應用 K 調和平均數演算法於決策過程中

An Application of K-Harmonic Means Algorithm in Decision Processes

吳信宏[1] 謝俊逸[2]

(Received: Feb. 8, 2005；First Revision: Aug. 23, 2005；Accepted: Nov. 3, 2005)

摘要

使用常見 p 值的 K 調和平均數演算法之特性，以及針對 p 等於無限大的 K 調和平均數演算法之數學性質都加以討論。我們也利用一個範例說明如何使用不同的 p 值於不同的方案，在各種不同的準則當中將方案的重要性做排序。當 p 等於無限大的 K 調和平均數法是找出任何一個準則對於比較標準的距離最短。因此，對於方案的排序是依照其中一個表現最好的準則來決定。K 調和平均數法的特點與缺失都在本論文當中做詳細的討論。最後，如何使用 K 調和平均數法於決定方案的重要性之建議也一併的提供。

關鍵字：K 調和平均數、決策、排序

Abstract

The property of the K-harmonic means (KHM) algorithm with some commonly seen p values is discussed, and the KHM with $p = \infty$ is also discussed mathematically. An example of applying KHM method with different p values to prioritize alternatives under a variety of criteria is illustrated. The KHM method with $p = \infty$ is to look for any value in a criterion that has the highest similarity to that of the referential series. Thus, the priorities of alternatives would be decided based upon one of the criteria with the best performance. The advantages and disadvantages of applying KHM method is presented in this paper in detail. Finally, the recommendations of using the KHM method in prioritizing the alternatives are provided.

***Keywords*:** K-harmonic means, Decision, Priority

1. Introduction

Zhang, Hsu, and Dayal (1999) have proposed the K-harmonic means (KHM) algorithm, a center-based iterative clustering algorithm, which is insensitive to the initialization of the centers. This algorithm takes the sum over all data points of the harmonic average of the squared distance from a data point to all the centers as its performance function. The

[1]國立彰化師範大學企業管理學系副教授
[2]亞洲大學資訊科學與應用學系助理教授

generalized KHM algorithm can be expressed by Equation (1) where the p value is typically an integer, such as 1, 2, and 3. In Equation (1), $A = (a_1, a_2, a_3, \ldots, a_i, \ldots, a_n)$, and $B = (b_1, b_2, b_3, \ldots, b_i, \ldots, b_n)$, and C is the weight, presented as $C = (c_1, c_2, c_3, \ldots, c_i, \ldots, c_n)$, where $\sum_{i=1}^{n} c_i = 1$, and $c_i > 0$ for all $i \in \{1, 2, \ldots, n\}$.

$$d_p^H(A, B) = \left(\sum_{i=1}^{n} c_i \left| a_i - b_i \right|^{-p} \right)^{-1/p}. \tag{1}$$

Studies conducted by Zhang (2000, 2001) and Zhang, Hsu, and Dayal (1999) have proven that the KHM method is essentially insensitive to the initialization of the centers especially compared with K-means method. More importantly, the KHM method could significantly improve the quality of clustering results compared with the K-means method in certain cases. Obviously, the KHM algorithm can be applied in a decision-making process to select the best alternative(s) under a variety of criteria. Specifically, this algorithm can be used to separate alternatives with appropriate clusters such that the best alternative(s) can be distinguished from with the others. To further evaluate the KHM algorithm in decision processes and exploit the property of the algorithm, an example is used to examine the KHM algorithm. In addition, the KHM algorithm with $p = \infty$ is also discussed mathematically.

This paper is organized as follows: The KHM algorithm is reviewed in Section 2 along with the mathematical property when $p = \infty$. An example is illustrated in Section 3 by applying the KHM algorithm in Equation (1) in a decision-making process. Finally, conclusions are drawn in Section 4.

2. The K-Harmonic Means Algorithm

The K-harmonic means algorithm was proposed to replace the winner-take-all strategy of the K-means algorithm, which makes the association between data points and the nearest center so strong that the membership of a data point is not changed until it is closer to a different center (Zhang, 2001). This KHM algorithm, on the other hand, has a “built-in” dynamic weighting function that boosts the data that are not close to any center by giving them a higher weight in the next iteration (Zhang, 2001). Therefore, the KHM algorithm becomes insensitive to initialization and performs better than the K-means algorithm with bad initialization (Zhang, 2000).

