Prediction of Regional Goods Demand Incorporating the Effect of Weather

Takuya Watanabe, Hiroaki Muroi, Motoki Naruke
Edirium K.K.
1-8-21 Ginza, Chuo, Tokyo, Japan
{sodium, muroi, motoki}@edirium.co.jp

Kyoto Yono
Nowcast, Inc.
2-8 Kojimachi, Chiyoda, Tokyo, Japan
yono@nowcast.co.jp

Gen Kobayashi
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki, Japan
s1330178@u.tsukuba.ac.jp

Masanori Yamasaki
Acroquest Technology Co., Ltd.
3-17-2 Shinyokohama, Kohoku, Yokohama, Kanagawa, Japan
yamasaki@acroquest.co.jp

***Abstract*—Precise prediction of the goods demand is an important element of the supply chain management because we can optimise the level of stock based on predicted demand. Demand of goods may vary influenced by numerous factors, including price elasticity and weather, which we focus in this paper. We analysed daily sales data of consumer goods collected from Point of Sale (POS) systems of Japanese retailers, mostly supermarkets, which consist of records of price and quantity sold for each item, spanning several years. Demand may change according to regional preferences, so we built prediction models for each region by estimating demand curve of each item by employing linear regression and neural networks. We show that there are regional differences of the demand itself and also regional differences of the effect of weather.**

***Keywords*-goods demand; effect of weather; linear regression; neural network;**

I. Introduction

Sales forecasting is a key component of the organisational management because it affects various functions of the organisation[1]. Manufacturing processes and supply chains are no exceptions. If a company is able to make the sales forecasting more accurate, the amount of stock accumulating in production lines, warehouses and retailers and also lead time to end users can be reduced. To forecast the sales, it is necessary to store past sales data and to make prediction models based on the stored data.

There are many studies that try to forecast fresh food sales using sales record collected from Point of Sale (POS) systems. Doganis et al. employed a specially designed radial basis function neural network method in combination with a genetic algorithm to predict daily sales of fresh milk in Athens, Greece[2]. Their algorithm is designed for time series prediction and performs better than linear time series methods. Lee et al. compared logistic regression, moving average and back-propagation neural network methods and concluded that logistic regression could perform better than moving average and back-propagation neural network methods to predict whether discarding food happens[3]. Their data are collected from a POS system of a single convenience store in Taiwan.

In this study, we estimate demand curves of items sold in supermarkets and analyse the effect of weather to the demand in order to forecast the daily sales. We also take into account regional differences of the effect of weather. For this objective, we use sales data collected from POS systems of supermarkets located in various regions in Japan. We employ both linear and nonlinear methods, namely linear regression and neural network respectively, to build prediction models from the sales data.

II. Data

We collected and examined two datasets, one is a collection of daily sales records gathered from Japanese retailers, and the other is the history of weather related metrics distributed by a governmental agency. Our goal is to estimate the price elasticity, or the demand curve more specifically, of each item, considering the effect of weather by combining these two datasets.

A. Daily sales data

In this study, we analyse daily sales data of consumer goods collected by Nikkei Digital Media, Inc. from the POS systems of Japanese retailers, mostly supermarkets. The sales data comprises several number of tables, including the sales table storing daily sales and quantity sold per each store and item. Each record of the sales table has store code, date, item code, sales and quantity sold fields, as shown in Table I. The item code is the Japanese Article Number (JAN), which is a globally unique identifier of manufacturers and their items. The sales table has approximately 3 billion records, and they are tied to the information on 1,500 stores and 3 million items stored in other tables. All the data are stored in a single instance of the PostgreSQL database, and the size of the database, including indices for several fields, is nearly 1 TB.

Table I
RECORD EXAMPLES IN THE SALES TABLE. ITEM CODES ARE ANONYMISED.

Store Code	Date	Item Code	Sales	Quantity Sold
1	2009/10/01	**********713	3048	1
1	2009/10/01	**********132	1029	1
1	2009/10/01	**********550	760	2
1	2009/10/01	**********635	872	4
1	2009/10/01	**********026	284	1
...	...	...	...	...
2	2009/10/01	**********550	1140	3
2	2009/10/01	**********635	436	2
2	2009/10/01	**********924	3048	1
2	2009/10/01	**********940	2858	1
2	2009/10/01	**********430	2858	1
...	...	...	...	...

