



High Sensitivity Layer Feature Analysis in Food Market

Yoshio Matsuyama^(✉) and Yumi Asahi

School of Information and Telecommunication Engineering,
Department of Management System Engineering,
Tokai University, Shibuya, Japan

6bjm2110@mail.u-tokai.ac.jp, asahi@tsc.u-tokai.ac.jp

Abstract. It is not uncommon to conduct test marketing for the purpose of market research when dropping new products to the market. However, if you actually drop it you will need a lot of money. This study, we pay attention to innovation theory. In Japan, the study reported a long sales period. However, the study didn't report a short sales period. Therefore, we report of a short sales period, especially food. This study, we call "High Sensitivity Layer" the innovators and the early adopters in innovation theory in term of to be interested in the innovation of products, sensitive to trends and constantly collecting new information by themselves and to have greater influence on other consumers. We think that those that collect a lot of empathy in the "High Sensitivity Layer" are diffusive in the innovators and the early adopters, and grab the characteristics of highly sensitive consumers who gather many empathies. I think that it may be able to fulfill the purpose of test marketing by seeing the response of new products of food to this consumer. We prepare a generalized model with a deep learning model and report features of highly sensitive consumers, visually and numerically clearly, using decision tree analysis from that model. From the analysis results, attached more images, and the older, the better it got a report that empathizes with sensitive consumers. When conducting test marketing, it is predicted that high-sensitivity consumers will be able to obtain preferable results by targeting people with this characteristic. Also, it was found that gender and emotion are not related to the characteristics of the person who writes the report sympathized with the consumer. In the future, I would like to further accurate classification by text mining of posted characters and analysis of posted images.

Keywords: Innovation theory · Test marketing · High sensitivity layer · Deep learning · Decision tree analysis · Foods marketing

1 Introduction

In this study, we analyze characteristics of highly sensitive consumers who write reports diffusive in the food market. You will be wondering why this analysis was done. We focused on innovation theory. In the first place, innovation theory was proposed by Professor Rogers in 1962. This theory has been successful in a wide range of fields including communication, agriculture, public health, criminal justice, and marketing. In Japan, Morio (2012) has reported that "the appropriateness of companies

in introducing new kinds of unique vegetables” is researched. In this study, long sales periods are assumed. Therefore, we target foods that can develop short ones. From innovation theory, if it can be disseminated to Innovators and Early Adopters, it is known that leading to Early Majority and Late Majority (Fig. 1), the market share will rise significantly (Fig. 2).

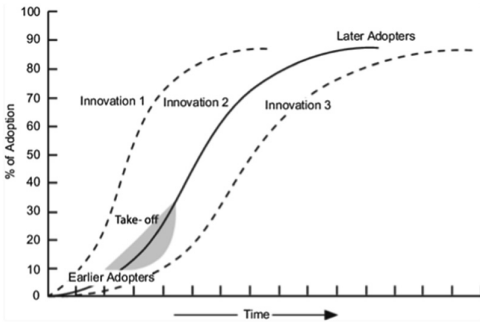


Fig. 1. Diffusion process

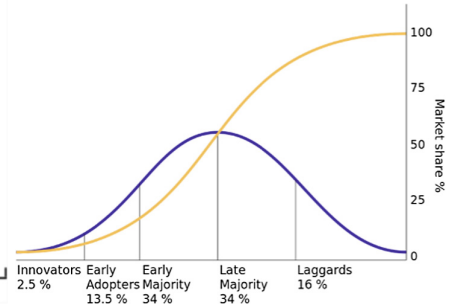


Fig. 2. Category classification

The characteristics of the Innovator and Early Adopter are that they are interested in product innovation and are sensitive to trends and constantly collect new information themselves and have a great influence on other consumers. In this study, these two layers are called “High Sensitivity Layer”. We think that “High Sensitivity Layer” with keeping them eyes open, and products that them sympathize will diffusion from the innovation theory to the Early and Late Majority and the market share is expanded to expand. We think that those that collect a lot of empathy in the “High Sensitivity Layer” are diffusive in the innovators and the early adopters, and grab the characteristics of highly sensitive consumers who gather many empathies. I think that it may be able to fulfill the purpose of test marketing by seeing the response of new products of food to this consumer. In this study, we report characteristics of highly sensitive consumers.

