Engineering Applications of Artificial Intelligence 43 (2015) 102–110

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Brief paper

Fuzzy regression for perceptual image quality assessment

Kit Yan Chan [a,*], Ulrich Engelke [b]

[a] Department of Electrical and Computer Engineering, Curtin University, Perth, WA, Australia
[b] Digital Productivity Flagship of The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Hobart, TAS, Australia

ARTICLE INFO

Article history:
Received 19 December 2014
Received in revised form
10 April 2015
Accepted 13 April 2015

Keywords:
Fuzzy regression
Mean opinion scores (MOS)
Subjective image quality assessment
Objective image quality metric

ABSTRACT

Subjective image quality assessment (IQA) is fundamentally important in various image processing applications such as image/video compression and image reconstruction, since it directly indicates the actual human perception of an image. However, fuzziness due to human judgment is neglected in current methodologies for predicting subjective IQA, where the fuzziness indicates assessment uncertainty. In this article, we propose a fuzzy regression method that accounts for fuzziness introduced through human judgment and the limitations of widely-used psychometric quality scales. We demonstrate how fuzzy regression models provide fuzziness information regarding subjective IQA. We benchmark the fuzzy regression method against the commonly used explicit modeling method for subjective IQA namely statistical regression by considering three real situations involving subjective image quality experiments where: (a) the number of participants is insufficient; (b) an insufficient amount of data is used for modelling; and (c) variant fuzziness is caused by human judgment. Results indicate that fuzzy regression models achieve more effective data fitting and better generalization capability when predicting subjective IQA under different types and levels of image distortion.

1. Introduction

The number of digital images we take each year has soared. In 2013, 27,800 digital images were uploaded to Instagram every minute and 208,300 digital images were uploaded to Facebook every minute. Yahoo expected that 880 billion digital images would be captured in 2014 (Popphoto, 2013). In the past decade, we have experienced great technological advancements of new devices for capture, storage, compression, transmission, and display of digital images, mostly resulting in significant increases of image quality. The raw visual information typically passes through multiple steps in an imaging pipeline, each of which affects the quality of the image at the receiver . With these increasingly complex multimedia applications, perceived image quality evaluation has been receiving significant attention as a means of ensuring certain levels of quality of service. Given the abundancy of visual data, it is essential to develop efficient computational prediction models to automatically evaluate image quality and to control the perceptual quality of the visual content by tuning the multi-parameters of the imaging pipeline.

However, it is challenging to develop prediction models that accurately represent image quality perceived by a human. Subjective image quality assessment (IQA) is typically used as a ground truth to develop computational image quality prediction models (Engelke et al., 2009) as humans are considered to be the observers and consumers of most systems and products involving digital images. In subjective IQA, a group of interviewers typically scores the quality of a number of images. Subjective IQA provides a useful tool for evaluating the visual effect of a wide range of artifacts which are imposed on digital images for image acquisition, processing, transportation, compression, and storage (Miyahara et al., 1998). However, it is not possible to implement subjective IQA in real time or as a systematic evaluator for image enhancement. It cannot be incorporated into the design and optimization of image processing algorithms in order to enhance image quality. For this reason, there has been an increasing interest in correlating subject IQA with the objective IQA metrics in order to automatically predict or estimate the perceived image quality (Engelke et al., 2009), where the objective IQA metrics range from simple numerical measures (Eskicioglu and Fisher, 1995) such as the signal-to-noise ratio and the bit error rate (Molisch, 2005) to highly complex models incorporating those characteristics of the human visual system that are considered crucial for visual quality perception (He et al., 2009; Martens, 2002; Rix et al., 1999; Wandell, 1995; Wu et al., 2013). These prediction models aim to automatically predict perceived image quality as obtained in subjective experiments. Currently, there are no image quality prediction models that work well across a wide range of visual content and distortion types; typically, they perform well only on subsets of the above (Engelke et al., 2009).

* Corresponding author. +618 9266 2945.
E-mail address: kit.chan@curtin.edu.au (K.Y. Chan).

http://dx.doi.org/10.1016/j.engappai.2015.04.007
0952-1976/

To develop quality prediction models, implicit modelling methods based on artificial intelligence have been used based on experimental data of subjective image quality experiments, which are typically based on n-point psychometric scales, such as the Likert (1932) scale, to record responses from a number of participants who judge the opinion scores of images presented to them. The opinion scores are then combined into Mean Opinion Scores (MOS) for all images, which in turn are instrumental in the training and validation of computational image quality prediction models. Neural networks (Engelke et al., 2007; Gu et al., 2014) have been used to develop models for predicting subjective IQA, but these approaches lack transparency since they are black-box in nature. Explicit information cannot be indicated in the neural networks. Also, the training time required by the neural networks is much longer compared with the statistical regression, when the network size is large. Fuzzy modeling-based approaches have also been applied to develop prediction models for IQA (De and Sil, 2015). However, compared with statistical regression methods, more explicit information can be found in statistical regression models which are in a polynomial form. Hence, variable significances and variable interactions can be determined in the polynomial of the regression models (Seber, 2003).

