Non-Linear Weighted Averaging for Multimodal Information Fusion by Employing Analytical Network Process*

Turgay Yilmaz† and Adnan Yazici
Middle East Technical University, Department of Computer Engineering
{*turgay—yazici*}*@ceng.metu.edu.tr*

Masaru Kitsuregawa
University of Tokyo, Institute of Industrial Science, Center for Information Fusion
kitsure@tkl.iis.u-tokyo.ac.jp

Abstract

Linear combination is a popular approach in information fusion due to its simplicity. However, it suffers from the performance upper-bound of linearity and dependency on the selection of weights. In this study, we introduce a 'simple' alternative for linear combination, which is a non-linear extension on it. The approach is based on the Analytical Network Process, which is a popular approach in Operational Research, but never applied for fusion before. The approach benefits from two major ideas; interdependency between classes and dependency of classes on the features. Experiments conducted on CCV dataset demonstrate that proposed approach outperforms linear combination and other simple approaches, moreover it is less-dependent on the selection of weights.

1. Introduction

Combining the information gathered from multiple modalities is an empirically validated approach to increase the retrieval accuracy [1]. Among the various combination methods that have been proposed, most frequently utilized approach is the Linear Weighted Fusion (or Linear Combination) [3, 11, 12], due to its simplicity and reasonable performance despite its simplicity. Some other well-known methods are as follows: Majority Voting, Support Vector Machines, Bayesian Inference, Dempster-Shafer, Neural Networks, Decision Templates and Borda Count [1].

When compared with the linear combination, these approaches are; (a) either has a simple design as the linear combination but worse/equal in performance, (b) or better in performance but require complex training setups in order to obtain an adequate performance. Moreover, the approaches in the latter group are usually not limited to linear approximations. So, it can be argued that the use of linearity in combiner design causes a performance upper bound on retrieval accuracy. A detailed analysis on the performance limits of linear combiners can be found in [12]. Besides, another important drawback with the linear combiners is the high dependency of the combiner performance on the selection of the weights. However, the selection of the optimal weights is one of the important issues that have not been adequately addressed yet in the fusion domain [1, 7].

Aligned to above given issues, we would like to investigate for a combination approach which (i) is as simple as the linear weighted fusion, (ii) can achieve the performance upper bound of linear weighted fusion, and (iii) is less-dependent on the selection of the weights. Through this study, we resemble the multimodal fusion problem to the real-life multi-criteria decision making problem in Operations Research domain and would like to introduce two popular approaches, Analytical Hierarchy Process (AHP) [9] and Analytical Network Process (ANP) [10]. AHP is a linear solution approach having the same principles with the linear weighted averaging method. However, ANP is a quite different solution that extends the linear weighted averaging method into a non-linear one, and has never been applied in the information fusion domain before. Thus, in this study, we adapt and extend the calculation approach and parameters of ANP for multimodal fusion. We show that it can be utilized as a 'simple', 'non-linear' and 'less-weight-dependent' way of fusion, which overcomes the problems listed above. We evaluate the approach by using the Columbia Consumer

*This work is supported in part by a research grant from TÜBİTAK EEEAG with grant number 109E014.

†Turgay Yilmaz is currently with the Center for Information Fusion, Institute of Industrial Science, University of Tokyo, Japan.

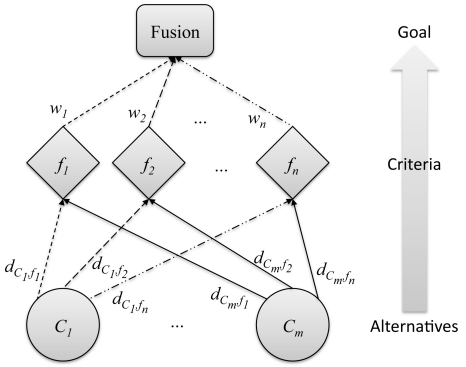

Figure 1. AHP Decision Hierarchy

Video dataset against several different approaches and obtain convincing results. Moreover, we empirically show that non-linear weighted averaging makes the accuracies less dependent on the selection of weights.

2. Linear Weighted Averaging and AHP

In linear weighted fusion methods, the information obtained from multimodal features is combined by assigning some particular weight for each modality and performing a summation or product operation to combine. Considering a summation preference, the final decision is calculated by;

$$\mathbf{S}_L = \mathbf{D}\mathbf{W}_D \tag{1}$$

where $\mathbf{D}$ is a $m \times n$ matrix, containing the output scores of classifier in each column; $\mathbf{W}_D$ is a n-sized vector, containing the weights of each feature; and $\mathbf{S}_L$ is a m-sized vector, containing the linearly combined decision scores for each retrieval class.