2 Data Summary

In this study, we used “Minrepo Data” from INTAGE Inc. distributed by IDR Dataset Service of National Institute of Informatics. We analyzed the report posted on the living post & enterprise co-creation SNS (Social Networking Service) application “Minrepo”. In this analysis, the data shown in Table 1 was mainly used among the data included in the report. In this study, 6,367 reports with genre “Foods report” and type “Bought or Received” were targeted. The used data period is from January 1, 2016 to June 30, 2016. The Minrepo official website says (2018), “High quality word of mouth contributions gather, which expresses products more attractive with high sensitivity consumers from photos and comments”, “real life actuality data”. That is, the Minrepo is an SNS where realistic living conditions data of high sensitivity consumers gather. The

Table 1. Mainly used variables

Variables name	Data type1	Data type2	Details
Image file1	Text	FA	Name of image file
Image file2	Text	FA	Name of image file
Image file3	Text	FA	Name of image file
Genre	Text	SA	Kinds of report (Objects report, Places report, Foods report, Note report)
Sub Genre	Text	SA	Detail of genre
Foods report type	Text	SA	Detail of Foods report. Set only Foods report. (Eaten out, Cooked, Bought or Received)
Feelings	Text	SA	Report Feelings (Excitement, Pleasure, Favorite, Puzzle, Reglet, Angry)
Gender	Text	SA	User set up
Age	Integer	SA	User set up
Number of favorite	Integer	–	Represent empathy to the report
Number of wish	Integer	–	Number entered in other user's wish list
Number of activity	Integer	–	Number tapped by people who did similar actions

high sensitivity sensitive consumer's post "Reports" of "Minrepo users" will receive responses from the same highly sensitive consumers.

The posting of this SNS resemble Tweet of the famous Twitter post, and it is posted when there is something the submitter is interested in or what he/she wishes to teach. The point different from Twitter is that there is no limit on the number of characters, and emotion at posting can be added. Figure 3 shows the number of characters for each report. From this Fig. 3, even if there is no limit on the number of characters, it is understood that about 50% of people contribute with the number of characters of 100 characters or less. Of the remaining people, about 45% of the people have posted within 300 characters, and as a whole, about 95% of people can confirm that they post within 300 characters. Figure 4 shows the gender of each report. From this Fig. 4, it is found that posting about food is about 20%, more women than men. Figure 5 shows the age group for each posted report. While SNS such as Twitter and Instagram has many young people, this Minrepo accounts for about 75% at the age of 35 to 50, and about 15% at 15 to 30 years old shows that there are few young people. Figure 6 shows the percentage of emotions when posting a report. It is understood that the report which posted about food just like a soliloquy without emotion accounted for about a quarter. Looking at the proportion of the remaining emotions, we can see that positive emotions account for about 70% of the total. Conversely, only 3% of the negative emotions are present, and when you come up with something happy, you can confirm that the report is very often posted.

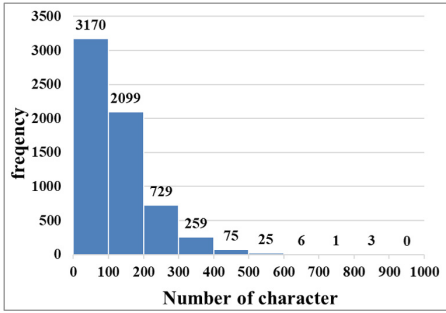


Fig. 3. Number of characters

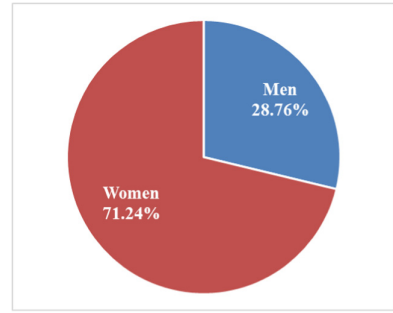


Fig. 4. Gender proportion

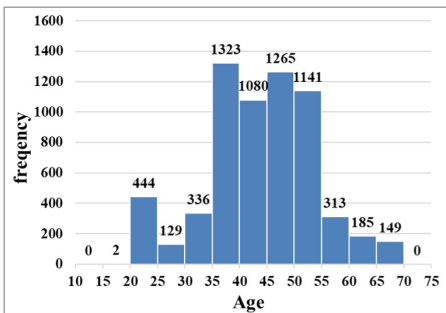


Fig. 5. Age distribution

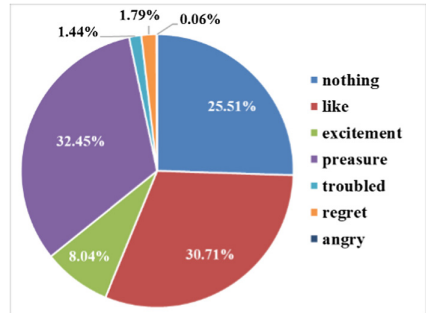


Fig. 6. Feelings proportion

The data used in this study included image files, but we created a variable (Number of images) simply counting the number of attached images. Figure 7 shows the ratio of the number of attached images in all reports. There are almost no reports with 0 attached images, and only one report accounts for about 55%. About 22% each report attached 2, 3 images, it turned out that many people attached multiple images.

