Extracting Important Knowledge
from Multiple Markets Using Transfer
Learning

Tokuhiro Kujiraoka[(✉)], Fumiaki Saitoh, and Syohei Ishizu

Aoyama Gakuin University, 5-10-1 Fuchinobe, Shagamihara,
Kanagawa 252-5258, Japan
`c56l6l77@aoyama.jp`, `{saitoh,ishizu}@ise.aoyama.ac.jp`

Abstract. The aim of this study is to extract a customer's needs from their reviews of an electronic commerce (EC) site using transfer learning. Transfer learning involves retaining and applying the knowledge learned from one or more tasks to efficiently develop an effective hypothesis for a new task. Recently, with the spread of EC sites, customer reviews have become a beneficial information source, as they include customers' opinions or product reputations, and can attract attention. However, this information is too huge to browse conveniently. Moreover, to develop new products with a competitive advantage, it is necessary to incorporate customers' opinions. Therefore, it is necessary to extract the customers' opinions from the enormous amount of customer reviews.

In this research, we focus on markets where multiple products compete. With the spread of smartphones, multiple products, e.g., cameras, compete in the same market. We want to understand customers' needs efficiently by extracting com-mon requests by consumers from the information about these multiple products. Hence, we propose a method of extracting customers' needs using transfer learning to comprehensively handle multiple products' information for this market.

Keywords: Transfer learning · Random forest · Ensemble learning · Text mining · Customer review

1 Introduction

Recently, online shopping, which allows people to purchase products from the inter-net, has been spreading widely. The market size for electronic commerce (EC) sites has been increasing. The average growth rate for the past 10 years is as much as 7.3% and the sales from online shopping in Japan exceeded 58 billion dollars in 2014. Therefore, the market for EC sites has been expanding and is expected to in-crease in the future. In addition, through customer review services, purchasers can freely describe their evaluations, impressions, opinions, and satisfaction with the products. Online shoppers often use customer reviews when making a purchasing decision. Since these are a valuable information source, as they contain personal voices, re-searchers have been trying to analyze the relevance of EC sites to marketing.

How-ever, these are huge quantities of information, which require considerable time and effort to review. To effectively utilize this enormous review data, researchers have investigated extracting only the necessary information and classifying the customer reviews by some criteria. In this research, the aim is to efficiently process huge amounts of customer reviews and extract customer needs. In this paper, we focus on smartphone reviews posted in customer reviews of EC sites. Smartphones are spreading quickly and their ownership rate exceeded 70% in 2016 in Japan. However, some markets are shrinking due to the spread of smartphones. This is because products with multiple functions, e.g., smart phones, and products with single functions, e.g., digital cameras, coexist in the same market. Products with multiple functions infringe on the market share of single-function products. As an example of this situation, we chose the market around the smartphone. Since smartphones can fulfill multiple functions, consumer demand is different for each product or function. Therefore, it is effective to extract the customer's needs for each smartphone function and demand. Thus, it is necessary to consider not only smartphone information but also information on competing products. Transfer learning is a comprehensive method for handling extensive information. It utilizes knowledge obtained from different resources to efficiently solve a given task. We consider that customer needs can be efficiently extracted using competing product information about smartphones. Extracting the knowledge creates new value by transferring the knowledge learned about one product to another. In this study, we use the Random Forest (RF) classifier, which is an ensemble-learning method. RF uses the voting of multiple decision trees to predict or identify data. These trees are generated by bootstrap random sampling and are not allowed to overlap; thus, the risk of overfitting is reduced. Although RF can also be applied to high dimensional data, it cannot be applied to sparse data containing many zeroes. In this research, we use the text from customer reviews to create a word-frequency matrix. This matrix is likely to be large and sparse. The Random Forest method does not learn well with this matrix. Thus, it is necessary to convert the sparse matrix to a dense matrix, using non-negative matrix factorization (NMF). This method is often used for text data and can easily extract topics.