The data contain records of more than 5 years, from 1 October 2009 to 30 November 2015, but we selectively used records spanning 4 years, from 1 November 2011 to 31 October 2015, for some reasons. We excluded records belonging to first 2 years because Japan experienced a huge earthquake in March 2011, the Tohoku earthquake, that involved a major tsunami and subsequent nuclear power plant failures. The earthquake had significant impact on economic activities in Japan[4], not only of directly damaged areas, but also of surrounding broad areas, and the nation-wide impact was observed as more than 1 point downward revision of GDP growth estimates by the Bank of Japan. Considering this immediate effect to economic activities and a possibility of structural change after the earthquake, we opted to use records starting from more than a half year after the earthquake. We also excluded records of the last month of the data because it was necessary for the mode price calculation, which we will explain later.

We limit the scope of the analysis to well sold items in order to avoid the problems coming from data sparseness. We counted the number of records for each item in the sales table, whose top amounts to about 423,000 records in total, as shown in Table II, which means that this item is sold in nearly 290 stores daily on average.

Table II
TOP 10 ITEMS IN THE SALES TABLE.

Item Code	Number of Records
4902102072618	422693
4902705104167	422558
4902102069359	412273
45019517	412230
4902102084178	411119
4901411011523	409652
49722741	407743
4901411011547	406670
4903110063278	406483
4901340689213	404653

The geographic locations of stores nearly cover entire Japan, but the locations are not randomly selected. The sales data are collected from contracted supermarket companies, and the companies and their stores are not equally distributed among the areas. In this study, we select 4 areas among 47 in Japan, *prefectures* in the Japanese official terminology. We choose them from the top 5 of the area list sorted by the number of stores they have, as shown in Table III. This is because we want to avoid problems which arise from data sparseness when we compare characteristics of the areas, as in the case of the items. They are located more than 500 km apart each other.

Table III
TOP 5 AREAS IN THE SALES TABLE IN TERMS OF THE NUMBER OF STORES.

Prefecture	Number of Stores
Kanagawa	182
Hokkaido	116
Fukuoka	99
Tokyo	90
Osaka	68

B. Daily weather data

For analysing the effect of weather, we collected daily weather data from the web site of the Japan Meteorological Agency (JMA)[5]. All Japanese prefectures have JMA's observation facilities, so we can obtain representative weather records for every area we analyse. What we can retrieve from the site are string representation of weather, atmospheric pressure, precipitation, temperature, humidity, wind speed and direction, sunlight hours, snowfall and snow depth. Some of them are sub-categorised like maximum, minimum and average temperature. Slightly different weather characteristics of each area can be seen in the records of total precipitation and maximum temperature, smoothed by 7 days moving average, in Figures 1 and 2 respectively. Each area has different peak of rainfall, and the temperature of Hokkaido is significantly lower than other three.

We can see the relationships between weather related metrics by a correlation matrix, as shown in Table IV. It is clearly seen that variables belonging to the same category have strong correlations, and this can be confirmed by seeing a multi-dimensional scaling (MDS) plot of the correlation matrix, as in Figure 3, by taking 1 - correlation coefficient as distance. This indicates that selecting one variable from each category suffices for the analysis of the effect of weather. Considering this observation and other related factors like data sparseness, we selected following variables as the explanatory variables of the analysis:

- average atmospheric pressure at observatory
- total precipitation
- maximum temperature
- minimum humidity
- maximum wind speed
- sunlight hours
- snow depth

Table IV
CORRELATION MATRIX OF WEATHER METRICS.

