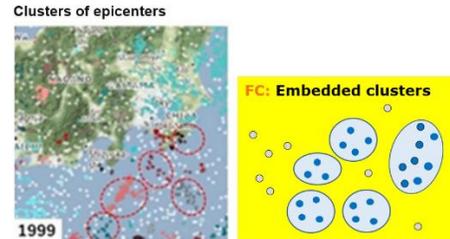


Fig. 1. Examples of feature concepts for three basic methods for data mining (left: from [Ohsawa 2018b]).

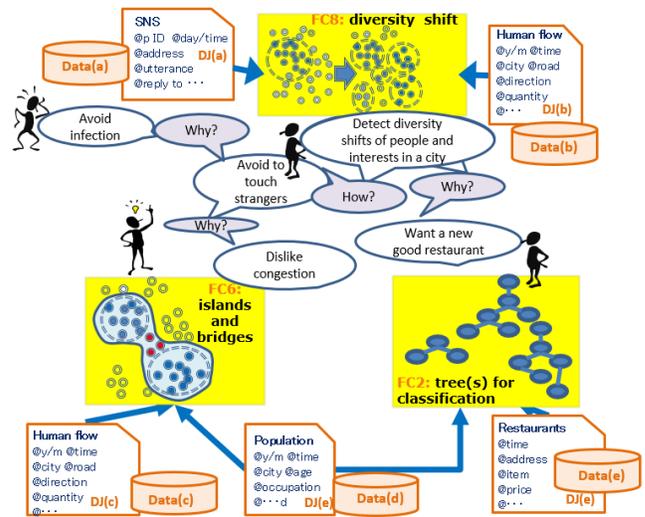


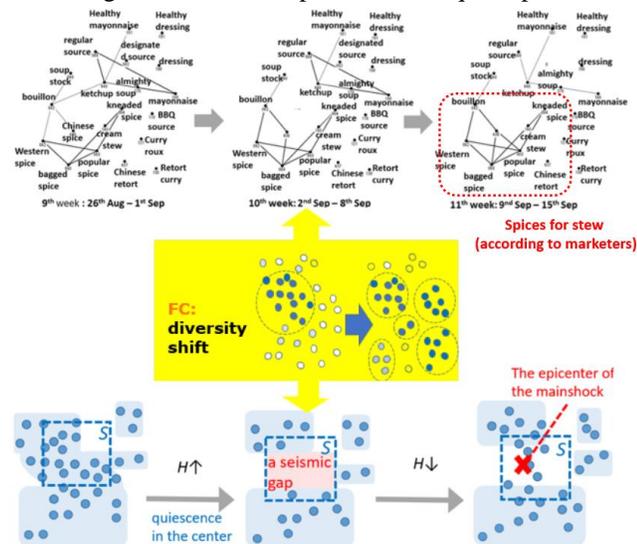
Fig. 2. The images and positions of feature concepts (FC#) in the communication to connect requirements to solutions and DJs.

In T. Kido, K. Takadama (Eds.), Proceedings of the AAAI 2022 Spring Symposium "How Fair is Fair? Achieving Wellbeing AI", Stanford University, Palo Alto, California, USA, March 21–23, 2022. Copyright © 2022 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

## Examples of feature concepts in data utilities

An example is the *change explanation* in businesses and sciences, which was elicited as a requirement for supermarkets. In comparison with the *detection* or *prediction* of changes using machine learning technologies (e.g., [Fearnhead and Liu 2007, Miyaguchi and Yamanishi 2017]), change explanation means linking the observed change in the data to human understanding of the dynamics in the real world. Thus, it is essential to create a feature concept for enabling data visualization that inspires humans to understand the underlying dynamics. Borrowing the idea of diversity shift proposed by Kahn [Kahn 1995] in a market, we drew an image corresponding to the feature concept in Fig.3 and invented graph-based entropy (GBE [Ohsawa 2018a]) which is an index of the diversity of events on their distribution to the clusters in the co-occurrence graph of items in the market. The change in GBE is a sign of structural change in the target real world and is informative in explaining changes if coupled with the graph shown in Fig.3, where the bridging edge between the two clusters is cut in the 10<sup>th</sup> week of the year, which is interpreted as the growth of the lower cluster corresponding to spices for cooking stew. The 10<sup>th</sup> week in the data was a hot period in August, but the frequency of the query “stew” in Google increased from August in Japan every year.

