Neurocomputing 194 (2016) 385–388

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Brief Papers

Category kappas for agreement between fuzzy classifications

Matthijs J. Warrens *

GION, University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 11 November 2014
Received in revised form
27 August 2015
Accepted 20 February 2016
Available online 10 March 2016

Keywords:
Cohen's kappa
Fuzzy classification
Category statistics
MRI
Brain tissue
Interobserver agreement

ABSTRACT

The kappa statistic is a widely used as a measure for quantifying agreement between two nominal classifications. The statistic has been extended to the case of two normalized fuzzy classifications. In this paper we define category kappas for quantifying agreement on a particular category of two normalized fuzzy classifications. The overall fuzzy kappa is a weighted average of the proposed category kappas. Since the value of the overall kappa lies between the minimum and maximum values of the category kappas, the overall kappa, in a way, summarizes the agreement reflected in the category kappas. The overall kappa meaningfully reflects the degree of agreement between the fuzzy classifications if the category kappas are approximately equal. If this is not the case, it is more informative to report the category kappas.

1. Introduction

In various fields of science, classification instruments are used to classify individuals or objects into categories. The instruments can be people or algorithms, and the categories are often unordered and mutually exclusive (nominal). Examples are, an educational psychologist that classifies the arithmetic strategies used by children to solve a mathematics problem [1,2], a developmental psychologist that classifies the attachment style of adults [3], a physical therapist that rates the type of injury [4], a radiologist that identifies tissues [5], an algorithm that classifies voxels of magnetic resonance images [6], or an algorithm that classifies the areas of a land cover map [7,8].

An important property of a classification instrument is its accuracy or reliability. The accuracy of an instrument can be quantified by assessing the agreement between two classifications made with the same instrument. High agreement between the classifications can then be taken as an indicator of the quality of the category definitions and the classification instrument in general.

A widely used statistic for quantifying agreement between two nominal classifications is the kappa statistic proposed by Cohen [5,9–11]. The value of kappa is 1 if there is perfect agreement between the two classifications, and 0 when agreement is equal to that expected under independence. Cohen's kappa was originally proposed as a measure of interobserver agreement in the context of psychological measurement, but the statistic is also used for map comparison in remote sensing [7,8] and content analysis [12]. Kappa depends on the boundary probabilities of the classifications. These quantities reflect how often a category was used in the classifications. Kappas from studies with different boundary probabilities are not comparable [13,14].

The popularity of the kappa statistic has led to the development of several extensions, e.g., kappas for three or more instruments [15], weighted kappas for classification instruments with ordinal categories [16–18], kappas for groups of observers [19] and kappas for clustered data [20,21]. In addition, extensions for the agreement between two fuzzy classifications [6] and two fuzzy categorical raster maps [7,8] have also been proposed.

The concept of a fuzzy classification is based on fuzzy set theory [22,23]. Fuzzy set theory is a well-developed mathematical system for addressing the degree of membership of objects to categories. In crisp classification objects are pigeonholed: an object either belongs or does not belong to a category. By contrast, fuzzy set theory permits the gradual assessment of the membership of an object into a category. The membership is described with the aid of a membership function valued in the real unit interval $[0,1]$.

Dou et al. [6] proposed a fuzzy kappa for quantifying agreement between two normalized fuzzy classifications. Their fuzzy kappa is an overall measure of agreement. If there is any disagreement between the classifications, the overall statistic does not tell us on which categories the agreement is almost perfect (if any) and on which categories agreement is less than perfect. Fuzzy kappas for the different categories are often more informative. The information provided by the category kappas can often be used to improve the accuracy of the classification instrument [24,25]. However,

* Tel.: +31 503636691.
E-mail address: m.j.warrens@rug.nl

http://dx.doi.org/10.1016/j.neucom.2016.02.038
0925-2312/

category kappas have not been defined for fuzzy classifications. This will be done in this paper.

The paper is organized as follows. In the next section the notation is introduced and the fuzzy kappa statistics are defined. The fuzzy kappas are illustrated on hypothetical brain tissue data in Section 3. In Section 4 we consider the case of two categories. It is shown that in this case the two category kappas and the overall kappa coincide. An alternative definition of the category kappas is presented in Section 5. Section 6 contains a discussion.

2. Fuzzy kappas

Let $X=\{x\}$ be a set of elements (voxels, individuals) of size $|X|=n$. In the classification process the objects are assigned to c unordered categories that are defined in advance, with $i \in (1,2,\ldots,c)$. The objects are classified twice, either with the same instrument or with different instruments. The two classifications are done independently.

