AN AGGREGATION OF SURVEY QUESTIONNAIRES IN A FUZZY ENVIRONMENT AND ITS APPLICATION

Dae Young Choi

Department of Management Information Systems, Yuhan Univ., South Korea
dychoi@yuhan.ac.kr

ABSTRACT

In the response process of survey questionnaires, respondents generally express their opinions (assessment) in a fuzzy environment, i.e., assessment with a degree of membership, and in many cases, an aggregation of survey questionnaires is generally computed mathematically. In this paper, we propose a new aggregation method using the FEV (Fuzzy Expected Value) in summarizing survey questionnaires. As an application of proposed aggregation method, to quantify accurately the current state of opinions and the reasons for national IT investment increasing the rate of economic growth, we decided to circulate a survey questionnaire on national IT investment to interested parties throughout the research institute, government, IT industry. Generally, as we have already known and experienced in most of survey questionnaires, there was a little response. Moreover, according to interested parties, the result was biased views on the government IT investment. So, we consider big data analysis on the selection of the proper IT items for national investment. Proposed aggregation method is consisted of 2-phases. In phase 1, keyword selection for the proper IT items based on big data such as Facebook, Twitter, blog, Google, etc. is achieved. This phase is not mandatory (optional and dependent on the characteristic of survey). In phase 2, aggregation by using the FEV is obtained. Proposed aggregation method on survey questionnaires is particularly useful to find current focal issues such as trend, what's new, etc., on big data such as Facebook, Twitter, blog, Google, etc. Moreover, in many cases, aggregation using the FEV is a better representative value than the arithmetic mean. Generally, the FEV is more suitable than the value of averaging computation in searching for the representative value of fuzzy set.

Keywords: *Big Data Analysis, 2-Phase Aggregation, FEV (Fuzzy Expected Value)*

1.   Introduction

A survey questionnaire is a set of questions used in a survey. The survey questionnaire is a type of data gathering method that is utilized to collect, analyze and interpret the different views of a group of people from a particular population. The survey questionnaire has been used in the different field such as research, marketing, politics views, psychology, and so on. People use survey questionnaire to gather information that is beneficial to a group of individuals. The survey questionnaire uses statistical analysis to collect data, and the result of it will be used in the development of an individual or to a community, as described in the Web site "*what is a survey questionnaire?*"

In this paper, for each survey questionnaire, the following question was asked: "what is the rate between 0 and 1 to which you agree, that 'necessity on national IT investment' is good". In general, we evaluate the appropriateness of this questionnaire and then answer the

questionnaire with a degree of agreement. That is, we generally answer the questionnaire fuzzily and subjectively with the rate between 0 (disagree) and 1(agree).

Aggregation of information is a major problem for all kinds of knowledge-based systems, from image processing to decision making, from pattern recognition to machine learning. The aggregation operators are mathematical objects that have the function of reducing a set of numbers into a unique representative (or meaningful) number. In a rather informal way, the aggregation problem consists in aggregating n-tuples of objects all belonging to a given set into a single object of the same set. In the case of mathematical aggregation operator, this set is all the real numbers. In this setting, an aggregation operator is simply a function, which assigns a real number y to any n-tuple $(x_1, x_2, \ldots, x_n)$ of real numbers: $y = Agg\ (x_1, x_2, \ldots, x_n)$. Aggregation plays a central role as a means of combining all opinions that are expressed by group members (Choi, 2008).

In this paper, we propose an aggregation method using the FEV (Fuzzy Expected Value) in summarizing survey questionnaires. As an application of proposed aggregation method, to quantify accurately the current state of opinions and the reasons for national IT investment increasing the rate of economic growth, we decided to circulate a survey questionnaire on national IT investment to interested parties throughout the research institute, government, IT industry. We consider big data analysis on the selection of the proper IT items for national investment. Proposed aggregation method is consisted of 2-phases. In phase 1, keyword selection for the proper IT items based on big data such as Facebook, Twitter, blog, Google, etc. is achieved. This phase is not mandatory (optional and dependent on the characteristic of survey). In phase 2, aggregation by using the FEV is obtained. Proposed aggregation method on survey questionnaires is particularly useful to find current focal issues such as trend, what's new, etc., on big data. Moreover, in many cases, aggregation using the FEV is a better representative value than the arithmetic mean (see Example 1 and 2). Generally, the FEV is more suitable than the value of averaging computation in searching for the representative value of fuzzy set (Choi, 1999, Friedman *et al*., 1997, Friedman *et al*., 1989, Kandel, 1981, Schneider *et al*., 1987). However, for the better aggregation, both aggregation methods are complementary rather than competitive.