The typical p values used by the K-means algorithm include 1, 2, and ∞ (Lai and Hwang, 1994). For $p = 1$, known as the Manhattan distance, it implies an equal importance (weights) for all these deviations, while the Euclidean distance, $p = 2$, implies that these deviations are weighted proportionately with the largest deviation having the largest weight. Finally, for $p = \infty$, the largest deviation completely dominates distance determination (Lai and Hwang, 1994). If $p = \infty$, Equation (1) can be simplified as follows:

$$d_{\infty}^{H}(A,B) = \min_{1\le i\le n}|a_i - b_i|. \qquad (2)$$

Interestingly, the philosophy of Equation (2) is to select the alternative(s) with the smallest difference between *A* and *B*, i.e., the minimum absolute value of ($a_1 - b_1$, $a_2 - b_2$, $a_3 - b_3$, …, $a_i - b_i$, …, $a_n - b_n$). If *A* is the referential series and *B* is the actual performance series, Equation (2) is to select the actual performance series with the greatest potential. The weights used in Equation (1) do not take into account when Equation (2) is applied. For the formula derivation of Equation (2), please refer to Appendix. To further discuss and exploit the property of the KHM algorithm, the selected *p* values used in this study include 1, 2, 3, 4, 5, and ∞.

3. An Example

To illustrate the K-harmonic means algorithm in decision processes, an example from Common Wealth magazine is used. In October issues of 2002, the Common Wealth magazine has revealed the current enterprises that have better business operations in Taiwan. According to the survey, twenty major industries, that may have significant impacts in Taiwan economic development, were investigated by the peers of the similar industries and the expertise. Each company was evaluated by the following ten indices: (1) Foresight, (2) Innovation, (3) Customer-oriented product and service quality, (4) Operational performance and organizational effectiveness, (5) Financial proficiency, (6) Ability to attracting and training employees, (7) Ability to applying information technologies to be competitive, (8) Internationalization, (9) Value of long-term investment, and (10) Social responsibility (The Common Wealth Magazine, 2002).

For each index, the highest and lowest scores a company can receive are 10 and 1, respectively. The original and normalized weights for these ten indices are summarized in Table 1. For information service industries, seven major companies were compared, including International Business Machines (IBM) in Taiwan, Motorola, Inc. in Taiwan, Philips Electronics in Taiwan, Agilent Technologies Taiwan Ltd., Panasonic Industrial Sales (Taiwan) Co. Ltd., Toshiba Electronics in Taiwan, and Samsung Electronics in Taiwan. The performance of these seven companies in each index is provided in Table 2.

The statement can be viewed as a decision-making process if a best company is to be chosen from these seven companies based upon the ten indices. For each index, the maximum score is 10. To apply Equation (1) with p = 1, 2, 3, 4, and 5, the referential series of *A* is set to *A* = (10, 10, 10, 10, 10, 10, 10, 10, 10, 10), where there are ten indices. On the other hand, the actual performance of *B* can be documented by each company. For instance, *B*(IBM) = (7.25, 7.01, 7.14, 7.07, 7.24, 7.58, 7.74, 7.96, 7.34, 6.65), and *B*(Samsung) = (6.66, 6.81, 6.51, 6.67, 6.63, 6.28, 6.98, 7.18, 6.60, 5.72). The weights of the ten indices are *C* = (0.2947, 0.2124,

Table 1 The Original and Normalized Weights of the Ten Indices

Index	Original Importance	Normalized Weight
1	2984	0.2947 (2984/10126)
2	2151	0.2124 (2151/10126)
3	1394	0.1377 (1394/10126)
4	1039	0.1026 (1039/10126)
5	637	0.0629 (637/10126)
6	696	0.0687 (696/10126)
7	325	0.0321 (325/10126)
8	557	0.0550 (557/10126)
9	176	0.0174 (176/10126)
10	167	0.0165 (167/10126)
Sum	10126	1.0000

