Information Sciences 202 (2012) 1–23

Contents lists available at SciVerse ScienceDirect

Information Sciences

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

Charles Ragin's Fuzzy Set Qualitative Comparative Analysis (fsQCA) used for linguistic summarizations

Jerry M. Mendel *, Mohammad M. Korjani

Signal and Image Processing Institute, Ming Hsieh Department of Electrical Engineering, University of Southern California, 3740 McClintock Ave., Los Angeles, CA 90089-2564, United States

ARTICLE INFO

Article history:
Received 6 December 2010
Received in revised form 7 February 2012
Accepted 24 February 2012
Available online 6 March 2012

Keywords:
Fuzzy Set
Fuzzy Set Qualitative Comparative Analysis
Linguistic summarization
Rule
Sufficient condition

ABSTRACT

Fuzzy Set Qualitative Comparative Analysis (fsQCA) is a methodology for obtaining linguistic summarizations from data that are associated with cases. It was developed by the eminent social scientist Prof. Charles C. Ragin, but has, as of this date, not been applied by engineers or computer scientists. Unlike more quantitative methods that are based on correlation, fsQCA seeks to establish logical connections between combinations of causal conditions and an outcome, the result being rules that summarize the sufficiency between subsets of all of the possible combinations of the causal conditions (or their complements) and the outcome. The rules are connected by the word OR to the output. Each rule is a possible path from the causal conditions to the outcome. This paper, for the first time, explains fsQCA in a very quantitative way, something that is needed if engineers and computer scientists are to use fsQCA.

1. Introduction

A *linguistic summarization* is a sentence or a group of sentences that describes a pattern in a database. There are different kinds of linguistic summarizations, ranging from a library of pre-chosen sentences from which the most representative one (or group) is chosen and is then declared to be *the* linguistic summarization, to a collection of if–then rules, some or all of which are chosen to be *the* linguistic summarization. Each of the different kinds of linguistic summarizations has its useful place; however, in this paper we are interested only in linguistic summarizations that are in the form of *if–then rules*.

Linguistic data (base) summaries using type-1 fuzzy sets were introduced by Yager [38–41], advanced by Kacprzyk and Yager [4], Kacprzyk et al. [6] and Zadrożny and Kacprzyk [48], implemented by Kacprzyk and Zadrożny (see, e.g., [5] and its many references by these authors), and extended to type-2 fuzzy sets by Niewiadomski [22,23]. Linguistic summarizations of time series that use type-1 fuzzy sets has been studied by Kacprzyk and Wilbik, e.g. [3] (see, also, 13 other references by these authors, including Zadrożny, that are in this article). Because all of these summarizations are for a library of pre-defined summarizers, and are not in the form of if–then rules, they are not elaborated upon in this paper; however, detailed comparisons of three different summarization methods are given in Section 4.

Linguistic summarization using if–then rules and type-1 fuzzy sets had its origins in Zadeh's classical 1973 paper [42]. Although these if–then rules are the foundation for the developments of many kinds of quantitative rule-based systems, such as fuzzy logic control and rule-based classification, until recently very few people, other than perhaps Zadeh (e.g., [43–45,47]), thought of them any longer as linguistic summarizations. This is because it is the mathematical implementations

* Corresponding author. Tel.: +1 213 740 4445; fax: +1 213 740 4651.
E-mail addresses: mendel@sipi.usc.edu (J.M. Mendel), korjani@usc.edu (M.M. Korjani).

0020-0255/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2012.02.039

of the set of rules that has become important in such applications, rather than the rules themselves. In essence, the rules have become the means to the end, where the end is a mathematical formula that produces a numerical output. Only since Zadeh's pioneering works on computing with words has there been a return to the understanding that a collection of if–then rules, by themselves, is indeed a linguistic summarization.