In recent years, social networking is enormously increasing. The rapid adoption of smartphones and SNS (Social Network Services) is driving up the usage of social networking (Manyika *et al*., 2011). 'Big data' refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse. Big data generally includes semi-structured data, unstructured data, etc. It does not conform with the formal structure of data models associated with existing relational databases or other forms of data tables (Manyika *et al*., 2011). Big data analytics is the process of examining large amounts of data of a variety of types to uncover hidden patterns, unknown correlations and other useful information. Such information can provide competitive advantages over rival organizations and result in business benefits such as more effective marketing and increased revenue. Big data sources may include credit card usage data, Web server logs, Internet clickstream data, social media activity, mobile phone call detail records, information captured by sensors, photos in SNS, etc (Manyika *et al*., 2011). In this paper, however, the domain of big data is restricted to Internet and SNS data such as Facebook, Twitter, etc.

2.   Existing Aggregation Operators and the FEV

Fuzzy set theory provides an attractive aggregation connectives for integrating membership values representing uncertain information. These connectives can be categorized into the following three classes: union, intersection and compensation connectives. In a fuzzy environment, existing aggregation operators are, in general, t-norm, t-conorm, mean operators, Ordered Weighted Averaging Operators (OWA) (Yager, 1993) and γ-operator (Zimmermann and Zysno, 1980). These aggregation operators have some problems in that they do not properly reflect the situation in the aggregation process. In order to solve this problem the fuzzy situation

assessment model (FSAM) is proposed (Choi, 2008). An overview of existing aggregation operators is described in (Choi, 1999, Klir and Folger, 1988).

2.1   Existing Aggregation Operators

2.1.1   Arithmetic Mean

It is often used since it is simple and satisfies the properties of monotonicity, continuity, symmetry, associativity, idempotence and stability for linear transformations (Choi, 1999). The mean always lies between the 'and' and 'or' operator. In many applications, especially multi-criteria decision making (MCDM), the union and intersection don't always capture the necessary aggregation of the fuzzy sets. In some of these cases, a mean-type aggregation is more appropriate.

2.1.2   Median

An aggregation operator that follows the idea obtaining "a middle value" is the median. It consists in ordering the arguments from the smallest one to the biggest one. And then takes the element in the middle. This aggregation operator satisfies the boundary conditions, monotonicity, symmetry, idempotence and evidently the compensation behavior. There exists a generalization of this operator: The k-order statistics, with which we can choose the element on the kth position on the ordered list (Choi, 1999).

2.1.3   Ordered Weighted Averaging Operators

The Ordered Weighted Averaging Operators (OWA) provide a means for aggregating scores associated with the satisfaction of multiple criteria, which unifies in one operator the conjunctive and disjunctive behaviors (Yager, 1993) :

$$\text{OWA}(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} w_j \, x_{\sigma(j)} \tag{1}$$

where σ is a permutation that orders the elements : $x_{\sigma(1)} \leq x_{\sigma(2)} \leq \ldots \leq x_{\sigma(n)}$.

The weights are all non-negative ($w_i \geq 0$) and their sum equals one ($\sum_{i=1}^{n} w_i = 1$). This operator has been proved to be very useful, because of its versatility as described in (Yager, 1993). The OWA operators provide a parametrized family of aggregation operators, which include many of the well-known operators such as the maximum, the minimum, the k-order statistics, the median and the arithmetic mean. In order to obtain these particular operators we should simply choose particular weights. For example, for all i, if $w_i = 1/n$, the OWA operator becomes the arithmetic mean. The OWA operators are commutative, monotone, idempotent. They are stable for positive linear transformations and have a compensatory behavior in that the aggregation done by an OWA operator always is between the minimum and the maximum. The properties of the OWA operators are explained in considerable details in (Yager, 1993).

2.1.4   T-norms and t-conorms

The t-norms generalize the conjunctive 'AND' operator and the t-conorms generalize the disjunctive 'OR' operator. This allows them to be used to define the intersection and union operations in fuzzy logic. T-norms and t-conorms play a notable role in fuzzy logic theory, but these operators do not admit a compensating behavior. Particular complete works are presented in (Klir and Folger, 1988).