Pressure at Observatory	1	0.49	-0.09	-0.11	-0.11	-0.14	-0.16	-0.13	-0.23	-0.13	0.04	-0.07	-0.10	0.08	-0.06	0.00
Pressure at Sea Level	0.49	1	-0.26	-0.27	-0.26	-0.48	-0.46	-0.50	-0.32	-0.29	-0.16	-0.21	-0.24	0.15	-0.00	0.00
Total Precipitation	-0.09	-0.26	1	0.85	0.73	0.08	0.04	0.12	0.41	0.39	0.09	0.15	0.17	-0.37	0.07	-0.00
1h Max Precipitation	-0.11	-0.27	0.85	1	0.92	0.14	0.10	0.17	0.38	0.36	0.07	0.14	0.17	-0.34	0.03	-0.02
10m Max Precipitation	-0.11	-0.26	0.73	0.92	1	0.16	0.13	0.20	0.38	0.35	0.05	0.13	0.16	-0.33	0.02	-0.01
Mean Temperature	-0.14	-0.48	0.08	0.14	0.16	1	0.99	0.99	0.28	0.29	-0.08	-0.13	-0.17	0.15	-0.24	-0.32
Max Temperature	-0.16	-0.46	0.04	0.10	0.13	0.99	1	0.95	0.21	0.18	-0.10	-0.13	-0.17	0.24	-0.25	-0.33
Min Temperature	-0.13	-0.50	0.12	0.17	0.20	0.99	0.95	1	0.35	0.38	-0.07	-0.12	-0.17	0.04	-0.22	-0.30
Mean Humidity	-0.23	-0.32	0.41	0.38	0.38	0.28	0.21	0.35	1	0.90	-0.18	-0.14	-0.17	-0.56	0.10	0.03
Min Humidity	-0.13	-0.29	0.39	0.36	0.35	0.29	0.18	0.38	0.90	1	-0.11	-0.13	-0.16	-0.61	0.05	0.01
Mean Wind Speed	0.04	-0.16	0.09	0.07	0.05	-0.08	-0.10	-0.07	-0.18	-0.11	1	0.85	0.80	0.00	0.05	0.05
Max Wind Speed	-0.07	-0.21	0.15	0.14	0.13	-0.13	-0.13	-0.12	-0.14	-0.13	0.85	1	0.93	-0.01	0.09	0.08
Instant. Wind Speed	-0.10	-0.24	0.17	0.17	0.16	-0.17	-0.17	-0.17	-0.17	-0.16	0.80	0.93	1	-0.04	0.10	0.11
Sunlight Hours	0.08	0.15	-0.37	-0.34	-0.33	0.15	0.24	0.04	-0.56	-0.61	0.00	-0.01	-0.04	1	-0.12	-0.08
Snowfall	-0.06	-0.00	0.07	0.03	0.02	-0.24	-0.25	-0.22	0.10	0.05	0.05	0.09	0.10	-0.12	1	0.46
Snow Depth	0.00	0.00	-0.00	-0.02	-0.01	-0.32	-0.33	-0.30	0.03	0.01	0.05	0.08	0.11	-0.08	0.46	1

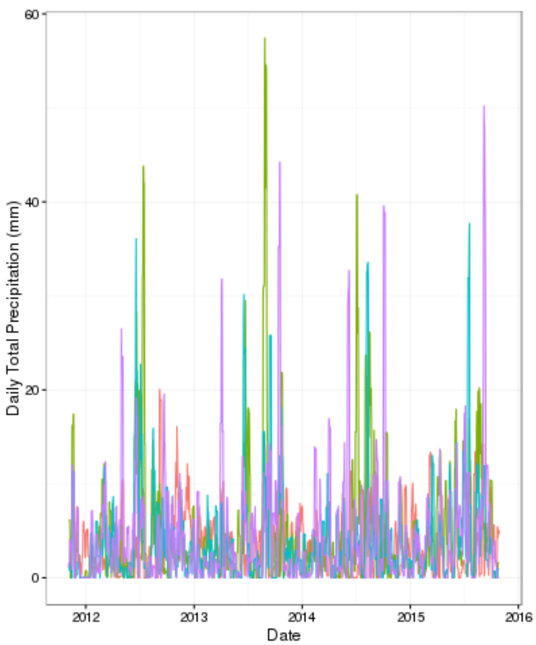


Figure 1. Record of daily total precipitation.

