ORIGINAL RESEARCH

Properly initialized Bayesian Network for decision making leveraging random forest

Yutaka Iwakami*[1], Hironori Takuma[2], Motoi Iwashita[3]

[1]*Department of Research & Analysis for IT Industry, Nork Research Co., Ltd., Tokyo, Japan*
[2]*Department of Project Management, Chiba Institute of Technology, Chiba, Japan*
[3]*Department of Management Information Science, Chiba Institute of Technology, Chiba, Japan*

ABSTRACT

Bayesian network is one of major methods for probabilistic inference among items. But if it contains particular targeting node and other explanatory nodes for decision making, for example how to select suitable appealing keywords to make customers like a product, edges around the target should be counted with more importance than those among others while constructing the network. In order to achieve this adjustment, this study proposes to configure initial state consisting of a few nodes and their edges connected with the target. The initial state is obtained by leveraging Random forest which is a proven method for decision making. Initial nodes are extracted by measuring mean decrease of Gini coefficient calculated with decision trees of Random forest. Initial edges are designated by comparing Wasserstein metric of conditional probability distribution among nodes which are corresponding to edge directions. Through an actual experiment, this method is confirmed to be effective for adjusting Bayesian network in decision making. This approach is especially useful for business scenes, such as selecting preferable keywords for appealing products over SNS.

Key Words: Decision making, Probabilistic inference, Bayesian network, Random forest, Decision tree, *Gini* Coefficient, Wasserstein metric

1. INTRODUCTION

In the era of TV and Radio, keywords for appealing a product had been unilaterally determined by its maker. SNS has now become popular and consumers are spreading impression and evaluation of various products with their own keywords across the Internet. Therefore, it is necessary for makers to understand which keywords should be used in the era of SNS. For this purpose, several useful studies are proposed.[1–3] In this study the authors are focusing on applying Bayesian network to select appropriate keywords.

As described in Section 2, the authors pick up an existent ex-ample of a beer maker and its representative product, which are noted "the maker" and "the product" respectively in this paper. If a consumer posts a keyword related with beer such as "sharpness" in SNS, the maker needs to know how often this keyword occurs with the product, in order to determine whether they should include "sharpness" into their marketing messages in SNS. If the keyword more often occurs in SNS with the product, it is considered to be engaged with and closely related with the product. It is called "engaged keyword" in this paper. On the other hand, the keyword with less often occurrence with the product, it is called "non-engaged

keyword". The business success for the maker depends on how to select engaged keywords effectively.

For extracting engaged keywords, several methods such as leveraging Co-Occurrence network analysis[4, 5] or Word2Vec[6–8] are already proposed. Even after applying these methods, there is still an issue of how to select actually used keywords from obtained keywords. As the budget for marketing and advertising is often limited in business, the maker needs to pick up especially effective engaged keywords among them. For the purpose above, the authors apply Bayesian network analysis, it has advantage such as relationships between the product and engaged keywords are visualized in the form of DAG (directed acyclic graph).[9, 10]

As described in Section 2, in Bayesian network analysis of this study, two sets of tweet records are retrieved from Twitter. One is "engaged tweets", which contain the product. The other is "non-engaged tweets", which only contain an ordinal word "beer". These two sets are combined into a dataset. Each record of the dataset has columns indicating whether contains each engaged keyword or not, and also has engaged/non-engaged flag as the last column. By applying Bayesian network analysis, the inference can be performed such as, tweets with keyword A are likely to be engaged tweets, but with the combination of keyword B, it does not. However, as mentioned in Section 2, the situation occurs such as, even when tweets with the combination of keyword A and keyword B are likely to be engaged tweets based upon the inference, the actual search result on Twitter shows the opposite result.

The cause of such disagreement would reside in the characteristic of Bayesian network. In the example above, the engaged/non-engaged flag is a kind of target node (explained variable), which has different role from nodes of engaged-keywords (explanatory variables). But Bayesian network usually handles all these nodes equally. Therefore, some adjustment will be required to apply Bayesian network for decision making task like this case.

On the other hand, Random forest is a proven method for analyzing the influence of explanatory variables upon explained variable.[11, 12] As described in Section 3, the authors propose to configure initial state of Bayesian network leveraging the result of Random forest analysis. The initial state consists of a few nodes around the target node and several edges between these nodes and the target. The former is called "initial nodes" and the latter is called "initial edges" in this paper. Initial nodes are extracted by measuring mean decrease of *Gini* coefficient calculated with decision trees of Random forest, because explanatory variables with much influence on explained variable show significant decrease of the coefficient. Directions of edges correspond to condi-

tional probability among nodes connected with those edges. Therefore, directions of initial edges are designated based on likelihood measured by similarity of conditional probability distribution between actual data and predicted result of Random forest. The similarity is calculated with Wasserstein metric. Initial nodes and initial edges are given as an initial state for the construction of Bayesian network.