2 Background and Related Works

2.1 Customer Reviews

There are several papers directly in the customer's review. There are many studies on customer reviews, because customers' opinions are directly reflected in customer reviews and it is a useful source of information. Studies of customer review include research that extracts only necessary information and research that classifies reviews. Both studies are being studied to efficiently process enormous amounts of review data in order to grasp the needs of customers. We introduce some research. Okada et al. [3] use SVM for automatic review of documents for review of travel sites. Firstly, they define evaluation sentence pattern using sentence patterns of Japanese, and the effectiveness of classification when using it was investigated experimentally. As a result, it was found that it is difficult to improve classification accuracy simply by incorporating

evaluation expressions. Přichystal [1] measures the quality of products and services from reviews. The author considers that the evaluation is based on human emotions. However, handling a large amount of data is almost impossible to process manually, it takes time to read all product reviews. They aim to automatically find human emotions hidden in customer reviews.

2.2 NMF(Non-negative Matrix Factorization)

In this research, since we used NMF when contracting the dimension of text data, so its effectiveness is shown from related research. SAWADA [5] proves that NMF is valid for document data. NMF is a method of analyzing a matrix composed of positive values including many zeros, and the result obtained by NMF utilizes a feature in which data elements are clustered based on the frequent pattern of the data. As a result of the experiment, it became possible to cluster news articles. Kimura et al. [12] proposes transfer learning by NMF. Kimura et al. pro-posed a transfer learning method based on conservativity of feature space based on the metastatic hypothesis that the feature space used for approximate expression between domains is similar. The pro-posed method is applied to transfer learning in document clustering and shows the effect of the proposed method. It is known that the proposed method has a wide applicability because it has an advantage that it does not require a label for data given as the original domain.

2.3 Machine Learning(Random Forest and Transfer Learning)

We introduce the research related to RF adopted in this research and the research related to metastasis learning. Transfer learning is used for a fairly wide range of machine learning frameworks, and the definite definitions cannot be said clearly. This idea is widely used that in order to efficiently find an effective hypothesis of a new task, it is necessary to obtain knowledge learned by one or more different tasks and to apply it [14]. In other words, it is to solve the problem efficiently by utilizing knowledge learned from different information source for a problem to be solved. Currently, as more data is available due to the spread of the Internet, it is required to use these information efficiently and effectively. There is transfer learning as a means for that. Fukumoto et al. [15] classifies documents using random forests. We proposed a method to generate co-occurrence matrix to extract useful information from large-scale document data, and document classification using the result. Compared with SVM and Bagging, it indicates that RF has high classification accuracy. A related study in transfer learning is TrBagg by Kamishima et al. [8]. TrBagg is a technique applied Bagging to transfer learning. Dai et al. [16] proposes TrAdaBoost applying AdaBoost to transfer learning. We explain the two studies that are the basis of our research. Kamishima et al. [8] proposes transfer learning using Bagging which is a type of ensemble learning. The algorithm is simple, and the transfer learning of this research is based on this method. Based on the idea that a weak learner that reduces the prediction error is not used. Kumagae et al. [10] proposes an algorithm (OptTrBagg) that points

out and solves the problem of transfer learning proposed by Kamishima [8]. The authors use proposal method to predict the purchase of products using information on multiple EC sites. Experimental results showed that this proposed method is effective. Since the proposed method has no fixed framework, several proposed methods have been proposed.

3 Method

In this research, we extract the customer's reviews from a market where multiple products compete, using transfer learning by random forest. First, we applied a text mining technique to the collected data. Many languages, including English, are generally divided between word and word by space. However, in the case of Japanese, since between word and word are not divided by space, it is necessary to separate the words using a morphological analysis. We then create a word-frequency matrix; i.e., a matrix expressing the word-appearance frequency for one review.