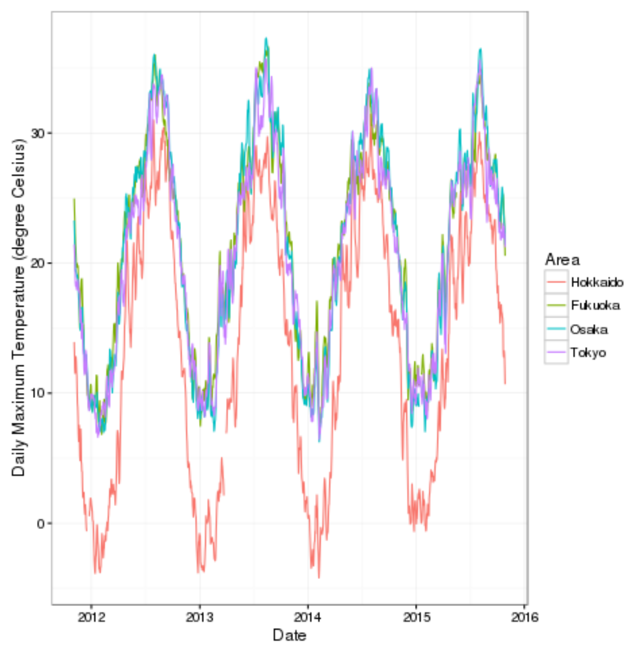


Figure 2. Record of daily maximum temperature.

III. METHODS

To estimate demand curves, we built prediction models of the quantity sold by employing two well-known methods, namely linear regression and neural network. The former is probably one of the most simple prediction methods, but it is easy to manipulate, and the resulting parameters are easily interpretable. The latter can incorporate nonlinear characteristics of the dataset into the model and are expected to fit well for such dataset that has strong nonlinearity, but the parameters hardly have interpretable meanings generally. In addition, we tested some configurations of deep neural networks to improve the prediction accuracy.

After building prediction models, we can obtain demand curves by fixing explanatory variables other than price and yielding predicted values of the quantity sold in response to price change. We adopted following variables as explanatory variables:

- quantity sold in the previous day
- day of the week
- season
- weather related variables listed in the previous section
- price
- sale flag, which we will explain in this section

A. Mode price calculation and sale flag

The sales data do not contain direct indicators of special sales where some items are sold at lower prices than usual during a limited period of consecutive days. In Figure 4, it can be seen that the price of an item is not continuously changed but split up into normal and sale prices. Many

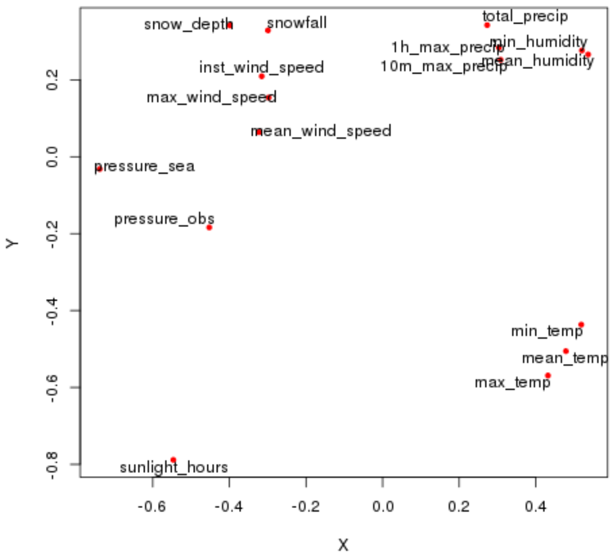


Figure 3. MDS plot of the relationships between weather metrics. Axes have no particular meanings.

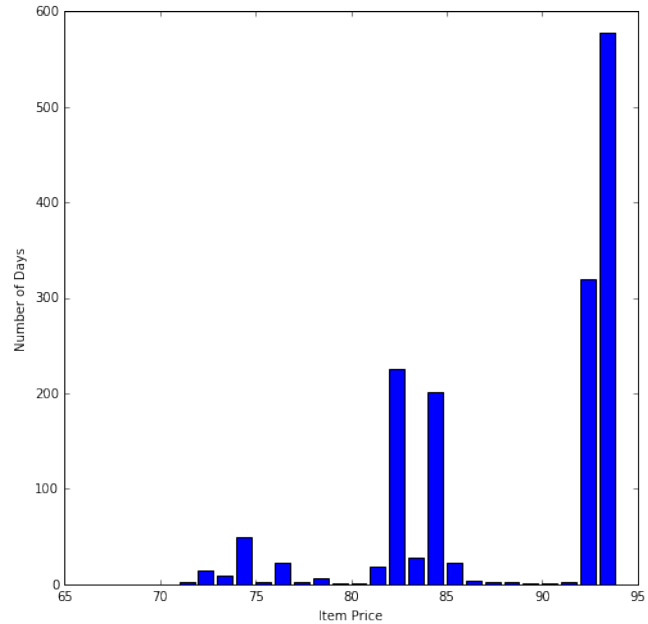


Figure 4. Histogram of price showing the number of days in which an item is sold at each price bin.

