

PREFERENCE SIMILARITY NETWORK CLUSTERING CONSENSUS GROUP DECISION MAKING MODEL IN ANALYSING CONSUMERS' REVIEWS AND SELECTING SAMPLES OF PRODUCT

Nur Syahera Ishak¹ and Nor Hanimah Kamis²

^{1, 2}Department of Mathematics, Faculty of Computer and Mathematical Sciences, Universiti
Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia.

¹syeeeraishak@gmail.com, ²norhanimah@fskm.uitm.edu.my

ABSTRACT

In recent years, the integration of notions from Social Network Analysis (SNA) into decision making context is rapidly increased. One of the feasible procedures is Preference Similarity Network Clustering Consensus Group Decision Making model, where it is capable to improve the effectiveness and efficiency of decision making process. We utilize this approach in analysing consumers' reviews and selecting the best sample of laboratory products. This is the first effort of applying this model in real life situation. The referred approach is capable of measuring the similarity of consumers' reviews, visualize their similarities in the form of network structure, partition them into subgroups, measure their group consensus level and select the best sample of product. The obtained results provide essential information to the laboratory, manufacturer or a company to improve the quality of product and further plan on the marketing strategy, advertisement and research development. Generally, this model can be used as an alternative tool in solving decision making problems, especially in analysing reviews and selection of alternatives.

Keywords: Preference similarity, Social Network Analysis (SNA), Clustering algorithm, Consensus Group Decision Making (CGDM), Product reviews, Selection problems.

Received for review: 15-09-2020; Accepted: 20-10-2020; Published: 28-10-2020

1. Introduction

Making decision is a common skill for most people in solving daily life problems. For the purpose of making a great decision, the problem, situation, goal, risk and options are among factors that should be taken into account. When more people or alternatives (criteria) involve in decision making process, there are situations where this group of decision makers face problems due to different opinions or inappropriate consideration of individual preferences. Lack of consensus in group decision making may lead to an irrational solution (Tang *et al.*, 2019).

Therefore, consensus among group members must be in a sufficient state before the final decision is made, thus an overall agreed solution can be achieved. Consensus group decision making (CGDM) models can be used as procedures to solve the inconsistency decision among decision makers or difficulty in achieving full agreement of the whole group. Recent studies

discussing on CGDM can be found in Liu *et al.* (2017), Dong *et al.* (2018), Ma *et al.* (2019) and Xu *et al.* (2020).

Current development of CGDM procedures merge relevant concepts and theories from other research areas in decision making context. One of the well-known collaborated knowledge is Social Network Analysis (SNA), where basically focuses on the analysis of relation between social entities in a network structure. Several studies that integrate notions or concepts of SNA in decision making perspective are appeared in Zhang *et al.* (2018), Kamis *et al.* (2018a, 2018b, 2019), Liu *et al.* (2019), Tian *et al.* (2019) Urena *et al.* (2019).

Other than notions from SNA, clustering algorithms become necessary procedures involved in decision making process. Large number of decision makers may cause a process of achieving consensus and final solution complicated, tedious and time consuming (Zhong & Xu, 2020). Clustering algorithms are able to partition or reduce large size of data into subgroups based on common characteristics, such as similarity of decision makers' opinions and trust relationships (Ma *et al.*, 2019; Khedmati & Azin, 2020). It also proved that clustering-similarity based method can be utilized to improve the consensus state in CGDM problems (Du *et al.*, 2020).

In this study, we aim to apply the Preference Similarity Network Clustering Consensus Group Decision Making model (Kamis *et al.*, 2018) in solving real life problem. For the purpose of analysing the product reviews from consumers, this approach is able to provide informative results including the visualisation of similarity network, optimal clustering solution based on their similarity of preferences and group consensus and ranking of the preferred product samples. The best samples of product can be selected and this information might help the laboratory or manufacturer to produce and market the product. The outline of this paper begins with Section 1 (Introduction), followed by Section 2 (Methodology), Section 3 (Results and Discussion) and Conclusion in Section 4.