On the popularity tab of Minrepo, reports posted within 3 days are sorted in descending order of “favorites” and introduced. Reports that are loaded at one time occupy more than 10 favorites. Figure 8 shows the distribution of Number of favorites in the overall report. It is worth noting here that the number of favorites shown in Fig. 8 is the total value since posting. The size of that number has no meaning. It is important that Number of favorites exceeds 10 within 3 days. In this study, we define “empathy report” for favorite number of 10 or more, and “non-empathetic report” for less than 10 favorites.

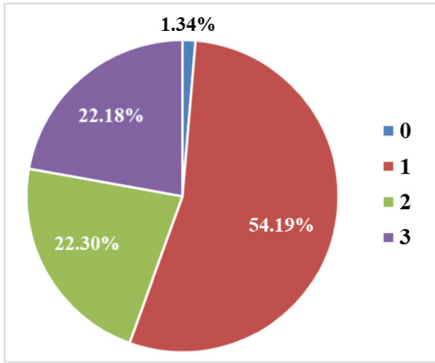


Fig. 7. Proportion of the number of attached images

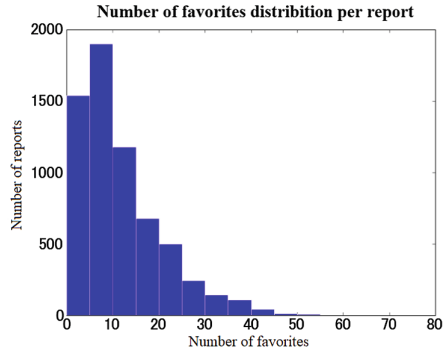


Fig. 8. Gender proportion

3 Build a Deep Learning Model

3.1 Outline of Deep Learning Model

For the purpose of creating a generalized model, we prepare a deep learning model using the data in Sect. 2. Deep learning is “a type of parametric function approximation”. The special thing compared to others is that it has the property that “arbitrary functions can be approximated with arbitrary precision by using a large parameter set”. This property is a property not found in the linear model. Learning is conducted with the “empathy report” and “non-empathy report” defined in Sect. 2 as objective variables. The deep learning model built in this study is a classified model of binary categories in supervised learning. From the data we handle, we use a network structure called Multi-layer perceptron (Fig. 9) to construct a deep learning model. The perceptron was developed by Rosenblatt (1958). For the explanatory variable, select “number of images”, “emotion”, “gender”, “age”.

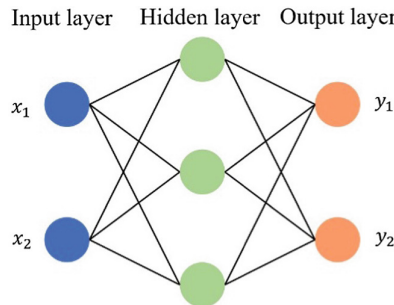


Fig. 9. Architecture of multi-layer perceptron

The purpose of this research is to predict the probability (1) of which category from the binary classification by logistic regression.

$$P(\hat{y}^{(i)} = \{0, 1\} | x^{(i)}) = \mathbf{softmax}(Wx^{(i)} + b) \# \tag{1}$$

At this time, $x^{(i)}$ is the input, the parenthesized superscript is the data number, $\hat{y}^{(i)}$ is the output. Also, the coupling weight W is a (C, D) matrix and the bias term b is a C dimensional vector (C : number of categories, D : dimension of each data point). The accuracy of the deep learning model is improved by adjusting the parameter $\{W, b\}$ to be close to supervised data. The softmax function is a function that returns d probability values with the d -dimensional vector as an argument, and is represented by (2).

$$\mathbf{softmax}(v)_i = \frac{\exp(v_i)}{\sum_{i=1}^d \exp(v_i)} \# \tag{2}$$

The probability value thus obtained and the observation value of the actual category $\{c^{(i)}\}_{i=1}^{N_{data}}$ are measured by the smallness of the cross entropy loss function (3), and adjusts parameters. We aim to minimize loss function by stochastic gradient descent method.