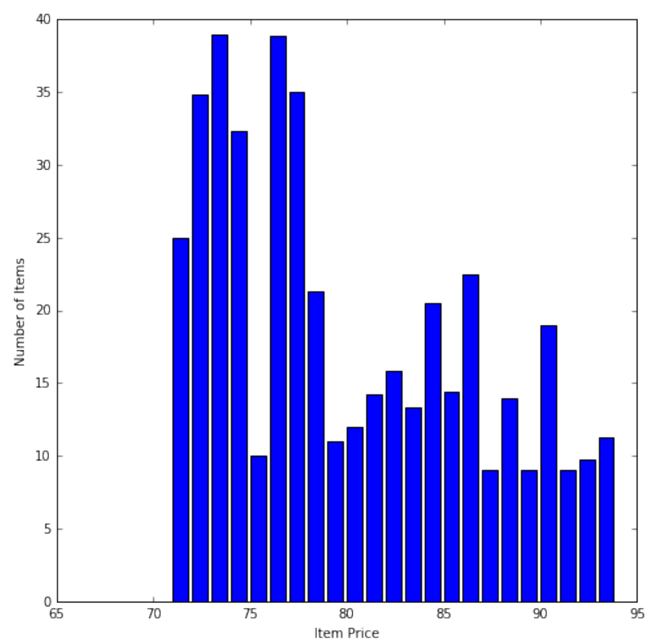


Figure 5. Average daily quantity sold per price bin.

supermarkets in the west of Japan customarily fix their sale date at a particular day of week. Normally, items sold at a sale price are expected to be sold better than usual, and the amount of increase of the quantity sold can be, at least partly, explained by the price elasticity. As an example, increasing tendency of the average quantity sold according to the level of price down can be observed in Figure 5. Nevertheless, there would be another factor that affects the amount of increase of the quantity sold: the mere fact that a sale is underway.

We tried to incorporate this factor, which indicates whether a sale is underway, into our models. At first, we calculate the mode price of each item for each day, based on recorded prices of the period of 28 days before and after the target day. Because the sale data have solely the daily sales field, which is the sum of prices of each instance of an item sold during a day, and an item price is calculated by dividing the daily sales by the quantity sold, it can take a real value. This makes the simple calculation of the mode price untrustworthy, so we make histograms of the price for each time window, thus obtaining the mode prices by picking the tallest bins.

Next, we calculate the standard deviation of the mode price, and the days where an item is sold at a price below 1σ of the mode price of the time window are considered sale days of that item. In Figure 6, a weekly pattern of price drop can be seen, and sale days indicated by dots are effectively identified. This sale day information is incorporated into the models as the sale flag.

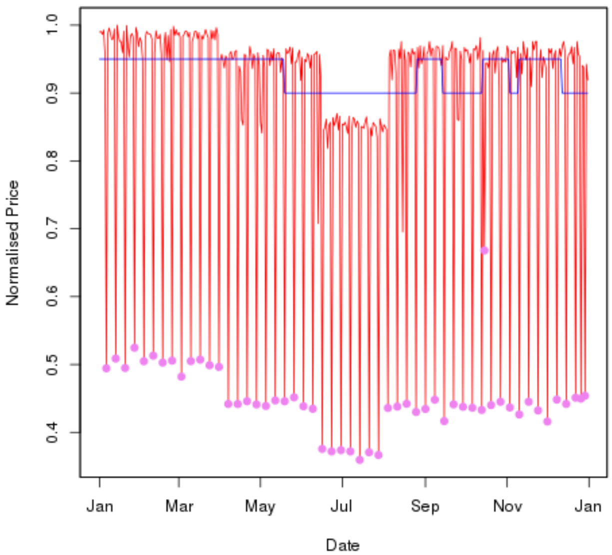


Figure 6. Price record example with mode price and sale flags. Red and blue lines represent price and mode price respectively, and magenta dots indicate days where the sale flag is on.

