 https://doi.org/10.1051/itmconf/20246701043

Probabilistic Soft Set Theory for Decision-Making in Choosing Korean Dramas

Era Setya Cahyati[1,*] and *Dedi* Rosadi[1]

[1]Universitas Gadjah Mada

Abstract. Decision-making involves subjective elements influenced by individual preferences and uncertainties. A flexible framework to handle these complications is offered by probabilistic soft set theory. Probabilistic soft sets is an extension of soft set. This mathematical framework deals with ambiguity and uncertainty in the process of making decisions. Recent results had been devoted for development the theoretical aspect of the method and only a little study has been done for the usage of the theory for decision-making. In this study, we provide two algorithms as the application of the probabilistic soft set theory for decision-making. For real application we use the algorithms to determine the Korean dramas to be watched. To apply this method, a set of Korean drama titles, which has release dates from 1 July 2023 until 31 December 2023, has been collected. Subsequently, using this method, the data is processed and analyzed to determine the chosen drama.

1 Introduction

Decision-making is a crucial part of every stage of life. In daily life, individuals are often confronted with various choices that require evaluation, consideration, and decisions. Whether it's small everyday decisions or significant ones with long-term impacts, the methods and processes of decision-making play a crucial role in determining the next steps.

The importance of decision-making is increasingly felt in the context of business and organizations, where decisions can significantly influence performance, strategy, and the existence of an entity. Effective decision-making requires a deep understanding of available information, critical thinking, and analytical skills. Moreover, in the era of rapid information technology, the ability to use data becomes essential in making optimal decisions.

The theory of soft sets is one approach to solving decision-making problems, with parameters serving as benchmarks for decision-making. Soft set parameters can be numbers, words, sentences, functions, or other forms, making them flexible in representing individual decision-making needs. The theory of soft sets was first introduced by [5]. Subsequently, properties related to soft set operations were later discussed in [7, 9, 11]. This theory has also been applied to decision-making algorithms, as seen in the work of Maji et al. in [6].

In everyday life, sets of data in the form of probabilities and rankings are encountered. However, research related to soft sets for decision-making only addresses binary assessments

*e-mail: era.setya.cahyati@ugm.ac.id

and the interval [0,1] but cannot be used for probability data. Therefore, further research combines soft sets with probability theory, known as probabilistic soft sets, introduced by [12]. Properties of set operations related to probabilistic soft sets have been studied by Gunduz et al. in [10].

Similar to soft sets, probabilistic soft sets can also be used for decision-making. The defined set operations' properties in probabilistic soft sets are beneficial in formulating decision-making algorithms. Then, in 2017, Fatimah et al. in [8] presented the theory and algorithm for making decisions using probabilistic soft sets. This algorithm will be employed in this paper to determine which Korean drama will be chosen.

The rest of this paper describes as follows. In section 2, we provide a short review of probabilistic soft set theory and algorithms for decision-making using this theory. In section 3, we provide a simple application of the theory for choosing the Korean dramas. Last section concludes the discussion.

2 Probabilistic Soft-Set

In this section we begin with the definition of soft set over a universe set. Let U be the universal set and E be the set of parameters. Let 2^U denotes the power set of U and $A \subset E$.

Definition 2.1 *[7] A pair (F, A) is called a soft set over U, where F is a function defined by*

$$F : A \to 2^U.$$

A soft set over U is a parameterized subsets of the universe U.

Example 2.2 *[7] Suppose that Mr.X consider to buy a house. Suppose that $U = \{h_1, \ldots, h_6\}$ be the set of houses under Mr.X consideration and E = {expensive, beautiful, cheap, wooden, in the green surroundings, in bad repair, in good repair} is the set of parameters. In this case, to define a soft set means to point out beautiful houses, expensive houses, and so on. The soft set (F, E) describes the "attractiveness of the houses" which Mr.X is going to buy.*

Next we give the definition of probabilistic set as a combination of soft sets theory and probability theory. Let $\mathcal{D}(U)$ be the set of all probability distribution on U.

