

The Basic Trend of Media Reports on Residents' Return in Fukushima: In the Realms of Text Mining Analysis

Daisuke Sasaki (Corresponding author)

Graduate School of Frontier Sciences, The University of Tokyo

5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8563, Japan

Tel: 81-90-6718-9295 E-mail: k-sasaki@mail.ecc.u-tokyo.ac.jp

Received: January 10, 2017 Accepted: February 5, 2017 Published: February 28, 2017

doi:10.5296/jad.v3i1.10572 URL: <http://dx.doi.org/10.5296/jad.v3i1.10572>

Abstract

In general, people's concerns appear to be represented through the mass media. In the Great East Japan Earthquake and the nuclear power station accident in Fukushima prefecture, the many reports made by the mass media provide enough information that we can understand the overall situation. In this context, examining reports by the mass media concerning this issue can be thought of as an analysis of stakeholders' concerns. However, it is of great significance to verify whether reports by the mass media correctly reflect people's concerns or not. I analyzed Japanese text data from the Fukushima local newspaper using a text mining methodology, and have confirmed the general accuracy of media reports with respect to evacuees' concerns. I also found that in media reports the biggest issue related to residents returning to their homes seemed to be their livelihoods, whereas matters related to the Fukushima Daiichi Nuclear Power Station were rarely reported in connection with residents' return. Therefore, the most effective measure for having evacuees willingly return to their homes may be to address their anxieties about daily life including the employment environment with high priority so that evacuees can return to their homes willingly.

Keywords: evacuees, return, text mining, word2vec, Fukushima

1. Introduction

The big earthquake off the Pacific coast of Tohoku and TEPCO's Fukushima Daiichi Nuclear Power Station (FDNPS) accident devastated Fukushima Prefecture on March 11, 2011. Due to the terrible nuclear power station accident, an evacuation order was issued within a three-kilometer radius of the FDNPS on the same day, followed by an expansion of the evacuation order zone to a 20-kilometer radius on the next day. In the following month, the planned evacuation zone and the emergency evacuation preparation zone were established in the neighboring area (Fukushima Prefecture, 2016).

Hirono Town is located on the east coast of Fukushima Prefecture. Its population, based on the Basic Resident Register, is about 5,000 as of November 2016 (Hirono Town, 2016). The town was part of the emergency evacuation preparation zone until September 30, 2011. As of June 2016, although almost five years had passed since the town was removed from the emergency evacuation preparation zone, about 2,300 people—equivalent to nearly half of its population—still lived outside the town (Hirono Town, 2016). Obviously, this is a very serious issue that the town is currently facing. The town has conducted two surveys on the townspeople's intentions concerning this matter, in November 2013 and November 2014. The results showed that the main reasons why people living outside of the town were not willing to return were as follows: inconvenience of living environment and infrastructure; anxiety about the FDNPS; influence by radioactive contamination; concerns about the study environment of children and so on (Hirono Town, 2013, 2014). These reasons indicated the respondents consider the conditions in their present settlements to be superior to those in their home town. On this point, there seems to be a similarity between the concerns arising in this case and those that are often identified by people displaced by international disputes who are not willing to return home.

In general, people's concerns are represented through the mass media. In other words, the mass media can function as a spokesperson for the public. In the case of the Great East Japan Earthquake and the FDNPS accident, the many reports made by the mass media provide enough information that we can understand the overall situation. In this context, examining reports by the mass media concerning the issue of former residents of Hirono Town returning to their homes can be thought of as an analysis of stakeholders' concerns. However, it is also true that there may exist some intervening factors such as subjective interpretations of social conditions by the mass media. Therefore, it is of great significance to verify whether reports by the mass media correctly reflect people's concerns or not.

2. Study Objective

The objective of this study is twofold: first, to verify whether reports by the mass media correctly reflect people's concerns concerning the Great East Japan Earthquake and the nuclear power station accident in Fukushima. As mentioned above, we cannot rule out the possibility of subjective interpretations by the mass media. There is no guarantee that the content of media reports coincides with people's concerns. Thus, I used a text mining methodology to analyze the concordance between them. Second, to confirm the underlying structure of the issue concerning evacuees who are not willing to return. If reports by the

mass media correctly represent people’s concerns, we can use those reports as data sources for more detailed research. In this study, I conducted further text mining analysis using *word2vec* to clarify the relationships between key elements concerning residents’ return and to confirm bottlenecks that are delaying the return of evacuees to their home town.