We first define a fuzzy kappa for a single category. Let $u_i : X \to [0,1]$ be the membership function of category i for the first classification, and let $v_i : X \to [0,1]$ be the membership function of category i for the second classification. It is assumed that both classifications are normalized:

$$\sum_{i=1}^{c} u_i(x) = 1, \quad x \in X, \tag{1}$$

and

$$\sum_{i=1}^{c} v_i(x) = 1, \quad x \in X. \tag{2}$$

Normalized classifications describe a certain perfect case. In reality the requirement may be violated if the classification is done by human experts.

Let p_i and q_i denote the boundary probabilities of the first and second classification, respectively. The boundary probabilities are defined as

$$p_i := \frac{1}{n}\sum_{x \in X} u_i(x), \tag{3}$$

and

$$q_i := \frac{1}{n}\sum_{x \in X} v_i(x). \tag{4}$$

Quantities (3) and (4) reflect how often category i was used in the first and second classification, respectively. Following Dou et al. [6], the observed agreement between the classifications on category i is defined as

$$O_i := \frac{1}{n}\sum_{x \in X} \min(u_i(x), v_i(x)). \tag{5}$$

Let $p(u_i)$ and $q(v_i)$ denote the probability distribution of $u_i(x)$ and $v_i(x)$, respectively. Assuming that $u_i(x)$ is independent of $v_i(x)$ for all $x \in X$, the expectation of random agreement for category i is defined as

$$E_i := \int_{u_i=0}^{1}\int_{v_i=0}^{1} p(u_i)\, q(v_i) \min(u_i, v_i)\, du_i\, dv_i. \tag{6}$$

If u_i and v_i are binary functions, i.e., if u_i and v_i take on the values 0 and 1 only, then the expectation of random agreement becomes

$$E_i = p_i q_i. \tag{7}$$

A fuzzy kappa for category i can now be defined as

$$\kappa_i := \frac{O_i - E_i}{(p_i+q_i)/2 - E_i}. \tag{8}$$

If u_i and v_i are binary functions, statistic (8) becomes the category kappa for crisp classifications (see, e.g., [24–26]).

It turns out that the overall fuzzy kappa introduced in Dou et al. [6] is a weighted average of the category kappas in (8). Define for category i the weight

$$w_i := \frac{p_i+q_i}{2} - E_i. \tag{9}$$

An overall fuzzy kappa can be defined as a weighted average of the fuzzy category kappas:

$$\kappa := \frac{\sum_{i=1}^{c} w_i \kappa_i}{\sum_{i=1}^{c} w_i}. \tag{10}$$

Since

$$\sum_{i=1}^{c} \frac{p_i+q_i}{2} = 1 \tag{11}$$

statistic (10) is identical to

$$\kappa = \frac{\sum_{i=1}^{c} (O_i - E_i)}{1 - \sum_{i=1}^{c} E_i}, \tag{12}$$

which is the fuzzy kappa proposed in Dou et al. [6]. If u_i and v_i are binary functions, fuzzy kappa (12) becomes the ordinary kappa for crisp classifications [24,25]. Several properties of fuzzy kappa (12) are discussed by Dou et al. [6].

3. An illustration

Brain images can be decomposed into tiny cubes called voxels. A typical size of a voxel is $1 \times 1 \times 1\ \text{mm}^3$ and an image may consist of tens of thousands of voxels. To illustrate the fuzzy category kappas and the overall fuzzy kappa we consider a hypothetical example that consists of ten voxels. Table 1 presents two hypothetical fuzzy classifications of ten voxels into three tissue categories: gray matter (GM), white matter (WM) and cerebral spinal fluid (CSF). Each row of Table 1 corresponds to a voxel. The six membership functions of the categories take on values from the set $(0, 0.2, 0.4, 0.6, 0.8, 1)$. It can be verified that the fuzzy classifications are normalized, i.e., the values in each row of either classification sum up to 1.

Table 2 presents the values of the observed agreement, expectation of random agreement, and the kappa statistics for the data in Table 1. The value of the overall kappa is 0.77, indicating a good level of agreement between the two fuzzy classifications. The values of the three category kappas are quite different. The values for WM and CSF are 0.75 and 0.66, respectively, indicating a good level of agreement. However, the value for GM is 0.92, indicating an excellent level of agreement. The category kappas show that there is virtually no disagreement on GM, while there is some

Table 1
Two hypothetical fuzzy classifications of ten voxels.

Classification 1			Classification 2		
GM	WM	CSF	GM	WM	CSF
0.2	0.4	0.4	0.2	0.6	0.2
0.6	0.4	0	0.6	0.4	0
0.4	0.4	0.2	0.4	0.6	0
0.2	0.4	0.4	0.2	0.4	0.4
0.4	0.4	0.2	0.4	0.2	0.4
0	0.8	0.2	0	0.8	0.2
0.6	0.4	0	0.6	0.4	0
0.2	0.2	0.6	0	0.2	0.8
0	0.8	0.2	0	0.8	0.2
0.2	0.8	0	0.2	0.8	0