Definition 2.3 *[8] A pair (F, A) is called a probabilistic soft set over U where F is a function from the set A to the set of all probability distributions on U, i.e., $F : A \to \mathcal{D}(U)$.*

In other words, probabilistic soft set is parameterized family of probability distributions for U. It means, for every $u \in U$ and $e \in E$, $F(e)(u) \in [0, 1]$. Therefore, for every index i, j, can be defined $P(a_{ij}) = F(e_j)(u_i)$ provided $0 \leq P(a_{ij}) \leq 1$ and $\sum_i P(a_{ij}) = 1$. The tabular representation can be found in Table 1 where the rows are the objects in U and the columns are the parameters in A.

Next we give the algorithm on decision-making based on Maji's algorithm.

Algorithm 2.1 *[8] Using choice value method, the steps of the method is given as follows:*

Step 1. *Consider the universal set U and the parameters set E.*

Step 2. *Consider $A \subseteq E$, which is the set of choice parameters.*

Step 3. *Consider (F, A) and denotes $P(a_{ij})$ as the entries of the tabular representation, where $\sum_i P(a_{ij}) = 1$.*

	e_1	e_2	$\cdots$
u_1	$P(a_{11})$	$P(a_{12})$	$\cdots$
u_2	$P(a_{21})$	$P(a_{22})$	$\cdots$
$\ldots$	$\ldots$	$\ldots$	$\ldots$
Total	$\sum_i P(a_{i1}) = 1$	$\sum_i P(a_{i2}) = 1$	$\cdots$

Table 1. Tabular Representation

Step 4. Find k for which $c_k = \max c_i$*, where* $c_i = \sum_j P(a_{ij})$ *for each i. Then* u_k *is the optimal choice object.*

If c_i *contain more than one values, then the decision-maker can choose any one of them as the optimal value.*

Next we give the definition of dominated options.

Definition 2.4 *[8] Define* (F, A) *as the probabilistic soft set with tabular representation with entries* $P(a_{ij})$*. Let* $U = \{u_1, \ldots, u_m\}$ *be the non-empty finite universe set. Options* u_i *is called dominates option* u_l *or* u_l *is dominated by* u_l *if* $P(a_{lj}) \leq P(a_{ij})$ *for all* j *and* $P(a_{lj}) < P(a_{ij})$ *for at least one parameter* j*.*

In the following lemma, we give the property that the value of the dominated options choice is smaller, as stated in the following lemma.

Lemma 2.5 *[8] If* u_1 *dominates* u_2*, then the choice value of* u_1 *is greater than the choice value of* u_2*.*

In this following definition, we provide the definition of dominant eigenvalue.

Definition 2.6 *[1] The dominant eigenvalue of a matrix is the one having the largest absolute value.*

In this case, if the eigenvalues contained complex numbers, we can compare their absolute value or modulus.

Next we provide the Perron-Frobenius Theorem, ensuring that a square matrix with positive entries has an eigenvalue with the largest absolute value and is a real number.

Theorem 2.7 *[4] A real square matrix with positive entries has a unique eigenvalue of largest absolute value and that eigenvalue is real.*

Theorem 2.7 guarantee that our choice of eigenvalue in step 3 of Algorithm 2.2 is a unique real number.

Next we give the Algorithm 2.2, which is similar to Algorithm 2 in [8] based on the properties of positive matrices. In this algorithm, slight modifications are made in steps 3 and 4. Since the dominant eigenvalue has been defined in Definition 2.6, step 3 in this algorithm can be performed with complex numbers. Similarly, in step 4, we modify it to be executable with complex numbers.