Table 2 The Original Data for Each Company from the Common Wealth Magazine

Index	IBM	Motorola	Philips	Agilent	Panasonic	Toshiba	Samsung
Foresight	7.25	7.28	7.04	6.57	6.61	6.60	6.66
Innovation	7.01	7.36	6.98	6.83	6.65	6.58	6.81
Customer-oriented product and service quality	7.14	7.02	6.90	6.63	6.75	6.72	6.51
Operational performance and organizational effectiveness	7.07	6.99	6.70	6.63	6.53	6.38	6.67
Financial proficiency	7.24	7.17	6.92	6.76	6.95	6.78	6.63
Ability to attracting and training employees	7.58	7.21	7.04	6.84	6.60	6.69	6.28
Ability to applying information technologies to be competitive	7.74	7.73	7.25	7.13	7.04	7.01	6.98
Internationalization	7.96	7.97	7.67	7.35	7.43	7.19	7.18
Value of long-term investment	7.34	7.17	7.10	6.98	6.82	6.57	6.60
Social responsibility	6.65	6.55	6.68	7.23	6.40	6.21	5.72

0.1377, 0.1026, 0.0629, 0.0687, 0.0321, 0.0550, 0.0174, 0.0165), where $\sum_{i=1}^{10} c_i = 1$ from Table 1. The numerical results of applying Equation (1) are summarized in Table 3, where the number in the parenthesis is the priority. The graphical presentation is illustrated in Figure 1.

Table 3 The Numerical Results and Priorities of Applying KHM Algorithm in Equation (1)

		$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
IBM	Total Score	2.7384	2.7224	2.7047	2.6853	2.6643
	Rank	(2)	(2)	(2)	(2)	(2)
Motorola	Total Score	2.7166	2.7022	2.6864	2.6691	2.6503
	Rank	(1)	(1)	(1)	(1)	(1)
Philips Electronics	Total Score	2.9826	2.9737	2.9638	2.9529	2.9409
	Rank	(3)	(3)	(3)	(3)	(3)
Agilent Technologies	Total Score	3.2358	3.2274	3.2184	3.2087	3.1984
	Rank	(4)	(4)	(4)	(4)	(4)
Panasonic Industrial Sales (Taiwan) Co.	Total Score	3.2744	3.2648	3.2542	3.2424	3.2293
	Rank	(5)	(5)	(5)	(5)	(5)
Toshiba Electronics	Total Score	3.3430	3.3370	3.3307	3.3239	3.3169
	Rank	(7)	(7)	(7)	(7)	(7)
Samsung Electronics	Total Score	3.3187	3.3115	3.3042	3.2970	3.2897
	Rank	(6)	(6)	(6)	(6)	(6)

