Chapter III

Combinatorial Fusion Analysis: Methods and Practices of Combining Multiple Scoring Systems

D. Frank Hsu, Fordham University, USA

Yun-Sheng Chung, National Tsing Hua University, Taiwan

Bruce S. Kristal, Burke Medical Research Institute and Weill Medical College of Cornell University, USA

Abstract

Combination methods have been investigated as a possible means to improve performance in multi-variable (multi-criterion or multi-objective) classification, prediction, learning, and optimization problems. In addition, information collected from multi-sensor or multi-source environment also often needs to be combined to produce more accurate information, to derive better estimation, or to make more knowledgeable decisions. In this chapter, we present a method, called Combinatorial Fusion Analysis (CFA), for analyzing combination and fusion of multiple scoring. CFA characterizes each Scoring system as having included a Score function, a Rank function, and a Rank/score function. Both rank combination and score combination are explored as to their combinatorial complexity and computational efficiency.

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.
Information derived from the scoring characteristics of each scoring system is used to perform system selection and to decide method combination. In particular, the rank/score graph defined by Hsu, Shapiro and Taksa (Hsu et al., 2002; Hsu & Taksa, 2005) is used to measure the diversity between scoring systems. We illustrate various applications of the framework using examples in information retrieval and biomedical informatics.

Introduction

Many problems in a variety of applications domains such as information retrieval, social/welfare/preference assignments, internet/intranet search, pattern recognition, multisensor surveillance, drug design and discovery, and biomedical informatics can be formulated as multi-variable (multi-criterion or multi-objective) classification, prediction, learning, or optimization problems. To help obtain the maximum possible (or practical) accuracy in calculated solution(s) for these problems, many groups have considered the design and integrated use of multiple, (hopefully) complementary scoring schemes (algorithms or methods) under various names such as multiple classifier systems (Ho, 2002; Ho, Hull, & Srihari, 1992, 1994; Melnik, Vardi, & Zhang 2004; Xu, Krzyzak, & Suen, 1992; Kittler & Alkoot, 2003), social choice functions (Arrow, 1963; Young, 1975; Young & Levenglick, 1978), multiple evidences, Web page scoring systems or meta searches (Aslam, Pavlu, & Savell, 2003; Fagin, Kumar, & Sivakumar, 2003; Diligenti, Gori, & Maggini, 2004), multiple statistical analysis (Chuang, Liu, Brown, et al., 2004; Chuang, Liu, Chen, Kao, & Hsu, 2004; Kuriakose et al., 2004), cooperative multi-sensor surveillance systems (Collins, Lipton, Fujiyoshi, & Kanade, 2001; Hu, Tan, Wang, & Maybank, 2004), multi-criterion ranking (Patil & Taillie, 2004), hybrid systems (Duerr, Haettich, Tropf, & Winkler, 1980; Perrone & Cooper, 1992), and multiple scoring functions and molecular similarity measurements (Ginn, Willett, & Bradshaw, 2000; Shoichet, 2004; Yang, Chen, Shen, Kristal, & Hsu, 2005). For convenience and without loss of generality, we use the term **multiple scoring systems (MSS)** to denote all these aforementioned schemes, algorithms, or methods. We further note the need for the word “hopefully” above — there are limited practical means of predicting which combinations will be fruitful — the problem we address in the remainder of this report.

The main purpose in constructing multiple scoring systems is to combine those MSS’s in order to improve the efficiency and effectiveness or increase the sensitivity and specificity of the results. This purpose has been met; it has been demonstrated that combining MSS’s can improve the optimization results. Combination of multiple scoring systems has been studied under different names such as classification ensemble (Ho, 2002; Ho et al., 1992, 1994; Kittler & Alkoot, 2003; Tumer & Ghosh, 1999; Xu et al., 1992), evidence combination (Belkin, Kantor, Fox, & Shaw, 1995; Chuang, Liu, Brown, et al., 2004; Chuang, Liu, Chen, et al., 2004), data/information fusion (Dasarathy, 2000; Hsu & Palumbo, 2004; Hsu et al., 2002; Hsu & Taksa, 2005; Ibraev, Ng, & Kantor, 2001; Kantor, 1998; Kuriakose et al., 2004; Lee, 1997; Ng & Kantor, 1998, 2000), rank aggregation (Dwork, Kumar, Naor, & Sivakumar, 2001; Fagin et al., 2003), consensus scoring (Ginn et al., 2000; Shoichet, 2004; Yang et al., 2005), and cooperative surveillance (Collins et
Related Content

State-of-the-Art GPGPU Applications in Bioinformatics
[www.igi-global.com/article/state-of-the-art-gpgpu-applications-in-bioinformatics/105596?camid=4v1a](www.igi-global.com/article/state-of-the-art-gpgpu-applications-in-bioinformatics/105596?camid=4v1a)

Mining Protein Interactome Networks to Measure Interaction Reliability and Select Hub Proteins
Young-Rae Cho and Aidong Zhang (2012). *Computational Knowledge Discovery for Bioinformatics Research* (pp. 222-238).
[www.igi-global.com/chapter/mining-protein-interactome-networks-measure/66713?camid=4v1a](www.igi-global.com/chapter/mining-protein-interactome-networks-measure/66713?camid=4v1a)
Methods for the Evaluation of Right Ventricular Volume Using Ultrasound on a Catheter, in Intensive Care Unit
[www.igi-global.com/article/methods-evaluation-right-ventricular-volume/78391?camid=4v1a](www.igi-global.com/article/methods-evaluation-right-ventricular-volume/78391?camid=4v1a)

Facilitating and Augmenting Collaboration in the Biomedical Domain
Nikos Karacapilidis, Manolis Tzagarakis, Spyros Christodoulou and Georgia Tsiliki (2012). *International Journal of Systems Biology and Biomedical Technologies* (pp. 52-65).
[www.igi-global.com/article/facilitating-augmenting-collaboration-biomedical-domain/63046?camid=4v1a](www.igi-global.com/article/facilitating-augmenting-collaboration-biomedical-domain/63046?camid=4v1a)