Algorithm 2.2 *[8] Using positive comparison matrices, the steps of the method is given as follows. Let us denote probabilistic soft sets contains n parameters and m options. Let* C_0 *be the matrix that contains the original probabilistic soft set.*

Step 1 *Delete the dominated options in C_0. Consider the probabilistic soft set with m options, in C_0 whose entries is $P(a_{ij})$.*

Step 2 *Construct a matrix with $m \times m$ entries and denoted by $C = (c_{ij})_{m\times m}$ where:*

a. *When $i \neq j$ is the number of parameters for which the probability value of u_i is strictly greater than the probability value of u_j. Thus c_{ij} is the number of parameters m for which $P(a_{im}) > P(a_{jm})$.*
b. *When $i = j$, $c_{ij} = n(m-1) - t_j$ where $t_j = \sum\{c_{qj} : q \neq j, q = 1, 2, \ldots, m\}$ is the sum of the non-diagonal elements in column j of C. This means that we define c_{ii} as the number such that column i in C sums up to $n(m-1)$.*

Step 3 *Compute the eigenvector $H = (H_1, \ldots, H_k)$ associated with the dominant eigenvalue of the matrix, which is $n(m-1)$.*

Step 4 *The decision is any object u_k such that u_k is the maximize of absolute value of the component of H. If there were objects repeated from a maximal object u_k, then they can also be selected.*

The detail of Algorithm 2.2 is given in [8].

3 Case Study

This paper will explore recommendations for Korean dramas to be chosen within a specific time frame.

3.1 Data Description

The data in this study was all Korean dramas that have release date from 1 July 2023 until 31 December 2023 and the voters data taken in 12 January 2024 from www.imdb.com. There are 40 Korean dramas with 7 genres and in total there are 115,101 voters.

3.2 Result and Discussion

Let U represent the set of Korean dramas released from 1 July 2023 to 31 December 2023. Let E be the collection of the seven genres, namely E = {action, comedy, crime, history, horror, mystery, romance}. Subsequently, consider A = {action, comedy, romance} as a subset of E, which is an example of the preference of the user. To obtain $P(a_{ij})$, the number of voters who watched i-th drama in j-th genre is divided by the total number of voters in j-th genre. The tabular representation is in Table 2. For computation of the algorithms we use Python 3.7.

3.2.1 Algorithm 2.1

Step 1. Input the universal U as the set of all Korean dramas and E be the set of 7 genres.

Step 2. Input the set of choice $A = \{action, comedy, romance\}$.

Step 3. Input the probability $P(a_{ij})$.

Step 4. Find k for which $c_k = \max c_i$, where $c_i = \sum_j P(a_{ij})$ for each i. The tabular representation is presented in the Table 2. We can easily find that for $k = 2$, $c_2 = 0.3550$ is the largest value of c_i. So, we can choose Moving.

3.2.2 Algorithm 2.2

Step 1. We eliminate the dominated alternatives and find that the remaining Korean dramas are Moving, Strong Girl Nam-soon, Mask Girl, Doona!, and Destined with You. We can redefine the set U as u_1 = Moving, u_2= Strong Girl Nam-Soon, u_3 = Mask Girl, u_4 = Doona!, u_5 = Destined with You. The matrix obtained is provided below.

$$\begin{pmatrix} \text{Title} & \text{Action} & \text{Comedy} & \text{Romance} \\ u_1 & 0.355 & 0 & 0 \\ u_2 & 0.0755 & 0.0667 & 0.0626 \\ u_3 & 0 & 0.2533 & 0 \\ u_4 & 0 & 0.1297 & 0.1217 \\ u_5 & 0 & 0 & 0.1239 \end{pmatrix}$$

Step 2. We obtained matrix C as follows.

$$\begin{pmatrix} 6 & 1 & 1 & 1 & 1 \\ 2 & 7 & 2 & 1 & 2 \\ 1 & 1 & 7 & 1 & 1 \\ 2 & 2 & 1 & 8 & 1 \\ 1 & 1 & 1 & 1 & 7 \end{pmatrix}$$

The set of all eigenvalues of C is $\{12, 5, 6, 6 + 0.00000005i, 6 - 0.00000005i\}$. Therefore, the dominant eigenvalue of C is 12.

