



Fuzzy numbers from raw discrete data using linear regression

J. Moreno-Garcia^{a,*}, L. Jimenez Linares^{b,1}, L. Rodriguez-Benitez^{b,1}, E. del Castillo^{b,1}

^a Uni. Castilla-La Mancha, E.I. Industrial, Avda Carlos III s/n, Toledo, Spain

^b Uni. Castilla-La Mancha, E.S. Informatica, Paseo de la Universidad 4, Ciudad Real, Spain

ARTICLE INFO

Article history:

Received 29 April 2011

Received in revised form 12 January 2013

Accepted 19 January 2013

Available online 1 February 2013

Keywords:

Fuzzy numbers

Data summarisation

Linear regression

Expected value

ABSTRACT

This paper focuses on modelling fuzzy numbers with meaningful membership functions. More precisely, it proposes a method to construct trapezoidal fuzzy number approximations from raw discrete data. In many applications, input information is numerical, and therefore, particular fuzzy sets, such as fuzzy numbers, hold great interest and relevance in managing data imprecision and vagueness. The proposed technique provides an efficient way to obtain trapezoidal numbers using linear regression. The technique is simple, fast, and effective. Preliminary tests are performed using different types of input data: a Gaussian function, a Sigmoidal function, three datasets of synthetic discrete data, and an histogram obtained from a colour satellite image.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

In many applications, input data are represented as raw discrete data. Such data are difficult to manage and use due to their discrete nature and, sometimes, the large amount of available information. Representing data that allow for applications with which they can be efficiently managed is thus necessary. An example of such data can be found in the definitions of classes that depend on different data distributions. Uncertainty and vagueness could appear in the definitions of characteristic class functions. Fuzzy logic allows for easily extracting information from data and modelling data uncertainty.

Fuzzy numbers with either trapezoidal or triangular membership functions can represent this kind of data: they can be defined easily, and their representation, i.e., a 3- or 4-tuple of real numbers, guarantees low memory consumption and high processing speed. Fuzzy number operations can also be defined using linguistic information. Transforming raw data to fuzzy numbers can thus be considered an important research issue.

Our work focuses on obtaining the membership function of the fuzzy set representing the fuzzy number using a linear regression-based technique. To simplify fuzzy number manipulation, a trapezoidal membership function can be used, as Grzegorzewski et al. indicate [18]. Although the accuracy of the results depends directly on the shape of the fuzzy number membership function, because we are working in a fuzzy domain, we must consider that fuzzy numbers with simpler shapes have more intuitive interpretations and require less computational time [33]. The proposed method contains an iterative procedure that represents raw discrete data using fuzzy numbers with a complexity of $O(n \log n)$. Fuzzy numbers are then obtained automatically at execution, and the method can be used in some application domains that require real-time computing, including video segmentation using fuzzy techniques [29].

First, we review some previous works related to our approach. Ralescu et al. [38] present an approach that uses the mass assignment theory as a framework to unify fuzzy sets and probability distributions. A recent study [40] considers the prob-

* Corresponding author. Tel.: +34 902 204100; fax: +34 925 268849.

E-mail addresses: Juan.Moreno@uclm.es (J. Moreno-Garcia), Luis.Jimenez@uclm.es (L. Jimenez Linares), Luis.Rodriguez@uclm.es (L. Rodriguez-Benitez), Ester.Castillo@uclm.es (E. del Castillo).

¹ Tel.: +34 902 204100.

lems encountered when input data sets are not balanced or when the classes to learn are rare. Grzegorzewski et al. [18] present a set of criteria for approximation operators and their nearest trapezoidal approximation operator, a work that has been extended [19] to always produce a well-formed trapezoidal fuzzy number. Another work proposes improvements for Grzegorzewski methods [44,45] and studies the properties of the trapezoidal approximation operator, preserving the expected interval [2,4]. Domanska and Wojtylak [11] present an algorithm for transforming a sequence of real numbers into a fuzzy number. Their method creates a special fuzzy number, and they use it in a model used to forecast pollution concentrations. Nasibov and Peker [30] study the parameter formulas for an exponential membership function based on a minimisation problem. Their method seeks to obtain an exponential membership function that assumes a data histogram shape. Au et al. [1] present a method to determine the membership functions of fuzzy sets directly from data. It maximises the class-attribute interdependence to improve the classification results. Choi and Rhee [5] suggest three novel interval type-2 fuzzy membership function (IT2 FMF) generation methods based on heuristics, histograms and interval type-2 fuzzy C-means. Hong et al. [22] present a multi-level ACS-based mining algorithm to extract an appropriate set of membership functions in fuzzy data mining. They use a multi-stage graph and adopts multi-layer processing such that the precision of the final membership functions may be incrementally improved. Their approach is suitable for solving problems with potentially large maximum quantities of an item value in the transactions. Grzegorzewski [20] proposes an approach to simplify fuzzy numbers called shadowed set approximation showing that its approximation might be useful for granular computing.