2.1.5   γ-operator

In Zimmermann and Zysno (1980), they suggest the compensatory behavior seems crucial in the aggregation process. However, t-norms and t-conorms lack of compensation behavior. They

discover that in a decision making context humans do not follow exactly the behavior of a t-norm (nor of a t-conorm) when aggregating. In order to get closer to the human aggregation process, they proposed an operator on the unit interval based on t-norms and t-conorms (Zimmermann *et al*., 1980):

$$\mu_\theta = (\Pi\mu_i)^{1-\gamma} (1 - \Pi(1 - \mu_i))^\gamma \qquad (2)$$

where i = 1, 2, …, n, n = number of sets to be connected and $0 \leq \mu_i \leq 1$, $0 \leq \gamma \leq 1$.

Here the parameter γ indicates the degree of compensation. If $\gamma = 0$, then $\mu_\theta = \Pi\mu_i$. This equals the product and provides the truth values for the connective 'AND'. If $\gamma = 1$, then $\mu_\theta = 1 - \Pi(1 - \mu_i)$. This formula equals the generalized algebraic sum and provides the truth value for the connective 'OR'. The γ-operator is pointwise injective, continuous, monotonous, commutative and in accordance with the truth tables of dual logic (Klir *et al*., 1988).

In this paper, however, the domain of aggregation operators is restricted to averaging function, which assigns a real number y to any n-tuple $(x_1, x_2, …, x_n)$ of real numbers: $y = Agg (x_1, x_2, …, x_n)$.

2.2 Fuzzy Expected Value (FEV)

Definition1. Let B be a Borel field of subsets of the real line $\Re$. A set function $\mu(\cdot)$ is defined on B is called a fuzzy measure if it has properties as follows:

  (1) $\mu(\emptyset) = 0$ ($\emptyset$ is the empty set),
  (2) $\mu(\Omega) = 1$ (Ω is the entire sample space),
  (3) if $\alpha, \beta \in B$ with $\alpha \subset \beta$, then $\mu(\alpha) \leq \mu(\beta)$,
  (4) if $\{\alpha_j \mid 1 \leq j < \infty\}$ is a monotone sequence, then $\lim_{j\to\infty} [\mu(\alpha_j)] = \mu [\lim_{j\to\infty} (\alpha_j)]$.

In Definition 1, (1) and (2) means that the fuzzy measure is bounded and nonnegative, (3) means monotonicity, and (4) means continuity. It should be noted that if Ω is a finite set, then the continuity requirement can be deleted. Let χ_A be a B-measurable function such that $\chi_A \in [0, 1]$. The FEV of χ_A over the set A, with respect to the fuzzy measure $\mu(\cdot)$ is defined as follows (Friedman *et al*., 1989):

$$\text{FEV}(\chi_A) = \sup_{T \in [0,1]} \{\min[T, \mu(\xi_T)]\} = \sup_{T \in [0,1]} \{\min[T, f_A(T)]\}, \qquad (3)$$

where $\xi_T = \{x \mid \chi_A(x) \geq T\}$, and $\mu\{x \mid \chi_A(x) \geq T\} = f_A(T)$ is a function of the threshold T.

The ability to summarize data provides an important method for getting a grasp of the meaning of a larger collection of data. It enables humans to understand the environment in a manner amenable to future useful manipulation. It also provides a starting point for the ability to make useful inferences from large collections of data. The mean does help in understanding the content of data, but in some respects, it may provide too terse as a summary (Choi, 2008, Choi, 1999). The FEV is usually used for evaluating the most 'representative' or 'typical' value of a fuzzy set as a measure of general tendency (Choi, 2008, Choi, 1999, Friedman *et al*., 1997, Friedman *et al*., 1989, Kandel, 1981, Schneider *et al*., 1987).

2.3 Examples of FEV

Example 1. In a certain decision situation, for computational simplicity, we assume that decision-maker consider 100 situation factors to solve his/her decision problem. For each situation factor, the following question was asked (Choi, 1999): "what is the rate between 0 and 1 to which you agree, that 'situation factor' is good". The results obtained as follows:

45 situation factors are rated: 0.0
40 situation factors are rated: 0.2
15 situation factors are rated: 0.25

We thus have 3 different thresholds (0.0, 0.2, 0.25). The first thing we have to do is to check how many answers are above each threshold (percentage-wise). Obviously, 100 answers are above or equal to 0.0, 55 answers are above or equal to 0.2, and 15 answers are above or equal to 0.25. Pairing these data and rearranging them by increasing order of the value of threshold, we obtain the following three [T, μ] pairs:

(0.0, 1.0)
(0.2, 0.55)
(0.25, 0.15)

