Aggregating and Weighting Expert Knowledge in Group Decision Making

Nipat Jongsawat and Wichian Premchaiswadi

Graduate School of Information Technology in Business, Siam University
Bangkok, Thailand
nipatj@yahoo.com, wichian@siam.edu

Abstract. In this paper, we proposed a methodology based on group decision making for aggregating and weighting expert knowledge or opinions and identifying the final agreement for a group of experts. We use a case study on causes of Alzheimer's Disease (AD) for an illustration of the procedural steps in the proposed methodology. We begin by mapping nine most commonly discussed possible causes or risk factors of Alzheimer's disease that we obtain from the reports into the alternatives. Then, we asked medical professionals or experts to sort the alternatives by means of a fixed set of linguistic categories; each one has associated with a numerical score. We average the scores obtained by each alternative and we consider the associated preference. A method of weighting individual expert opinions is applied in order to arrange the sequence of alternatives in each step of the decision making procedure. We calculate the collective scores after we weight the opinions of the experts with the overall contributions to agreement. The sequential decision procedure is repeated until it determines a final subset of experts where all of them positively contribute to the agreement for group decision making.

Keywords: group decision making, agreement measure, collective preference, collective ranking, Euclidean distance, weak order.

1 Introduction

In group decision making, different experts often think about the same problem in quite different ways. They frequently have different opinions for decision making about the same situation. In a typical group decision making situation, each expert in a group is asked to identify as percentages what his/her beliefs are about each cause or solution of the problem. Let's consider a simple situation with three experts and one cause of the problem. For example, expert-1 may give his opinion that "strongly believe that cause-A has a great impact on the problem". But expert-2 may state that "believe that cause-A may have less impact on the problem". Likewise, expert-3 may state that "believe that cause-A has no impact on the problem". Even a simple case of decision making, expert opinions obtained from different experts may be quite different. It is typically not possible to avoid contradictions among different expert's opinions in group decision making. We can observe the different levels of expert's

D.-S. Huang et al. (Eds.): ICIC 2010, CCIS 93, pp. 138–146, 2010.

belief based on their knowledge or experience from the three experts in indentifying the degree of impact on each cause of the problem. Obviously; expert-3 has a negative opinion about the agreement of the group. The questions that arise are: "What happen if we include expert 3's solution as the final group solution?", "Should we exclude expert 3's opinion from the group?", "Should we weight expert 3's opinion with other opinions for the group?", "How can we weight all the expert's opinions in the group and represent a final agreement in a group of experts?". All of these questions seem to be possible. This paper presents a methodology to aggregate and weight expert opinions or knowledge in group decision making.

2 Related Work

Baral et al [1] present algorithms to combine knowledge present in multiple knowledge base systems into a single knowledge base. They consider the knowledge bases to be logic programs with integrity constraints associated with the program. They define what it means for an integrity constraint to violate a disjunctive knowledge base and defined the notion of strong equivalence of logic programs, and used them in their algorithms and proofs. Palomo et al [2] present different models combining opinions of two experts given at specific time points when forecasting with dynamic models. They describe a Bayesian approach for inference and prediction in such situations, using both historical data and current expert's opinions. They consider cooperation cases in which experts' opinions are merged into a class of priors or an expert provides information to other one, and compare these cases with the non-cooperative, independent one. Druzdzel et al [3], [4] present guidelines for detecting when different sources of data can be safely combined. They show that combining probabilistic knowledge that originates from different sources requires utmost care. They demonstrate the risks of such a combination, even when this knowledge encompasses such seemingly population-independent characteristics as sensitivity and specificity of medical symptoms. Wouter et al [5] present combining expert advice efficiently. They introduce prior distributions on sequences of experts, which allow unified description of many existing models. They show how models for prediction with expert advice can be defined concisely and clearly using Hidden Markov Models (HMM).

For consensus measurement in group decision making, Bosch [6] introduced a general concept of consensus measurement within the class of linear orders by assuming three axioms: Unanimity, Anonymity (symmetry with respect to decision makers) and Neutrality (symmetry with respect to alternatives). Cook, Kress and Seiford [7] introduce the technique to find the marginal contribution to consensus of voters with respect to a profile. It is used for prioritizing the decision makers in order of their contribution to consensus. Cook and Seiford [8], [9] introduce a system for codifying linear and weak orders by means of vectors which represent the relative position of each alternative in the corresponding order. Jose [10] considers similar procedures in the generalization of scoring rules from linear orders to weak orders. He proposed a multi-stage decision making procedure by measuring distances among individual and collective scoring vectors by means of different metrics.

3 Weighting Expert Opinions Scheme for Decision Making Procedure

In this section we present the sequence of steps in the decision making procedure using the weighting expert opinions scheme. The sequence of decision procedure is described as follows.

3.1 Sort the Alternatives and Assign a Score

Experts $\{v_1,\ldots,v_m\}$ sort the alternatives of $X = \{x_1,\ldots,x_n\}$ according to the linguistic categories of $L = \{l_1,\ldots,l_q\}$. Then, we obtain individual weak orders $R_1,\ldots,R_m$ which ranks the alternatives within the fixed set of linguistic categories. Next, taking into account the scores $s_1,\ldots,s_p$ associated with $l_1,\ldots,l_q$, a score is assigned to each alternative for every expert: $S_i(x_u)$, $I = 1,\ldots m$; $u = 1,\ldots,n$.

3.2 Calculate the Euclidean Distance

In order to have some information about the agreement in each subset of experts, we first calculate a distance between pairs of preferences (scoring vector). Since the arithmetic mean minimizes the sum of distances to individual values with respect to the Euclidean metric, it seems reasonable to use this metric for measuring the distance among scoring vectors. Let $(S(x_1),\ldots,S(x_n))$, $(S'(x_1),\ldots,S'(x_n))$ be two individual or collective scoring vectors. The distance between these vectors by means of the Euclidean metric is derived by (1).

$$d(S,S') = \sqrt{\sum_{u=1}^{n}(S(x_u) - S'(x_u))^2} \quad (1)$$

3.3 Aggregate the Expert Opinions

We aggregate the expert opinions by means of collective scores which are defined as the average of the individual scores. There are several steps in this procedure.

1) Calculate the Overall Agreement Measure

We calculate a specific agreement measure which is based on the distances among individual and collective scoring vectors in each subset of experts. The overall agreement measure is derived by (2).

$$M(C,I)=M(C,I) = 1 - \frac{\sum v_{\in I}\, d(S_i,S)}{|I|S_1\sqrt{n}} \quad (2)$$

We note that $S_1\sqrt{n}$ is the maximum distance among scoring vectors, clearly between $(S(x_1),\ldots, S(x_n)) = (s_1,\ldots,s_1)$ and $(S'(x_1),\ldots, S'(x_n)) = (0,\ldots,0)$; $d(S, S') = \sqrt{n * S_1{}^2} = S_1\sqrt{n}$. $M(C, I)$ is equal to 0 if $I = \phi$. Then, $M(C, I) \in [0,1]$, for every $(C, I) \in C \times P(V)$. It is easy to see that the overall agreement measure satisfies the other axioms of Bosch [6], Anonymity and Neutrality.

2) Calculate the Overall Contribution to the Agreement

We now calculate an index which measures the overall contribution to agreement by each expert with respect to a fixed profile, by adding up the marginal contributions to the agreement in all subsets of experts. The overall contribution to the agreement of expert v_i with respect to a profile is defined by (3).

$$w_i = \sum_{I \in V}(M(C,I) - M(C, I\backslash\{v_i\})) \quad (3)$$

If $w_i > 0$, we can conclude that expert v_i positively contributes to the agreement; and if $w_i < 0$, we can conclude that that expert v_i negatively contributes to the agreement.

3) Calculate the Weak Order

We now introduce a new collective preference by weighting the score which experts (indirectly) assign to alternatives with the corresponding overall contribution to the agreement indices. The collective weak order associated with the weighting vector w = $(w_1,\ldots,w_m)$, R^w, is defined by (4) and (5).

$$x_u R_w x_v \leftrightarrow S^W(x_u) \geq S^W(x_v) \quad (4)$$

where

$$S^W(x_u) = \frac{1}{m}\sum_{i=1}^{m} w_i * S_i(x_u) \quad (5)$$

Consequently, we prioritize the experts in order of their contribution to agreement [7].

4 A Case Study

We illustrate the procedural steps in aggregating and weighting expert knowledge by using a case study on causes of Alzheimer's Disease (AD). Although researchers do not know exactly what causes Alzheimer's, there are several theories that are being studied. Many studies are exploring the factors involved in the cause and development of AD. The nine most commonly discussed possible causes or risk factors [13],[14] are 1) risk gene and deterministic genes, 2) general health, 3) heart health and brain health, 4) beta-amyloids, 5) age, 6) dental fillings, 7) family history, 8) aluminum, and 9) aspartame. In the experiment, we asked four medical professionals to evaluate about the levels of evidence in each cause mentioned above that develops Alzheimer's disease. They can evaluate each of the possible causes by using statistical data or their own experiences. The six-levels of evaluation criteria consist of strongest evidence, very strong evidence, strong evidence, some evidence, less evidence, and no evidence. Each level has associated with a numerical score.

We have a group of four medical professionals or experts, labeled v_1, v_2, v_3, and v_4 and nine causes or risk factors of Alzheimer's disease mentioned above. For simplicity, every cause we consider is mapped into a set of $X = \{x_1, \ldots, x_n\}$ when n is the number of causes. See Table 1.

Table 1. Mapping causes into a set of X

Causes	X	Causes	X	Causes	X
Risk/deterministic genes	x_1	Beta-amyloids	x_4	Family history	x_7
General health	x_2	Age	x_5	Aluminum	x_8
Heart/brain health	x_3	Dental fillings	x_6	Aspartame	x_9

Consider four experts who sort the $x_1,\ldots,x_9$ according to a set of linguistic categories $L = \{l_1,\ldots,l_6\}$ and the associated scores given in Table 2.

Table 2. Linguistic categories

L	Meaning	Score	L	Meaning	Score
l_1	Strongest Evidence	8	l_4	Some Evidence	2
l_2	Very Strong Evidence	5	l_5	Less Evidence	1
l_3	Strong Evidence	3	l_6	No Evidence	0

Table 3 contains the way these experts rank the causes $x_1,\ldots,x_9$, which from now on $x_1,\ldots,x_9$ are called the alternatives. In Table 4, we present each individual expert and collective scores obtained by each alternative. Clearly, a group of expert agrees that x_1, x_2, and x_3 are the most important risk factors for developing Alzheimer's disease because these three major risk factors obtain the highest average score, 4.25.

We calculate the distances among the individual opinions and the collective preference by using (1) in the previous section. For example, $d(s_1, s)$ is the Euclidean distance between s_1 and the average score, s. The results are $d(s_1, s) = 4.91 < d(s_4, s) = 6.56 < d(s_2, s) = 6.97 < d(s_3, s) = 10.08$.

Table 3. Sorting alternatives

	l_1 (8)	l_2 (5)	l_3 (3)	l_4 (2)	l_5 (1)	l_6 (0)
v_1	x_3	x_1, x_2, x_4	x_5	x_6, x_7	x_8, x_9	-
v_2	x_1	x_4	x_5	x_7, x_8	x_2, x_3, x_6	x_9
v_3	x_2, x_6, x_9	x_3, x_7, x_8	x_1	-	x_4	x_5
v_4	x_5	x_4	x_2, x_3, x_7, x_8	x_9	x_1	x_6

Table 4. Scores

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
s_1	5	5	8	5	3	2	2	1	1
s_2	8	1	1	5	3	1	2	2	0
s_3	3	8	5	1	0	8	5	5	8
s_4	1	3	3	5	8	0	3	3	2
s(average)	4.25	4.25	4.25	4	3.5	2.75	3	2.75	2.75

Table 5. Collective order

Order	1	2	3	4	5	6
X	x_1, x_2, x_3	x_4	x_5	x_7	x_6	x_8, x_9

Table 5 includes the collective preference provided by the weak order R. We calculate the overall contributions to agreement introduced in (3). We obtain $w_1 = 0.39$, $w_2 = 0.15$, $w_3 = -0.20$, and $w_4 = 0.19$. We apply these weights in the collective decision procedure of (5), then the opinion of the first expert counts as $w_1/ w_2 = 2.56$ times the opinion of the second one; $w_1/ w_3 = 1.89$ times the opinion of the second one; $w_1/ w_4 = 1.96$ times the opinion of the second one. In Table 6, we show the initial collective scores given in Table 4 and the new collective scores after we weight the opinions of the experts with the overall contributions to agreement. The ratio between the new collective score and the initial collective scores (s_w/ s) is calculated. These differences are due to the individual contributions to agreement.

Table 6. New collective scores

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
s	4.25	4.25	4.25	4	3.5	2.75	3	2.75	2.75
s_w	0.68	0.26	0.70	0.86	0.79	-0.17	0.16	0.06	-0.21
s_w /s	0.16	0.06	0.16	0.21	0.22	-0.06	0.05	0.02	-0.07

According to the obtained weights, the new version of the decision procedure linearly orders the alternatives, by means of R_w, as shown in Table 7. After we calculate the new collective scores, the top position in this initial round becomes x_4. The second and third positions are x_5 and x_3, respectively.

Table 7. New order of alternatives

Rank (s_w)	1	2	3	4	5	6	7	8	9
Alternative	x_4	x_5	x_3	x_1	x_2	x_7	x_8	x_6	x_9

When we observe that the third expert negatively contributes to agreement, then his or her associated scores are multiplied by a negative weight. In order to avoid this undesirable effect, we will consider non negative weights for w_3 so that the weights are $w_1 = 0.39$, $w_2 = 0.15$, $w_3 = 0$, and $w_4 = 0.19$. Applying again the decision procedure, we obtain a new s_w (see Table 8) and a new linear order on the set of alternatives (see Table 9).

We clearly see that x_3 is ranked in the third position in Table 7 ($w_3 = -0.20$) and now it becomes the first alternative in Table 9 ($w_3 = 0$). Since in Table 4, $S_3(x_3) = 5$ has been multiplied by the negative weight $w_3 = -0.20$, thus this alternative has been penalized. However, in Table 9 the opinion of the third expert has not been considered. This fact joint with the first expert, with the highest weight $w_1 = 0.39$, ranks x_3 at the first alternative, induce that this alternative reaches the top position.

Table 8. New collective scores ($w_3 = 0$)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
s	4.25	4.25	4.25	4	3.5	2.75	3	2.75	2.75
s_w	0.83	0.66	0.95	0.91	0.79	0.23	0.41	0.32	0.19
s_w /s	0.19	0.15	0.22	0.22	0.22	0.08	0.13	0.11	0.07

Table 9. New Order of Alternatives ($w_3 = 0$)

Rank (s_w)	1	2	3	4	5	6	7	8	9
Alternative	x_3	x_4	x_1	x_5	x_2	x_7	x_8	x_6	x_9

Although the new ranking in Table 9 is more appropriate than the ranking in Table 7 for reflecting each expert's opinion, it is important to note that all the calculations have been made taking into account the opinions of the third expert who has divergent opinions with respect to the global opinion. If we think that the third expert's judgments should not be considered, we can start a new step of the decision procedure where only the opinions of the rest of the experts are taken into account. Table 10 shows the individual and collective scores obtained by each alternative. Applying the decision procedure again, we obtain a new s_w (see Table 11) and a new linear order on the set of alternatives (see Table 12).

Table 10. Scores (S_1, S_2, S_4 only)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
s_1	5	5	8	5	3	2	2	1	1
s_2	8	1	1	5	3	1	2	2	0
s_4	1	3	3	5	8	0	3	3	2
s(average)	4.66	3	4	5	4.66	1	2.33	2	1

Table 11. New collective scores (not include w_3)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
s	4.66	3	4	5	4.66	1	2.33	2	1
s_w	2.68	1.72	2.31	2.86	2.66	0.578	1.33	1.14	0.571
s_w /s	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57

Table 12. New order of alternatives (not include w_3)

Rank (s_w)	1	2	3	4	5	6	7	8	9
Alternative	x_4	x_1	x_5	x_3	x_2	x_7	x_8	x_6	x_9

The new overall contributions to agreement for this round are $w_1^{(2)} = 0.58 > w_2^{(2)} = 0.57 > w_4^{(2)} = 0.56$ while $w_1^{(1)} = w_1 = 0.38 > w_4^{(1)} = w_4 = 0.19 > w_2^{(1)} = w_2 = 0.15$. These differences are due to the fact that in the second iteration of the decision procedure the divergent opinions of the third expert have not been considered.

According to the new weights $w_1^{(2)}$, $w_2^{(2)}$, $w_4^{(2)}$ calculated in this round, the new stage of the decision procedure linearly order the alternatives as shown in Table 12. Table 13 shows the order of alternatives in the different iterations of the decision procedure. We see that x_3 is ranked in the top position in Table 9 (w3 = 0) and now it becomes the forth alternative in Table 12 (not consider the third expert's judgments). The top position in the second iteration of the decision procedure becomes x_4 again. The second, third, and forth positions are x_1, x_5, and x_3, respectively.

Table 13. Order of alternatives in the different iterations

Iteration: Rank (s_w)	1	2	3	4	5	6	7	8	9
Initial: (s_1,s_2,s_3,s_4; w_1,w_2,w_3,w_4)	x_4	x_5	x_3	x_1	x_2	x_7	x_8	x_6	x_9
s_w:	0.8	0.7	0.7	0.6	0.2	0.1	0.0	-0.17	-0.21
1st: (s_1,s_2,s_3,s_4; $w_1,w_2,w_3=0,w_4$)	x_3	x_4	x_1	x_5	x_2	x_7	x_8	x_6	x_9
s_w:	0.9	0.9	0.8	0.7	0.6	0.4	0.3	0.23	0.19
2nd: (s_1,s_2,s_4; w_1,w_2,w_4)	x_4	x_1	x_5	x_3	x_2	x_7	x_8	x_6	x_9
s_w:	2.8	2.6	2.6	2.3	1.7	1.3	1.1	0.57	0.57

Clearly, there exist important differences among the linear orders shown in Table 13. In fact, the initial step of the decision procedure takes into account the divergent opinions of the third expert. The first iteration of the decision procedure does not consider the opinions of the third expert (w_3=0), but the collective ranking and, consequently, all the weights are based on the opinions of all the experts, including that of the divergent third expert. The second iteration of the decision procedure totally excludes the opinions of the third expert.

Finally, in this illustrative example we can classify the results into two groups and conclude that, in the first group the order of alternatives in the fifth rank to ninth rank has never been changed. The levels of evidence in the general health (x_2), family history (x_7), aluminum (x_8), dental fillings (x_6), and aspartame (x_9) may approximately be evaluated as some evidence to no evidence for developing Alzheimer's disease. We may roughly conclude that these five causes might be considered as the minor causes for developing Alzheimer's disease by the agreement of the group. There are no contradictions among different expert's opinions in ranking the order of alternatives in the fifth rank to ninth rank. On the other hand, in the second group, the order of alternatives in the first rank to forth rank in each step of decision procedure has been changed. The main reasons come from the bias opinions from an expert in the group. The bias opinions reflect the overall contribution to the agreement of all experts in the group.