As confirmed in Section 4, the inference result of the Bayesian network with initial state coincides well with the actual search result on Twitter. Configuring initial state leveraging the result of Random forest analysis is considered to be a kind of adjustment of Bayesian network to perform decision making with explained/explanatory variable as nodes. In the following Section 5 the summarization of this study is described. At last in Section 6, the authors discuss the future work such as the case having multiple explained variables. As a tool for retrieving and analyzing of data, one of the representatives of statistical computing language and environment "R" is used in this study.

2. EXAMPLE CASE AND CURRENT ISSUE

2.1 Example case and current issue

For clarification and demonstration, the authors pick up an existent example of a beer maker and its representative product. In this paper they are called "the maker" and "the product" respectively. Prior to this study, the authors extract 18 engaged keywords for the product leveraging Word2Vec.[13] At first two sets of tweets are searched and retrieved on Twitter. One is the set of tweets which include the product. The other consists of tweets including ordinal keyword in terms of the business domain such as "beer" in this case. Then Word2Vec analysis is applied for the mixture of the two sets. If a keyword shows closer direction to the product than to the ordinal keyword in vector space obtained by Word2Vec, the keyword is considered to be more closely related with the product than other keywords. With this procedure, engaged keywords for the product are obtained as shown in Table 1.

The purpose is to pick up engaged keywords more related with the product from Table 1 with Bayesian network analysis. For performing Bayesian network analysis, a dataset is retrieved on Twitter. As tweet retrieving tool, "rtweet" library of R is used in this study. The dataset consists of two types of tweets. One is "engaged tweets", which contain the product. The other is "non-engaged tweets", which only contain an ordinal word "beer". Each record of the dataset has columns indicating whether contains each engaged keyword (contain=1/ not contain=0) and also has engaged/non-engaged flag as the last column by the name of "engaged"(engaged=1/not engaged=0). The total number of tweets is 1046 (engaged: 357, non-engaged: 689). The structure of the dataset is shown in Figure 1.

		Word1	Word2	- - - -	Word18	engaged
	Tweet1	0	0		1	1
	Tweet2	1	1		1	1
	:					
	:					
	:					
	Tweet357	1	0		1	1
	Tweet358	0	1			0
	:					
	:					
	Tweet1046	0	1		0	0

Figure 1. Structure of the dataset

Table 1. Engaged keywords of the product

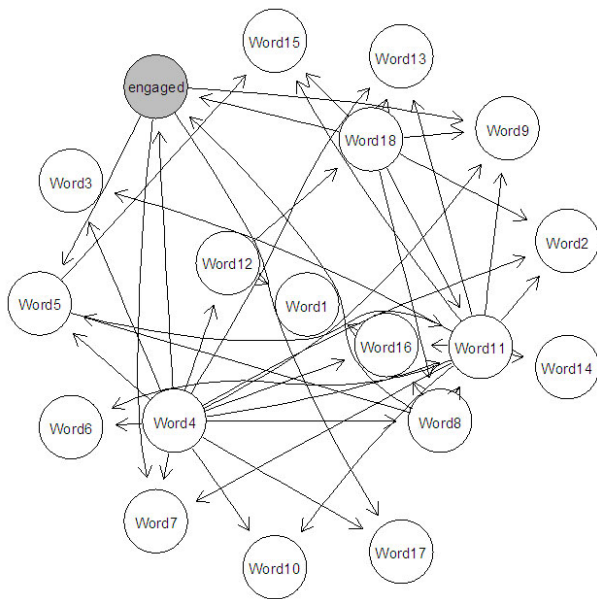| Word1	rich	Word6	alcohol percent	Word11	cheers	Word16	solid
Word2	guzzle	Word7	craft beer	Word12	bitterness	Word17	thick
Word3	chilled	Word8	dry	Word13	refreshing	Word18	taste
Word4	drinkable	Word9	Belgium	Word14	brisk		
Word5	lager	Word10	fruity	Word15	strongest		

Figure 2. Result of ordinal Bayesian network

2.2 Current issue

The obtained Bayesian network is shown in Figure 2. Tabu search algorithm[14, 15] in "bnlearn" library of R is used for learning the structure of Bayesian Network. The target node ("engaged") is marked as gray. It is directly connected with engaged keywords such as Word4, Word5, Word7, Word8, Word9, Word17 and Word18.