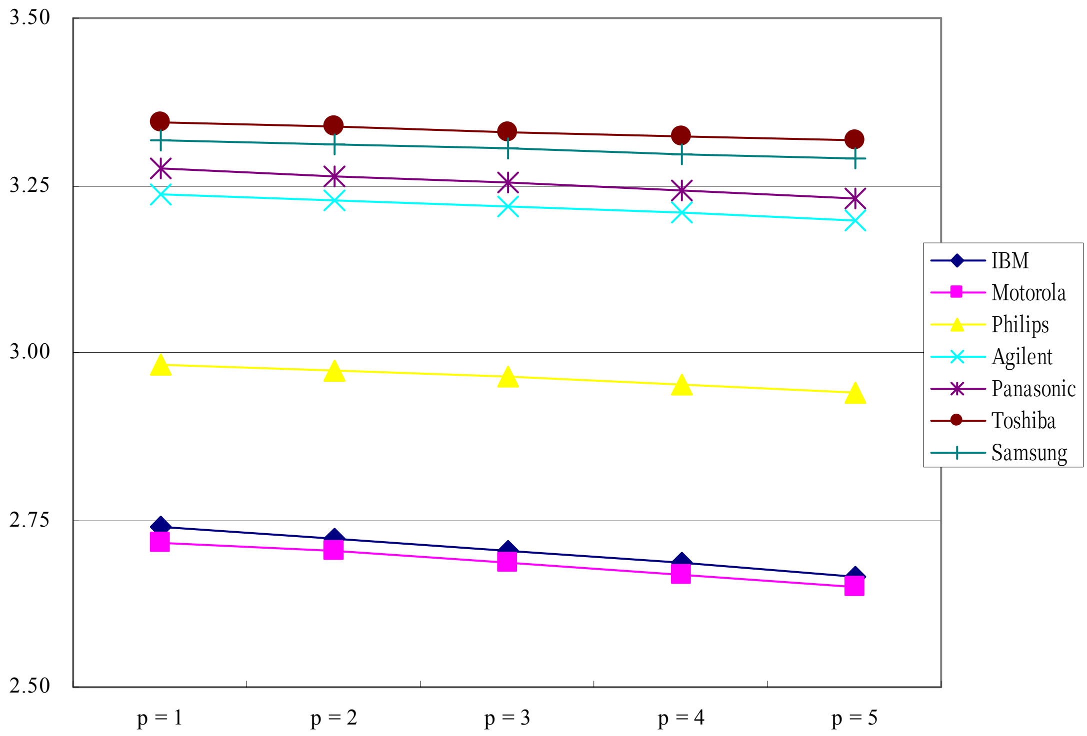

Figure 1 The Graphical Presentation of Using KHM Method

Any alternative with the smallest value to the referential series is considered to be the best alternative. Obviously, Motorola, Inc. in Taiwan has the smallest value to the referential series, i.e., the best company among these seven companies. Besides, the priority is consistent: Motorola > IBM > Philips > Agilent > Panasonic > Toshiba > Samsung. Interestingly, as p increases, the computed value for each company becomes smaller. Therefore, we might expect that the computed value for each company would converge to a particular value when $p = \infty$.

When $p = \infty$ is used, Equation (2) is used to replace Equation (1) in computing the dissimilarity between the referential series and the compared series. The numerical results of applying Equation (2) are depicted in Table 4. It is worth to note that the priority is somewhat different to that in Table 3. In Table 4, the importance is Motorola > IBM > Philips > Panasonic > Agilent > Samsung > Toshiba. In addition, the property of Equation (2) is to discuss the greatest potential of each alternative (company) without considering the weights when the value for each criterion is set to 10. Clearly, the highest similarity for each company falls in Index 8. In fact, the minimum value in each column (company) was selected to represent the highest similarity. Therefore, the priority of using Equation (2) is dependent upon the criterion that has the highest similarity to the referential series. Obviously, 9 criteria except for the eighth criterion are not used when Equation (2) is applied for priority.

Table 4 The Computational Results of Applying Equation (2)

Index	IBM	Motorola	Philips	Agilent	Panasonic	Toshiba	Samsung
Foresight	2.75	2.72	2.96	3.43	3.39	3.40	3.34
Innovation	2.99	2.64	3.02	3.17	3.35	3.42	3.19
Customer-oriented product and service quality	2.86	2.98	3.10	3.37	3.25	3.28	3.49
Operational performance and organizational effectiveness	2.93	3.01	3.30	3.37	3.47	3.62	3.33
Financial proficiency	2.76	2.83	3.08	3.24	3.05	3.22	3.37
Ability to attracting and training employees	2.42	2.79	2.96	3.16	3.40	3.31	3.72
Ability to applying information technologies to be competitive	2.26	2.27	2.75	2.87	2.96	2.99	3.02
Internationalization	**2.04**	**2.03**	**2.33**	**2.65**	**2.57**	**2.81**	**2.82**
Value of long-term investment	2.66	2.83	2.90	3.02	3.18	3.43	3.40
Social responsibility	3.35	3.45	3.32	2.77	3.60	3.79	4.28

Equation (2) can be further represented by Figure 2 pictorially, where $d_p^H(A,0)$ = $\left(\sum_{i=1}^{2} c_i |a_i - 0|^{-p}\right)^{-1/p}$ and $A = (a_1, a_2)$. When $p = \infty$, the scenario is a right-angle area, whereas the philosophy is similar to the idea of geometric means when p = 1 and 2, $\{A \in R^2 | d^G(A,0) = 1\}$ = $\{(x, y) | xy = 1 \text{ or } xy = -1\}$ where x, y $\in$ R and x and y represent horizontal axis and vertical axis, respectively. The philosophy of geometric means is as follows: For $A = (a_1, a_2)$, if $d^G(A,0)$ = $|a_1 - 0|^{\omega_1} |a_2 - 0|^{\omega_2}$ when ω_1 + ω_2 = 1, the case belongs to weighted geometric means. If $d^G(A,0)$ = 1 and we assume that it is equal weight, then $|a_1 - 0|^{1/2} |a_2 - 0|^{1/2}$ = 1. That is, $|a_1 * a_2| = 1$, i.e., $a_1 * a_2$ = 1 or $a_1 * a_2$ = −1.