AHP presents 1 with a more concrete representation. First the multi-criteria decision making problem is modelled with a simple hierarchical model consisting of a *goal*, *criteria* and *alternatives* nodes. 1 presents a hierarchy for the multimodal information fusion problem with m number of classes and n number of features. Here, it should be noted that the edges between nodes are unidirectional, as a result of being a 'hierarchy'. In order to find the combined decisions, the total of alternative path lengths from each *alternative* to *goal* is calculated, where a path length is the product of the values on the edges along the path. A detailed description of AHP can be found in [9].

A crucial step in this approach is the determination of weights, which directly affects the fusion performance. An optimal solution is not guaranteed without an exhaustive search in the feature space. However, several heuristic solutions can be applied. As the most simplistic case, the weights of features can be selected equally ($w_i = 1/N$) which is also called *Simple Averaging*. Furthermore, some well-known heuristics are RELIEF [6], Information Gain [4] and Gain Ratio [8]. In this study, we utilize RELIEF and exhaustive search for experimental purposes. We also use a random weight selection approach to show the effect of weight selection.

3. Non-linear Weighted Averaging and ANP

ANP is a generalization of AHP and created with a consideration that many decision problems cannot be modelled with a simple hierarchy because they can involve interactions/dependencies of the included nodes [10]. Thus, ANP proposes to model the decision problem with a network which allows to define bidirectional transitions between the nodes. A network model, which is designed for the multimodal fusion problem with m number of classes and n number of features, is given in 2. Combined decision calculation is similar with AHP. However in ANP, the number of alternative paths is more than AHP, even indefinitely many, considering the possible bidirectional transitions between the nodes.

Considering the ANP approach, we can extend the linear weighted averaging approach into a non-linear approach by employing an additional weight factor.

$$\begin{aligned} \mathbf{S}_N &= \mathbf{W}_I \mathbf{S}_L \\ &= \mathbf{W}_I(\mathbf{D}\mathbf{W}_D) \end{aligned} \tag{2}$$

where $\mathbf{S}_N$ is a m-sized vector, containing the non-linearly combined decision scores for each retrieval class and $\mathbf{W}_D$ represents the direct weights, which are the traditional feature weights as used in linear weighted averaging. Besides, $\mathbf{W}_I$ is a $m \times m$ matrix used for the indirect weights, which can be described by incorporating two crucial ideas, in a multimodal fusion problem: (i) interdependency between the retrieval classes, and (ii) class-specific feature selection. The former idea provides exploiting the interdependencies between classes and benefit from the correlation as a weighting factor. In order to obtain the correlation between the

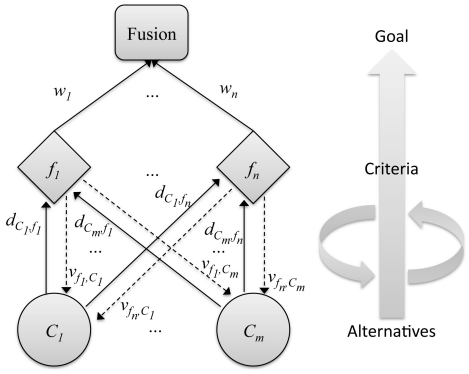

Figure 2. ANP Decision Network

classes, outputs of the classifiers are utilized. The correlation between the classifier outputs are usually ignored by many of the late fusion approaches, and only the corresponding output score of each classifier with the retrieval class is used during combination. For instance, in linear weighted averaging, the fusion result for C_1 is calculated by using only the scores for C_1 of each classifier. To exploit the interdependency, we incorporate all score outputs of all classifiers while performing fusion. Furthermore, the latter idea is based on the dependency of classes on the features. Although feature weighting methods usually propose solutions such that the resulting feature set is selected independent of the classes, defining feature weights that are specific to each class is a intuitive and promising approach [13]. For instance, in a multimodal scenario of multimedia data, the audial features are more useful for a $MusicPerformance$ class, whereas it is better to utilize visual modality for detecting a $Beach$ occurrence. In order to obtain class-specific feature weights, the feature weight calculation methods can be used separately for each feature, in a one-against-all fashion.

Considering these two ideas, the indirect weights $\mathbf{W}_I$ are calculated as;

$$\mathbf{W}_I = (\mathbf{DV})^i \tag{3}$$

where $\mathbf{D}$ is a $m \times n$ matrix, containing the output scores of classifier in each column; and $\mathbf{V}$ is a $n \times m$ matrix, containing the class-specific weights. In $\mathbf{V}$, each column holds the feature weights for a retrieval class. Considering that the product $\mathbf{DV}$ provides a square matrix, any power of this term is applicable. It should be noted that having $\mathbf{D}$ in the calculation of $\mathbf{W}_I$ and using powers provide 'non-linearity' into the solution. In addition they provide an implicit feature weighting estimation capability and make the solution 'less-dependent' on the weights $\mathbf{W}_I$ and $\mathbf{V}$. The resulting $\mathbf{W}_I$ contains linear combination results by using own class-specific features on the diagonal and linear combination results by using the class-specific weights of other classes as the rest. Thus, the final non-linear weighted averaging formulation is as follows:

$$\mathbf{S}_N^i = (\mathbf{DV})^i(\mathbf{DW}_D) \tag{4}$$

In order to obtain the most appropriate value of i, we focus on three solutions: First one (NWA-CB) is based on the converging characteristic of 4. Solution is converting 4 into a general eigenvalue problem at the convergence point. However, it is not guaranteed to obtain the best fusion performance for the converged $\mathbf{S}_N$ value. Second solution (NWA-BCo) is searching for the i value between 1 and convergence-based i value, which gives the best accuracy, via a training set . For the third one (NWA-BCl), the class-specific approach is mentioned again and it is argued that it is most likely to see the i value being different for each class. Thus, the i value is optimized for each class separately, similarly with the second approach.

4. Experiments

The experiments are carried out on the Columbia Consumer Video (CCV) Database [5], based on the semantic retrieval of classes. The dataset contains multimodal features –visual (SIFT), audial (MFCC), motion (STIP)– of 9,317 videos for 20 semantic classes listed on 1. The dataset is equally divided into training and test sets. Feature details can be found in [5]. To measure the retrieval accuracy, Average Precision (AP) and Mean Average Precision (MAP) metrics are used.

As the first test, non-linear weighted averaging method (NWA) is compared against; (i) Single features, (ii) Simple combination; Simple Averaging (AVG), Minimum Selection (MIN), Maximum Selection (MAX), (iii) Learning based combination; Naive Bayes (NB), Support Vector Machines (SVM), (iv) Linear weighted averaging (LWA) methods. For the feature weight selection of LWA and NWA, a RELIEF based feature weighting is used. For the NWA calculation, the 'best class accuracy' based approach is preferred. During all tests, first a classification process is performed with SVM classifiers, then the results of these classifications are combined. The multi-class classification with SVM is performed with a one-against-all approach. When needed, Naive Bayes implementation of MatLab Statistics Toolbox and LibSVM [2] are used. In 1, the APs of each class and the MAPs are presented for each combination approach.

As a secondary test, LWA and three NWA approaches, which are convergence-based (NWA-CB), best common accuracy (NWA-BCo) and best class accuracy (NWA-BCl), are compared against three different feature weighting methods: Random, RELIEF and Exhaustive Search (2). For the 'Random' weighting approach, a random weighting process is repeated 1000 times. The minimum (Rand-Min) and the mean (Rand-Avg) values obtained is presented in the table.

Considering the results given in 1 and 2, NWA achieves the performance upper-bound of linearity and outperforms all other approaches. Simple methods like MIN and MAX seems not adequate for fusion, since they lack the advantage of combining multiple features; though they perform better than the best of the single features. Besides, the AVG method, which is a linear approach with equal weights, is more accurate than LWA. This is the result of a probable deficiency of RELIEF method to assign weights. However, NWA eliminates such deficiency and obtains the best accuracy values despite the use of RELIEF weights. Thus, the most

Table 1. Accuracy comparisons. The best result for each category is highlighted in bold.

	SIFT	STIP	MFCC	AVG	MIN	MAX	NB	SVM	LWA	NWA
Basketball	66.95%	63.37%	44.65%	73.11%	67.16%	70.04%	22.87%	72.55%	69.35%	**75.89%**
Baseball	40.30%	18.38%	9.17%	43.15%	31.18%	39.00%	24.39%	46.37%	46.01%	**48.91%**
Soccer	49.29%	39.18%	17.59%	53.68%	49.65%	47.63%	25.40%	54.98%	53.98%	**58.26%**
IceSkating	81.18%	65.82%	16.18%	81.37%	71.27%	79.79%	73.63%	83.77%	82.90%	**85.32%**
Skiing	76.85%	60.27%	29.73%	74.31%	68.47%	72.03%	64.77%	**78.50%**	75.27%	78.18%
Swimming	68.84%	53.80%	15.35%	68.89%	56.70%	65.65%	57.30%	70.65%	71.30%	**72.61%**
Biking	36.85%	23.52%	11.36%	39.52%	29.90%	36.75%	32.68%	41.35%	38.73%	**42.76%**
Cat	34.24%	23.82%	17.40%	39.37%	33.94%	34.19%	41.65%	**41.75%**	35.27%	40.02%
Dog	25.48%	27.64%	22.10%	37.80%	35.07%	31.03%	9.92%	39.00%	28.81%	**42.99%**
Bird	17.40%	14.12%	17.63%	26.60%	22.81%	22.97%	16.34%	26.21%	19.86%	**28.80%**
Graduation	31.58%	22.09%	12.44%	36.23%	36.66%	28.28%	26.80%	40.05%	35.34%	**44.94%**
Birthday	33.32%	15.38%	35.94%	49.43%	41.39%	41.27%	45.92%	47.04%	40.54%	**55.53%**
Wed.Reception	18.65%	22.54%	12.41%	24.15%	**27.65%**	20.29%	2.98%	22.39%	17.37%	26.22%
Wed.Ceremony	35.20%	32.88%	35.04%	50.79%	**58.64%**	40.83%	37.86%	54.39%	38.74%	55.63%
Wed.Dance	56.68%	47.61%	28.01%	61.19%	54.52%	54.95%	46.45%	61.17%	59.53%	**66.62%**
MusicPerf.	48.20%	37.75%	56.71%	65.74%	60.51%	61.27%	61.68%	67.90%	53.77%	**68.87%**
NonMusicPerf.	45.21%	53.23%	29.78%	59.61%	51.77%	54.50%	11.79%	53.22%	53.31%	**64.60%**
Parade	48.71%	39.19%	25.62%	58.85%	56.82%	51.26%	46.13%	58.58%	55.17%	**65.33%**
Beach	69.99%	47.49%	37.34%	71.41%	64.16%	67.97%	3.83%	74.02%	71.83%	**75.43%**
Playground	44.59%	30.26%	23.83%	51.30%	49.62%	43.72%	51.11%	52.28%	49.51%	**57.90%**
MAP	46.48%	36.92%	24.91%	53.32%	48.39%	48.17%	35.18%	54.31%	49.83%	**57.74%**