In this research, since we deal with approximately 3000 customer reviews, the created frequency matrix is likely to be sparse. RF does not work well on sparse matrices, so they need to be converted to dense matrices. Therefore, we per-form non-negative matrix factorization, which transforms the sparse matrix into a dense matrix and reduces the dimension. We perform transfer learning by RF on the reduced data and extract common knowledge. Next, we describe a non-negative matrix factorization and RF outline for transforming a large-frequency sparse matrix to a dense matrix. We subsequently describe transfer learning as the process of the proposed method.

3.1 Non-negative Matrix Factorization

A decision tree is difficult to create if the target data is a large sparse matrix. Therefore, RF does not work well. Document data, e.g., customer reviews, are most likely to be sparse, because the documents are not very long; however, we deal with a vast number of reviews and many words appear. To avoid this, we use non-negative matrix factorization (NMF). If we use NMF, we can transfer from a sparse matrix to a dense matrix and reduce the dimensions without losing information.

NMF is a method of decomposing one non-negative matrix into two non-negative matrices and approximating them. The dimension can be reduced without losing the latent meaning of the original matrix. NMF is often used for document, sound, and image data. Several frequent patterns are obtained by implementing NMF. That is, highly similar variables are reduced to synthetic variables. Moreover, since similar words affect common latent variables, it is relatively easy to determine the meaning of the reduced variables. From this property, we can expect that the analysis accuracy will be improved and the interpretation can be expanded in the knowledge extracted from the data.

A schematic diagram of the non-negative matrix factorization is shown in Fig. 1. NMF can approximate the original data matrix X ($I \times J$ matrix) as the product of two
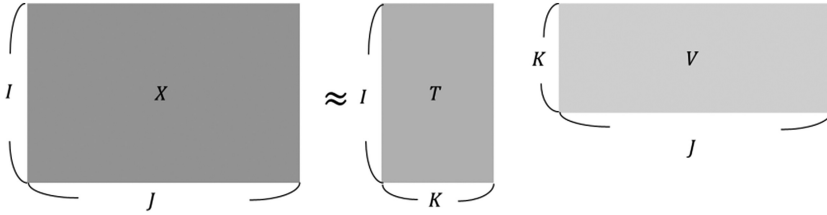
Fig. 1. NMF schematic

matrices $(X \simeq TV)$. K is the basis and indicates the number of dimensions we want to reduce. In this algorithm, each element t_{ik}, v_{kj} of the matrix T, V is first initialized with a non-negative random number. The specified number of updates is determined using Eqs. (1) and (2).

$$t_{ik} \leftarrow t_{ik} \frac{\sum_j x_{ij} v_{kj}}{\sum_j \hat{x}_{ij} v_{ik}} \tag{1}$$

$$v_{kj} \leftarrow w_{kj} \frac{\sum_i x_{ij} v_{ik}}{\sum_i \hat{x}_{ij} v_{ik}} \tag{2}$$

NMF has already been proven to be able to extract document topics from text data []. In this research, the knowledge is extracted learning a matrix T by RF

3.2   Random Forest (RF)

In this research, we use a RF which is one of machine learning methods to learn product information. RF is one of ensemble learning methods, and a transition learning method in bagging, which is a type of ensemble learning, has al-ready been proposed. Transfer learning using a random forest has also been pro-posed, but if the number of data is proven to improve accuracy when the number of target data is few, it is not effective when the number of data is many.

Decision Tree
Since the Random Forest is ensemble learning with the decision tree as a weak learner, we first outline the decision tree. Decision tree analysis is a data-mining method that uses classification and prediction. As shown in Fig. 2, it has a tree structure and a directed graph that is not closed. A leaf (color) represents a classification, and a branch (no color) is a tree structure representing a collection of features up to that classifica-tion. Various algorithms have been proposed for constructing decision trees. In the Random Forest, learning by the most representative Classification and Regression Tree (CART) analysis is used.
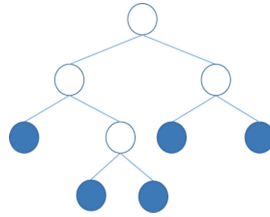
Fig. 2. Decision tree schematic

Random Forest

Random Forest is a machine-learning algorithm proposed by Leo Breiman in 2001. Its advantage is that it can cope with high dimensional data, while reducing the risk of over-learning. On the other hand, the disadvantage is that when the data is a sparse matrix, it cannot be implemented well. We construct multiple decision trees using samples generated with ensemble learning, and a random sampling overlap by boot-strap sampling, and perform identification, regression, and clustering using the results.

