Proceedings of the 2007 IEEE Symposium on Computational
Intelligence in Multicriteria Decision Making (MCDM 2007)

A Two-level Information Filtering Model in Generating Warning Information

Jun Ma, Jie Lu, Guangquan Zhang
Faculty of Information Technology, University of Technology, Sydney
PO Box 123, Broadwary, NSW 2007, Australia
Email: {junm, jielu, zhangg}@it.uts.edu.au

***Abstract*— Information filtering is an important component in warning systems. This paper proposes a two-level information filtering model for generating warning information. In this model, information is represented by n-tuple, whose elements are values of information features. The features of information are divided into critical and uncritical features. Within this model, the collected information is filtered in two stages by users at different levels. At the first stage, exceptions are separated from normal information. And at the second stage, critical exceptions are separated from uncritical information. To illustration the proposed model, an example is discussed.**

I. Introduction

Warning systems can help users to find out potential risk timely and make a better decision in emergency through generating warning information. That to generate warning information involves a series of information processing, including information collecting, information checking, information filtering, and information fusion. Among them, information filtering mainly focuses on how to extract abnormal information effectively and accurately from a large amount of collected information.

Information filtering technique, as a kind of techniques for solving information overload problems, has been applied in many application fields, such as web information search/retrieval [1]–[3], medical information classification [4], [5], dynamic data rectification in industrial process [6], Websites design [7], recommender systems [8], and emergency management [9], [10]. Generally speaking, information filtering can be treated as information classification, in which information is divided into different classes, and therefore machine learning methods, such as evolutionary computation [11], artificial neural networks [1], and probabilistic learning [12], are widely used in information filtering. Basic information filtering strategies include three main categories, which are cognitive filtering, social filtering, and economic filtering [13]. In cognitive filtering, information is selected mainly based on users' interests and requirements. In social filtering, the filtering processing is normally established on the collaboration between users. In economic filtering, the cost for getting profit from filtering processing is more emphasized and concerned. These strategies have been integrated in some application systems [14].

In order to design efficient information filtering process mechanism for a warning system, some characteristics of the organization structures and responsibilities of users and the information forms they face to in a real situation are considered. First, users of a real situation can always be grouped into three main decision levels, namely, the operators (at the lower level), the managers (at the middle level), and the decision makers (at the top level). In general, a decision target is divided by the decision makers into many sub-targets and each sub-target is assigned to a manager group. Then managers will divide the received sub-target into several tasks and each task will be send to an operator group. To complete the received task, operators are mainly responsible for extracting exceptions from the collected information by using information processing techniques which are similar to cognitive filtering. For the managers, their responsibilities include evaluating exception reports from operators and extracting and reporting critical exceptions to the decision makers. Hence, the managers need to synthesize exceptions reports for different tasks and extract critical exceptions by using information processing technique which resembles collaborative filtering. For the decision makers, their main responsibilities are evaluating critical exceptions, predicating possible risk, and making decision. Since the information process of decision makers is mainly integrating information from multiple manager groups, the main information process of decision makers may be treated as information fusion rather than information filtering. Therefore, information filtering is mainly carried out at two levels, i.e., by the operators and by the managers. Secondly, the information processed by the operators is collected from real information sources and is often in concrete forms such as numbers, texts, or images. While the information processed by managers is in abstract forms, for instance trends analysis. Therefore, the information filtering in a warning system is hierarchically organized and realized in different categories.

Based on above analysis, this paper proposes a two-level information filtering model in generating warning information. The rest of the paper is organized as: Section II introduces some related works on information filtering technique. Section III proposes a two-level information filtering model which can be used in warning systems. Section IV illustrates the proposed model by an example. Our future work is discussed in Section V.

II. Related Works

Information filtering has been an important research issue in information science, computer sciences, management science and so on for many years [15]. As a crucial type of information requirements and corresponding information services, information filtering has a close relationship with information retrieval. Information retrieval systems aim at meeting the need for once-off information, but information filtering systems fulfill the need for continuing information on long-term interest [16].