Wang and Mendel [33] developed the first method to extract if–then rules from time-series data (the WM method). Many improvements to the WM method have been published since the original method. All of these works use the if–then rules as a *predictive model*, which according to Hand et al. [2] "... *has the specific objective of allowing us to predict the value of some target characteristic of an object on the basis of observed values of other characteristics of the object.*"

Recently, Wu and Mendel [36,37] developed a different way to extract rules from data in which the rules use interval type-2 fuzzy sets to model the words in both their antecedents and consequents. Their rules construct a *descriptive model*, which according to Hand et al. [2] "... *presents, in convenient form, the main features of the data. It is essentially a summary of the data, permitting us to study the most important aspects of the data without them being obscured by the sheer size of the data set.*[1]

The linguistic summarization method that is described in this paper also leads to a descriptive model, and is called *Qualitative Comparative Analysis* (QCA). It is not a method originated by the authors of this paper, but is a method discovered by the first author that has been used mainly in the fields of social and political sciences and that does not seem to have been used (prior to this work) in engineering or computer science.[2] Consequently, this paper should be viewed as a *conduit* for fsQCA from the less mathematically oriented social and political sciences literatures into the more mathematically oriented engineering and computer science literatures.

According to Ragin [28, p. 183]: "The goal of QCA is to derive a logically simplified statement describing the different combinations of conditions linked to an outcome." Each combination of conditions and same outcome is sometimes referred to [1] as a *type* or a *typological configuration*. According to Rihoux and Ragin [31, p. 33 and p. 66]:

> Crisp set Qualitative Comparative Analysis (csQCA) was the first QCA technique, developed in the late 1980s, by Professor Charles Ragin[3] and programmer Kriss Drass. Ragin's research in the field of historical sociology led him to search for tools for the treatment of complex sets of binary data that did not exist in the mainstream statistics literature. He adapted for his own research, with the help of Drass, Boolean algorithms that had been developed in the 1950s by electrical engineers to simplify switching circuits, most notably Quine[4] [24] and McCluskey [14]. In these so-called minimization algorithms, he had found an instrument for identifying patterns of multiple-conjunctural causation and a tool to simplify complex data structures in a logical and holistic manner [26, p. viii]. ... csQCA is based on Boolean algebra, which uses only binary data (0 or 1), and is based on a few simple logical operations[5] [union, intersection and complement]. ... [In csQCA,] it is important to follow a sequence of steps, from the construction of a binary data table to the final 'minimal formulas.' ... Two key challenges in this sequence, before running the minimization procedure, are: (1) implementing a useful and meaningful dichotomization of each variable, and (2) obtaining a 'truth table' (table of configurations) that is free of 'contradictory configurations.' ... The key csQCA procedure is 'Boolean minimization.'

csQCA was extended by Ragin to fuzzy sets, because he realized that categorizing social science causes and effects as black or white was not realistic. Fuzzy sets let him get around this. According to [31, p. 120]:

> fsQCA retains key aspects of the general QCA approach, while allowing the analysis of phenomena that vary by level or degree. ... The fsQCA procedure ... provides a bridge between fuzzy sets and conventional truth table analysis by constructing a Boolean truth table summarizing the results of multiple fuzzy-set analyses. ... Fuzzy membership scores (i.e., the varying degree to which cases belong to sets) combine qualitative and quantitative assessments. ... The key set theoretic relation in the study of causal complexity is the *subset relation*; cases can be precisely assessed in terms of their degree of consistency [subsethood] with the subset relation, usually with the goal of establishing that a combination of conditions is sufficient for a given outcome.

Both csQCA and fsQCA are set-theoretic methods. They differ from conventional quantitative variable-based methods (e.g., correlation and regression) in that they [1] "... do not disaggregate cases into independent, analytically separate aspects but instead treat configurations as different types of cases." Additionally [1], "The basic intuition underlying QCA[6] is that cases are best understood as configurations of attributes resembling overall types and that a comparison across cases can allow the researcher to strip away attributes that are unrelated to the outcome in question."

According to [28, p. 183], "... QCA summarizes the truth table in a logically shorthand manner." This is *linguistic summarization*.

[1] The linguistic summarizations mentioned above, due to Yager, Kacprzyk, Zadrożny, Niewiadomski and Wilbik also fall into the class of descriptive models.

[2] An exception to this are [34,35], but they only discuss a few aspects of fsQCA and not all of it.

[3] He is now a professor of sociology and political science at the University of Arizona. In the 1980s he was a professor of sociology and political science at Northwestern University.

[4] Actually, Quine is a famous logician and is not an electrical engineer.

[5] Bracketed phrases, inserted by the present authors, are meant to clarify quoted materials.

[6] It is quite common to refer to both csQCA and fsQCA as "QCA" letting the context determine which QCA it is. More recently, the phrase *Configurational Comparative Methods* is used to cover all QCA methods, e.g. [31].