$$L = - \sum_{i=1}^{N_{data}} \log(P(\hat{y}^{(i)} = \{0, 1\} | x^{(i)})) \# \tag{3}$$

In the logistic regression described above, there was only one pair of coupling weight W and bias term b for performing linear transformation as a parameter. In a multilayer perceptron with k hidden layers, there are $(k + 1)$ combining weights and biases, and they convert the input $h^0 = x$ to (4).

$$\begin{aligned} h^1 &= \text{activate}_1(W^1 h^0 + b^1) (h^0 = x) \\ h^2 &= \text{activate}_2(W^2 h^1 + b^2) \\ &\vdots \\ h^k &= \text{activate}_k(W^k h^{k-1} + b^k) \\ h &= Wh^k + b \end{aligned} \tag{4}$$

h^0 is the input layer, h^k is the hidden layer k , h is the output layer.

There are “wish count” and “number of activities” as responses to the report, but these two are attached to the “favorite” report and are not explanatory variables. One of the things to keep in mind when asking explanatory variables is the problem of multiple collinearity. This is a problem especially in multiple regression analysis. However, according to Yoshida (1987), multicollinearity is said to be an essential problem that the correlation matrix is not regular when there is a correlation between explanatory variables, so that the solution cannot be obtained or is unstable. In this study, regularization by Dropout is carried out, prevention of over learning and prevention of multiple collinearity are also done Srivastava et al. (2014) has shown the prevention of over-learning by regularization..

As a precaution, look at the relationship between explanatory variables here. From the scale of explanatory variable (Table 2), use Spearman’s rank correlation coefficient for ordinal scale and ratio scale, correlation ratio on nominal scale, ordinal scale nominal scale, ratio scale, and Cramer’s coefficient of association for nominal scale. The results are shown in Table 3.

Table 2. Scale chart

Explanatory variable	Scale
Number of images	Ordinal scale
Feelings	Nominal scale
Gender	Nominal scale
Age	Ratio scale

Table 3. Relationship between explanatory variables

variable 1	variable 2	Rank correlation coefficient
images	age	0.15

variable 1	variable 2	Correlation ratio
feelings	images	0.04
feelings	age	0.06
gender	images	0.00
gender	age	0.13

variable 1	variable 2	Coefficient of association
feelings	gender	0.07

3.2 Hyper Parameter Setting

In order to create a deep learning model, it is necessary to set a large number of hyperparameters. First of all, for learning of this model, 80% (5,094 cases) of the total 6,367 reports were used for learning and the remaining 20% (1,237 cases) were used for verification.

Next, we set the number of hidden layers in a multi-layer perceptron. According to Asakawa (2014), if there is only one hidden layer, we can solve only linearly identifiable problems, so we know that at least two or more hidden layers are needed. Also, according to Nielsen (2014), as the number of hidden layers increases, the stochastic gradient descent method is said to cause gradient disappearance problem/instability of gradient, so increase the number of hidden layers too much. For these reasons, we set two hidden layers in this model.

In each hidden layer, it is necessary to set three hyperparameters of “output dimension number”, “activation function”, and “dropout ratio”. In this model, three parameters are set by Bayesian optimization for Accuracy in confusion matrix.

Finally, set up learning. In the multi-layer perceptron, it is necessary to set parameters for proceeding learning of “optimization function”, “learning rate”, “epochs” and “mini-batch size”.

We compare accuracy with representative “Adam”, “SimpleSGD”, “AdaDelta”, “AdaGrad”, “RMSprop”. This is because the optimum optimization function differs depending on the data to be used. The optimization function selects an optimization function with a high rate of accuracy (Needless to say, the other parameters are the same). From Table 4, the optimization function of this model sets “AdaDelta”.

According to Taguchi (1998), if the learning rate is increased, the convergence of the error becomes faster, but it becomes unstable. When it is decreased, the stability of error increases but it is said that it takes time to converge. However, in this model, since “AdaDelta” was set as the optimization function, it became unnecessary to set the initial value learning rate. The details are described in detail by developer Zeiler (2012). In fact, even if the learning rate was greatly changed and compared, the change in the correct answer rate was not recognized (Table 5). In order to do learning it is necessary to set the learning rate, so the default value of 0.0001 is set in this model.