The network shows several visual insights. For example, the connection of three nodes, "engaged", "word4" and "word8" includes a tail-to-tail relationship (there are two edges from "Word4" to both "engaged" and "Word8"), which is one of three basic connections composing Bayesian Network. In a tail-to-tail, after "Word4" is determined, "engaged" is conditionally independent of "Word8". But the actual network in Figure 2 is much more complicated, such as "Word8" has an edge directly connected to "engaged". A few examples of probabilistic inference for those three nodes as shown in Table 2.

Table 2. Example of probabilistic inference

P(engaged = 1 | Word8 = 0)	31.2%
P(engaged = 1 | Word8 = 1)	78.6%
P(engaged = 1 | Word8 = 0, Word4 = 0)	33.4%
P(engaged = 1 | Word8 = 1, Word4 = 0)	82.3%
P(engaged = 1 | Word8 = 0, Word4 = 1)	22.3%
P(engaged = 1 | Word8 = 1, Word4 = 1)	24.5%

Table 2 shows that "engaged" is strongly affected by "Word8", but not under condition "Word4" =1 is given. The result of the inference tells that "Word8" is not preferable for being used with "Word4" for making marketing messages be effectively engaged with the product. As "Word4" has similar relationships among other nodes, the network suggests, the maker should pay attention in using "Word4".

However, some of the probabilistic inference with the network in Figure 2 do not coincide with the actual search result on Twitter. For example, as "Word18" ("taste") is a commonly used word, the combination with other taste-related words, such as "Word8" ("dry") would be a candidate for appealing the product on SNS. Actually, the result of the inference is shown in Table 3.

Table 3. A case of the inference not correspondent with the actual data

| Case1 | P(engaged = 1 | Word8 = 0, Word18 = 0) | 78.1% |
| Case2 | P(engaged = 1 | Word8 = 0, Word18 = 1) | 80.8% |

Table 3 shows, tweets are likely to be engaged with the product when "Word8" is used with "Word18". On the other hand, actual search on Twitter is performed under these two search words,

Conditon1: correspondent to Case1: the product and Word8

Conditon2: correspondent to Case2: the product and Word8 and Word18

The emerging ratio of tweets that match condition 1 is 58.0%. As for condition 2 is 9.0%. The actual search result on Twitter tells "Word18" should not be used with "Word8" in order to make "Word8" engaged with the product.

The cause of this disagreement would reside in the characteristic of Bayesian network. In this case, the node "engaged" is a kind of explained variable, which has different role from other nodes acting as explanatory variables. But Bayesian network usually handles all these nodes equally. Therefore, some adjustment will be required.

3. PROPOSED METHOD

To resolve the issue in Section 2, the authors propose to optimize Bayesian network by configuring initial state which reflects and emphasizes relationships between the target nodes and others. The authors propose to configure initial state to Bayesian network leveraging the result of Random forest, as it is a proven method for analyzing the influence of explanatory variables upon explained variable. The initial state consists of a few nodes around the target node and several edges between these nodes and the target. The former is called "initial nodes" and the latter is called "initial edges"

in this paper. In this chapter, detailed steps of the proposed method are described. The outline of the proposed method is shown in Figure 3.
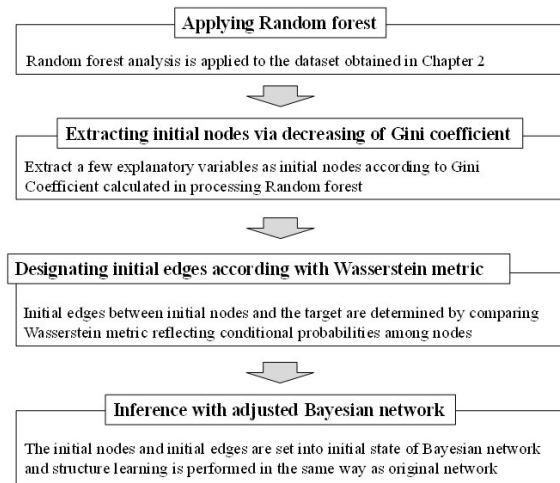

Figure 3. Outline of proposed method

3.1 Applying Random forest

At first Random forest analysis is applied to the dataset, in which "engaged" is explained variable and 18 engaged keywords ("Word1" - "Word18") are explanatory variables. As an analyzing tool, "randomForest" library of R is used in this study. Random forest is an advanced algorithm based on decision trees,[17] in which a lot of trees are generated according with randomly selected explanatory variables and the result is obtained as major vote of those trees. Therefore, these two parameters should be given properly in advance.

Parameter1: Number of explanatory variables selected while generating trees

Parameter2: Total number of trees generated

Along with the result of grid searching approach,[18] parameter1 is set to 4 and parameter 2 is set to 500 in this study. The dataset is split into train data (80% of 1046 records) and the remining is left for out of bag check. The estimated error ratio in out of bag check is 29.2%, which is higher than usual task for decision making. Because it is not decisively determined whether a tweet is engaged or not in this example.