Step 3. The eigenvector associated with 12 is

$$H = \begin{pmatrix} 0.3090 \\ 0.5322 \\ 0.3605 \\ 0.6009 \\ 0.3605 \end{pmatrix}.$$

Step 4. The maximal of absolute value of every component in H is 0.6009 and it is associated with u_4 = Doona!. Therefore, we can choose Doona!.

3.2.3 Decision

From Algorithm 2.1, it is determined that 'Moving' is the Korean drama that should be preferred. However, Algorithm 2.2 suggests that 'Doona!' is the preferable choice. This difference arises because Algorithm 2.1 only sums the probabilities within each genre but does not consider whether the probability for a particular genre is zero or not. As a result, 'Moving,' which has the highest overall probability, is chosen even though it does not fall under the genres of comedy and romance.

On the other hand, when using Algorithm 2.2, even though u_1 = 'Moving' has the highest probability in the action genre, it is still loses by other Korean dramas whose probabilities in genres other than action are not zero. The Korean drama u_2 = 'Strong Girl Nam-Soon,' even though the probabilities in all three genres are not zero, is significantly smaller when compared to the probability of 'Doona!'. Therefore, it makes sense that 'Doona!' is the recommended choice.

From these results then the best decision is depend on the preferences of the user. If the user doesn't restrict the number of the genre to make a decision, then Algorithm 2.1 will provide the best decision. On the other hand, if the number of genres is more than two, then Algorithm 2.2 will provide the best decision.

4 Conclusion

In conclusion, the two algorithms, Algorithm 2.1 and Algorithm 2.2, provide different recommendations for the preferred Korean drama to watch. While Algorithm 2.1 favors 'Moving' due to its highest overall probability, it overlooks the fact that 'Moving' lacks probabilities in the comedy and romance genres. On the contrary, Algorithm 2.2 suggests 'Doona!' as the preferable choice, taking into account probabilities across all genres and considering whether the probabilities are non-zero.

Algorithm 2.2's preference for 'Doona!' aligns with a more comprehensive evaluation, where even though 'Moving' has the highest probability in the action genre, it loses out to other Korean dramas with non-zero probabilities in genres beyond action. 'Strong Girl Nam-Soon,' while having non-zero probabilities in all three genres, still falls significantly short when compared to the probability of 'Doona!'. Consequently, the recommendation from Algorithm 2.2, favoring 'Doona!,' is more sensible in the context of considering probabilities across all relevant genres.

Additionally we may say that the best decision is depend on the preferences of the user. If the user doesn't restrict the number of the genre to make a decision, then Algorithm 2.1 will provide the best decision. On the other hand, if the number of genres is more than two, then Algorithm 2.2 will provide the best decision.

References

[1] Bronson, R., & Costa, G. B. (2020). Matrix Methods. Academic Press.

[2] Fatimah, F., & Andriyansah. (2020). Analysis of Tourism Facilities Using N-Soft Set Decision Making Procedures. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 4(1), 135–141.

[3] Fatimah, F., Rosadi, D., Hakim, R. B. F., & Alcantud, J. C. R. (2017). N-soft sets and their decision making algorithms. Soft Computing, 22(12), 3829–3842.

[4] Samelson, H. (1957). On the Perron-Frobenius theorem. Michigan Mathematical Journal, 4(1).

[5] Molodtsov, D. (1999). Soft set theory—First results. Computers & Mathematics With Applications, 37(4–5), 19–31.

[6] Maji, P., Roy, A., & Biswas, R. (2002). An application of soft sets in a decision making problem. Computers & Mathematics With Applications, 44(8–9), 1077–1083.

[7] Maji, P., Biswas, R., & Roy, A. (2003). Soft set theory. Computers & Mathematics With Applications, 45(4–5), 555–562.

