Exploring multi-dimensional concept of well-being thorough geometry of culture approach

Keywords: well-being; text as data; word embedding model; word2vec; sociology of culture

Extended Abstract

Well-being is of great concern to researchers, policymakers, and citizens alike. Nevertheless, it is difficult to clearly define what well-being is, as the CDC states that there is "no consensus around a single definition of well-being" (CDC 2020). This is possibly because of the multidimensional nature and historical variation of the conception of well-being.

In traditional research, there have been two approaches to investigate well-being. On the one hand, in the field of philosophical research, from Aristotle to Bentham and Rawls, there have been philosophical analyses of what happiness or good means to human beings. More recently, Sen and Nussbaum's capability approach (Sen 1993; Nussbaum 2007) has attracted attention. While it is necessary to examine these theoretical studies in the light of people's actual thoughts and ideas of what well-being is. On the other hands, empirical studies of well-being often use surveys to ask people about their thoughts and subjective sense of well-being. However, surveys can only explore a limited number of pre-selected items.

Here, we adopt “text as data” approach to not only identify the dimensions that constitute well-being in an exploratory and heuristic way, but also track their historical changes. Compared to other approaches, the text as data approach has the following advantages. First, it allows us to examine the extent to which the various dimensions of well-being proposed by philosophical studies capture the thoughts of real people as embodied in texts. Second, it allows us to identify the dimensions and word schemas that constitute well-being in a more informative and heuristic way than other empirical methods such as surveys.

Specifically, based on the previous study (Koslowski et al. 2019), we investigate the multiple dimensions of well-being and how they evolve over time with a word embedding model applying the massive Japanese library corpus.

Data and Method

The National Diet Library (NDL) of Japan recently launched the digital library project, which provides a large amount of digitalized text dataset comprising over two million magazines and books published from mid-19th century to the 20th century. This project is ongoing and is expected to add more data in the future, but at present it is the most extensive and long-term dataset on Japanese language publications, and is ideal for exploring the multidimensional meaning and historical changes of well-being in the Japanese-speaking world.

We employ what we call “geometry of culture” approach to NDL dataset. The basic idea is to identify specific semantic and cultural dimensions in the semantic space obtained by the word embedding model and to examine the relationships and historical changes among these multiple dimensions. The analysis will proceed as follows. First, we use Word2vec algorithm to train the word embedding model based on the text corpus of each year. Specifically, to avoid any biases that may arise from the comparison of semantic meaning between *Kyujitai* (old character form) and *Shinjitai* (simplified Japanese), this study specifically focuses on text data from the 1950s, when simplified Japanese became mainstream in language use. Then, we specified semantic dimensions related to well-being. Here, based on Sen and Nussbaum's theory, we chose nine semantic dimensions related to well-being in addition to subjective well-

being dimension, that amounts to total ten semantic dimensions. we establish the semantic dimension by taking the average of differences between numerous pairs of antonym:

$$\frac{\sum_{p_1}^{|P_1|}\overrightarrow{p_1}}{|P_1|} - \frac{\sum_{p_2}^{|P_2|}\overrightarrow{p_2}}{|P_2|}$$

Here, P_1 and P_2 implies the word vectors of antonym word pairs relatd to relevant concepts. For example, to construct the semantic dimension of subjective well-being, we prepare a word list of subjective happiness like "glad（嬉しい）" and a word list of subjective unhappiness like "uncomfortable（不快）". Leveraging the property that Word2Vec is able to capture word analogy, the concept of subjective well-being can be characterized by this antonym pair. In this term, we also construct the semantic dimensions for the components of well-being, such as health, education, freedom, peace, security, democracy, play, money and affiliation. Finally, based on the constructed semantic dimensions, we estimate the cosine similarity of subjective well-being dimension and each of other semantic dimensions. In general, a higher cosine similarity value implies a stronger association between subjective well-being and the corresponding components.

Results and Summary

Table 1 demonstrates the cosine similarity between the subjective well-being dimension and other semantic dimensions, which allows us to observe which component plays a more important role in the construction of well-being. Throughout the years, the peace dimension is the most highly correlated with well-being, followed by health, freedom, education, and safety.

