Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Sets and Systems 270 (2015) 25–52
FUZZY sets and systems
www.elsevier.com/locate/fss

Dynamic classifier aggregation using interaction-sensitive fuzzy measures

David Štefka [a,*], Martin Holeňa [b]

[a] *Faculty of Nuclear Science and Physical Engineering, Czech Technical University, Trojanova 13, 120 00 Prague 2, Czech Republic*
[b] *Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 2, 182 07 Prague 8, Czech Republic*

Received 20 September 2013; accepted 8 September 2014

Available online 16 September 2014

Abstract

In classifier aggregation using fuzzy integral, the performance of the classifier system depends heavily on the choice of the underlying fuzzy measure. However, little attention has been given to the choice of the fuzzy measure in the literature; usually, the Sugeno λ-measure is used. A weakness of the Sugeno λ-measure is that it cannot model the interactions between individual classifiers. That motivated us to develop two novel fuzzy measures and a modification of an existing fuzzy measure which are *interaction-sensitive*, i.e., they model not only the confidences of classifiers, but also their mutual similarities. The properties of the measures are first studied theoretically, and in the experimental section, the performance of the proposed measures is compared to the traditionally used additive measure and Sugeno λ-measure. Experiments on 23 benchmark datasets and 3 different classifier systems show that the interaction-sensitive fuzzy measures clearly outperform their non-interaction sensitive counterparts.

Keywords: Fuzzy integral; Fuzzy measure; Dynamic classifier aggregation

1. Introduction

Classifier combining methods are a popular tool for improving the quality of classification. Instead of using just one classifier, a team of classifiers is created, and the predictions of the team are combined into a single prediction [1–3]. There are two main approaches to classifier combining: *classifier selection* (where a single classifier from the team is selected for prediction according to some criterion) and *classifier aggregation* (where the outputs of the classifiers are aggregated into a single prediction). Classifier combination can be either *static*, i.e., the combining process is the same for all patterns, or *dynamic*, where the combination process is adapted to the currently classified pattern [4–8].

One of the popular aggregation operators is the *fuzzy integral* [1,2,9–13]. The fuzzy integral aggregates the outputs of the individual classifiers in the team with respect to a fuzzy measure, representing the classification confidences.

* Corresponding author.
E-mail addresses: david.stefka@gmail.com (D. Štefka), martin@cs.cas.cz (M. Holeňa).

http://dx.doi.org/10.1016/j.fss.2014.09.005

Fuzzy measure is a generalization of the additive probabilistic measure, where the additivity is replaced by a weaker condition, monotonicity – this gives us a tool which can model interactions between different elements of the fuzzy measure space. However, due to the lack of additivity, the fuzzy measure needs to be defined on all subsets of the fuzzy measure space, resulting in 2^r defining values in finite cases, where r is the size of the universe. There are several approaches to overcome this weakness, e.g., *additive measures*, *symmetric fuzzy measures* [9], for which the value of the measure depends only on the number of elements in the argument, and $\perp$-*decomposable* fuzzy measures, including *Sugeno* λ-*measure* [9,10], for which the fuzzy measure values are computed from the fuzzy measure values for the singletons (called *fuzzy densities*) using a fixed t-conorm $\perp$.

As we will show in this paper, none of these approaches can take the similarity of the elements in the set into account, and therefore the ability to model interactions between different elements of the fuzzy measure universe is limited. For example, in the literature of classifier aggregation, fuzzy integral is usually used with Sugeno λ-measure. There is typically no explicit reason for the choice of this measure other than its simplicity. Sugeno λ-measure is a special case of a $\perp$-decomposable fuzzy measure, and as such, it cannot model similarities between the individual classifiers (cf. Section 3.2.3), and thus the contribution of using fuzzy integral is unclear.

In classifier aggregation, we usually try to create a team of classifiers that are not similar. This property is called *diversity* [14]. There are many methods for building a diverse team of classifiers [3,15–17]. If we use the fuzzy integral with a symmetric or $\perp$-decomposable fuzzy measure, we are not able to incorporate the diversity into the measure (and thus to the aggregation process), because the fuzzy measure of a union of two sets is a function only of the fuzzy measures of the two sets, regardless of the similarity of the elements in the sets.

This weakness motivated us to develop an *Interaction-Sensitive Fuzzy Measure* (ISFM) [18], which is defined using the fuzzy measure values for the singletons (fuzzy densities), and the similarities of the elements in the universe. If the fuzzy measure space corresponds to a team of classifiers, the fuzzy measure incorporates both the classification confidence (fuzzy densities), and the diversity of the team of classifiers (similarities of the classifiers). Using ISFM in fuzzy integral as an aggregation operator in classifier aggregation, the aggregation process involves all the important properties: the predictions of the classifiers, the classification confidences, and the diversity of the team. Our preliminary experiments with ISFM used with the Choquet integral in Random Forest ensembles have shown that ISFM outperforms Sugeno λ-measure if used in the Choquet integral [18].

However, as we will show in this paper, the ISFM is a fuzzy measure only for a fixed integrand, and therefore it cannot be viewed as a fuzzy measure on the whole universe. This is the reason why we reformulate the theoretical results from [18], and we will call this fuzzy measure *Induced Interaction-Sensitive Fuzzy measure* (I-ISFM). We will also define *Global Interaction-Sensitive Fuzzy Measure* (G-ISFM), which follows similar ideas as the I-ISFM, but which is a fuzzy measure on the whole fuzzy measure space. We will examine special cases of the I-ISFM and G-ISFM to show how they handle extreme cases of totally similar and totally dissimilar classifiers.

The idea of aggregating different information sources while considering mutual interactions (or similarities) between the individual sources was also introduced in the Cho-k-NN classifier [19], where the author uses the Choquet integral based on a non-additive fuzzy measure to improve the performance of a k-nearest neighbor classifier. The fuzzy measure considers the similarities of individual neighbors, and therefore in the resulting aggregation scheme, the redundant information is reduced, while the complementary information is stressed out. However, as we will show later, the measure is not directly applicable to classifier combining, and we therefore present a modified variant, called *Modified Hüllermeier Measure* (MHM), to be used as an interaction-sensitive fuzzy measure in dynamic classifier systems.

The novelty of the approach lies in the idea that instead of maximizing the diversity of the classifier team, we can focus on the aggregation itself and handle the diversity directly in the aggregation process (i.e., in the fuzzy measure in the context of fuzzy integral). The approach is not intended as another method for classifier combining; we just suggest that if the aggregation of the classifiers is done by the fuzzy integral (or its special cases, e.g., weighted mean), the underlying fuzzy measure should consider the classifier similarities, which is a natural (and easy) improvement of the aggregation.

In the experimental section, we compare the performance of I-ISFM, G-ISFM, and MHM with the traditionally used Sugeno λ-measure in combination with the Choquet and Sugeno integrals on three different classifier systems, namely Random Forests [17], ensembles of k-Nearest Neighbor classifiers [20] created by bagging [15] and ensembles of Quadratic Discriminant Classifiers [20] created by the Multiple feature subset method [21]. The objective of the experiments was to thoroughly investigate the added value of incorporating similarity into the definition of the fuzzy