Table 2. LWA, NWA vs. Weigh. Methods

	Rand-Min	Rand-Avg	RELIEF	Exh.Search
LWA	30.135%	47.618%	49,829%	57.783%
NWA-CB	55.139%	56.944%	57.734%	57.734%
NWA-BCo	56.031%	57.082%	57.740%	57.783%
NWA-BCl	56.242%	57.287%	57.741%	57.966%

crucial evaluation is the superiority of NWA solutions on LWA, independent from the feature weights. In addition, particularly focusing on 2, NWA seems less dependent on the selection of weights than the LWA method and can provide reasonably good results even with a worse selection of feature weights. A last comment on this table can be the slight but robust increase in the accuracy by the extensions made on the NWA-CB.

5. Conclusion

In this paper, an ANP-based non-linear weighted averaging method is introduced for the multimodal fusion problem. The method extends linear weighted fusion with two crucial ideas; interdependency between classes and dependency of classes on the features. The approach is tested on CCV dataset in a multimodal fusion scenario. The results demonstrate that introduced non-linear weighting approach is superior than linear combination as well as the other basic approaches and is less-dependent on the selection of weights.

References

[1] P. Atrey, M. Hossain, A. Saddik, and M. Kankanhalli. Multimodal fusion for multimedia analysis: a survey. *Multimedia Systems*, 16:345–379, 2010.

[2] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[3] G. Fumera and F. Roli. A theoretical and experimental analysis of linear combiners for multiple classifier systems. *IEEE TPAMI*, 27(6):942–956, June 2005.

[4] E. B. Hunt, P. J. Stone, and J. Marin. *Experiments in induction / Earl B. Hunt, Janet Marin, Philip J. Stone*. Academic Press, New York :, 1966.

[5] Y.-G. Jiang, G. Ye, S.-F. Chang, D. Ellis, and A. C. Loui. Consumer video understanding: A benchmark database and an evaluation of human and machine performance. In *Proc. of ACM ICMR*, 2011.

[6] K. Kira and L. A. Rendell. A practical approach to feature selection. In *Proc.of the 9th Int. Workshop on Machine Learning*, ML '92, pages 249–256, San Francisco, CA, USA, 1992. Morgan Kaufmann Publishers Inc.

[7] N. Poh and J. Kittler. *Multimodal Information Fusion: Theory and Applications for Human-Computer Interaction*, chapter 8, pages 153–169. Academic Press, 2010.

[8] J. R. Quinlan. Induction of decision trees. *Mach. Learn.*, 1:81–106, March 1986.

[9] T. Saaty. How to make a decision: The Analytic Hierarchy Process. *European Journal of Operational Research*, 48:9–26, 1990.

[10] T. Saaty. *Decision Making with Dependence and Feedback: The Analytic Network Process*. RWS Publications, Pittsburgh, 1996.

[11] K. Tumer and J. Ghosh. Linear and order statistics combiners for pattern classification. *CoRR*, cs.NE/9905012, 1999.

[12] R. Yan and A. G. Hauptmann. The combination limit in multimedia retrieval. ACM MULTIMEDIA '03, pages 339–342, NY, USA, 2003.

[13] T. Yilmaz, A. Yazici, and Y. Yildirim. Exploiting class-specific features in multi-feature dissimilarity space for efficient querying of images. In *FQAS*, pages 149–161, 2011.