3.2 Extracting initial nodes via decreasing of *Gini* coefficient

The second step of the proposed method is to extract a few explanatory variables as initial nodes. Initial nodes should be explanatory variables which have more influence on the target (explained variable). While processing Random forest,

Gini coefficient[19] as defined in Equation 1 is calculated for a node in each tree.

$$Gini(i) = \sum_k p(i, engaged = k) \times (1 - p(i, engaged = k))$$

(1)

$Gini(i)$: *Gini* coefficient of node i

$p(i, engaged = k)$: frequency ration of record within node i, of which value of "engaged" is k (=0 or 1)

Furthermore, *Gini* decrease according to an explanatory variable is defined as in Equation 2.

$$GiniDec(w) = \sum_{i \in Node(w)} N(i) \times Gini(w) - N(left(i)) \times Gini(left(i)) - N(right(i)) \times Gini(right(i)) \quad (2)$$

$GiniDec(w)$: *Gini* decrease according to explanatory variable w

$Node(w)$: a set of nodes split by w

$N(i)$: number of records in node i

$left(i)$: left child node of node i split by w

$right(i)$: right child node of node I split by w

Gini coefficient represents impurity of records within a node according to classification via the target. *Gini* decrease represents how much impurity is improved by generating trees according with an explanatory variable. That means explanatory variables with higher *Gini* decrease have more influence on the target. By calculation the mean of *Gini* decreases across the trees generated in Random forest, importance of explanatory variables can be compared. The mean decrease *Gini* values of engaged keywords are shown in Figure 4.
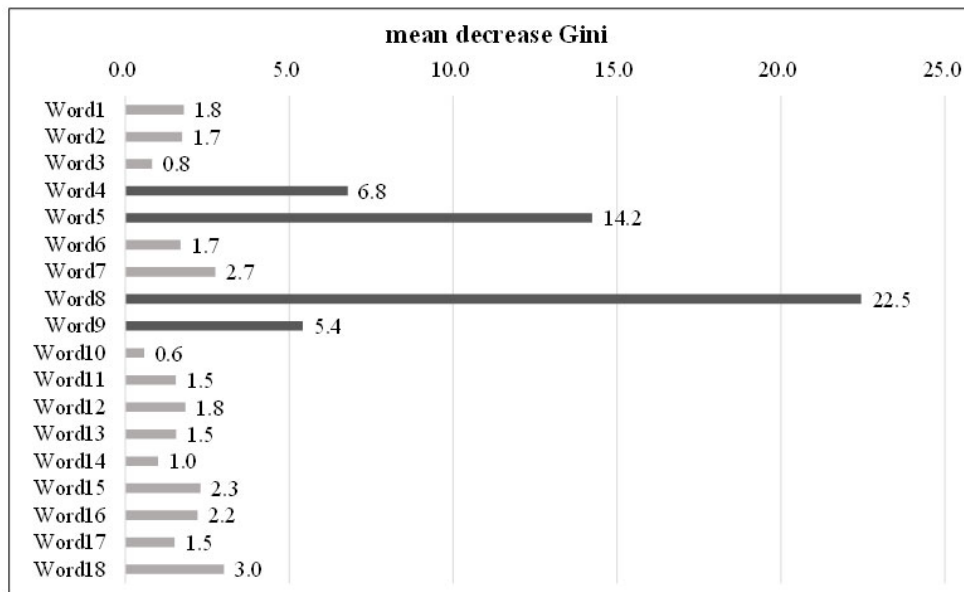

Figure 4. The values of mean decrease *Gini* of engaged keywords

As highlighted in Figure 4, Word4, Word5, Word8 and Word9 have relatively more influence on the target. Therefore, these four explanatory variables are selected as initial nodes.

3.3 Designating initial edges according with Wasserstein metric

The latter half of configuring initial state is to designate initial edges between initial nodes and the target. As shown in Figure 5, in Bayesian network, an edge from Node A to Node B represents that the conditional probability of Node

A given Node B is defined and vice versa.

In order to properly reflect the influence of initial nodes on the target, initial edges also should be defined leveraging the result of Random forest analysis. If one direction is more appropriate than the other in terms of stochastic model, its conditional probability is also more similar to the distribution of the dataset than the other. As described in Section 3.1, 20% of the dataset is remained for out of bag check. By using this remained data, predicted result of the target

node can be obtained. Then two values of the conditional probability for initial nodes are obtained. One is from the remained data (true distribution) and the other is from predicted data (learned distribution). The purpose here is to determine which is closer between the conditional probability in the left of Figure 5 or in the right when true distribution and learned distribution are compared. Kullback-Leibler divergence[20] is often used to compare several probability distributions, but it can't be applied in this case. Because it is a measure for bringing learned distribution closer to a specific true distribution and does not provide the distance between two different distributions. Wasserstein metric,[21]

which is one of methods in the area of optimal transport, is commonly used for defining the distance between two different probability distributions. Given two probability distributions A and B, Wasserstein metric of p-th order is described as in Equation 3.

$$WS(A, B) = \inf E(d(A, B))^p \tag{3}$$

d: the distance defined in the domain of the two probability distributions.