Regarding the mini-batch size, it is preferable to divide it into 2^n for memory access efficiency. The smaller the size is, the higher the accuracy rate increases, but the overfitting becomes more likely to occur. In this model, we change the value from 128 to 64, 64 to 32 with the default value of 64 as a reference, and set 64 which was high effect (Table 6).

The epochs is the number that learns one training data repeatedly. In deep learning, parameters are learned by learning training data repeatedly, so it is necessary to take many epochs to the extent that over learning does not occur. Santos (2007) reports on the error caused by taking too many epoch numbers. Therefore, we set epochs using Early stopping method. Early stopping is a method proposed by Prechelt (1998). Early Stopping is to stop learning when errors begin to converge. This makes it possible to stop learning so as not to overfitting. In this model, we set 100 which the error began converging to the epochs.

In this model, we made the above setting and finally got a model that gains 73.3% accuracy.

Table 4. Accuracy for each optimization function

Optimization function	Adam	Simple SGD	AdaDelta	AdaGrad	RMSprop
Accuracy	0.72536	0.68	0.73263	0.64	0.72

Table 5. Accuracy for each learning rate

Learning rate	0.001	0.0001	0.00001
Accuracy	0.73263	0.73263	0.73263

Table 6. Accuracy for each mini-batch size

Mini-batch size	32	64	128
Accuracy	0.73380	0.73263	0.72556

4 Decision Tree Analysis

In Sect. 3, a deep learning model was created. In this section, we report features of highly sensitive consumers, visually and numerically clearly, using decision tree analysis from that model. A positive feature of decision tree analysis is that it is easy to

interpret the output result, it is possible to handle various scales for explanatory variable/objective variable. The primary reason for using decision tree analysis is that the classification process can be visualized in an easy-to-understand manner. Of course, there are not only good things but also negative features. The analysis is likely to over-adapt and the generalization performance may be lowered. Overfitting refers to excessive learning of the observed data, which also sensitively reacts to noise, which means that the fitness to unknown data worsens. As mentioned earlier, since it is top priority to visualize the classification process in an easy-to-understand manner, we will make efforts to minimize excessive adaptation. The reason for creating a deep learning model is also to make observation data closer to true category and to improve generalization performance.

The explanatory variables used in this model have values as shown in Table 7. Apply this model created as a generalized model, predict which “empathy report” or “Non-empathetic report” will be written and analyze the binary features predicted by decision tree analysis. Table 8 shows the settings used for the decision tree analysis. After creating the decision tree model, we pruned for the purpose of improving generalization performance.

Table 7. Value to be taken by explanatory variable

Images	Feelings	Gender	Age
0	Nothing	Men	16
1	Like	Women	17
2	Excitement		•
3	Preasure		•
	Troubled		•
	Regret		68
	Angry	69	

Table 8. Setting of decision tree analysis.

Explanatory variable	Max buranch	Parameter	Contents
Number of images	5	Objective variable	Empathy or non-empathetic
Feelings	Unlimited	Branching method	Gini coefficient
Gender	Unlimited	Min number of data	0.80%
Age	5	Max branch (common)	5
		Node impurity	0.01
		Height limit	Unlimited
		Weighting	Nothing

The Gini coefficient of the branching method is defined by (5).

$$\text{Gini coefficient} = 1 - \sum_{i=1}^c p_i^2 \# \tag{5}$$

Figure 10 shows the results of the decision tree analysis. In this model, it is important to see the state of branching and the proportion of each class. For example, when the number of attached images is 1 or less, it is found that the ratio of Non-empathetic report is 92.7% and if the number of attached images is 1 or less, it is almost written a Non-empathetic report. In the other hand, when the number of attached

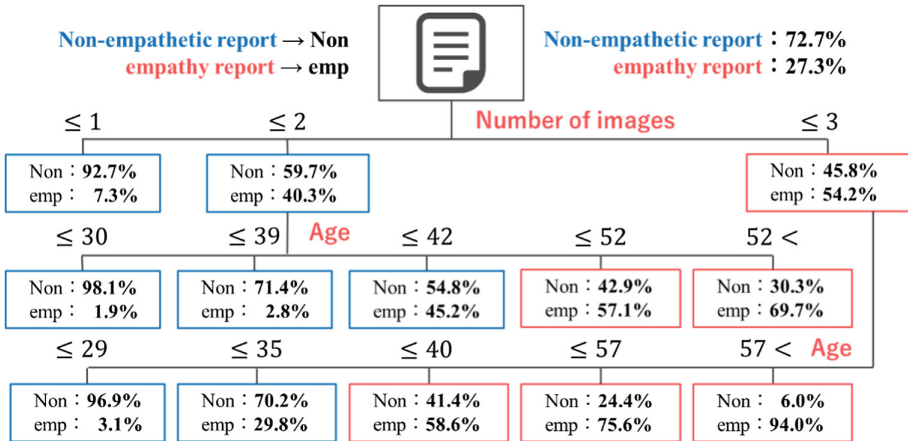


Fig. 10. Result of decision tree analysis

images is 3, and when the age exceeds 57 years old, you can find that you write almost empathy report.