With the development of Internet. Information filtering technology becomes more and more important for using on-line information efficiently and many application systems have been developed [17]–[19]. Pollock [20] reported a message filtering system (ISCREEN). This system uses formal rules to drive message filtering. Filtering is carried out through rules matching. Bell and Moffat [16] discussed the design of a high performance information filtering system based on a vector-space-model. Palme [13] made an overview of methods and problems in information filtering on the Internet. Belkin et al [15] compared information filtering and information retrieval based on their characteristics. Because information filtering can be treated as a kind of information classification, many information classification methods are applied to information filtering. Sebastiani [21] made an overview about issues in document representation, classifier construction, and classifier evaluation. To reduce the volume of terms set for representing the user's interest in information filtering system, Kuflik et al [22] described a terms selection technique based on dimensionality-reduction mechanism.

Matching is one of main strategies in information filtering, which depends on the user's interest (profile). For the simplicity of processing, the user's interest is represented as collections of terms. Hence, the matching established on terms cannot sufficiently explored the meanings (known as senses) behind those terms. Kehagias et al [23] compared the categorization accuracy of classifiers based on words to that of classifiers based on senses. Through their experiments, they thought that the use of senses cannot result in significant improvement in classification.

Considering the dynamics, domain specificity, and cause-effect relationship of information in an emergent situation, Atoji et al [24] proposed an information filtering model based on self-organizing map (SOM) technique. In their model, information filtering is treated as a kind of information classification. The input information is semi-structured messages and is classified into several categories under the supervision of its user's preferences and requirements. The major challenge for establishing an information filtering system is that the volume of collected information may be overwhelming and some of the information are not correct. Janeja et al [25] reported an alert management system (AMS), which can generate meaningful alerts from the collected alerts. The alert generation module of the system has the functions of information filtering, information integration, and dynamic flow identification.

Research has shown that information filtering plays more and more important role in a warning system. However, current methods mainly focus on how to improve the efficiency and accuracy of filtering processing and ignore the generation of warning information by filtering processing. In addition, information filtering in an emergency system is carried out at many levels and has different targets at different levels. But few works is reported on this topic. Fig. 1 gives a general framework of emergency risk management. From this figure, it is known that risk information is processed in a series of stages, such as identifying, analyzing and evaluating, before being accepted as real risk. At each stage, the risk information is classified (filtered) in the light of tasks at that level. Similarly, to generating warning information by using a warning system, the collected information should be processed at several stages. This course is accompanied by information filtering processing. This paper focuses on multiple-level information filtering technique for generating warning information.

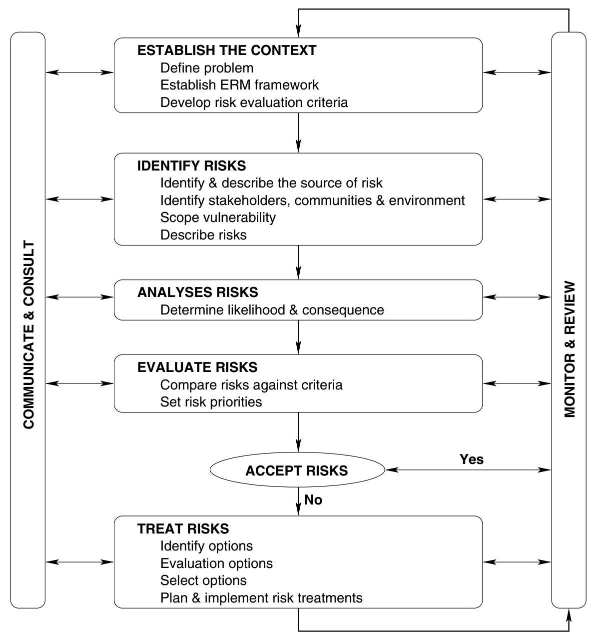

Fig. 1. A framework of emergency risk management [26].

III. A Two-level Information Filtering Model

In this section, we shall propose a two-level information filtering model in warning systems and discuss the features of information filtering at the lower and middle levels shown in the model.

A. Overview

Information filtering in a complex system, such as a warning system, is under a hierarchical processing. Based on this idea, we identify multiple-stage information processing issue and establish a hierarchical information filtering processing model (Fig. 2).