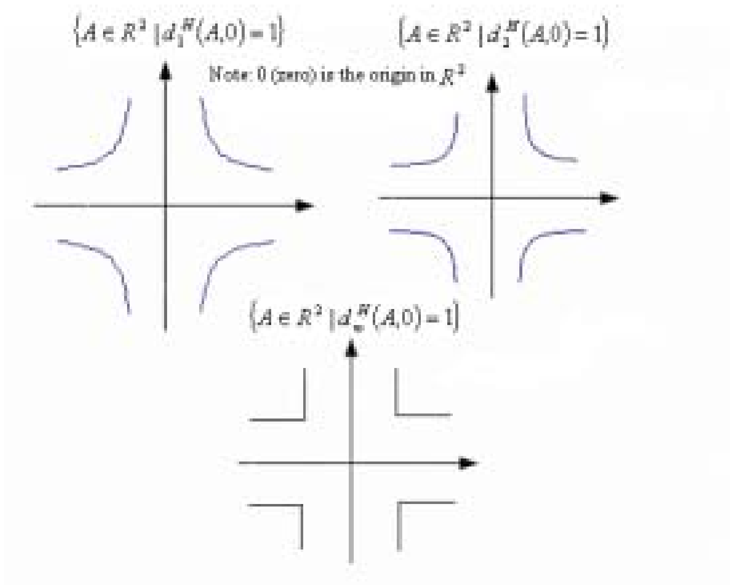

Figure 2 The Graphical Presentation of the KHM Algorithm with Equal Weights

To deal with a decision-making problem, the KHM algorithm with p = 1, 2, 3, 4, 5, and ∞ can be applied to choose the best alternative from several candidates. The entire computational procedure is quite straightforward. In this case as $p = \infty$, the computed value for each company is converged to a particular value without further taking into consideration the weight. For instance, the computed value of IBM in Taiwan will be reduced from 2.7384 with p = 1 to 2.04 with $p = \infty$. This type of phenomena is very unique when the K-harmonic means method is used.

When p = 1, 2, 3, 4, and 5, each criterion with its respective weight would be used to prioritize alternatives. In contrast, when $p = \infty$, the weights of all criteria are not taken into account when the alternatives are prioritized. Moreover, as long as the value in any criterion for each alternative has the highest similarity to the referential series, the rest of criteria will not be used for priority. That is, the priority of alternatives might be decided by only one of the criteria as illustrated in Table 4.

4. Conclusions

This study first discusses the property of the K-harmonic means algorithm and then uses an example to prioritize the importance of alternatives. Generally speaking, the KHM method with different p values can be used in a decision-making problem. Specifically, with p = 1, 2, 3, 4, and 5, all criteria with their respective weights are taken into consideration when KHM method is used. However, as $p = \infty$, the KHM method, which is irrelevant to the weights of all criteria, might only use one of the criteria in a decision-making process. That is, KHM method with $p = \infty$ is to look for any value in a criterion that has the highest similarity to that of the referential series. Therefore, the priorities of alternatives would be decided based upon one of the criteria with the best performance.

Appendix

Let A and $B \in R^n$, $A = (a_1, a_2, a_3, \ldots, a_i, \ldots, a_n)$ and $B = (b_1, b_2, b_3, \ldots, b_i, \ldots, b_n)$. Assume that $a_i \neq b_i$, $\forall\ i = 1, 2, \ldots, n$. $d_p^H(A,B) = \left(\sum_{i=1}^{n} c_i |a_i - b_i|^{-p}\right)^{-1/p}$, where $p \in R^+$, $0 < c_i < 1$, for $1 \le i \le n$. Let $d_\infty^H(A,B) = \lim_{p\to\infty} d_p^H(A,B)$, we want to show that $d_\infty^H(A,B) = \min_{1\le i\le n}\{|a_i - b_i|\}$.