B. Prediction by linear regression

We built prediction models of the daily quantity sold for each area and item with the aforementioned explanatory variables by linear regression. The formulation is very straightforward and without regularisation terms because explanatory variables are not so many, and highly correlated variables are already filtered out. The quantity sold is normalised to $[0, 1]$ real-valued interval in order to absorb differences of the size of stores and the expected quantity sold of items. As a by-product of this normalisation, we can compare the prediction performance of models for different items simply by root mean squared error (RMSE).

C. Prediction by neural networks

We employed the *nnet* package[6] for R to build and train neural network based models. *nnet* is an implementation of simple feed forward neural networks having only one hidden layer, like a network shown in Figure 7. Input and output variables are exactly the same as the explanatory and predicted variables in the linear regression case, and we used the linear activation function (or equivalently, no activation fuction).

We can optimise resulting models' prediction performance by tuning the controllable hyperparameters of *nnet*: number of nodes in the hidden layer, weight decay, skipping flag to allow direct connections between nodes of the input and output layers and maximum number of iterations. Hyperparameter optimisation can be automatically done by

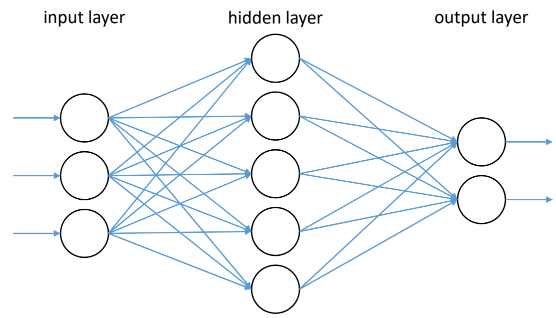


Figure 7. Neural network example. Each node is connected to all nodes in the adjacent layer.

the *caret* package[7] made for R, which searches the best hyperparameters in a grid search manner. We fixed skipping flag to true and maximum number of iterations to 10,000 and left optimisation of number of nodes in the hidden layer and weight decay to *caret*. As a cross validated optimisation result, using sales records of a specific store and item, we set number of nodes in the hidden layer to 3 and weight decay to 1.

D. Prediction by deep neural networks

We used the *h2o* package[8] for R to build and train deep neural networks. Its functionalities reflect recent advances in deep neural network research. The most notable differences to traditional neural networks are dropout functionality and activation functions. Dropout is a clever way to avoid overfitting during the training phase, by selectively ignoring some nodes in the network in each iteration[9], as shown in Figure 8. Target nodes of dropout are changed in each iteration. *h2o* offers a set of popular activation functions, which includes rectified linear unit (ReLU), whose shape is shown in Figure 9. ReLU is known to perform better than the traditional sigmoid function when it is applied to classification tasks on large and complex datasets[10]. Whether ReLU also performs well for regression tasks on a real world dataset has not been well documented as far as we know.

h2o's function for training deep neural networks has a substantial number of tunable hyperparameters, and it has a hyperparameter optimisation functionality by a kind of randomly selecting grid search like method. We optimised the hyperparameters by using this functionality and by manual inspection of model performance. The final hyperparameters we adopt are 3 hidden layers, 10 nodes in each hidden layer and the ReLU activation function.

IV. Results

To visualise regional and item-wise demand differences in a concise way, we integrate the area under demand curve and mapped resulting values to the colour scale. Figure 10 is an item by area matrix where each item and area pair's

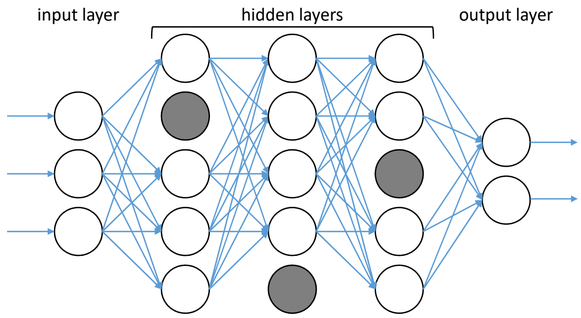
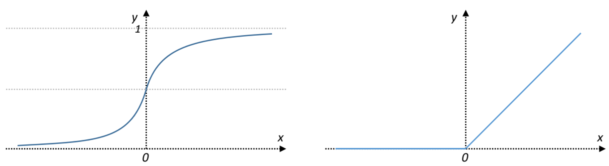


Figure 8. Deep neural network example with dropout. Dropped out nodes are coloured grey.