2. Methodology

In this study, we use a data from 30 consumers of 6 product samples for the laboratory test reviews. We assume that they are experienced in applying those samples in a specific duration and need to give their reviews based on a set of questionnaires.

We utilize the Preference Similarity Network Clustering Consensus Group Decision Making model, introduced by Kamis *et al.* (2018a) in order to analyse the reviews and select the best sample of products used by consumers. The consecutive steps are presented as follows:

(1) Identify the consumers, $E = \{e_1, e_2, \dots, e_{30}\}$ and the product samples, $A = \{a_1, a_2, \dots, a_6\}$.

(2) Evaluate the product samples based on the reciprocal preference relation.

Definition 1. A reciprocal preference relation on A is a fuzzy binary relation R where the preference intensity of alternative i over alternative j , $\mu_R(a_i, a_j) = r_{ij}$, verify that $\mu_R(a_i, a_i) = 0.5 \forall a_i \in A$ and $r_{ij} + r_{ji} = 1, \forall a_i, a_j \in A$.

(3) Extract the reciprocal preference relation entries into the intensity preference vector.

Definition 2. The intensity preference vector of a reciprocal preference relation $R = (r_{ij})_{n \times n} \in R_{n \times n}$ is the vector of dimension $n(n-1)/2$, U with components the elements above its main diagonal:

$$V = (r_{12}, r_{13}, \dots, r_{1n}, r_{23}, \dots, r_{2n}, \dots, r_{(n-1)n}) = (u_1, u_2, \dots, u_{n(n-1)/2}).$$

(4) Measure the similarity of consumer's preferences using cosine similarity function.

Definition 3. The measure of cosine similarity between preference of consumer, e^r and consumer, e^s is:

$$P^{rs} = P(U^r, U^s) = \frac{\sum_{i=1}^{n(n-1)/2} (u_i^r \cdot u_i^s)}{\sqrt{\sum_{i=1}^{n(n-1)/2} (u_i^r)^2} \cdot \sqrt{\sum_{i=1}^{n(n-1)/2} (u_i^s)^2}}.$$

- (5) Construct the undirected weighted consumer preference similarity network.

Definition 4. An undirected weighted preference similarity network is an ordered triple, $G = \langle E, T, P \rangle$ which consist of a set of nodes (consumers), $E = \{e_1, e_2, \dots, e_m\}$, a set of ties between consumers, $T = (t_{12}, t_{13}, \dots, t_{1m}, t_{23}, \dots, t_{2m}, \dots, t_{(m-1)m})$ and a set of weights based on cosine similarity degrees, $P = (p_1, p_2, \dots, p_{m(m-1)/2})$.

- (6) Cluster the consumers based on their similarity of preferences using the agglomerative hierarchical clustering with complete linkage function. The details of the algorithm can be found in Kamis *et al.* (2018a, 2018b, 2019).

- (7) Compute the internal cohesion degree.

Definition 5. The measure of internal cohesion degree, (δ_{intC}) of cluster K_{lt} at level α_l , $\delta_{intC}(K_{lt})$ is:

$$\delta_{intC}(K_{lt}) = \frac{\sum_{i \in K_{lt}} \sum_{j \in K_{lt}} P^{ij}}{(\#K_{lt})^2},$$

where $L = \{\alpha_l; l = 2, \dots, m - 1\}$ is the set of all distinct α -level of the agglomerative hierarchical clustering solution, $K_l = \{K_{lt}; t = 1, \dots, l\}$ is the set of clusters at α -level and $\#K_{lt}$ is the cardinality of K_{lt} .

- (8) Calculate the external cohesion degree.

Definition 6. The measure of external cohesion degree, (δ_{extC}) of cluster K_{lt} at level α_l , $\delta_{extC}(K_{lt})$, is:

$$\delta_{extC}(K_{lt}) = \frac{\sum_{i \in K_{lt}} \sum_{j \notin K_{lt}} P^{ij}}{\#K_{lt}(m - \#K_{lt})}.$$

- (9) Measure the cluster consensus degree.