Proof:

Note that $c_i \neq 0$, for $1 \le i \le n$. From the definition of $d_\infty^H(A,B)$, we have $d_\infty^H(A,B) \equiv$

$$\lim_{p\to+\infty}\left[\sum_{i=1}^{n} c_i |a_i - b_i|^{-p}\right]^{-1/p} = \lim_{p\to+\infty}\left[\sum_{i=1}^{n}\frac{c_i}{|a_i - b_i|^p}\right]^{-1/p}.$$

Let $u = \min_{1\le i\le n}\{|a_i - b_i|\}$ and i_0 be a number such that $u = |a_{i_0} - b_{i_0}|$, where $1 \le i_0 \le n$.

Note that $\dfrac{c_{i_0}}{\left|\min_{1\le i\le n}\{|a_i - b_i|\}\right|^p} \leqq \sum_{i=1}^{n}\dfrac{c_i}{|a_i - b_i|^p} \leqq \sum_{i=1}^{n}\dfrac{1}{|a_i - b_i|^p}$ (From the fact $0 < c_i < 1$ for each i)

$$\leqq \sum_{i=1}^{n} \frac{1}{\left|\min_{1\le i\le n}\{|a_i - b_i|\}\right|^p} = \frac{n}{\left|\min_{1\le i\le n}\{|a_i - b_i|\}\right|^p}.$$

In our case, p is a positive number. So $-\frac{1}{p}$ is a negative number, then we have the following:

$$\Rightarrow \left[\frac{n}{\left|\min_{1\le i\le n}\{|a_i - b_i|\}\right|^p}\right]^{-1/p} \le \left[\sum_{i=1}^{n} \frac{c_i}{|a_i - b_i|^p}\right]^{-1/p} \le \left[\frac{c_{i_0}}{\left|\min_{1\le i\le n}\{|a_i - b_i|\}\right|^p}\right]^{-1/p}.$$

$$\Rightarrow \min_{1\le i\le n}\{|a_i - b_i|\}\, n^{-\frac{1}{p}} \le \left[\sum_{i=1}^{n} \frac{c_i}{|a_i - b_i|^p}\right]^{-1/p} \le (c_{i_0})^{-1/p} \min_{1\le i\le n}\{|a_i - b_i|\}$$

Take limit as $p \to \infty$

$$\Rightarrow \lim_{p\to\infty}\left[\min_{1\le i\le n}\{|a_i - b_i|\} n^{\frac{-1}{p}}\right] \le \lim_{p\to\infty}\left[\sum_{i=1}^{n} \frac{c_i}{|a_i - b_i|^p}\right]^{-1/p} \le \lim_{p\to\infty}(c_{i_0})^{-1/p} \min_{1\le i\le n}\{|a_i - b_i|\}$$

$$\Rightarrow \min_{1\le i\le n}\{|a_i - b_i|\} = d_{\infty}^{H}(A,B) = \min_{1\le i\le n}\{|a_i - b_i|\}$$

$$\therefore d_{\infty}^{H}(A,B) = \min_{1\le i\le n}\{|a_i - b_i|\}.$$

References

1. Lai, Y.J. and C.L. Hwang (1994), *Fuzzy Multiple Objective Decision Making: Methods and Applications*, Springer-Verlag, Berlin.
2. The Common Wealth magazine (2002), http://www.cw.com.tw/Files/magazine/archive/frontend/MagCatalog.asp?My Issue=59
3. Zhang, B., M. Hsu and U. Dayal (1999), *K-Harmonic Means – A Data Clustering Algorithm*, Technical Report HPL-1999-124, Hewlett-Packard Research Laboratory. http://www.hpl.hp.com/techreports/1999/HPL-1999-124.html
4. Zhang, B. (2000), *Generalized K-Harmonic Means – Boosting in Unsupervised Learning*, Technical Report HPL-2000-137, Hewlett-Packard Laboratories. http://www.hpl.hp.com/techreports/2000/HPL-2000-137.pdf
5. Zhang, B. (2001), "Generalized K-Harmonic Means – Dynamic Weighting of Data in Unsupervised Learning," *First SIAM International Conference on Data Mining*, Section L2.