E: Expected value of probability distribution

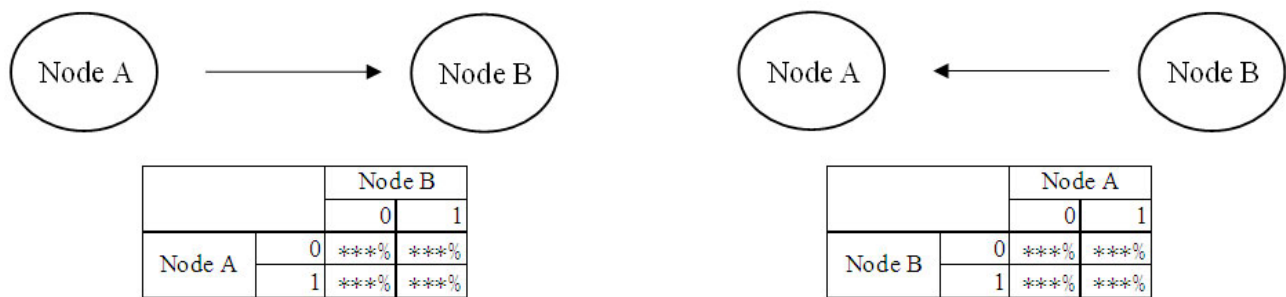inf: Operation to calculate the infimum

Figure 5. Relationship between direction of edge and conditional probability

As the domain of probability distributions are node values which are 0 or 1 in this case, it is sufficient to calculate the absolute value among each random variable. Therefore, the order of Wasserstein metric is configured as 1. For calculating Wasserstein metric, "wasserstein1d" library of R is used in this study. Here, let $WS*(nodeX, nodeY)$ be Wasserstein metric between true distribution and learned distribution in terms of the conditional probability according with an edge from nodeX to nodeY.

Smaller value of $WS(A, B)$ means the two probability distributions A and B are closer. Therefore, the direction of edge between the target ("engaged") and initial edges ("Word4", "Word5", "Word8", "Word9") are determined by comparing $WS*(engaged, word*)$ and $WS*(word*, engaged)$. The corresponding edge of smaller $WS*$ value should be designated as initial edges. The result is shown in Table 4.

Table 4. Wasserstein metric and designation of initial edges

Item	WS*(word*, engaged)	WS*(engaged, word*)	designated direction of edge
Word4	0.269	0.071	engaged ⇒Word4
Word5	0.304	0.240	engaged ⇒Word5
Word8	0.217	0.260	Word8 ⇒engaged
Word9	0.010	0.006	engaged ⇒Word9

3.4 Inference with adjusted Bayesian network

By the procedure in Section 3.3, the initial nodes and initial edges are obtained. In the final step, these two conditions are set into initial state of Bayesian network and structure learning is performed in the same way in Section 2. The result of Bayesian network analysis with initial state is shown in Figure 6.

The target node ("engaged") is directly connected with engaged keywords such as Word4, Word5, Word7, Word8, Word9, Word17. Different form ordinal Bayesian network in Section 2, the edge between Word4 and the target is opposite and the edge between Word18 and the target is eliminated. The initial state leveraging the result of Random Forest analysis makes this difference. As described in the next chapter 4, Bayesian network with initial state coincides well with actual search result on Twitter. Therefore, Bayesian network in Figure 6 is considered to be adjusted for decision making as for the target node.

4. RESULT IN EXAMPLE CASE

For confirming effectiveness of adjusted Bayesian network in Figure 6, probabilistic inference as for "Word8" and "Word18" in the same ways as in chapter 2 is performed. The result is shown in Table 5.
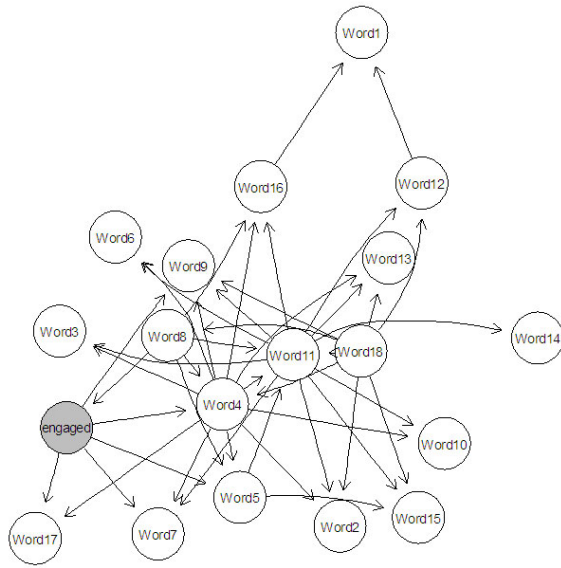
Figure 6. Result of Bayesian network with initial state