Now, the minimum value of each pair is:

min (0.0, 1.0) = 0.0
min (0.2, 0.55) = 0.2
min (0.25, 0.15) = 0.15

Thus, the FEV, which is the maximum of all these minima, is:

max (0.0, 0.2, 0.15) = 0.2

It means that the representative value of "good", in 'decision situation is good' is 0.2. i.e., FEV = 0.2. In this case, arithmetic mean of these data is 0.1175. We know that arithmetic mean skewed toward 0.0 more than the FEV. Note that the FEV is a better representative value than the arithmetic mean. Generally, the FEV is more suitable than the value of averaging computation in searching for the representative value of fuzzy set (Choi, 2008, Choi, 1999, Friedman *et al.*, 1997, Friedman et al., 1989, Kandel, 1981, Schneider et al., 1987).

Example 2. Let *X* consist of two populations {5, 1} with grades of membership {0.1, 1.0}, respectively. The arithmetic mean is 0.25 while FEV = 1/6.

min (0.1, 6/6) = 0.1
min (1.0, 1/6) = 1/6

Thus, the FEV, which is the maximum of all these minima, is:

max (0.1, 1/6) = 1/6

Note that the FEV is a better representative value than the arithmetic mean.

Many explanations and interpretations for FEV and fuzzy set are described in considerable detail in (Choi, 2008, Choi, 1999, Friedman *et al.*, 1997, Friedman *et al.*, 1989, Kandel, 1981, Schneider *et al.*, 1987).

3.   An Application of Proposed Aggregation Method

To quantify accurately the current state of opinions and the reasons for national IT investment increasing the rate of economic growth, we decided to circulate a survey questionnaire on national IT investment to interested parties throughout the research institute, IT industry. As we have already known and experienced in most of survey questionnaires, there was a little response. Moreover, according to interested parties, the result was biased views on the government IT investment. So, we consider big data analysis on the selection of the proper IT

items for national investment. This experiment was conducted under the following conditions: First, use open API, crawling tool, etc., to collect keyword (IT items)-related data. Second, boundary of crawling is domestic level. Third, crawling targets for big data are Facebook, Twitter, blog, Internet portal. After keywords selection process for national IT investment, we obtain some keywords for IT items such as 'AI' 'IoT', 'big data', etc. Now, these keywords for IT items can be implicitly used as a trend. Using these selected IT items, a survey questionnaire on national IT investment in Table 1 is obtained and it is circulated to interested parties such as the research institute, IT industry, etc. We try to find IT items' trend ranking for national IT investment. The following each question was asked: "What is the rate between 0 and 1 to which you agree, that 'necessity on national IT investment' is good", respectively, as in Table 1.

Table 1. Survey questionnaire on IT items

Questions	Necessity on national IT investment
Question 1	National IT investment on Mobile App.
Question 2	National IT investment on Big Data
Question 3	National IT investment on IoT
Question 4	National IT investment on AR
Question 5	National IT investment on Healthcare
Question 6	National IT investment on VR
Question 7	National IT investment on AI
Question 8	National IT investment on Info. Security
Question 9	National IT investment on Cloud
Question10	National IT investment on Platform

* Assessment 0 means disagree and 1 means agree

For each survey questionnaire, for computational simplicity, we assume that 100 different views of a group of people from a particular population such as the research institute, IT industry, etc., are collected. For example, on question 1 ("National IT investment on Mobile App.") in Table 1, the results obtained as aforementioned in Example 1.

45 opinions are rated: 0.0
40 opinions are rated: 0.2
15 opinions are rated: 0.25

In this case, aggregation results using the average and the FEV on question 1 are shown respectively in Table 2 and 3. Similarly, aggregation results using the average and the FEV on questions 2-10 are obtained as shown respectively in Table 2 and 3.

Table 2. Aggregation Results Using the Average on Survey Questionnaires

Rank	IT items	Aggregation Results
1	Big Data	0.73
2	IoT	0.72
3	AI	0.63
4	VR (Virtual Reality)	0.62
5	AR (Augmented Reality)	0.61
6	Healthcare	0.53
7	Cloud	0.43
8	Info. Security	0.41
9	Mobile App.	0.17
10	Platform	0.16

Table 3. Aggregation Results Using the FEV on Survey Questionnaires

Rank	IT items	Aggregation Results
1	AI	0.80
2	Big Data	0.63
3	IoT	0.61
4	VR (Virtual Reality)	0.53
5	AR (Augmented Reality)	0.52
6	Healthcare	0.50
7	Platform	0.49
8	Mobile App.	0.20
9	Cloud	0.19
10	Info. Security	0.17

According to aggregation methods, we may get different aggregation results as shown in Table 2 and 3. Proposed aggregation method in an aggregation of survey questionnaires can be explained by 2-phases as follows:

Phase 1 (Keyword selection based on big data): Based on big data such as Facebook, Twitter, blog, Google, keyword selection for the proper IT items is achieved as described in this Section. They become trend keywords such as what's new, real-time keywords on big data. This phase is not mandatory (optional and dependent on the characteristic of survey).