To generate explicit models, statistical regression (Seber, 2003) is commonly used. Engineers, in general, prefer to use statistical methods because more explicit information can be found than using the fuzzy modeling-based approaches or the neural networks. However, subjective image quality experiments involve human opinion judgments which are inherently imprecise, inconsistent over time, and often non-consensual when involving a group of individuals (Dubois and Prade, 1980). Hence, the assessment represents a source of uncertainty that is typically neglected in the development of quality prediction models that correlate subjective IQA and objective IQA. Therefore, the judging process inherently represents a source of uncertainty and bias that is neglected in statistical regression used to match predicted quality with MOS (Engelke et al., 2009; Mittal et al., 2012). Also, the regression models may not be performed accurately, as they can only be applied accurately within the range for which they are developed (Jobson, 1991). They can be applied only if the given experimental data is normally distributed according to the developed regression model. They can represent a crisp relationship only between the objective image quality metrics and subjective image quality measure, while the uncertainty of the relationship cannot be addressed. Instead, in this paper, we propose to use fuzzy regression to overcome these shortcomings.

In new product development, fuzzy regression (Chen et al., 2013; Jiang et al., 2013; Karsak, 2008; Kwong et al., 2010; Sener and Karsak, 2010, 2011) has commonly been used to model correlations between subject customer satisfaction and objective engineering characteristics of new products, where settings of engineering characteristics can affect customer satisfaction with the product. Based on the correlation models, the engineering characteristics can be specified by optimizing customer satisfaction. Fuzzy regression has a distinct advantage over statistical regression as it can address the fuzziness of subjective judgments and it can perform effectively using a small or even incomplete data set (Tanaka and Watada, 1998). In this article, we propose a novel image quality assessment technique based on fuzzy regression that attempts to account for the 'fuzziness' of human judgment introduced through subjective IQA. Indeed, the approach of fuzzy regression is the first time to be developed in order to model the relationship between objective image quality metrics and subjective image quality measure, where the fuzzy regression model attempts to address the fuzziness caused by the subjective IQA. Three validations with three conditions in MOS data sets were performed in order to evaluate whether fuzzy regression outperforms statistical regression in term of generalization capability: (1) varying number of participants; (2) varying data sizes; and (3) varying amount of fuzziness. These three conditions simulated the real situations in subjective image quality experiments where: insufficient numbers of participants are involved; insufficient amount of MOS data is used for modelling; and different amounts of fuzziness are caused by human evaluation of MOS. Experimental results shows that the proposed method overcomes the shortcomings of more widely adopted statistical regression techniques which disregard fuzziness of human judgment and require large data sets with normal distribution assumption.

The rest of the article can be organized as follows: Sections 2 and 3 discuss fuzziness in IQA and fuzzy regression in developing prediction models for MOS. Section 4 validates the fuzzy regression on an extensive image quality database and benchmarks it against statistical regression. Section 5 concludes the article.

2. Fuzziness in subjective IQA

Perception of subjective IQA is inherently imprecise as, typically, only an approximate judgment is made. The widely used N-point psychometric scales, however, usually map qualitative judgments onto opinion scores (Likert, 1932). It has been shown in Thu et al. (2011) that on a continuous rating scale with N opinion scores, people t to judge quality around the integers with some degree of uncertainty. One may refer to judgments 'about' a particular integer X on psychometric scales. For instance, one may judge quality to be 'about 2' on a 5-point scale when one feels that the image quality is 'Poor' and 'about 3' when one feels that the image quality is 'Fair'. The question then arises: what does 'about X' actually mean in opinion scores for image quality?

Based on the fuzzy set theory (Tanaka and Watada, 1998), the linguistic term 'about y^c' can be explained by a fuzzy number, $\tilde{y} = (y^c, y^r, y^l)$, with a fuzzy membership function, $\mu_{\tilde{y}}(y)$:

$$\mu_{\tilde{y}}(y) = \begin{cases} 1 & y = y^c \\ \frac{y - y^c}{y^c - y^l} & y^l \le y < y^c \\ \frac{y^r - y}{y^r - y^c} & y^c \le y < y^r \\ 0, & \text{otherwise} \end{cases}$$

where $\mu_{\tilde{y}}(y)$ indicates the membership function of the linguistic term 'about y^c'; y^c is the center of the fuzzy member of which the center indicates the degree of opinion score for 'about y^c'; y^l and y^r are the left and right spreads respectively which indicates the fuzziness of opinion score. When y is exactly equal to y^c, the membership, $\mu_{\tilde{y}}(y)$, is 1 and thus y is a full membership of 'about y^c'. When y is within the value between y^l and y^r, y is a membership of 'about y^c'. When y^l is close to y^r, the fuzziness of 'about y^c' is low. Here, a fuzzy triangle function is used as it requires uncomplicated fuzzy arithmetical operations compared with a Gaussian or trapezoidal function (Dubois and Prade, 1980).

Fig. 1 shows two fuzzy numbers for the opinion scores namely 'about 2', $\tilde{y}_1 = (2, 0.25, 0.25)$, and 'about 3', $\tilde{y}_2 = (3, 1, 0.5)$. In this example, the fuzziness of 'about 2' is smaller than that of 'about 3',

Fig. 1. Two fuzzy numbers, 'about 2' and 'about 3'.