3. Methodology

Word2vec is a tool that applies the methodology of neural networks to text mining; it was proposed by Tomas Mikolov and his colleagues at Google in 2013 (Mikolov et al., 2013a, 2013b, 2013c). More specifically, it adopts the skip-gram model as shown in Figure 1 (Mikolov et al., 2013c). In short, it optimizes the conditional appearance probabilities of nearby words (i.e., $W(t-2)$, $W(t-1)$, $W(t+1)$, and $W(t+2)$ in the case of Figure 1) for an input word $W(t)$. A full explanation of the algorithm adopted in *word2vec* is detailed in Goldberg and Levy (2014).

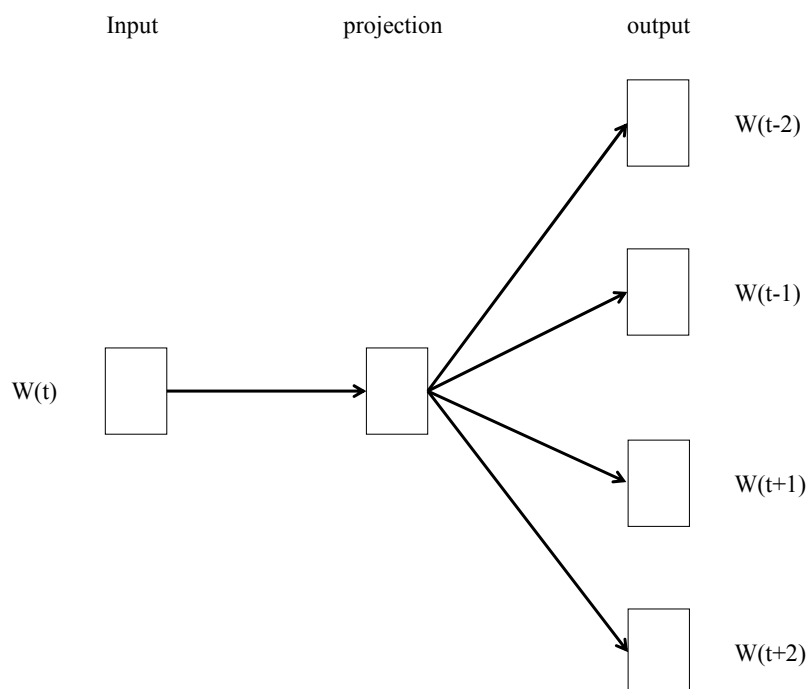


Figure 1. The skip-gram model architecture

Source: Mikolov et al. (2013c)

One of the most characteristic features of *word2vec* is that every word is represented as a vector in higher dimensions such that all vectors are obtained by performing the optimization stated above. This enables the addition and/or subtraction of vectors which, in turn, enables operations involving semantic representations. As an example, Mikolov et al. (2013c) show: “ $\text{vec}(\text{“Russia”}) + \text{vec}(\text{“river”})$ is close to $\text{vec}(\text{“Volga River”})$, and $\text{vec}(\text{“Germany”}) + \text{vec}(\text{“capital”})$ is close to $\text{vec}(\text{“Berlin”})$.” I assumed that this kind of calculation was applicable to text concerning the Great East Japan Earthquake and the FDNPS accident and used this methodology to examine the case of Hirono Town in this study.

In addition to *word2vec*, I used the technique of *t-SNE* (Stochastic Neighbor Embedding) to visualize the high-dimensional vectors representing respective words derived from *word2vec* on a two-dimensional map. This technique was first proposed by van der Maaten and Hinton in 2008, followed by improvement of acceleration (van der Maaten, 2014). One of the most characteristic features of this technique is that Euclidean distances between high-dimensional points (vectors) correspond to conditional probabilities newly defined with a Gaussian function. Thus, *t-SNE* enables an efficient reduction in dimensionality and clear visualization. This methodology follows previous studies concerning word embedding using *word2vec* and vector visualization with *t-SNE* such as Li et al. (2016) and Heuer (2015).