In this model, the users are classified into three levels, i.e. operators, managers, and decision makers. The users at the lower lever are operators, who are responsible for information collecting, information features extracting and representing,

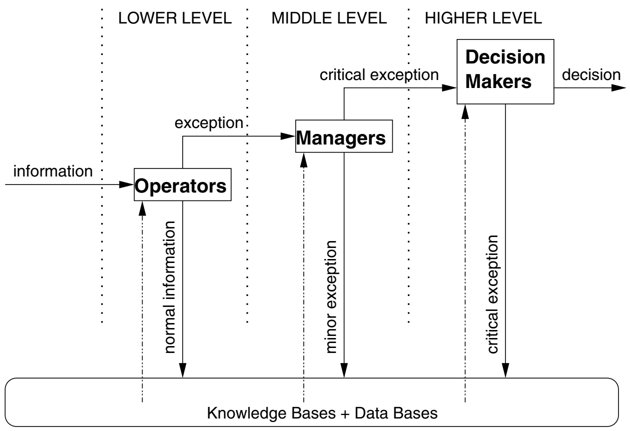

Fig. 2. A hierarchical information filtering processing model.

and information classifying. The users at the top level are decision makers. Decision makers send information filtering requirements to and collect reports of emergency from the users at the middle level (managers), and make decisions. The managers' duties include spreading decision-makers' information filtering requirements to and analyzing exception reports from operators, and reporting emergency to decision-makers.

The collected information will be processed at all levels. This model mainly focuses on the filtering at the lower and middle levels. The two levels are linked together and the output of lower level is the input of the middle level. Hence, the collected information will be filtered via two stages. The first stage is carried out by the operators, and the second stage is taken by the managers.

At the first stage, the information is collected from related information sources. This information may have various forms, such as text, image, and video. For the convenience of information processing and to meet the requirement of information filtering, the related features of this information will be extracted first. Then these features will be embedded into a flexible structure for information representation, which is usually a semi-structure message composing of pairs of feature and feature value. Next, the collected information will be classified into normal information and abnormal information through comparing the extract feature value of the collected information with that of normal criterion feature by feature. The abnormal information will be reported to managers as exceptions for further analysis and the normal information will be stored for other using aim in the future. After this stage, exceptions are separated from normal information.

At the second stage, the reported exceptions will be further evaluated in order to separate exceptions caused by critical features from those caused by uncritical features. In general, exceptions caused by critical features imply higher possibility of potential disaster, which are primary emergency warning. Firstly, some features will be selected as critical features according to the information filtering requirement. Then the reported exceptions will be classified into critical exceptions and uncritical exceptions. The critical exceptions are caused by at least one critical feature and will be reported to decision-makers as emergency for consideration. The uncritical exceptions are caused totally by uncritical features, which can be treated as minor exceptions and will be stored in a database for further using. After the second stage, emergency is separated from the uncritical exceptions.

In an emergency situation, a concrete objective is given such as earthquake, tsunami, bushfire, which will be called a decision target and will be denoted by G. Correspondingly, some features can be selected as the basis for collecting and filtering information. Suppose the selected features are

$$\mathcal{F} = \{F_1, F_2, \ldots, F_n\}$$

Because each feature has individual contribution to the decision target, a degree α to measure the contribution can be associated with each feature. Without loss of generality, suppose the all degrees are real numbers in $[0, 1]$. Moreover, each feature may have many possible values, which are denoted by S_i, $i = 1, 2, \ldots, n$.

The information filtering requirements are related to the warning grades. The requirement in general is enhanced with the increase of warning grade and in turn more critical features of information should be selected and considered. For convenience, a set of warning grades is assumed to be $D = \{d_1, d_2, \ldots, d_m\}$ and $d_1 < d_2 < \cdots < d_m$.

Let g_F be a mapping from D to $\mathscr{P}(\mathcal{F})$ (power set of $\mathcal{F}$) such that

$$g_F(d_i) \subseteq g_F(d_j) \text{ if } d_i < d_j.$$

For each d_i in D, each element in $g_F(d)$ is called a critical feature under warning grade d_i. In the following, $g_F(d)$ will be written as $\mathcal{F}^{(d)}$.

For each feature F and a given warning grade d, some values can be selected to describe normal situation under d. In this stage, normal values of feature F under a warning grade d will be denoted by $S^{(d)}$. Normal values indicate the acceptable changing of the feature under the given warning grade. Hence, it is rational to assume that $S^{(d_i)} \subseteq S^{(d_j)}$ if $d_i > d_j$.

In the following, the detailed process of the two-level information filtering is given under the assumption that the warning grades D is finite.

B. Information Filtering at the Operator Level

Information is collected from multiple sources by operators for initial filtering. The main tasks of operators include:

TASK 1: Feature extracting and information representation.
TASK 2: Feature verifying and information classification.