(a) Sigmoid (b) ReLU

Figure 9. Activation functions.

demand is obtained by linear regression. Larger values are mapped to lighter colours in the figure. We can clearly see that there are area dependent item preferences.

If we change the value of weather related variables when we estimate the demand curve, we can observe corresponding changes of the estimated demand curve. In Figure 11, the effect of temperature change to the demand for a specific item is visualised as the same way as in the item by area case. This item is sold better under warmer conditions, but this effect of temperature is not apparent in Hokkaido, using

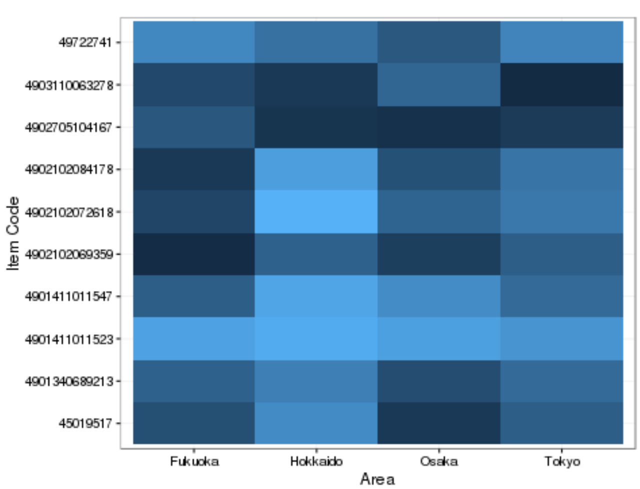


Figure 10. Integrated demand in each region for each item modelled by linear regression. Lighter colour indicates more demand.

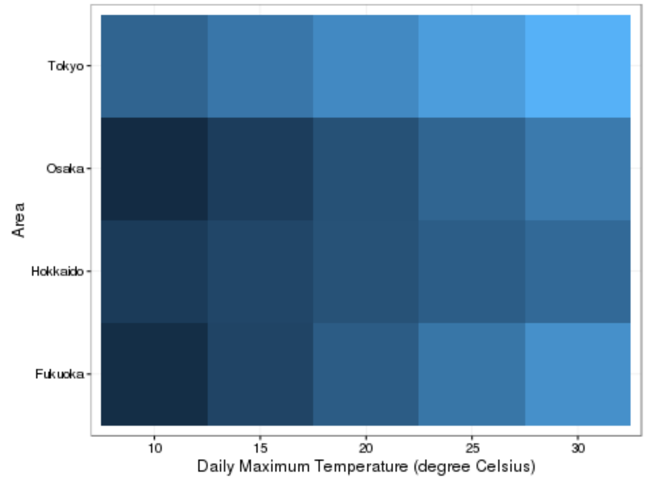


Figure 11. Integrated demand in each region in response to the temperature change modelled by linear regression.

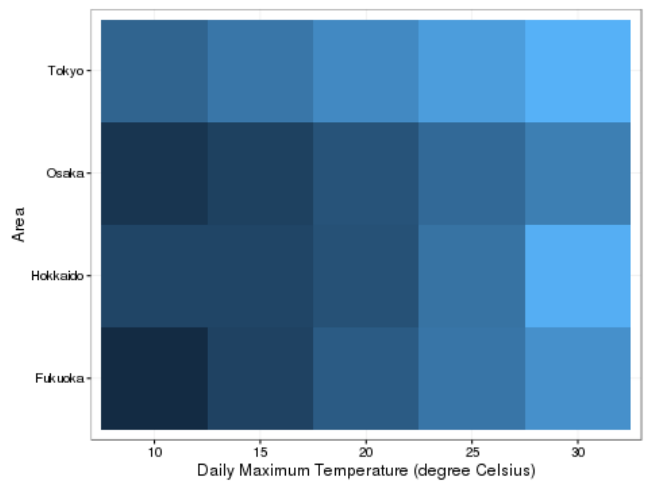


Figure 12. Integrated demand in each region in response to the temperature change modelled by neural network.

the model built by linear regression. On the other hand, response of the demand to temperature change in Hokkaido is as easily observable as in other areas by using a neural network based model, as shown in Figure 12.