As described in Section 2, the actual search result on Twitter shows "Word18" should not be used with "Word8". Original

network falsely suggests that a tweet is more likely to be engaged with the product when "Word8" and "Word18" are used together. Adjusted network does not recommend this combination for improving engaged ratio. Although it would be best if the value of Case2 is much lower than that of Case1 in Adjusted network, the adjusted one provides better result than the original.

Table 5. Improvement in problematic case

Cases		Original network	Adjusted network
Case1	P(engaged = 1 | Word8 = 1, Word18 = 0)	78.1%	78.9%
Case2	P(engaged = 1 | Word8 = 1, Word18 = 1)	80.8%	78.9%

In the same way, as for nodes which are directly connected to the target ("engaged") with the same directions, such as "Node5","Node7", "Node8", "Node9" and "Node17", the comparison of the inference between original and adjusted network and the emerging ratio in actual search on Twitter are shown in Table 6.

Table 6. Comparison between original and adjusted network with actual search result

	Original network			Adjusted network			Emerging ratio on Twitter search	
	Word18 =0 (*1)	Word18 =1 (*2)	Diff (*2) - (*1)	Word18 =0 (*3)	Word18 =1 (*4)	Diff (*4) – (*3)	Without Word18	With Word18
Word5	62.7%	61.8%	-1.0	64.1%	59.6%	-4.4	39.0%	3.0%
Word7	15.8%	14.2%	-1.6	16.7%	12.8%	-3.9	5.0%	1.0%
Word8	78.1%	80.8%	2.7	78.9%	78.9%	0.0	58.0%	9.0%
Word9	0.4%	1.2%	0.8	0.4%	1.1%	0.7	1.0%	1.0%
Word17	1.8%	2.4%	0.6	2.0%	2.0%	0.1	0.0%	0.0%

The search conditions used for "Emerging ratio on Twitter search" column are as follows. "Word*" is "Word5", "Word7", "Word8", "Word9" or "Word17".

Conditon1: correspondent to the item "without Word18": the product and Word*

Conditon2: correspondent to the item "with Word18": the product and Word* and Word18

The actual search result shows, "Word18" is not appropriate for combined use with other words for making marketing messages engaged with the product. The Diff column in Table 6 of adjusted network shows less value than that of original network. That means adjusted network less recommend the use of "Word18" with other words than original one. In this way, Bayesian network gives suggestions of combination use of multiple keywords by probabilistic inference.

Furthermore, adjusted Bayesian network leveraged with the result of Random forest matches better than ordinal one in terms of decision making concerned with particular target node.

5. CONCLUSION

SNS has now become popular and consumers are spreading impression and evaluation of various products with their own keywords across the Internet. Therefore, it is necessary for makers to understand which keywords should be used on SNS for appealing their products. For extracting these keywords, several methods such as leveraging Co-Occurrence network analysis or Word2Vec are already proposed. Even after applying these methods, there is still an issue of how to select actually used keywords. As the budget for marketing and advertising is often limited in business, makers need to pick up especially effective keywords.

For this purpose, the authors apply Bayesian network analysis. Bayesian network analysis has advantage to visualize relationships among keywords, which is useful for considering combination of several keywords in appealing products. However, it sometimes occurs, even when the network suggests keyword A and keyword B should be used together, the actual search result on SNS shows the opposite result. In an example of this study, the network suggests the probability of a tweet being related with the product becomes 80.8%, when a keyword "dry" is used with "taste", increasing from 78.1% in which "dry" is solitarily used. But the emerging ratio of tweets including both "dry" and "taste" on Twitter is 9.0%, much smaller than 58.0%, which is the case of solitary use of "dry". The result of the inference does not coincide with the actual data on SNS.

This case includes particular relationships between explained variable (a node which indicates whether a tweet is related with the product) and explanatory variables (keywords as "dry" or "taste"). But Bayesian network usually handles all these nodes equally. The cause of disagreement above would reside in this difference. The authors leverage Random forest to solve this issue. Random forest is a proven method for analyzing the influence of explanatory variables upon explained variable. The authors propose to configure initial state which consists of a few nodes around explained variable and several edges between them. These nodes are extracted by measuring mean decrease of *Gini* coefficient calculated with decision trees of Random forest, because explanatory variables with much influence on explained variable show significant decrease of the coefficient. On the other hand, edges are designated based on likelihood measured by similarity of conditional probability distribution between actual data and predicted result of Random forest. The similarity is calculated with Wasserstein metric. The Initial state are given as a start point for the construction of Bayesian network. This is called adjusted Bayesian network in this study.