In the identification case, the data is classified into the class with the majority decision of the output class of the decision tree. In the regression case, the value is determined by the average value of the output of the decision tree. Compared to other machine-learning methods, e.g., support vector machines and neural networks, the calculation speed is high and it is said to be highly accurate. A schematic diagram of a Random Forest is shown in Fig. 3. The Random Forest algorithm is shown below.

STEP 1   Extract bootstrap specimens that allow random duplication
STEP 2   Build decision tree from bootstrap specimen
             –Randomly extracted feature quantities without allowing duplication
STEP 3   Construct a repetition model with STEP 1, STEP 2 specified times
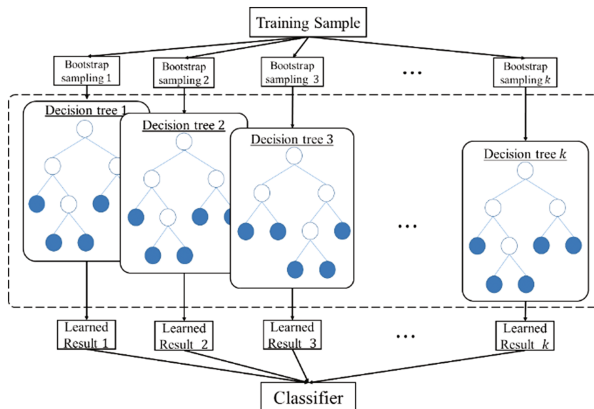**STEP 4**   Identify by the majority vote of the constructed decision tree

Fig. 3. Random Forest Schematic

3.3   Transfer Learning

In this section, we explain about transfer learning and the proposed transfer learning method using Random Forest. Transfer learning has not established much of a defini-tion but it is a learning method that efficiently solves for a certain task using knowledge learned from information on related or similar tasks. We call the domain to be predicted the transfer destination (target domain), and the data the transfer source (source domain).

In this research, we consider the problem of using the data about smartphones to gain knowledge about cameras. The desired knowledge is the common colored knowledge between product A and products B, C, or D. Product A is the target do-main (smartphone) and products B, C, and D (cameras, etc.) are the source domains; we extract the specialized needs by transferring the knowledge of the competing camera products.

First, we summarize the symbols to be handled in this research. $\mathcal{D}_T$ is the target domain that we want to predict, or improve the accuracy of the knowledge transfer. The source domain that transfers knowledge is $\mathcal{D}_S$. In this paper, the smartphone infor-mation is $\mathcal{D}_T$ and the camera information is $\mathcal{D}_S$. $\mathcal{D}$ is the combined data of $\mathcal{D}_T$ and $\mathcal{D}_S$.

When RF is learning, a parameter should be defined that determines the size of the forest. If a model is built from the target domain by setting the parameter to B, when the model is built from the combined target and source domain data, the parameter will be B'.

3.4   Proposed Method

The transition learning implemented in this paper refers to TrBagg, and a similar algorithm is implemented by Random Forest instead of Bagging. Transfer learning using Random Forest consists of two algorithms for learning and selecting the available decision trees. First of all, in the learning part, we obtain a set $\mathcal{F}_{T+S}$ of decision trees learned by combining the target domain and source domain, and a set $\mathcal{F}_T$ of decision trees learned from the target domain. In the filtering part, we combine the decision trees obtained by Algorithm 1, and arrange them in ascending order according to the pre-diction error for the target domain. Let e be the prediction error of the rearranged decision tree $\hat{f}_1$, and we add $\hat{f}_1$ to $\mathcal{F}'$ and $\mathcal{F}^*$. We repeat this $(B + B')$ times to predict by majority vote using a decision tree by adding a target domain $\mathcal{F}'$ in order from $\hat{f}_2$. At
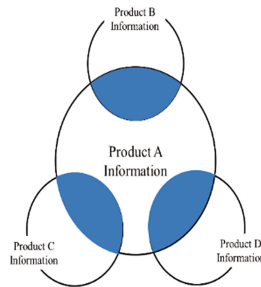