[8] Fatimah, F., Rosadi, D., Hakim, R. F., & R. Alcantud, J. C. (2017). Probabilistic soft sets and dual probabilistic soft sets in decision-making. Neural Computing and Applications, 31(S1), 397–407.

[9] Ali, M. I., Feng, F., Liu, X., Min, W. K., & Shabir, M. (2009). On some new operations in soft set theory. Computers & Mathematics With Applications, 57(9), 1547–1553.

[10] Gunduz(Aras), C., Posul, H. (2016). On some new operations in probabilistic soft set theory. European Journal of Pure and Applied Mathematics 9(3):333–339.

[11] Sezgin, A. and Atagu¨n, A. O. (2011). On operations of soft sets. Computers and Mathematics with Applications, 61(5), 1457-1467.

[12] Zhu, P., & Wen, Q. (2010). Probabilistic Soft Sets. 2010 IEEE International Conference on Granular Computing.

https://doi.org/10.1051/itmconf/20246701043

[13] Ashraf, S., Kousar, M., & Hameed, M. S. (2023). Early infectious diseases identification based on complex probabilistic hesitant fuzzy N-soft information. Soft Computing, 27(23), 18285–18310.

5 Appendix

 https://doi.org/10.1051/itmconf/20246701043

Title	Action	Comedy	Romance	Choice value
Gyeongseong Creature	0.2425	0	0	0.2425
Moving	0.3550	0	0	0.3550
The Worst of Evil	0.1368	0	0	0.1368
Vigilante	0.0605	0	0	0.0605
Strong Girl Nam-soon	0.0755	0.0667	0.0626	0.2048
Song of the Bandits	0.0892	0	0	0.0892
Korea-Khitan War	0.0041	0	0	0.0041
A Bloody Lucky Day	0.0058	0	0	0.0058
A Returner's Magic Should Be Special	0.0208	0	0	0.0208
Han River Police	0.0073	0.0064	0	0.0137
Shasha and Milo	0.0006	0.0006	0	0.0012
Hero Inside	0.0014	0.0012	0	0.0026
Dragon Force: Power of Dragon	0.0004	0.0004	0	0.0008
My Demon	0	0.1120	0.1051	0.2171
Twinkling Watermelon	0	0.1148	0.1078	0.2226
Welcome to Samdalri	0	0.0337	0.0317	0.0654
A Good Day to Be a Dog	0	0.0350	0.0329	0.0679
My Man is Cupid	0	0.0096	0.0090	0.0186
The Story of Park's Marriage Contract	0	0.0260	0.0244	0.0503
Castaway Diva	0	0.0643	0.0603	0.1246
Mask Girl	0	0.2533	0	0.2533
Doona!	0	0.1297	0.1217	0.2514
The Sand Flower	0	0.0031	0.0029	0.0060
The Matchmakers	0	0.0132	0.0124	0.0257
Behind Your Touch	0	0.0832	0.0780	0.1612
Once Upon a Boyhood	0	0.0099	0	0.0099
Zombieverse	0	0.0351	0	0.0351
Comedy Royale	0	0.0017	0	0.0017
My Dearest	0	0	0	0
Moon in the Day	0	0	0.0187	0.0187
Hi Cookie	0	0	0	0
The Escape of the Seven	0	0	0	0
Tell Me That You Love Me	0	0	0.0097	0.0097
Destined with You	0	0	0.1239	0.1239
Perfect Marriage Revenge	0	0	0.0325	0.0325
A Time Called You	0	0	0.1130	0.1130
Live Your Own Life	0	0	0.0007	0.0007
A Breeze of Love	0	0	0.0059	0.0059
Why R U?	0	0	0.0066	0.0066
My Lovely Boxer	0	0	0.0072	0.0072

Table 2. Tabular Representation of Korean dramas with choice values. $P(a_{ij})$ denotes the number of voters who watched i-th drama in j-th genre is divided by the total number of voters in j-th genre.