Definition 7. The measure of consensus degree of cluster K_{lt} at level α_l , $\delta_{KC}(K_{lt})$, is:

$$\delta_{KC}(K_{lt}) = \frac{\#K_{lt}(\delta_{intC}(K_{lt}) - \delta_{extC}(K_{lt}))}{m} + \delta_{extC}(K_{lt}).$$

- (10) Determine the level consensus degree.

Definition 8. The measure of level consensus degree, $\delta_{GC}(l)$, is:

$$\delta_{GC}(l) = \frac{\sum_{t=1}^l \delta_{KC}(K_{lt})}{l}.$$

- (11) Identify the group consensus degree based on the optimal agglomerative hierarchical clustering level, having the maximum degree of consensus. If the pre-determined consensus threshold is set up, then the optimal agglomerative hierarchical clustering level is chosen according to the threshold value.

- (12) Aggregate the individual consumer preferences into a collective one.

Definition 9. An IOWA operator of dimension n is a function $\Phi_W: (\mathbb{R} \times \mathbb{R})^n \rightarrow \mathbb{R}$ is associated with a set of weighting vector, $V = (v_1, \dots, v_m)$, verify that $v_i \in [0,1]$ and $\sum_{i=1}^n v_i = 1$, and aggregate the set of second arguments $\{\langle u_1, r_1 \rangle, \dots, \langle u_n, r_n \rangle\}$ with following expression:

$$\Phi_W\{\langle u_1, r_1 \rangle, \dots, \langle u_n, r_n \rangle\} = \sum_{i=1}^n v_i \cdot r_{\sigma(i)}.$$

The weighting vector for IOWA operator is determined by using:

$$v_i = Q\left(\frac{i}{m}\right) - Q\left(\frac{i-1}{m}\right), i = 1, \dots, m \tag{1}$$

and the collective consumer preferences is determined using:

$$r_{ij}^c = \Phi_W(\langle X(e_1), r_{ij}^1 \rangle, \dots, \langle X(e_m), r_{ij}^m \rangle). \tag{2}$$

- (13) Determine the Quantifier Guided Choice Degree (QGDD) and rank the samples of product.

Definition 10. Given a collective preference relation, $R^c = r_{ij}^c$ towards a set of alternatives $A = \{a_1, a_2, \dots, a_n\}$, the quantifier guided dominance degree, QGDD (a_i), quantifies the dominance of alternative a_i over other alternatives in a fuzzy majority with the following expression:

$$QGDD(a_i) = \Phi_Q(r_{ij}^c, j = 1, \dots, n, j \neq i),$$

where Φ_Q is an OWA operator guided by the linguistic quantifier Q which represents the fuzzy majority concept. The higher the QGDD value, the most preferred sample is chosen.

3. Results and Discussion

Referring to the consecutive steps in the Methodology section, the result from Step (1) until (5) can be visualised in the form of 2-dimensional scaling, as presented in Figure 1.

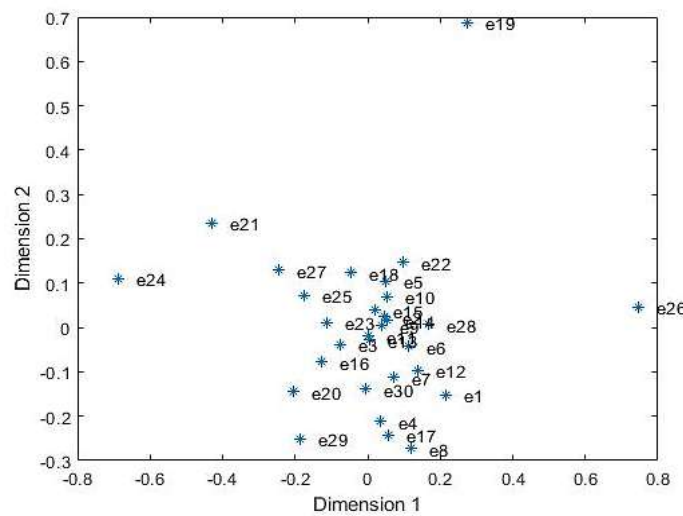


Figure 1. Visualisation of consumers' preferences similarities in 2-dimensional scaling.