Fig. 4. Common knowledge that can be extracted from transfer learning

this time, if a decision tree is added that improves the prediction error for the target domain, it is added to $\mathcal{F}^*$, and e is updated to the error e' at that time. Figure 5, Algorithms 1 and 2 show transfer learning using RF. (Figure 4)

Algorithm1 *Transfer Learning by Random Forest*

1: **INPUT** $\mathcal{D}_T, \mathcal{D}_S, B, B'$
2: $\mathcal{D}_{T+S} = \mathcal{D}_T \cup \mathcal{D}_S$ (union)
3: $\mathcal{F} = \{\emptyset\}, \mathcal{F}_T = \{\emptyset\}$
4: **for all** t = 1 to B **do**
5:     $\mathcal{D}'_{T+S} \leftarrow$ Bootstrap sample from $\mathcal{D}_{T+S}$
6:     $\hat{f}_t \leftarrow$ Clasiifire learnd from $\mathcal{D}'_{T+S}$
7:     $\mathcal{F}_{T+S} = \mathcal{F} \cup \hat{f}_t$
8: **for all** t = 1 to B' **do**
9:     $\mathcal{D}'_t \leftarrow$ Bootstrap sample from $\mathcal{D}_T$
10:     $\hat{f}_t \leftarrow$ Clasiifire learnd from $\mathcal{D}'_t$
11:     $\mathcal{F}_T = \mathcal{F}_T \cup \hat{f}_t$
12: **end for**
13: $\mathcal{F}^* = Filtering(\mathcal{F}, \mathcal{F}_T)$
14: **OUTPUT** $\mathcal{F}^*$: $\{\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n\}$

Algorithm2 *Filtering*

1: function Filtering
2: **INPUT** $\mathcal{F}$: $\{\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_B\}, \mathcal{F}_T$: $\{\hat{f}_{T_1}, \cdots, \hat{f}_{T_{B'}}\}$
3: $\mathcal{F} = \mathcal{F} \cup \mathcal{F}_T$
4: $\{\hat{f}_1, \cdots, \hat{f}_{B+B'}\}$: Sort $\mathcal{F}$ prediction error for $\mathcal{D}_T$(ascending)
5: $\mathcal{F}' = \{\emptyset\}; \mathcal{F}' \cup \hat{f}_1$
6: $\mathcal{F}^* = \{\emptyset\}; \mathcal{F}^* \cup \hat{f}_1$
7: $e \leftarrow$ Prediction error to $\mathcal{D}_T$ by $\hat{f}_1$
8: **for all** $t = 1$ to $B + B'$ **do**
9:     $\mathcal{F}' = \mathcal{F}' \cup \hat{f}_t$
10:     $e' \leftarrow$ Prediction error to $\mathcal{D}_T$ by $\mathcal{F}'$
11:     **if** $e' \leqq e$ **then**
12:         $\mathcal{F}^* = \mathcal{F}^* \cup \mathcal{F}'$
13:         $e = e'$
14:     **end if**
15: **end for**
16: **OUTPUT** $\mathcal{F}^* = \{\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n\}$

In this section explain the proposed method. The proposed method is divided into 4 stages and it is as following.