As for changes over time, there are two types of dimensions, the one type of dimensions that have been relatively stable in their correlations with well-being throughout the period, and the other type of dimensions whose relationship with well-being has changed significantly. The former includes peace, security, health, and freedom dimension. In the latter, the play and affliation dimensions are particularly noteworthy. As Figure 1 shows, both the cosine similarity between well-being dimension and "affiliation-isolation" dimension and the cosine similarity between well-being dimension and "play-work" has increased over the 50 years.

Our analysis attempted to extract the multidimensional concept of well-being, which has been theoretically proposed by philosophers and others, "in the wild" by means of text as data analysis and a geometry of culture approach. As a result, the complexity of the multidimensional axis of well-being was revealed. It was also suggested that while some dimensions of well-being are stable over time, others follow historical changes. The historical changes we captured indicate the growing importance of play and affiliation for well-being, consistent with the post-materialist trend advocated by R. Inglehart (2018).

References

Centers for Disease Control and Prevention quality of life (CDC), 2020, *Health-Related Quality of Life (HRQOL): Well-Being Concepts*, https://www.cdc.gov/hrqol/wellbeing.htm

Kozlowski, A.C., Taddy, M. and Evans, J.A., 2019, "The geometry of Culture: Analyzing the Meanings of Class through Word Embeddings.", *American Sociological Review*, 84(5), pp.905-949.

Nussbaum, M.C., 2007, *Frontiers of justice*, Harvard University Press.

Sen, A., 1993, "Capability and well-being", *The Quality of Life,* Oxford: Clarendon Press, pp.270-293.

Inglehart, R., 2018, *Cultural Evolution, People's Motivations are Changing, and Reshaping the World*, Cambridge University Press

Table 1. Cosine Similarity between the Well-being Dimension and Other Semantic Dimensions

Dimension	Mean	SD	Max	Min
Peace - War	0.415	0.0224	0.456	0.351
Healthy - Unhealthy	0.398	0.0245	0.456	0.348
Freedom - Unfreedom	0.386	0.0283	0.435	0.333
Educated - Uneducated	0.374	0.0358	0.472	0.304
Safety - Danger	0.362	0.0223	0.408	0.317
Play - Work	0.311	0.0822	0.432	0.061
Democracy - Despotism	0.255	0.040	0.325	0.167
Rich - Poor	0.233	0.0437	0.323	0.157
Affiliation - Isolation	0.186	0.0355	0.265	0.0987

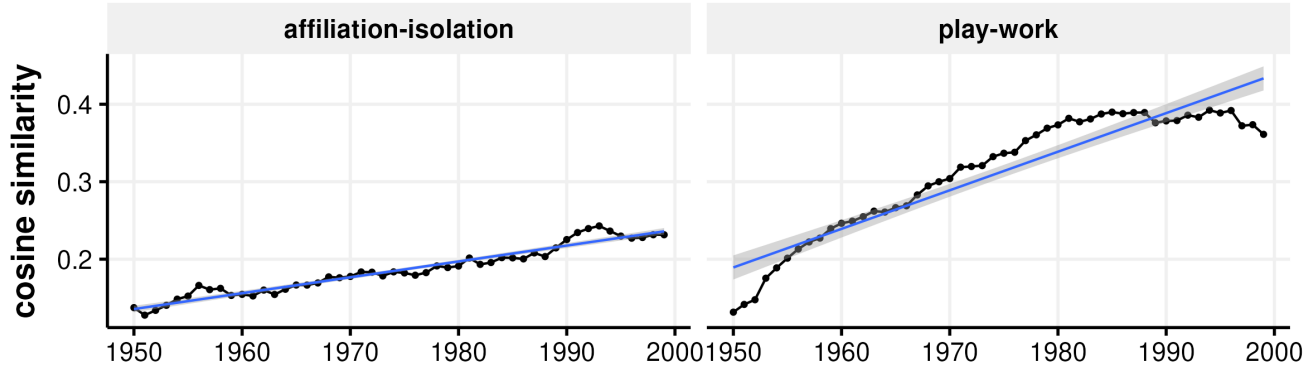

Figure 1.The Dynamics of Cosine Similarity between the Well-being Dimension and Other Semantic Dimensions over 50-years