Comparing RMSE of the resulting models obtained with 5-fold cross validation, neural network outperforms linear regression, as shown in Table V. Deep neural network seems slightly outperforming neural network, even under this relatively simple input and output setting. This result suggests that using deep neural network would possibly be beneficial to solve this kind of regression problems.

V. Discussion

The current models do not incorporate temporal changes of the characteristics of the consumer behaviour. Estimating mid-term (year-by-year) baseline changes of the price and

Table V
RMSE AND RMSE STANDARD DEVIATION OF THE MODELS WITH 5-FOLD CROSS VALIDATION.

Area	RMSE (Linear)	RMSE SD (Linear)	RMSE (Neural)	RMSE SD (Neural)	RMSE (Deep Neural)	RMSE SD (Deep Neural)
Fukuoka	0.11320	0.00089	0.11137	0.00170	0.11016	0.00104
Hokkaido	0.09364	0.00356	0.09259	0.00327	0.08527	0.00101
Osaka	0.09448	0.00151	0.08918	0.00174	0.08801	0.00098
Tokyo	0.10308	0.00095	0.09878	0.00296	0.09934	0.00160

the quantity sold, we could separate "inherent" price elasticity of each item from the fluctuation component caused by economic situations that should be examined from broader perspectives, thus making the prediction models more precise.

Incorporating item group and category information to the models could also improve prediction accuracy. We built the models from fine-grained item sales data in this study. For many items, however, there are numerous similar items in the inventory, like colour variations and quantity variations, each having distinct item code. If we could group these similar items and harness expected strong correlations between items in each group, models' predictive power would be improved. Each item also has a category label, and there will be correlations between items in each category, though weaker than that of within groups, which would be used to augment the models as well.

VI. CONCLUSIONS

We showed that the effect of weather to the demand of consumer goods would be different area by area, by estimating regional demand curves. This indicates that sales prediction models have better incorporate not only regional differences and effect of weather, but also their interactions as well.

We also showed that applying deep neural network methods for this class of regression problems would possibly improve the predictive performance of the model.

ACKNOWLEDGMENT

The authors would like to thank Tsutomu Watanabe, who granted us permission to use the daily sales data acquired for his research group.

REFERENCES

[1] J. T. Mentzer and C. C. Bienstock, *Sales Forecasting Management: Understanding the Techniques, Systems and Management of the Sales Forecasting Process*. Thousand Oaks, CA: Sage publications, 1998, iSBN-13 978-0761908227.

[2] P. Doganis, A. Alexandridis, P. Patrinos, and H. Sarimveis, "Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing," *Journal of Food Engineering*, vol. 75, no. 2, pp. 196–204, 2006.

[3] W.-I. Lee, C.-W. Chen, K.-H. Chen, T.-H. Chen, and C.-C. Liu, "Comparative study on the forecast of fresh food sales using logistic regression, moving average and bpnn methods," *Journal of Marine Science and Technology*, vol. 20, no. 2, pp. 142–152, 2012.

[4] M. K. Schnell and D. E. Weinstein, "Evaluating the economic response to japan's earthquake," *RIETI Policy Discussion Paper Series 12-P-003*, Feb. 2012.

[5] Japan meteorological agency. [Online]. Available: http://www.jma.go.jp/jma/indexe.html

[6] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed. New York: Springer, 2002, iSBN 0-387-95457-0. [Online]. Available: http://www.stats.ox.ac.uk/pub/MASS4

[7] M. Kuhn. The caret package. [Online]. Available: https://github.com/topepo/caret/

[8] S. Aiello, T. Kraljevic, and P. Maj. H2o.ai. [Online]. Available: http://www.h2o.ai/

[9] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.

[10] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 807–814.