STEP 1   Collect customer reviews and perform morphological analysis to create frequency matrix
STEP 2   Since the random forest does not function when it is a sparse matrix, it converts to a dense matrix and reduces the dimensions using nonnegative matrix factorization (NMF)
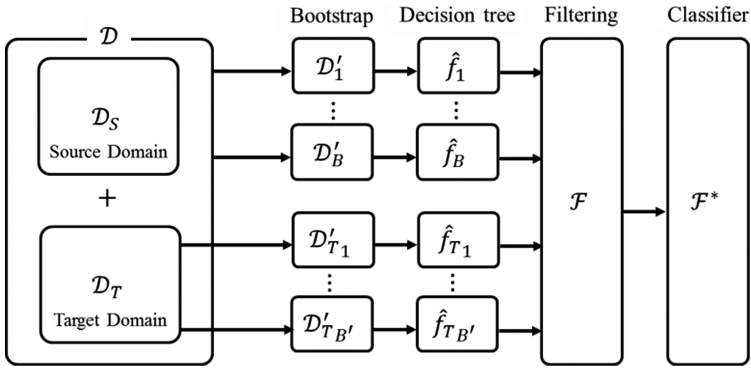
Fig. 5. Schematic transfer learning by RF

STEP 3   Learns the feature amount of the reduced data in chapter 2, class label as "good" or "bad" evaluation for products. Construct model by Transfer learning using random forest

STEP 4   Classify the review of smartphone with the constructed model and extract customer needs

4   Experiment and Result

4.1   Experiment Condition

We gathered 1324 reviews on smartphones and 1912 reviews on digital cameras, totaling 3236 reviews from review data provided by Rakuten, Inc. and are the subject of investigation. We performed morphological analysis on text data, created frequency matrix, and reduced dimensions using non-negative matrix factorization.

The number of decision trees in the random forest was 500. We summarized data and parameter using this experiment as following table

4.2   Result

We summarize the results predicted with the constructed model. For comparison, we summarize the results of the following experiments (Tables 4, 5 and 6):

(1)   Predicting the target domain with the model generated by the RF from the target domain,

(2)   Predicting when smartphone is transferred by transfer learning, and

(3)   Prediction when smartphone is target domain by transfer learning

Needs specialized for highly evaluated camera function

- I ordered it for my husband, it was unused and in a beautiful state and charged a little, it was very good. Shipping was also fast. The camera function is superior to

Table 1. Number of data

	Good (4 ~ 5)	Bad (1 ~ 3)	Total
Smart phone	643	681	1324
Camera	947	965	1912

Table 2. Parameter of NMF

Base	50
Number of update	100

Table 3. Parameter of RF

Number of decision tree	B	50
	B'	100
Number of selected feature		14

Table 4. Accuracy and prediction Table 1

(1)	Good	Bad
good	526	92
bad	39	464
	Accuracy	88.31%

Table 5. Accuracy and prediction Table 2

(2)	Good	Bad
good	512	87
Bad	131	391
	Accuracy	80.6%

Table 6. Accuracy and prediction Table 3

(2)	Good	Bad
good	405	206
Bad	223	287
	Accuracy	61.74%

my cell phone, the shutter is quickly turned off and the image is also good, so I wanted my part as well.

- I ordered it for easy tomorrow free, gift wrapping, but I am satisfied with the arrival of goods as desired. Calls and e-mails are main, so the items here are enough. Also, since the number of pixels of the camera is quite poor in the model before this item, I made it to this model. The opponent who gave it is also pleased.
- I received it in two days after placing an order. I could use it as soon as I inserted the SIM card (^ - ^). The camera is also beautiful and the motion is crisp and comfortable. Good shopping was done!

Needs specialized for lowly evaluated camera function

- The camera is really nice. Especially, I think that shooting at night cannot be imitated by other mobile phones. However, other than that, the touch panel is particularly bad.
- There are some difficulties to use such as not being able to directly write the photos taken with the camera application on the SD card, but was it a place that ordinary smartphone was said?