In this study, I used Japanese text data from *Fukushima Minpo*, the Fukushima local newspaper, as input data for *word2vec*. The target period of the analysis was set from November 2015 to October 2016. To prepare proper text data, I retrieved relevant articles using three keywords, namely 広野町 (Hirono Town), 帰還 (Return), and 復興 (Revitalization). The total number of Japanese characters in these articles reached nearly 65,000. I conducted the analysis with the *wordVectors* and *tsne* packages for R under a Windows 7 environment. The number of dimensions and the context window for *word2vec* were set at 300 and 30, respectively.

4. Results and Discussion

4.1 Accuracy of Media Reports

First, I verified whether reports by *Fukushima Minpo* correctly reflected people's concerns. The result of the analysis using *word2vec* showed that there were close relationships between some symbolic words. 帰還 (Return) is a very straightforward word concerning evacuees returning to their home town, and was shown to have close relationships with the words in Table 1.

Table 1. The words that have a close relationship with 帰還 (Return), 住民 (Residents), and 促進 (Promotion).

[帰還 (Return)]	[住民 (Residents)]	[促進 (Promotion)]
住民 (Residents)	帰還 (Return)	帰還 (Return)
促進 (Promotion)	充実 (Fulfillment)	再建 (Reconstruction)
生活 (Livelihood)	促進 (Promotion)	後押し (Endorsement)
基盤 (Foundation)	生活 (Livelihood)	場合 (Case)
再建 (Reconstruction)	進む (Proceed)	生活 (Livelihood)
充実 (Fulfillment)	後押し (Endorsement)	連携 (Cooperation)
状況 (Condition)	再建 (Reconstruction)	雇用 (Employment)
困難 (Difficulty)	基盤 (Foundation)	住民 (Residents)
首長 (Mayor)	状況 (Condition)	狙い (Aim)

For example, we can see that 帰還 (Return) has close relationships with 住民 (Residents) and 促進 (Promotion). Presumably, this means that there were not a few reports by the media concerning the promotion to residents of their returning home. Similarly, the appearance of 生活 (Livelihood), 基盤 (Foundation), and 再建 (Reconstruction) may imply that reconstruction of the foundation of residents' livelihoods is essential to their return. It should be noted that 帰還 (Return) also has a close relationship with 困難 (Difficulty). This may show that the difficulty residents may experience with their return is an issue taken up by the media. Both 住民 (Residents) and 促進 (Promotion) have close relationships with 帰還 (Return), therefore I also confirmed words with close relationships to these words, as shown in Table 1.

As we can see from Table 1, common words such as 生活 (Livelihood) and 再建 (Reconstruction) appear. This is quite reasonable because the three words, 帰還 (Return), 住民 (Residents), and 促進 (Promotion), are located closely in the vector space and, thus, the nearby words are picked up at the same time. It is interesting that 促進 (Promotion) has close relationships with 連携 (Cooperation) and 雇用 (Employment) that may be associated with concrete measures for residents' return.

In summary, many of the words that have a close relationship with 帰還 (Return) are in accord with the key elements identified in the surveys by Hirono Town (2013, 2014) concerning the issue of residents' return and in residents' concerns represented in the international forum held in Hirono Town in September 2015 (Mikiyasu Nakayama, personal communication, December 25, 2016). Therefore, it can be said that the accuracy of media reports regarding residents' concerns has been generally confirmed. However, I should point out that a close relationship between 帰還 (Return) and 原発 (Nuclear Power Station) was not observed. This may imply that the mass media dissociated the FDNPS from the issue of residents' return because of an assumption that the nuclear power station and related matters were understood to be major obstacles to residents' return. In this regard, attention should be paid to interpretation of further results of the analysis.

4.2 Vector Calculation

In the first part of my analysis, presented in the previous section, I found that 原発 (Nuclear Power Station) did not seem to have a close relationship with 帰還 (Return). Thus, I calculated the sum of the two words, namely 帰還 (Return) and 原発 (Nuclear Power Station), followed by retrieving the words that had a close relationship with the sum. The result is shown in Table 2. From this, I confirmed that the sum of 帰還 (Return) and 原発 (Nuclear Power Station) is located near 帰還 (Return) and/or 原発 (Nuclear Power Station). Furthermore, some unique words related to 原発 (Nuclear Power Station) such as 事故 (Accident) and 東京電力 (TEPCO) appeared as words close to the sum of 帰還 (Return) and 原発 (Nuclear Power Station). This may imply that the effect of the word 原発 (Nuclear Power Station) is much bigger.