1) Feature extracting and information representation: Feature extracting aims at finding the primary values of the referred features. The result of feature extracting is to form a semi-structure message for the input information. Notice that n features are selected according the decision target, a n-tuple

$$u = (u_1, u_2, \ldots, u_n)$$

is used to represent a piece of information. Each element of u represents values of the corresponding feature. In general, not all features must be existed in a piece of information. Hence, we use the following approach to construct the u corresponding to a piece of information.

For any $F_i \in \mathcal{F}$, if F_i can be extracted from the input information, then u_i is assigned the obtained values of F_i, otherwise, u_i is assigned as $\emptyset$.

2) Feature verifying and information classification: Each obtained u of the input information will be compared with that of normal information feature by feature. After the comparison, an exception report $e_u = (e_1, \ldots, e_n)$ will be generated and

$$e_i = \begin{cases} 1, & \text{if } u_i \neq \emptyset \text{ and } u_i \in S_i^{(d)}, \\ 0, & \text{otherwise.} \end{cases}$$

According to e_u, a piece of information is abnormal if and only if there exists some i such that $e_i = 1$. The abnormal information is exception we are concern about. Exception will be reported to managers in the form of (u, e_u).

C. Information Filtering at the Manager Level

Exceptions are reported to manager for further filtering. The main tasks of managers include:

TASK 1: Exceptions evaluation.

TASK 2: Emergencies evaluation.

1) Exceptions evaluation: Exceptions evaluation will separate critical exceptions from uncritical ones. According to the given warning grade d, some features can be selected as critical features which will be denoted by $\mathcal{F}^{(d)}$.

For each exception report e_u, if there is i such that $e_i = 1$ and $F_i \in \mathcal{F}^{(d)}$, then the exception corresponding to e_u is an critical exception.

2) Emergencies evaluation: An exception is critical means that the exception is caused by critical features, however, cannot indicate the degree of potential disaster. In order to present suitable recommendation to the decision makers for better decision making, the managers will present a synthesized emergency evaluation to the decision makers. The synthesized emergency evaluation is of form (u, e_u, e_p), where u is the input information, e_u is the exception report, and e_p is a predicated emergency degree.

For a critical exception e_u, a synthesized evaluation $\hat{u}$ will be obtained by $\mathrm{Agg}(\mathcal{F}^{(d)})$ and $\mathrm{Agg}(e_u)$, where Agg is a aggregation operator which can be selected according to real situation and selecting principles in [27]. The $\hat{u}$ is defined as

$$\hat{u} = \frac{\mathrm{Agg}(e_u)}{|\mathcal{F}^{(d)}|} \tag{1}$$

Here, the generalized weighted sum is used for example as follow:

$$\mathrm{Agg}(e_u) = \sum e_i \times \alpha_i = \sum_{F \in \mathcal{F}^{(d)}} e_F \times \alpha_F. \tag{2}$$

To obtain the potential emergency degree under the circumstance u, a mapping Q_d is defined according to the given warning grade d. Without loss of generality, suppose

$$Q_d : D \to [0, 1], \tag{3}$$

and $Q_d(d_i) \leqslant Q_d(d_j)$ if $d_i < d_j$. $Q_d(d_i)$ is called emergency degrees. Then e_p is determined by

$$e_p = \begin{cases} d_1, & \hat{u} < \frac{Q_d(d_2)+Q_d(d_1)}{2}, \\ d_i, & \frac{Q_d(d_{i-1})+Q_d(d_i)}{2} \leqslant \hat{u} < \frac{Q_d(d_i)+Q_d(d_{i+1})}{2}, \\ & \qquad 2 \leqslant i \leqslant m-1 \\ d_m, & \frac{Q_d(d_{m-1})+Q_d(d_m)}{2} \leqslant \hat{u} \leqslant 1, \end{cases} \tag{4}$$

Therefore, the emergence report (u, e_u, e_p) will be reported to decision makers for decision making.

IV. Illustration Example

Suppose in a situation, five warning grades are defined, which are d_1, d_2, d_3, d_4, d_5, and

$$d_1 < d_2 < d_3 < d_4 < d_5.$$

The current warning grade is designed as d_3. According to the warning grade d_3, nine features of information are selected for reference and three of them (F_2, F_3, and F_9, marked with (+) in Table I) are selected as critical features.