5   Discussion

We compared the model based on the Random Forest generated from the target domain, and the model using transition learning using two of the transfer targets and the source domain. The results showed that the accuracy decreased. However, since our aim is to extract camera-specific reviews from high-rated smartphone reviews, we can predict that the prediction accuracy will drop. We consider that classification using other product information may result in reviews closer to the needs of the camera. Looking at the actually extracted review, it can be seen from the review of the smartphone that the customer's voice with respect to the camera function is included. From this result, it can be said that the need for the camera function against that from the smartphone review can be extracted.

6   Conclusion

In our paper, we extracted the customers' needs from the smartphone reviews. Due to the diversity of smartphone demands, we tried to subdivide the needs by transferring competing product information. Through this experiment, we classified reviews specialized for cameras by selecting camera-review information from smartphone re-views and extracting customers' needs. We can extract camera function needs for smartphones from the classified reviews. However, since the method has not been verified, this paper was simply a proposal of the extraction method. The obtained results leave much room for verification, and we must also verify other products. Moreover, the suggested transfer learning with RF is simple, and we should consider a more sophisticated method in the future.

References

1. Přichystal, J.: Mobile application for customers' reviews opinion mining. In: 19th International Conference Enterprise and Competitive Environment 2016, vol. 220, pp. 373–381 (2016)
2. Review decision about the documents by the ratio of a review sentence by classifier a sentence review. Inf. Process. Soc. Jpn. **2012-NL-205**(2), 1–5 (2012)

 3. Okada, M., Takeuchi, K., Hashimoto, K.: An investigation of an effectiveness of estimation sentence patterns for a classification of customer reviews considering their conditions. In: The 28th Annual Conference of the Japanese Society for Artificial Intelligence, vol. 28, pp. 1–3 (2014)
 4. Kobayashi, T., Saitoh, F., Ishizu, S.: Important quality element and preference rule extraction in the case of competition among different types of products. In: The Japan Society for Management Information Autumn National Research Competition, pp. 331–334 (2015)
 5. Sawada, H.: Nonnegative matrix factorization and its applications to data/signal analysis. Inst. Electron. Inf. Commun. Eng. **95**(9), 829–833 (2012)
 6. Lee, D.D., Sebastian Seung, H.: Learning the parts of objects with nonnegative matrix factorization. Nature **401**, 788–791 (1999)
 7. Takaya, N., Kumagae, Y., Ichikawa, Y., Sawada, H.: Adopting transfer learning towards item purchase prediction on web marketing. Int. J. Inform. Soc. **7**(1), 19–27 (2015)
 8. Kamishima, T., Hamasaki, M., Akaho, S.: TrBagg: a simple transfer learning method and its application to personalization in collaborative tagging. In: Proceeding of the 9th IEEE International Conference on Data Mining, pp. 219–228 (2009)
 9. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and application: a survey. Ain Shams Eng. **5**, 1093–1113 (2014)
10. Kumagae, Y., Murata, M., Takaya, N., Uchiyama, T.: Transfer leaning for prediction of purchase items using multiple e-commerce sites' information. In: DEIM Forum 2012, C8–6 (2012)
11. Kotani, N.: Knowledge selection based on value in transfer learning. Inst. Syst. Control Inf. Eng. **28**(6), 275–283 (2015)
12. Kimura, K., Yoshida, T.: Topic graph based transfer learning via generalized KL divergence based NMF. Inf. Process. Soc. Jpn **2011-MPS-85**(3) (2011)
13. Asoh, H., Kobayashi, M., Kobayashi, I.: Transfer learning of language models based on combinatorial structure of topics. In: The 28th Annual Conference of the Japanese Society for Artificial Intelligence (2014)
14. Kamishima, T.: Transfer learning. Jpn. Soc. Artif. Intell. **25**(4), 1–9 (2010)
15. Fukumoto, S., Futcida, T.: Document classification by machine learning based on co-occurrence relations of words. Inf. Process. Soc. Jpn. **2014-DBS-160**(28) (2014)
16. Dai, W., Yang, Q., Xue, G.-R., Yu, Y.: Boosting for transfer learning. In: Proceedings of the 24th International Conference on Machine Learning, ICML 2007, pp. 193–200 (2007)