Similarly, I calculated the sum of the three words 帰還 (Return), 原発 (Nuclear Power Station), and 課題 (Issue). This result is also shown in Table 2, and shows that the above unique words and 原発 (Nuclear Power Station) itself have disappeared. That is to say, the

sum of 帰還 (Return), 原発 (Nuclear Power Station), and 課題 (Issue) seems to be located far from 原発 (Nuclear Power Station). This may be interpreted as follows: when the reports by *Fukushima Minpo* took up the issue of residents' return, matters related to 原発 (Nuclear Power Station) were rarely mentioned. Incidentally, this interpretation is consistent with the result of the previous section.

Table 2. The words that have a close relationship with the sum of the two words 帰還 (Return) and 原発 (Nuclear Power Station), and the sum of the three words 帰還 (Return), 原発 (Nuclear Power Station), and 課題 (Issue).

[帰還 (Return)+原発 (Nuclear Power Station)]	[帰還 (Return)+原発 (Nuclear Power Station)+課題 (Issue)]
原発 (Nuclear Power Station)	課題 (Issue)
帰還 (Return)	帰還 (Return)
避難 (Evacuation)	協議 (Discussion)
事故 (Accident)	避難 (Evacuation)
加速 (Acceleration)	加速 (Acceleration)
考え (Thought)	示し (Indicate)
東京電力 (TEPCO)	取り組み (Activity)
協議 (Discussion)	当たり (With)
対応 (Correspondence)	向け (Toward)
伴い (Due to)	策 (Measure)

4.3 Visualization of Vector Space

I visualized the high-dimensional vectors representing respective words derived from *word2vec* in a two-dimensional map with the technique of *t-SNE*. As stated above, the algorithm converts Euclidean distances between high-dimensional vectors to conditional probabilities that were used to plot the vectors on a two-dimensional map.

The result is shown in Figure 2. This figure shows that there are several relevant words described in the previous section such as 住民 (Residents) and 生活 (Livelihood) that are found to be close to 帰還 (Return). Similarly, 再建 (Reconstruction) and 課題 (Issue) are also plotted close to 帰還 (Return). Based on these facts, we can say that in the media reports the reconstruction of the livelihoods of residents and/or the livelihoods themselves are regarded as the biggest issue connected with residents' return. Conversely, the words relevant to 原発 (Nuclear Power Station), such as 事故 (Accident), 東京電力 (TEPCO), and 原発 (Nuclear Power Station) itself, are plotted far from 帰還 (Return) and its related words. The same interpretation may also hold true here: that is to say, the media reports rarely took up matters related to 原発 (Nuclear Power Station) as an issue of residents' return.