As for the case including keywords "dry" and "taste" above, the probability of a tweet being related with the product does not increase regardless of the combination with them, when applying the inference of adjusted Bayesian network. That means, the result of inference with adjusted Bayesian network coincides well with the actual data on SNS. The authors compare other result between original and adjusted network and the effectiveness of adjusted network is confirmed.

6. DISCUSSION FOR FUTURE WORK

In this study there is only one target node to be considered. But in real business scenes, there are a lot of situations in which several target nodes exist. For example, in product development management, several KGIs such as annual sales, market share and the number of patents should be considered simultaneously.[21–24] Further experiments are expected to apply the proposed method to cases with multiple target nodes.
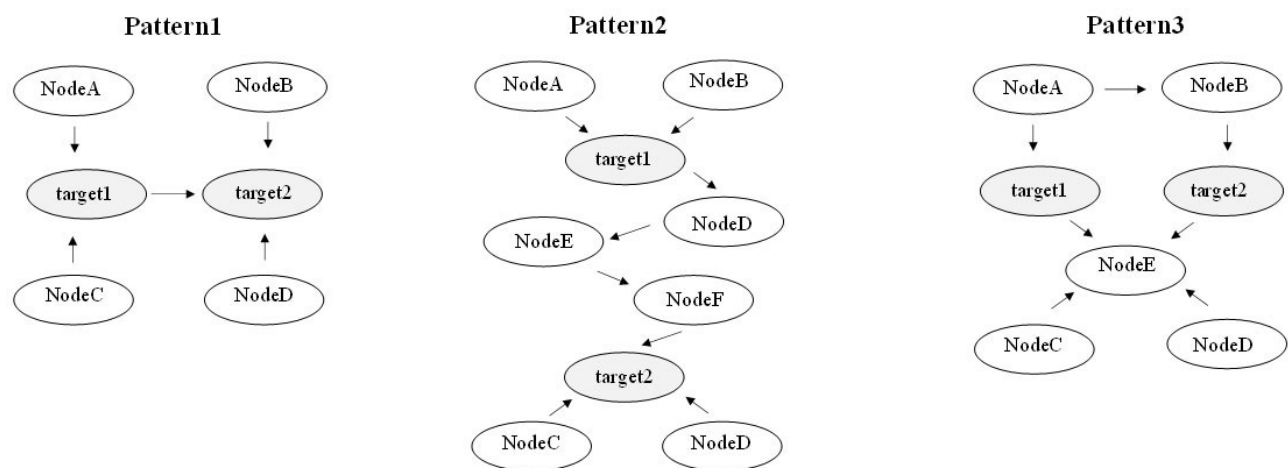

Figure 7. Three patterns of relationship between two target nodes

There are three patterns in relationship between two target nodes shown in Figure 7. Pattern 1 in Figure 7 is the case in which two targets are directly connected. In this case, it would be one of practical solutions to merge these two targets into one node and apply the method in this study. Contrary to Pattern 1, two nodes are separated by efficiently multiple nodes in Pattern 2. In this case, the proposed method can be applied to each target node separately, because there is no concern that initial conditions for each node affect each other. But in Pattern 3, in which two targets share a

few nodes as neighbors, some problems might occur. For example, in Pattern3 of Figure 7, if the result of Random forest analysis for target 1 suggests opposite directions for "NodeA" and "NodeE", cyclic direction including "NodeE", "target1", "NodeA", "NodeB", and "target2" is generated. In other cases, such as there are two edges from NodeE to both target1 and target2 respectively, the same situation might occur. In Bayesian network, cyclic direction is not allowed. Therefore, further adjustment will be required.

CONFLICTS OF INTEREST DISCLOSURE

The authors declare that they have no competing interests.