A. Filtering by Operators

For convenience, suppose three pieces of information u, v, and w are collected by the operators, which are shown in Table I. Suppose the normal information is $(0, 0, 0, 0, 0, 0, 0, 0, 0)$.

TABLE I
Collected information.

	u	v	w
F_1	1	1	1
F_2(+)	2	0	0
F_3(+)	0	0	1
F_4	1	1	1
F_5	2	1	3
F_6	0	0	0
F_7	0	0	0
F_8(+)	0	0	1
F_9	0	0	1

By Table I, the exception reports are generated as Table II. Obviously, these three pieces of information are all exceptions.

TABLE II
Exception reports

e	u	v	w
e_1	1	1	1
e_2	1	0	0
e_3	0	0	1
e_4	1	1	1
e_5	1	1	1
e_6	0	0	0
e_7	0	0	0
e_8	0	0	1
e_9	0	0	1

Hence three exception reports (u, e_u), (v, e_v), (w, e_w) are reported to the managers.

B. Filtering by Managers

According to the reports, it is easy to know that the information v is caused by uncritical features. So v is treated as uncritical exception and stored. For the information u and w, synthesized evaluation $\hat{u}$ and $\hat{w}$ are obtained as below.

Let $\alpha_2 = 0.9$, $\alpha_3 = 0.7$, and $\alpha_9 = 0.6$. By Eq. (1),

$$\hat{u} = \frac{0.9 \times 1}{0.9 \times 1 + 0.7 \times 1 + 0.6 \times 1} = 0.41$$

$$\hat{w} = \frac{0.7 \times 1 + 0.6 \times 1}{0.9 \times 1 + 0.7 \times 1 + 0.6 \times 1} = 0.59.$$

Suppose Q_{d_3} is defined as

$$Q_{d_3}(d_1) = 0.2,$$
$$Q_{d_3}(d_2) = 0.4,$$
$$Q_{d_3}(d_3) = 0.5,$$
$$Q_{d_3}(d_4) = 0.6,$$
$$Q_{d_5}(d_1) = 0.65.$$

So the emergency degrees for u and w are d_2 and d_4. Notice that the current warning grade is d_3, the information u indicates that the possibility of potential disaster is decreasing. While, the information w shows that the possibility is increasing.

The emergency reports (u, e_u, e_{p_u}) and (w, e_w, e_{p_w}) will be generated and reported to the decision makers.

Through this example, we can see the proposed model can efficiently generate warning information from the collected information and can also predict potential emergency degree through the filtering processing.

V. CONCLUSION

This paper proposes a two-level information filtering model for generating warning information. In this model, the collected information is filtered in two stages. At the first stage, exceptions are separated from normal information, and at the second stage, critical exceptions are separated from uncritical information. An example is given to illustrated the proposed model.

Notice that in the information filtering, the details of the collected information play important role to affect decision making, the following issues will be considered as our further research tasks: (1) efficient methods for information feature extraction and representation, and (2) appropriate methods for information fusion after filtering.

ACKNOWLEDGMENT

This work is supported by the Australia Research Council (ARC) Discovery Grant with grant number DP0559213.

REFERENCES

[1] Z. Bogar, T. Kuflik, P. Shoval, and B. Shapira, "Automatic keyword identification by artificial neural networks compared to manual identification by users of filtering systems," *Information Processing and Management*, vol. 37, pp. 187–198, 2001.

[2] S. Singh and L. Day, "A rough-fuzzy document grading system for customized text information retrieval," *Information Processing and Management*, vol. 41, pp. 195–216, 2005.

[3] B. L. D. Bezerra and F. de A. T. de Carvalho, "A symbolic approach for content-based information filtering," *Information Processing Letters*, vol. 92, pp. 45–52, 2004.

[4] J. Mostafa and W. Lam, "Automatic classification using supervised learning in a medical document filtering application," *Information Processing and Management*, vol. 36, pp. 415–444, 2000.

[5] Y. Quintana, "Intelligent medical information filtering," *International Journal of Medical Informatics*, vol. 51, pp. 197–204, 1998.

[6] M. Kong, B. Chen, and B. Li, "An integral approach to dynamic data rectification," *Computers and Chemical Engineering*, vol. 24, pp. 749–753, 2000.

[7] A. Kohrs and B. Merialdo, "Creating user-adapted websites by the use of collaborative filtering," *Interacting with Computers*, vol. 13, pp. 695–716, 2001.