A two dimensional reduction of the vector space model using t-SNE

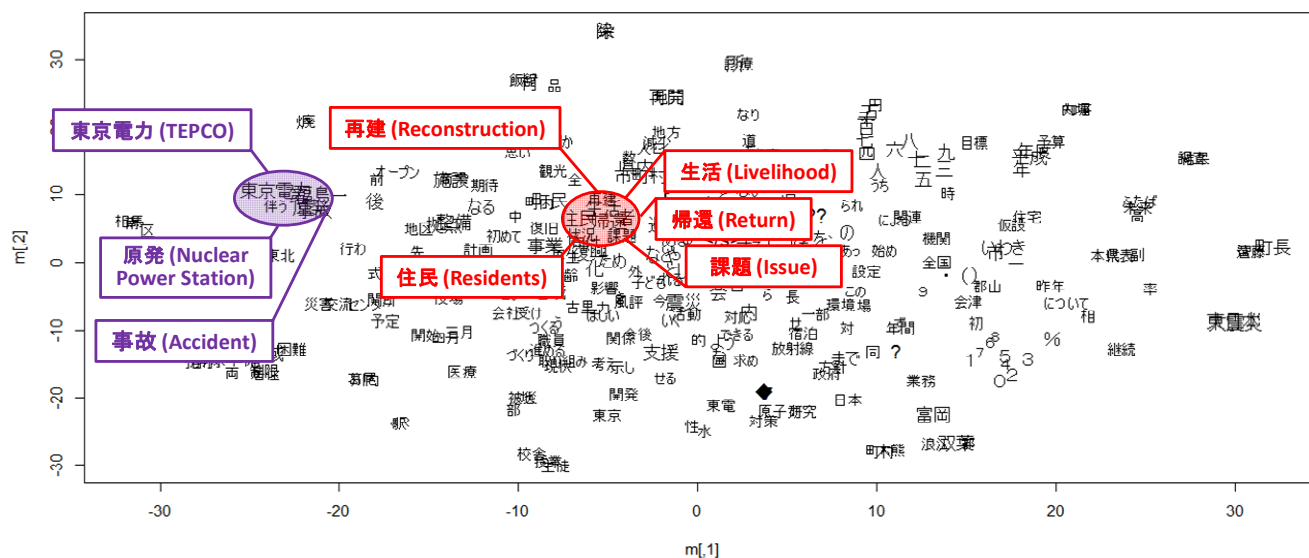


Figure 2. The visualization of vector space

5. Conclusion

Because of my analysis, I have confirmed the general accuracy of media reports on the return of evacuees to Hirono Town with respect to the evacuees' concerns. Also, I have found that the reconstruction of the residents' livelihoods and/or the livelihoods themselves seemed to be regarded in the media reports as the biggest issue concerning residents' return, whereas the media reports rarely included matters related to the FDNPS as an issue associated with residents' return. Therefore, the most effective measure for having evacuees willingly return to their homes may be to remove their anxieties about daily life including the employment environment with high priority so that evacuees can return to their homes willingly.

There seem to be two key next steps that follow from this study: the first is to expand the target area of analysis to a wider range. In this study, I used the media reports related to Hirono Town as a model case for the larger area suffering from the issue of residents not returning to their previous homes. It is important to confirm whether similar results can be obtained for other areas. The second is to compare the structure of text data of media reports through time. The target period of this study was set at one year. However, if multi-year data sets are prepared, it is possible to analyze data of this type as time series. With such analyses, we could identify and assess trends in media reports concerning the issue of residents' return.

Acknowledgments

This study was funded by the Mitsui & Co., Ltd, Environmental Fund, JSPS KAKENHI Grant Number JP15H02864, and the Graduate School of Frontier Sciences at the University of Tokyo. Support was also provided by the Reconstruction Agency of the Japanese Government through Hirono Town in the Fukushima Prefecture.

References

- Fukushima Prefecture. (2016). Transition of evacuation instruction zones. Retrieved from <http://www.pref.fukushima.lg.jp/site/portal-english/en03-08.html> (January 1, 2017).
- Goldberg, Y., & Levy, O. (2014). Word2vec explained: Deriving Mikolov et al.'s negative-sampling word-embedding method. *arXiv:1402.3722 [cs, stat]*. *arXiv: 1402.3722*.
- Heuer, H. (2015). Text comparison using word vector representations and dimensionality reduction. *Proceedings of the 8th European Conference on Python in Science (EuroSciPy 2015)*, 13-16.
- Hirono Town. (2013). *The results of the survey on townpeople's intentions for developing the Second Hirono Revitalization Plan*. Fukushima, Japan: Hirono Town. (Japanese).
- Hirono Town. (2014). *The report on the survey on the development of Hirono Town by Hirono Town Development Board*. Fukushima, Japan: Hirono Town. (Japanese).
- Hirono Town. (2016). *Census registration information of Hirono Town*. Retrieved from <http://www.town.hirono.fukushima.jp/chominhoken/kosekinomado.html> (January 1, 2017) (Japanese).
- Li, Y., Zheng, R., Tian, T., Hu, Z., Iyer, R., & Sycara, K. (2016). Joint embedding of hierarchical categories and entities for Concept Categorization and Dataless Classification. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2678–2688.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. ICLR Proceeding of the *International Conference on Learning Representations Workshop Track*. arXiv:1301.3781 [cs.CL]
- Mikolov, T., Yih, W., & Zweig, G. (2013b). Linguistic regularities in continuous space word representations. *Proceedings of North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 746–751.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013c). Distributed representations of words and phrases and their compositionality. *Proceedings of Advances in Neural Information Processing Systems*, 3111–3119.
- Van der Maaten, L., & Hinton, G. (2008). Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.
- Van der Maaten, L. (2014). Accelerating t-SNE using tree-based algorithms. *Journal of Machine Learning Research*, 15, 3221–3245.

Copyright Disclaimer

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).