Figure 1 displays the position of consumers to each other based on their similarity of reviews. It is clearly shown that the position of e_{19} , e_{21} , e_{24} and e_{26} are isolated and farther from other consumers. It is because these four consumers have different opinions and low preference similarity degree between other consumers in the network. Conversely, other consumers' positions are close respectively, representing the degree of similarity of each of them are higher. Meaning that they have similar reviews over samples of product and group together at the centre of the dimensional scaling diagram.

For this case study, we pre-determined the consensus threshold at 0.8. This value represents the sufficient level of group agreement before the final solution is obtained. We are not considering maximum level of group consensus because this selection problem focuses on the optimal number of clusters and the preferred alternatives, rather than the consensus measure itself.

By referring to Table 1, we choose the optimal cluster at level 13, with a sufficient consensus degree at 0.8. This shows that the total number of clusters for 30 consumers are 13, where 6 consumers are grouped in cluster 1 with internal cohesion index of 0.97, which more than the external cohesion index of 0.86. This result presents the closeness of 6 consumers' preferences within the group members are higher compared to the members outside cluster 1. The same situation happens to cluster 2 until 7.

As shown in Figure 1, e_{19} , e_{21} , e_{24} and e_{26} are isolated, while consumer e_{20} and e_{29} are positioned farther from the others. This situation is supported by the result in Table 1, where these consumers are individually partitioned in different groups, cluster 8 until 13.

Table 1. The Optimal Clustering Solution with The Consensus Threshold = 0.8.

α -levels	Clusters	Cluster members	Internal Cohesion	External Cohesion	Cluster Consensus	Group Consensus
13	1	$e_3, e_5, e_9, e_{11}, e_{13}, e_{15}$	0.97	0.86	0.88	0.80
	2	$e_4, e_6, e_7, e_{17}, e_{30}$	0.94	0.85	0.87	
	3	e_2, e_{10}, e_{14}	0.99	0.84	0.86	
	4	e_{22}, e_{28}	0.98	0.85	0.86	
	5	e_1, e_8, e_{12}	0.95	0.81	0.83	
	6	e_{16}, e_{18}, e_{23}	0.95	0.86	0.87	
	7	e_{25}, e_{27}	0.99	0.85	0.86	
	8	e_{20}	1	0.82	0.83	
	9	e_{29}	1	0.79	0.80	
	10	e_{21}	1	0.71	0.72	
	11	e_{24}	1	0.70	0.71	
	12	e_{19}	1	0.64	0.65	
	13	e_{26}	1	0.69	0.70	

The consensus degree of each cluster is determined and the group consensus for the optimal cluster level is computed (0.8). This value expresses sufficient level of agreement of the group and all individual consumers' reviews are appropriately taken into consideration throughout the decision making process. The consensus measure is necessary to preserve the satisfaction of each decision maker on the final solution, thus post-argument can be avoided.

The clustering solution can be used by the laboratory or manufacturer to further analyse the characteristic of consumers for each cluster, such as skin type, age, gender and many more. This information is important for them to focus on their targeted consumers and marketing strategy.

In Table 2, the ranking of preferred samples of product is presented. Based on the QGDD values, sample a_5 is the best sample of product chosen by consumers, followed by a_3 , a_2 , a_6 , a_1 and a_4 . From here, the laboratory or manufacturer can start to work together on how to further explore the selected sample in terms of quality, packaging, advertisement etc.

Table 2. The ranking of preferred product samples according to the QGDD values.

Samples, A	a_1	a_2	a_3	a_4	a_5	a_6
QGDD (i)	0.525	0.534	0.565	0.524	0.576	0.532
Ranking	5	3	2	6	1	4

This result reflects all consumers' reviews on the samples of product and agreed by them since the consensus level is sufficient.