5 Discussion

Based on the results of the decision tree analysis, it is found that the report is more likely to be empathized as the number of images is larger. Reports that only 0 and 1 images are attached are classified as non-empathetic reports by 92.7%. Therefore, 2 or more images are necessary as the first condition of the empathy report. It is understood that visual information has more information amount than simple text data, and it is easy to acquire empathy.

Next, in the group with a large proportion of “empathy report”, the report attached with 2 images shows that the age is 43 years old or more, and the report attached with 3 images shows that the age is 36 years old or over. Under the condition that there are two or more attached images, it turns out that people who are years old can acquire empathy more easily. Also, from the result of the decision tree analysis, it is understood that the more the age, the more the proportion of the empathy report is increasing.

Conversely, 97% or more is classified as non-empathetic report at the age of 30 or under even if the number of images is 2 or 3. If the number of images is 2 and 39 years old, if the number of images is 3, 70% or more are classified as non-empathy report

even under 35 years old. Therefore, writing an easy-to-empathy report is understood from the “Arafor” (Arafor: Referring to people, aged around 40).

In addition, the explanatory variable contains emotion and gender in addition to the number of images and age. Emotions and gender were not recognized as such important conditions from the model created in this study.

When developing a new product of food, it is necessary to care about the sensitive layer from innovation theory. From the results of this study, we believe that we can withstand the role of test marketing (although verification is necessary) by attaching a lot of images in the high sensitivity layer and seeing the reactions for people who are years old.

6 Future Work

In this time, we have constructed a deep learning model that predicts binary values of “non-empathetic report” and “empathy report”, but its explanatory variables are limited to those that can be numerically interpreted (including the nominal scale). It is conceivable that accuracy will be further improved if we build a deep learning model from text mining as to what kind of words are easily sympathized to people. Also, it is also interesting to search empathize words. Since I searched whether reports are easy to sympathize for each report this time, I think that interesting knowledge can be obtained by searching by associating users.

In addition, this time data included image files. Since the importance of the image is indicated also from the result of the decision tree analysis, I want to verify what type of image is easily sympathized by carrying out image tagging with AI or the like and performing text mining.

Finally, we decided to use the decision tree analysis with the highest priority for ease of interpretation. There is a logistic regression analysis as a binary classification model, so I would like to verify this as well.

References

- Rogers, E.M.: *Diffusion of Innovations*, 3rd edn. Macmillan Publishers, London (1962)
- Morio, A.: Suitability of companies when introducing distinctive new vegetables varieties—a case study of new varieties of sweet potato. *Bull. Naro Agric. Res. Center* (17), 39–48 (2012)
- INTAGE Inc. Minrepo Campaign. <https://www.intage.co.jp/solution/process/product-development/minrepopc>. Accessed 20 Oct 2018
- INTAGE Inc. Minrepo SNS analysis. <https://www.intage.co.jp/solution/process/market/minreposns>. Accessed 20 Oct 2018
- Rosenblatt, F.: The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**(6), 386–408 (1958)
- Yoshida, M.: Multicollinearity and ridge regression in regression analysis. *Fac. Hum. Sci. Bull.* **13**, 227–242 (1987)
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014)

- Asakawa, S.: Deep learning and the meanings of hidden layer. In: The 12th Annual Convention of the Japanese Society for Cognitive Psychology Bibliography, p. 28 (2014)
- Nielsen, M.A.: Neural Networks and Deep Learning. Determination Press (2014)
- Taguchi, I.: The convergence condition and learning acceleration in the backpropagation neural network. Keiai Univ. Int. Res. (2), 77–109 (1998)
- Zeiler, M.D.: Ada delta-an adaptive learning rate method. Cornell University (2012)
- Prechelt, L.: Early Stopping – but when? Fakultät für Informatik; Universität Karlsruhe, pp 55–69 (1998)
- Santos, J.M.: Data classification with neural networks and entropic criteria. Universidade da Beira Interior Tese de doutoramento (2007)