REFERENCES

[1] Iwashita M. A Proposal of Matching Algorithm for New Type of Advertisement Business Model. Procedia Computer Science. 2019; 159: 1966-1975. https://doi.org/10.1016/j.procs.2019.09.369

[2] Haan W, Kaltenbrunner G. Anticipated growth and business cycles in matching models. Journal of Monetary Economics. April 2009; 56(3): 309-327. https://doi.org/10.1016/j.jmoneco.2009.03.003

[3] Iwashita M, Tanimoto S, Tsuchiya K. Framework of highly secure transaction management for affiliate services of video advertising. Procedia Computer Science. 2018; 126: 1802-1809. https://doi.org/10.1016/j.procs.2018.08.097

[4] Garg M, Kumar M. Identifying influential segments from word co-occurrence networks using AHP. Cognitive Systems Research. January 2018; 47: 28-41. https://doi.org/10.1016/j.cogsys.2017.07.003

[5] Angelo L, Stefan P, Fratocchi L, et al. An AHP-based method for choosing the best 3D scanner for cultural heritage applications. Journal of Cultural Heritage. November-December 2018; Volume 34, Pages 109-115. https://doi.org/10.1016/j.culher.2018.03.026

[6] Mkolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space. Computation and Language. September 2013; arXiv:1301.3781 [cs.CL].

[7] Fukui K, Miyazaki T, Ohira M. Suggesting Questions that Match Each User's Expertise in Community Question and Answering Services: from 2019. 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). July 8-11 2019. https://doi.org/10.1109/SNPD.2019.8935747

[8] Jing X, Wang P, Rayz J. Discovering Attribute-Specific Features From Online Reviews What Is the Gap Between Automated Tools and Human Cognition? Software Science and Computational Intelligence. 2018: 24. https://doi.org/10.4018/IJSSCI.201804010

[9] Kita E, Zuo Y, Harada M, et al. Application of Bayesian Network to stock price prediction. Artificial Intelligence Research. December 2012; 11(2). http://dx.doi.org/10.5430/air.v1n2p171

[10] Velasco F. A Bayesian Network approach to diagnosing the root cause of failure from trouble tickets. Artificial Intelligence Research. December 2012; 11(2). http://dx.doi.org/10.5430/air.v1n2p75

[11] Harris T. Default definition selection for credit scoring. Artificial Intelligence Research. September 2013. http://dx.doi.org/10.5430/air.v2n4p49

[12] Khalilia M, Chakaborty S, Popescu M. Predicting disease risks from highly imbalanced data using random forest. July 2011; Article number 51. https://doi.org/10.1186/1472-6947-11-51

[13] Iwakami Y, Takuma H, Iwashita M. Improving Matching Process with Expanding and Classifying Criterial Keywords leveraging Word Embedding and Hierarchical Clustering Methods. Submitted to The Review of Socionetwork Strategies. May 2020.

[14] Glover F. Tabu Search-Part I. ORSA Journal on Computing. August 1989; 11(3). https://doi.org/10.1287/ijoc.1.3.190

[15] Bai X, Padman R. Tabu Search Enhanced Markov Blanket Classifier for High Dimensional Data Sets: Part of the Operations Research/Computer Science Interfaces Series. 1993-2018; 29. https://doi.org/10.1007/0-387-23529-9_22

[16] Heckerman D, Breese J. Causal independence for probability assessment and inference using Bayesian networks: IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans. November 1996; 26(6). https://doi.org/10.1109/3468.541341

[17] Ali J, Khan R, Ahmad N, et al. Random Forests and Decision Trees. IJCSI International Journal of Computer Science Issues. September 2012; 9(5).

[18] Jimenez A, Lazaro J, Dorronsoro J. Finding Optimal Model Parameters by Discrete Grid Search: Innovations in Hybrid Intelligent Systems. 2007:120-127. https://doi.org/10.1007/978-3-540-74972-1_17

[19] Yizhaki S. Relative Deprivation and the Gini Coefficient. The Quarterly Journal of Economics. May 1979; 93(2). https://doi.org/10.2307/1883197

[20] Shelens J. Notes on Kullback-Leibler Divergence and Likelihood Theory: Cornell University. August 2014; arXiv:1404.2000

[21] Vallender SS. Calculation of the Wasserstein Distance Between Probability Distributions on the Line: Theory of Probability & Its Applications. 2006; 18(4): 784-786. https://doi.org/10.1137/1118101

[22] Takuma H. Consideration of feasibility to support function for value indicator management by mathematical analysis for implementation of IoPM. J. Intern. Assoc. of P2M. 2018; 13(1): 249-259.

[23] Takuma H, Hiyama M. Discussion of the value indicators for associating projects with programs. J. Intern. Assoc. of P2M. 2015; 10(1): 23-34.

[24] Takuma H, Iwakami Y. Extraction of fundamental KPIs in new product development using Bayesian network analysis. Journal of International Association of P2M. 2019; 14(1): 446-455. https://doi.org/10.20702/iappmjour.14.1_446

[25] Takuma H, Iwakami Y. Consideration of Fundamental KPIs and Their Relationship with Environmental Protection in New Product Development Using Bayesian Network Analysis. From 2019 International Conference on Information Management and Technology (ICIMTech). August 2019. https://doi.org/10.1109/ICIMTech.2019.8843762