The author then diverted the *diversity shift* to an analysis of earthquakes [Ohsawa 2018b]. Here, a model was introduced to explain the dynamics of earthquakes in two phases: (1) the increase in the diversity of epicenter clusters, and (2) the coupling of the clusters due to new activity in the seismic gap, followed by a large one. The entropy defined in the distribution of the epicenters increases in phase (1) and decreases in phase (2). Thus, the FC *diversity shift* used for marketing was reused to explain the earthquake precursors.



**Fig. 3.** The FC for explaining the changing point in the market, analogically used for earthquakes.

## Feature concepts and pattern language

Feature concepts may be regarded as a customized pattern language, initially proposed in urban planning [Alexander et al 1977] and diverted so far to other systems design. Here, a set of patterns composed of urban elements, such as parks, bridges, houses, etc. were used to explain and design structures of urban areas. Each pattern with an illustration is linked to a context, problem, and solution to the problem. Individual thoughts and communication toward consensus within a team engaged in a task of design or other collaborations can be smoothed by using the patterns as a common language for expressing contexts, problems, and solutions. Furthermore, the patterns can be connected via relationships from/to each other, which may be hierarchical relations or likeliness to be combined. Similarly, once a feature concept is created and shared with others, it becomes a tool for innovators who think and communicate to a federate and/or use data. In addition, the relationships among feature concepts can be, similarly to patterns in a pattern language, the hierarchical structure (e.g., “diversity shift” over “diversity,” etc.), the connectivity (e.g., *diversity shift* can be connected with *clusters* or with *networks*), etc. Thus, the links between feature concepts or from feature concepts to the real-world should come from the communication between data scientists or data scientists with others.

## Acknowledgement

This study is partially supported by JSPS 20K20482

## References

- Alexander, C., Ishikawa, S., Silverstein, M. 1977. *A Pattern Language: Towns, Buildings, Construction*. Oxford Univ. Press, USA.
- Fearnhead, P, Liu, Z. 2007. Online Inference for Multiple Change-point Problems. *J. Royal Statistical Soc. B69(4)* ISSN 1369-7412
- Fränti P., Yang J. 2018. Medoid-Shift for Noise Removal to Improve Clustering. In: Rutkowski L., et al (eds), *Artificial Intelligence and Soft Computing*, LNCS10841. Springer
- Kahn, B.K. 1995. Consumer variety seeking among goods and service, *J. Retailing and Consumer Services* 2, 139-148
- Miyaguchi, K., and Yamanishi, K. 2017. Online detection of continuous changes in stochastic processes, *Int J. Data Science and Analytics* 3 (3), 213–229
- Ohsawa, Y., Kido, H., Hayashi, T., Liu, C. 2013. Data Jackets for Synthesizing Values in the Market of Data, *Procedia Computer Science* 22, 709-716, doi.org/10.1016/j.procs.2013.09.152
- Ohsawa, Y. 2018a. Graph-Based Entropy for Detecting Explanatory Signs of Changes in Market. *Rev Socionetwork Strat* 12, 183–203 (2018). https://doi.org/10.1007/s12626-018-0023-8
- Ohsawa, Y. 2018b. Regional Seismic Information Entropy for Detecting Earthquake Activation Precursors, *Entropy* 20(11), 861.
- Quinlan, J. R. 1986. Induction of Decision Trees. *Mach. Learn.* 1, 1, 81–106