Pedrycz is a leading author in this research field. He presents [32] a technique that automatically extracts the optimal fuzzy rules using an auto-tuning algorithm and the weighting factors of an object function. He uses the improved complex method for auto-tuning of parameters of the premise membership functions in consideration with the overall structure of fuzzy rules. Pedrycz and Vukovich [35] introduce a hybrid user-driven and data-validated approach to membership function elicitation. The introduced algorithm presents some advantages. Furthermore, Pedrycz [34] introduces the concept of fuzzy equalization as “a process of building information granules that are semantically and experimentally meaningful”, and an algorithm developed for triangular fuzzy sets. This study elaborates on the impact the equalization effect has on system design. Nobuhara et al. [31] use the concept of fuzzy equalization to design a motion compression/reconstruction method by fuzzy relational equations (MCF). Uniform coders and non-uniform coders compress the motion sequence. The design method of non-uniform coders is based on an overlap level of fuzzy sets and a fuzzy equalization. Pedrycz presents some works about information granularity. For example, in [36] discusses the role of information granulation and the ensuing information granules in description of time series. A detailed algorithm produces information granules (fuzzy sets) used to the description of numeric time series. Pedrycz [10] proposes a granular algorithm for communicating between granular worlds. This algorithm allows communication between the subset of granular worlds thereby transferring the information contained in one granular world into the granularity of another granular world. Finally, Pedrycz [37] presents a study to the principle of justifiable granularity and the concept of optimal allocation of granularity in the design of intelligent systems, and ensuing principles addressing the ways of constructing and processing information granules in spite of a way they are formalized.

Other approaches attempt to simplify the task of handling fuzzy numbers by introducing real indices to capture information contained in the fuzzy number. Delgado et al. [8,9] thus introduce parameters such as value and ambiguity. Using these parameters, a canonical representation of non-discrete fuzzy numbers is obtained. This representation is used in decision problems. Woxman [41] extends this canonical representation to discrete fuzzy numbers. A more recent paper that discusses such problems is given in [3]. They obtain the nearest trapezoidal approximation and nearest symmetric trapezoidal approximation to a given fuzzy number using the average Euclidean distance, preserving the value and ambiguity. They offer a less sophisticated method to avoid the laborious calculus associated with the Karush–Kuhn–Tucker theorem. They present algorithms to compute the approximations as well as many examples.

Moreover, there is a relationship between the attainment of approximations of fuzzy numbers and the use of these approximations in concrete real applications [23]. There is thus great interest in constructing meaningful approximations with efficient methods.

Finally, there must be a measure of how good the fuzzy number approximation is. The expected value is used in some works with different aims [15–17,42,43]. We use the expected value because it allows for representing the characteristics of the possibility distributions in a unique value as described in [21,26]. The expected value [21] is used to verify the representativity of the fuzzy number compared to the distribution of possibilities. Both represent the same concept if their expected value is close. The qualitative verification of the fit is then performed by comparing both expected values. Expected values are not used in the fit process. This data fit is made by obtaining the left and right regression lines that minimise the expected error. The experiments described in this work indicate that our approach preserves the expected value.

1.1. Contributions of this work

This paper presents a new approach for obtaining the fuzzy set membership function representing raw discrete data. To simplify the fuzzy number manipulation, the trapezoidal membership function is used. This work differs from others in three aspects. The presented method is simple, as our mechanism is based on linear regression to obtain the two lateral straight lines of the trapezoid. Our technique is also fast, with an order of complexity $O(n \log n)$, which allows for its use in applications with real-time constraints, with the capacity to handle large amounts of input data. Our approximation is further effective: because the error is low, it satisfies the expected value for the input data, and the obtained trapezoid has a good fit to the shape of the graphic input data.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات