A Probabilistic Approach to Group Decision Making

Sujoy Chatterjee
Department of Computer Science and Engineering
University of Kalyani
Nadia – 741235, India
sujoy@klyuniv.ac.in

Malay Bhattacharyya
Department of Information Technology
Indian Institute of Engineering Science and Technology, Shibpur
Howrah – 711103, India
malaybhattacharyya@it.iiests.ac.in

 CHI'17 Extended Abstracts, May 06-11, 2017, Denver, CO, USA
ACM 978-1-4503-4656-6/17/05. http://dx.doi.org/10.1145/3027063.3053226

Abstract
Large-scale judgment analysis from multiple opinions is a challenging job in terms of time and cost involved. Over the last few years, with the popularity of crowd-powered models, the process of decision making is efficiently getting done by using the knowledge of crowd. In management science, a closely related class of problems, popularly known as group decision making, is often addressed. Unfortunately, majority of the algorithms developed for this purpose work for binary or multiple opinions without taking care of the semantic meaning of the options. Moreover, group decision considers a feedback set comprising range of continuous values unlike the judgment analysis problem. In this paper, we address this problem, hereafter termed as multi-opinion group decision making, with a probabilistic approach taking into account the annotator accuracy, annotator bias and question difficulty. The effectiveness of the approach is demonstrated by applying this on a benchmark group decision making dataset.

Author Keywords
Group Decision Making; Probabilistic Graphical Model; Crowdsourcing.

ACM Classification Keywords
F.1.2 [Modes of Computation]: Probabilistic computation; H.4.2 [Types of Systems]: Decision support (e.g., MIS);

H.5.3 [Group and Organization Interfaces]: Collaborative computing

Introduction

Crowdsourcing and human annotation have drawn a lot of research attention as large-scale datasets can be annotated very easily in a rapid rate as well as in very limited cost by effective use of common people. Numerous algorithms from different areas of data mining, machine learning, bioinformatics, etc. have already been proposed employing crowdsourcing for the annotation of large-scale datasets [7, 8, 10]. Group decision making is one of the important application areas where the human annotation has been used very effectively.

Group decision making deals with the diverse opinions obtained from the crowd in order to produce better consensus. It also has a wide application in the domains of management science, pattern recognition, behavioral science, strategic planning, etc. [3, 5, 6, 9]. It can be seen that the problem of group decision making closely matches with the problem of judgment analysis [4], which is popular in applications involving human annotators. In fact, there can be a loose correspondence between the cases, experts and feedback in group decision making problem to the questions, annotators and opinions in judgment analysis [4], respectively. However, the major difference lies in the way we define the feedback (opinion) set. Unlike judgment analysis, where the opinion set contains discrete values [14], group decision has a feedback set that comprises range of continuous values. Therefore, in group decision making the final decision is not necessarily an element from the representative opinions. Additionally, no semantic relation can be defined between the feedback values, which was possible in judgment analysis [2].

In group decision making process, the main decision problem might have multi-attribute interval data [17]. To make agreement between the attributes, appropriate measures should be taken such that the total information loss for multiple decision makers become minimal [16]. To distinguish the quality of the decision makers, weight has been used as an important criterion in a wide range of applications. Investigations of different weighting schemes (like similarity to ideal solutions, projection based weight, etc.) for the decision makers is one of the main focus of some studies [15]. Subjective preference information, objective preference relations and fuzzy set based application on strategic planning have also been known to be effective to the researchers to reach to an ideal decision from multi-criteria group decision making process [9]. All of these group decision making problems can be framed in the crowdsourcing environment where the individual crowd workers can act as the decision makers in the group decision process to achieve a robust model for aggregating multiple diverse alternatives. In this paper, the multi-opinion group decision making problem is solved with a probabilistic graphical model (PGM) considering the factors like expert accuracy, bias and case difficulty over their feedback.

Basic Terminologies

In this section, we introduce the basic terminologies that will be used throughout the paper. We involve four basic terms, namely case, expert, feedback and decision for the background description. These are detailed hereunder.

- **Case:** A case is simply a labeling or annotation task. E.g., it can be providing scores to the students based on their performances, annotation of a text, etc.
- **Expert:** An expert is basically a crowd worker who gives the feedback (annotates a label) over the given

case. An expert may be good in one type of case and can be bad for some other types of cases.

- **Feedback:** A feedback is the annotation score given by an expert for a particular case. Generally, the feedback is chosen from a range of feedback values.
- **Decision:** A decision is the best consensus of the feedback provided by the experts. E.g., if five of the experts provide the feedback as '1' and only two say '2', then the final decision could be '1'.

Problem Formulation

Let us formalize the group decision making problem for the crowdsourced scenario. We consider a set of cases $\{c_1, c_2, \ldots, c_m\}$ that are the labeling tasks and a set of experts $\{e_1, e_2, \ldots, e_n\}$ who are the crowd workers. The set of feedback values $F = \{f_1, f_2, \ldots, f_z\}$, for any particular cases, can take either some positive integer values (e.g., 1, 2, etc.) or some categorical values (e.g., Yes, No, Possibly, etc.). Note that, in this problem there may not be any specific and predefined feedback set rather a range of feedback values available.

An annotation process is a 4 tuple (C, E, F, τ) consisting of (i) a set of cases $C = \{c_1, c_2, \ldots, c_m\}$, (ii) a set of experts $E = \{e_1, e_2, \ldots, e_n\}$, (iii) a set of feedback $F = \{f_1, f_2, \ldots, f_z\}$ and a mapping function τ: $(C \times E) \rightarrow F$. We have to obtain the final decision for all the cases in C.

Note that, the cardinalities of the sets C, E and F (m, n and z, respectively) are not necessarily the same. We define a response matrix $\mathcal{R}$ as a matrix of dimension $n \times m$ whose elements $\mathcal{R}_{ij}$ denote the feedback provided by the i^{th} expert for the j^{th} case such that $\mathcal{R}_{ij} \in F$, for all i, j. We further denote a decision matrix $\mathcal{J}$ as a matrix of dimension $m \times z$ whose elements $\mathcal{J}_{ij}$ denote a weight of the j^{th} feedback for the i^{th} case. Note that, there can be multiple entries in a row of the decision matrix, however, the maximum value denotes the most appropriate feedback for a particular case.

Proposed Model

Consider a response matrix with m experts giving their feedback from the feedback set $\{1, 2, \ldots, z\}$ over *n* cases. We have to predict the original label z_j of every case *j* taking feedback from *m* experts. As mentioned earlier, the observed feedback is dependent on several factors: (i) the accuracy of expert, (ii) the difficulty of case, and (iii) the biasness of expert. The accuracy of an expert i is denoted by a parameter α_i. Here, α_i = 0 means the expert gives all the feedback incorrectly whereas α_i= 1 means the expert has given all the feedback correctly. The biasness of an expert i is denoted by a parameter γ_i. The difficulty of case *j* is denoted by $\beta_j \in (0, \infty)$.

Suppose, λ_{ij} denotes the feedback given by a particular expert *i* for a given case *j*. Then the probability that the given feedback matches with the true feedback is given by following equation. Here, *z* be the original true feedback that need to be estimated.

$$P(\lambda_{ij} = z | \alpha_i, \beta_j, \gamma_i) = \frac{1}{1 + \exp(\frac{-\alpha_i}{\beta_j \gamma_i})}. \quad (1)$$

The rationale behind using Eqn. (1) is that more the expertise of an expert (i.e., the experts having high accuracy) denotes it has a higher chance to give the correct feedback. Again, more difficult the case is highlights the probability that it will be annotated correctly is lesser. If the biasness gets higher then there is also a lesser probability of annotating the cases correctly.

E Step:

Assume the set of feedback given by all the experts who attempted the case *j* is δ_j and σ_i be the set of cases that have been annotated by the i^{th} expert. Initially, the different feedback (obtained from different experts) are counted for a particular case. We need to estimate the original true value of this particular case. Now, we compute the posterior distribution of each feedback value (obtained after binning) and in this context binomial distribution is used. As the experts give their feedback independently (one feedback is not affected by other) and for a single case multiple opinions are solicited, therefore binomial distribution has been used. So, the posterior distribution of each feedback for any particular case is defined as follows.

$$\mathcal{J}_j(z) = \prod_{i \in \delta_j} P(z)^{I(R_{ij}=z)}(1-P(z))^{I(R_{ij} \neq z)} \quad (2)$$

Here, $\mathcal{R}_{ij}$ is the response of i^{th} expert for the j^{th} case and $\mathcal{J}_j(z)$ is the posterior probability of the j^{th} case when the feedback is *z*. The indicator function I returns 1 if the expert's response $\mathcal{R}_{ij}$ matches with the corresponding true feedback *z* for which the posterior distribution is calculated.

M Step:

Expert accuracy and biasness (i.e., α_i and γ_i) are updated in the following M step.

$$\alpha_i = \frac{1}{|\sigma_i|} \sum_{i \in \sigma_i} \mathcal{J}(R_{ij}) \quad (3)$$

$$\gamma_i = \frac{1}{|\sigma_i|} \sum_{i \in \sigma_i} \mathcal{B}(R_{ij}, \arg\max_z \mathcal{J}_j). \quad (4)$$

Here, $\mathcal{B}$ is the bias score matrix for a particular dataset, $\mathcal{R}_{ij}$ is the response of i^{th} expert for the j^{th} case and z be the feedback obtained from majority voting over the posterior distribution values of various feedback for the same case in decision matrix.

Application to Group Decision Making Problem

Let us illustrate the applicability of the proposed model in the field of group decision making problem. In this problem [1], there are several decision makers and they need to reach into a decision from multiple alternatives. Here, basically the model needs to find out the final aggregated decision from multiple alternatives. There are numerous study available in the literature dealing with this Multiple attributes decision making problem (MADM) problem [13, 15] and in depth study have also been carried out to quantify the expertise of decision makers. Ranking of experts based on their accuracies, using correlation concept to find expertise of experts, utilizing eigen vector method to quantify relative importance of the decision makers within a group are few important strategies among them. In order to compute the weights of the experts, most of the approaches incorporate the Saaty's multiplicative preference relation [11]. But due to substantial drawback of this preference relation, recently TOPSIS method was developed as an important scheme to find out the weight of the experts [11, 12, 13, 16, 18]. In this paper, we adapted an instance of group decision making problem from [15] to demonstrate the utility of the proposed method in the domain of group decision making also. The human resources selection example mentioned in [15] is explained as follows.

A company needs to hire an on-line manager and there are 17 candidates qualified in the second round of the selection process. Now there are four experts who have the responsibility to select the final candidate based on the performance in two different situations namely, panel-interview and 1:1 interview. As there are two different situations (i.e., panel-interview and 1:1 interview) based on which the scores are given to the candidates, it can be considered as eight experts selecting 17 candidates. In this current experiment, we have used the same weighted normalized decision matrix described in Table 5 of [15], but it has been transformed into a decision matrix using the binning procedure. The binning procedure is used to quantize the weighted normalized decision matrix by replacing the values of a particular range with a corresponding representative value. The motivation of using the binning procedure is to make the group decision problem suitable for crowdsourcing based framework. Hence, in respect of group decision making problem each of the experts can be treated as annotator (crowd worker) and the candidates are termed similar like case (questions). The experts provide their feedback for each question and in order to produce the transformed decision matrix using binning procedure several threshold values are chosen. To perform this binning, initially the maximum and minimum elements of the weighted decision matrix are found. The difference between these two values defines the spread of opinions of the experts. So, if we need to transform the values of weighted normalized decision matrix (as described in Table 5 of [15]) into a matrix containing feedback from a set $\{1, 2, 3\}$, then three threshold values should be chosen for the binning process. This requires dividing the difference of minimum and maximum element of the weighted decision matrix by three. Here, higher the score means better the performance of the candidates. For example, let the experts need to choose one particular value from the feedback set $F = \{1, 2, 3, 4, 5\}$ and this requires to divide the total range into five intervals.

In this problem, there is a requirement to find out the expertise of the human annotators based on the ideal solution. According to the proposed approach, primarily the ideal solution is generated by applying majority voting on the experts' feedback for each case. Again, the accuracy of an expert is defined as the absolute deviation of expert's individual score from the majority score. To quantify accuracy, we cannot use the count of match or mismatch of annotator's individual opinion with the ideal opinion because there are a preference relation between the option values. Basically, here higher the feedback value means better the candidate so finding deviation from majority opinion is more appropriate here.

Experimental Results

To make the model simple, the bias of all the experts is chosen as 1 and we treated all cases having the same difficulty label. Finally, the proposed approach is executed on the transformed weighted normalized decision matrix (in transposed form to make it similar like a response matrix) and the final scores of the candidates are obtained (see Table 1). From this table it can be seen that the candidates can be categorized into several groups based on their performances. As here higher score means better the candidates therefore, 16^{th} candidate (denoted by #) is in elite group and it should have higher preference to get recruited. On the other hand, there are 5 candidates who have a least chance to get recruited as aggregated opinion is 2 for all of these 5 candidates. The majority opinions have also been provided here. It can be easily seen that for candidate 11 and 12 the aggregated score differ with the majority opinions. As in majority voting the expert accuracy, biasness is not considered therefore all the experts' feedback are treated equally and hence although the candidate is not so good (considering accuracy of expert) but it achives better score by majority. Thus the propose method produce accurate prediction over majority.

Table 1: Aggregated decision obtained by applying the proposed approach for 17 candidates. The symbols '#' and '*' are used to mark the candidate having the highest score and the 2^{nd} highest score, respectively.

Candidates	Majority Score	PGM based Score
1 (*)	4	4
2	3	3
3	3	3
4	2	2
5	3	3
6	3	3
7	2	2
8	3	3
9 (*)	4	4
10	3	3
11	3	2
12	1	2
13	3	3
14	3	3
15	3	3
16 (#)	5	5
17	2	2

Notion of Case Difficulty and Bias Score

In the proposed approach, standard deviation of the different feedback obtained from various experts (for a particular case) can be computed and we consider that a greater standard deviation means the case is harder.

Alongside finding the case difficulty, the expert biasness is also required to be minimized to remove the error generated due to the biasness of the expert towards a particular feedback value. Although in this group decision making problem the feedback are from a set containing {1,2,3,4,5}, but in some cases only the range 1-5 may be available. So proper binning procedure is required to define the feedback set. Moreover, it may be the situation that instead of these types of feedback, the feedback can be of complex type with some ambiguous semantic meaning. For example the feedback can be of 'yes', 'no', 'skip' or 'Unsure'. Now, suppose the true feedback is 'yes' but the feedback obtained from two crowd workers are 'no' and 'unsure'. Now it reflects that the first worker has pretended as an expert by saying 'no', whereas the other one admitted that he has lack of knowledge on it. Therefore, the penalty should be higher for the first crowd worker. So, proper scoring matrices are to be maintained to find the bias score.

Concluding Remarks

In this paper, a probabilistic graphical approach for the crowd group decision making problem has been proposed. Consideration of different important factors like annotator accuracy, annotator biasness and question difficulty and combination of their effects to conclude about the final aggregated judgment is one of the main contribution in this paper. The utility of the proposed method in any decision making problem has also been demonstrated by applying it on a benchmark group decision making dataset. As a future scope this method can be extended by developing suitable expression for biasness and case difficulty to apply on the large-scale unbalanced datasets where the opinions are more ambiguous. Moreover, to discretize the decision matrix in an efficient way the Bayesian binning approach can also be used to find the optimal number of bins.

Acknowledgments

The work of Malay Bhattacharyya is supported by the Visvesvaraya Young Faculty Research Fellowship 2015-16 of DeitY, Government of India.

References

[1] V. Belton and T. J. Stewart. 2002. Multiple Criteria

Decision Analysis: An Integrated Approach. *Springer.* (2002).

[2] M. Bhattacharyya. 2013. Opinion Ensembling: Learning from Dependent judgements of the Crowd. In *Proceedings of the AAAI HCOMP Workshop on Crowdsourcing at Scale (CrowdScale).* Palm Springs, USA, Shared Task Challenge Paper 01.

[3] U. Bose and D. B. Paradice. 1999. The Effects of Integrating Cognitive Feedback and Multi-Attribute Utility-Based Multicriteria Decision-Making Methods in GDSS. *Group Decision and Negotiation, Kluwer Academic Publishers* 8 (1999), 157–182.

[4] S. Chatterjee and M. Bhattacharyya. 2017. Judgment Analysis of Crowdsourced Opinions using Biclustering. *Information Sciences, Elsevier* 375 (2017), 138–154.

[5] Francineide de Morais Bezerra, Paulo Melo, and João Paulo Costa. 2014. Visual and Interactive Comparative Analysis of Individual Opinions: A Group Decision Support Tool. *Group Decision and Negotiation, Springer* 23 (2014), 101–125.

[6] Kihwan Kim, Nicole L. Cundiff, and Suk Bong Choi. 2015. Emotional Intelligence and Negotiation Outcomes: Mediating Effects of Rapport, Negotiation Strategy, and Judgment Accuracy. *Group Decision and Negotiation, Springer* 24 (2015), 477–493.

[7] A. Kumar and M. Lease. 2011. Modeling annotator accuracies for supervised learning. In *Proceedings of the Workshop on Crowdsourcing for Search and Data Mining (CSDM) at the Fourth ACM International Conference on Web Search and Data Mining(WSDM)*. 19–22.

[8] E. Law and L. von Ahn. 2011. Human computation. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 5(3) (2011), 1–121.

[9] R. O. Parreiras, P. Ya. Ekel, and D. C. Morais. 2012. Fuzzy Set Based Consensus Schemes for Multicriteria Group Decision making Applied to Strategic Planning. *Group Decision and Negotiation, Springer* 21 (2012), 153–183.

[10] A. J. Quinn and B. B. Bederson. 2011. Human computation: a survey and taxonomy of a growing field.. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1403–1412.

[11] T. L. Saaty. 1980. The analytic hierarchy process. *McGraw-Hill.* (1980).

[12] H. S. Shih, H. J. Shyur, and E.S. Lee. 2007. An extension of TOPSIS for group decision making. *Math. Comput. Model.* 45 (2007), 801–803.

[13] H. Theil. 1963. On the symmetry approach to the committee decision problem. *Manage. Sci* 9, 3 (1963), 380–393.

[14] M. Venanzi, J. Guiver, G. Kazai, and P. Kohli. 2013. Bayesian Combination of Crowd-Based Tweet Sentiment Analysis Jugdments. In *Proceedings of the AAAI HCOMP Workshop on Crowdsourcing at Scale (CrowdScale)*. Palm Springs, USA, Shared Task Challenge Paper 03.

[15] Z. Yue. 2012. Approach to group decision making based on determining the weights of experts by using projection method. *Applied Mathematical Modelling* 36, 7 (2012), 2900–2910.

[16] Z. Yue. 2013a. An avoiding information loss approach to group decision making. *Applied Mathematical Modelling* 37, 1-2 (2013), 112–126.

[17] Zhongliang Yue. 2013b. Group decision making with multi-attribute interval data. *Information Fusion, Elsevier* 14 (2013), 551–561.

[18] Z. L. Yue. 2011. A method for group decision-making based on determining weights of decision makers using TOPSIS. *Appl. Math. Model.* 35, 4 (2011), 1926–1936.