4. Conclusion

This is the first effort of utilizing the Preference Similarity Network Clustering Consensus Group Decision Making model, introduced by Kamis *et al.* (2018a) in a real-life problem. We proved that this approach is relevant to be used for the purpose of analysing reviews and solving selection problem(s). The analysis can be one of the informative ways for the laboratory, manufacturer or company to improve their products, marketing plan and research development.

The referred model integrated several Social Network Analysis (SNA) notions and clustering algorithm in decision making perspective. Consensus measure improved the effectiveness of decision by ensuring the satisfaction and agreement of all individual opinions throughout the decision-making process. This is necessary in order to prevent post-argument situation among group of decision makers.

References

- Dong, Y., Zha, Q., Zhang, H., Kou, G., Fujita, H., Chiclana, F., & Herrera-Viedma, E. (2018). Consensus reaching in social network group decision making: Research paradigms and challenges. *Knowledge-Based Systems*, 162, 3–13.
- Du, Z., Luo, H., Lin, X., & Yu, S. (2020). A trust-similarity analysis-based clustering method for large-scale group decision-making under a social network. *Information Fusion*, 63, 13–29.
- Kamis, N. H., Chiclana, F., & Levesley, J. (2018a). Preference similarity network structural equivalence clustering based consensus group decision making model. *Applied Soft Computing*, 67, 706–720.
- Kamis, N. H., Chiclana, F., & Levesley, J. (2018b). Geo-uniform consistency control module for preference similarity network hierarchical clustering based consensus model, *Knowledge-Based Systems* 162, 103-114.

- Kamis, N. H., Chiclana, F., & Levesley, J. (2019). An Influence-Driven Feedback System For Preference Similarity Network Clustering Based Consensus Group Decision Making Model, *Information Fusion* 52, 257-267.
- Khedmati, M., & Azin, P. (2020). An online portfolio selection algorithm using clustering approaches and considering transaction costs. *Expert Systems with Applications*, 159, 113546.
- Liu, Y., Liang, C., Chiclana, F., & Wu, J. (2017). A trust induced recommendation mechanism for reaching consensus in group decision making. *Knowledge-Based Systems*, 119, 221–231.
- Liu, B., Zhou, Q., Ding, R.-X., Palomares, I., & Herrera, F. (2019). Large-scale group decision making model based on social network analysis: Trust relationship-based conflict detection and elimination. *European Journal of Operational Research*, 275(2), 737–754.
- Ma, X., Zhao, M., & Zou, X. (2019). Measuring and reaching consensus in group decision making with the linguistic computing model based on discrete fuzzy numbers. *Applied Soft Computing*, 77, 135–154.
- Tang, X., Peng, Z., Zhang, Q., Pedrycz, W., & Yang, S. (2019). Consistency and consensus-driven models to personalize individual semantics of linguistic terms for supporting group decision making with distribution linguistic preference relations. *Knowledge-Based Systems*, 105078.
- Tian, Z., Nie, R., & Wang, J. (2019). Social network analysis-based consensus-supporting framework for large-scale group decision-making with incomplete interval type-2 fuzzy information. *Information Sciences*, 502, 446–471.
- Urena, R., Chiclana, F., Melancon, G., Herrera-Viedma, E. (2019). A social network based approach for consensus achievement in multiperson decision making. *Information Fusion*, 47, 72-87.
- Xu, W., Chen, X., Dong, Y., Chiclana, F. (2020). Impact of decision rules and non-cooperative behaviors on minimum consensus cost in group decision making, *Group Decision and Negotiation*, 1-22.
- Zhang, H., Palomares, I., Dong, Y., & Wang, W. (2018). Managing non-cooperative behaviors in consensus-based multiple attribute group decision making: An approach based on social network analysis. *Knowledge-Based Systems*, 162, 29–45.
- Zhong, X., & Xu, X. (2020). Clustering-based method for large group decision making with hesitant fuzzy linguistic information: Integrating correlation and consensus. *Applied Soft Computing*, 87, 105973.