Phase 2 (Aggregation): Aggregation using the FEV is obtained.

Proposed aggregation method on survey questionnaires is particularly useful to find current focal issues such as trend, what's new, etc., on big data such as Facebook, Twitter, blog, Google. Moreover, aggregation using the FEV is a better representative value than the arithmetic mean (see Example 1 and 2). Generally, the FEV is more suitable than the value of averaging computation in searching for the representative value of fuzzy set (Choi, 2008, Choi, 1999, Friedman *et al*., 1997, Friedman *et al*., 1989, Kandel, 1981, Schneider *et al*., 1987). However, for the better aggregation, both aggregation methods are complementary rather than competitive. We briefly summarize the differences between the proposed aggregation method and existing aggregation using the average in Table 4.

Table 4. Comparisons

	Proposed aggregation method	**Existing aggregation using the Average**
Big data analysis (optional)	Keywords selected from big data (Facebook, Twitter, blog, Google, etc.)	Not used
Aggregation method	FEV	Average
Computing representative value of fuzzy set	Aggregation based on the fuzzy measure $\mu(\cdot)$ and threshold T (See Eq.(3), Examples 1, 2)	Fuzzy measure $\mu(\cdot)$ in Eq.(3) is not used
− For the better aggregation, both aggregation methods are complementary rather than competitive		

4.   Conclusions

In this paper, we propose a new aggregation method using the FEV in summarizing survey questionnaires. It is consisted of 2-phases. In phase 1, keyword selection based on big data such as Facebook, Twitter, blog, Google, etc. is achieved. This phase is not mandatory (optional and dependent on the characteristic of survey). In phase 2, aggregation using the FEV is obtained. Proposed aggregation method on survey questionnaires is particularly useful to find current focal issues such as trend, what's new, etc., on big data such as Facebook, Twitter, blog, Google. Moreover, in many cases, aggregation using the FEV is a better representative value than the arithmetic mean. Generally, the FEV is more suitable than the value of averaging computation in searching for the representative value of fuzzy set (Choi, 2008, Choi, 1999,

Friedman *et al*., 1997, Friedman *et al*., 1989, Kandel, 1981, Schneider *et al*., 1987). However, in many cases, the use of the FEV as an aggregation method may also lead to improper results. For the better aggregation, both aggregation methods are complementary rather than competitive. There is also a need for some field works to test the value of the proposed approach. In addition, we need further research on different aggregation methods such as clustering FEV (CFEV), most typical value (MTV) (Friedman *et al*., 1997) for n-dimensional sets, etc.

Acknowledgement

References

Choi, D.-Y. (2008, Jan.). Aggregation of Preferences Based on FSAM in GDSS, *IEEE Trans. on systems, man and cybernetics – part A: systems and humans* 38(1), 2-8.

Choi, D.-Y. (1999, Feb.). A new aggregation method in a fuzzy environment, *Decision Support System* 25(1), 39–51

Friedman, M, Henne, M., Kandel, A. (1997). Most Typical Values for Fuzzy Sets, *Fuzzy Sets and Systems* 87, 27-37.

Friedman, M, Schneider, M., Kandel, A. (1989). The use of weighted fuzzy expected value in fuzzy expert systems, *Fuzzy Sets and Systems* 31, 37-45.

Kandel, A. (Sep. 1981). Fuzzy expectation and energy states in fuzzy media, *Fuzzy Sets and Systems* 6(2), 145–160.

Klir, G.J. and Folger, T.A. (Prentice Hall, 1988). *Fuzzy sets, uncertainty and information*.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute.

Schneider, M., Friedman, M., Kandel, A. (1987, March). On fuzzy reasoning in expert systems, *FSU-SCRI-87-09*.

What Is a Survey Questionnaire? | Examples. Retrieved from *https://www.examples.com/ education/what-is-a-survey-questionnaire.html*

Yager, R. R. (1993). Families of OWA Operators, *Fuzzy Sets and Systems* 59,125-148.

Zimmermann, H. J., Zysno, P. (1980). Latent Connectives in Human Decision Making, *Fuzzy Sets and Systems* 4, 37-51.