[8] S.-H. Min and I. Han, "Detection of the customer time-variant pattern for improving recommender systems," *Expert Systems with Applications*, vol. 28, pp. 189–199, 2005.

[9] Y. Atoji, T. Koiso, and S. Nishida, "Information filtering for emergency management," in *Proc. of the IEEE International Workshop on Robot and human Interactive Communication*, September 2000, pp. 27–29.

[10] Q. Gu, D. Mendonça, and D. Wu, "An exploration of information-seeking behavior in emergency management," in *IEEE International Conference on Systems, Man and Cybernetics*, 10 2003, pp. 1798–1803.

[11] D. Tauritz, J. N. Kok, and I. G. Sprinkhuizen-Kuyper, "Adaptive information filtering using evolutionary computation," *Information Sciences*, vol. 122, pp. 121–140, 2000.

[12] G. Amati and F. Crestani, "Probabilistic learning for selective dissemination of information," *Information Processing and Management*, vol. 35, pp. 633–654, 1999.

[13] J. Palme, "Information filtering," in *Proc. of ITS98 Conference, 1998.*, 1998. [Online]. Available: citeseer.ist.psu.edu/palme98information.html

[14] B. Shapira, P. Shoval, and U. Hanani, "Experimentation with an information filtering system that combines cognitive and sociological filtering integrated with user stereotypes," *Decision Support Systems*, vol. 27, pp. 5–24, 1999.

[15] N. J. Belkin and W. B. Croft, "Information filtering and information retrieval: two sides of the same coin?" *Communications of the ACM*, vol. 35, no. 12, pp. 29–38, 1992.

[16] T. A. H. Bell and A. Moffat, "The design of a high performance information filtering system," in *SIGIR'96: Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM Press, 1996.

[17] G. Paliouras, C. Papatheodorou, V. Karkaletsis, and C. D. Spyropoulos, "Discovering user communities on the Internet using unsupervised machine learning techniques," *Interacting with Computers*, vol. 14, pp. 761–791, 2002.

[18] O. Nouali and P. Blache, "A semantic vector space and features-based approach for automatic information filtering," *Expert Systems with Applications*, vol. 26, pp. 171–179, 2004.

[19] N. G. Shaw, A. Mian, and S. B. Yadav, "A comprehensive agent-based architecture for intelligent information retrieval in a distributed heterogeneous environment," *Deceision Support Systems*, vol. 32, pp. 401–405, 2002.

[20] S. Pollock, "A rule-based message filtering system," *ACM Transactions on Information Systems (TOIS)*, vol. 6, no. 3, pp. 232–254, 1988.

[21] F. Sebastiani, "Machine learning in automated text categorization," *ACM Computing Surveys*, vol. 34, no. 1, pp. 1–47, 2002.

[22] T. Kuflik, Z. Boger, and P. Shoval, "Filtering search results using an optimal set of terms identified by an artificial neural," *Information Processing and Management: an International Journal*, vol. 42, no. 2, pp. 469–483, 2006.

[23] A. Kehagias, V. Petridis, V. G. Kaburlasos, and P. Fragkou, "A comparison of word- and sense- based text categorization using several classification algorithms," *Journal of Intelligent Information Systems*, vol. 21, no. 3, pp. 227–247, 2003.

[24] Y. Atoji, T. koiso, M. Nakatani, and S. Nishida, "An information filtering method for emergency management," *Electrical Engineering in Japan*, vol. 147, no. 1, pp. 60–69, 2004.

[25] V. P. Janeja, V. Atluri, A. Gomaa, N. Adam, C. Bornhoevd, and T. Lin, "DM-AMS: employing data mining techniques for alert management," in *dg.o2005: Proceedings of the 2005 national conference on Digital government research*. Digital Government Research Center, 2005, pp. 103–111.

[26] E. M. Australia. (2003, January) Critical infrastructure emergency risk management and assurance. Emergency Management Australia. Mt Macedon Road, Mt Macedon, VIC 3441, Australia. [Online]. Available: http://www.ema.gov.au/agd/EMA/emaInternet.nsf/Page/Emergency_Management

[27] G. Beliakov and J. Warren, "Appropriate choice of aggregation operators in fuzzy decision support systems," *IEEE Trans. on Fuzzy Systems*, vol. 9, no. 6, pp. 773–784, 2001.