5 Conclusion and Future Work

We proposed a methodology based on group decision making for aggregating and weighting expert knowledge or opinions in identifying the final agreement in a group of experts. We illustrate the procedural steps in aggregating and weighting expert knowledge by using a case study on causes of Alzheimer's Disease (AD). We begin by mapping nine most commonly discussed possible causes or risk factors of Alzheimer's disease that we obtain from the reports into the alternatives. Then, we asked medical professionals to sort the alternatives by means of a fixed set of linguistic categories; each one has associated with a numerical score. We average the scores obtained by each alternative and we consider the associated preference. Then we obtain a distance between each individual preference and the collective one through the Euclidean distance among the individual and collective scoring vectors. Taking into account these distances, we measure the agreement in each subset of experts, and a weight is assigned to each expert. We calculate the collective scores after we weight the opinions of the experts with the overall contributions to agreement. Those experts whose overall contribution to the agreement is negative are excluded and we re-calculate the decision procedure with only the opinions of the experts which positively contribute to agreement. The sequential decision procedure is repeated until it determines a final subset of experts where all of them positively contribute to agreement for group decision making. Finally, we classify the results of the case study into two groups and draw a conclusion in each group.

For future work based on this study, we will introduce a criterion for combining knowledge from different experts or different sources in the decision models.

References

1. Chitta, B., Sarit, K., Jack, M.: Combining Multiple Knowledge Bases. IEEE Transactions on Knowledge and Data Engineering 3(2), 208–220 (1991)
2. Jesus, P., David, R.I., Fabrizio, R.: On Combining Expertise in Dynamic Linear Models. Statistical and Applied Mathematical Sciences Institute. Technical report#2005-6 (2005)
3. Druzdzel, M.J., Diez, F.: Criteria for Combining Knowledge from Different Sources in Probabilistic Models. In: 16th Annual Conference on Uncertainty in Artificial Intellgence, Standford, CA, pp. 23–29 (2000)
4. Druzdzel, M.J., Diez, F.: Combining Knowledge from Different Sources in Caual Probabilistic models. Journal of Machine Learning Research 4, 295–316 (2003)
5. Koolen, M., Rooij, S.D.: Combining Expert Advice Efficiently. Centrum voor Wiskunde en Informatica (CWI), Amsterdam, Netherlands
6. Bosch, R.: Characterizations of Voting Rules and Consensus Measures, Ph.D. Dissertation, Tilburg University (2005)
7. Cook, W.D., Kress, M., Seiford, L.M.: A General Framework for Distance-Based Consensus in Ordinal Ranking Models. European Journal of Operational Research, 392–397 (1996)
8. Cook, W.D., Seiford, L.M.: Priority Ranking and Concensus Formation. Management Science 24, 1721–1732 (1978)
9. Cook, W.D., Seiford, L.M.: On the Borda-Kendall Consensus Method for Priority Ranking Problems. Management Science 28, 621–637 (1982)
10. Jose, L.: Weighting Individual Opinions in Group Decision Making. In: Torra, V., Narukawa, Y., Yoshida, Y. (eds.) MDAI 2007. LNCS (LNAI), vol. 4617, pp. 92–103. Springer, Heidelberg (2007)
11. Herrera, F., Herrera-Viedma, E.: Linguistic Decision Analysis: Steps for Solving Decision Problems under Linguistic Information. Fuzzy Sets and System 115, 67–82 (2000)
12. Yager, R.R.: Non-Numeric Multi-Criteria Multi-Person Decision Making. Journal of Group Decision and Negotiation 2, 81–93 (1993)
13. `http://www.essortment.com/family/alzheimersdise_svvg.htm`
14. `http://alzheimers.about.com/od/whatisalzheimer1/a/causes.htm`