Applied Soft Computing 21 (2014) 38–56

Contents lists available at ScienceDirect

Applied Soft Computing

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

Compound Linguistic Scale

Kevin Kam Fung Yuen

Department of Computer Science and Software Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, China

ARTICLE INFO

Article history:
Received 10 June 2012
Received in revised form 4 February 2014
Accepted 9 February 2014
Available online 6 March 2014

Keywords:
Rating scale
Linguistic Variable
Mathematical linguistics
Fuzzy theory
Recommender system
Decision sciences

ABSTRACT

Rating scales are the essential interfaces for many research areas such as in decision making and recommendation. Some issues concerning syntactic and sematic structures are still open to discuss. This research proposes a Compound Linguistic Scale (CLS), a two dimension rating scale, as a promising rating interface. The CLS is comprised of Compound Linguistic Variable (CLV) and Deductive Rating Strategy (DRS). CLV can ideally produce 21 to 73 $((7\pm2)((7\pm2)-1)+1)$ ordinal-in-ordinal rating items, which extends the classic rating scales usually on the basis of 7 ± 2 principle, to better reflect the raters' preferences whilst DRS is a double step rating approach for a rater to choose a compound linguistic term among two dimensional options on a dynamic rating interface. The numerical analyses show that the proposed CLS can address rating dilemma for a single rater and more accurately reflects consistency among various raters. CLS can be applied to surveys, questionnaires, psychometrics, recommender systems and decision analysis of various application domains.

1. Introduction

Measurement consists of rules for assigning symbols to objects so as to (1) represent quantities of attributes numerically (scaling) or (2) define whether the objects fall into the same or into a different category with respect to a given attribute (classification) [31]. A rating scale is essential for measurement and receives many attentions from scholars as its applications including social and engineering sciences are vast.

Rating is usually referred to choose the most appropriate item among a set of items according to rater's perception toward an attribute of object. The Likert scale [23] of five or seven points is widely used. Others, or quasi-Likert Scales, which can be regarded as minor variations of Likert scales, include numerical scale, graphic rating scale, visual analogue scale [42], and semantic differential scale [32].

The fundamental issues regarding the appropriate number of the rating scale items, and rating accuracy are still open to discuss [34,12,9,41,21,19,20]. Miller (1956) [28] indicated that a rater could usually manage a set with (7 ± 2) items. The qualitative scales such as ordinal or nominal are referred to as being 'soft' or 'weak scales' [29], as Likert-Style scales are ordinal. [9] found evidence suggesting that as scale granularity increases, recommendation accuracy increases, whilst [36] found that users work harder with more granular rating scales. In views of their findings, this study proposes the Hedge-Direction-Atom Linguistic Representation Algorithm (Algorithm 1) to form two dimensional scale items to strengthen the softness and weakness of the classical scales, and a Deductive Rating Strategy (Algorithm 3) is proposed to handle the rating process of the items.

Research of sematic representation for linguistic terms is limited. [37] indicated that linearity of an ordinal scale is open to discuss. The classical research from [14] which compared with [35] investigated twenty modifying words like usually, often, sometimes, occasionally, seldom, rarely and commonly, and the findings conclude that such words do not mean the same thing to all people. [6] argued that the commonly used Likert categories are not necessarily evenly spaced along this level of agreement continuum, although researchers frequently assume that they are. There are many situations where observations cannot be described accurately as, for instance, when they depend on environmental conditions or on individual responses [38]. As lacking theoretical approaches develop a pattern for a collection of linguistic labels represented by number, this research proposes a Fuzzy Normal Distribution Algorithm (Algorithm 2) for this issue.

As to the development of the fuzzy set theory, the fuzzy set was firstly introduced by Zadeh (1965). A number of notions extend the concept of this basic notion. Zadeh introduced the idea of type-2 fuzzy set (T2 FS) and computing with words paradigm [45–49]. T2 FS has also been investigated by recent studies (e.g. [26,27,10,1,33,24,50,51]). Atanassov [2] introduced Intuitionistic Fuzzy Set, and later

E-mail address: kevinkf.yuen@gmail.com

http://dx.doi.org/10.1016/j.asoc.2014.02.012
1568-4946/

Interval valued fuzzy set [3] have received many attentions by recent studies (e.g. [39,8,22,43,40,13,5,30]). Refs. [15–18,25,11] introduced 2-tuple fuzzy linguistic representation model. The concept of compound linguistic terms computation, however, appears not to receive any attentions from the literature.

The rest of this paper is organized as follows. Section 2 introduces the concept of Compound Linguistic Variable and its syntactic rule. The semantic rules, i.e. "Computing with CLV", are discussed in Section 3. That is how to map the linguistic form of CLV to the numerical form by Fuzzy Normal Distribution. Section 4 presents the Compound Linguistic Scale (CLS), which is formed by CLV and Deductive Rating Strategy (DRS). Section 5 demonstrates the numerical analyses of CLS, which include examining four parameters' behaviors, and two comparisons for a single user rating and a group of user ratings by using a single dimension rating scale and CLS respectively. Conclusions are drawn in Section 6. This article is essentially derived and revised from the author's thesis [44], especially for reduction of notations' complexity, and corrections of some presentations. Mathematical proofs are shown in Appendix I, whilst the axiomatic structures are shown in Appendix II.

2. Compound Linguistic Variable and syntaticl rule

Developing an appropriate schema for the combination of linguistic terms is challenging. It seems a lack of literature to build a system of compound linguistic terms and to compute the mapping between linguistic terms and numbers. To address this, the concept of Compound Linguistic Variable (CLV) is introduced. The established popular concept of linguistic variable is reviewed as below.

Definition 1 *((Linguistic Variable) [45]).* A linguistic variable is characterized by a quintuple $(\aleph, T(\aleph), X, G, M)$ in which $\aleph$ is the name of a linguistic variable; $T(\aleph)$ (or simply T) denotes the term-set of $\aleph$, that is, the set of names of linguistic values of $\aleph$, with each value being a fuzzy variable denoted generally by α and ranging over a universe of discourse X which is associated with the base variable x; G is a syntactic rule (which usually has the form of a grammar) for generating the names, α, of values of $\aleph$; and M is a semantic rule for associating with each α meaning, $M(\alpha)$, which is a fuzzy subset of X.

CLV extending the above concept is revised as below.

Definition 2 *((CLV)).* A Compound Linguistic Variable is characterized by a quintuple $(\aleph, T_\aleph, X_\aleph, G_\aleph, M_\aleph)$. $\aleph$ is the name of the Compound Linguistic Variable. A syntactic rule G is the function to generate a term-matrix $T_\aleph$, a matrix of linguistic terms of $\aleph$. A semantic rule $M_\aleph$ is the function to map the linguistic term domain $T_\aleph$ into real number domain $X_\aleph$. Each term α in $T_\aleph$, i.e. $\alpha \in T_\aleph$, is associated with a prototype element x, i.e. $x \in X$, generated by $M_\aleph$.

Definitions of notations in Definition 2 replace ones in Definition 1 in this article. To further illustrate CLV, term-matrix is described as follows.

Definition 3 *(($G_\aleph$ and $T_\aleph$)).* A term-matrix of a Compound Linguistic Variable $\aleph$, i.e. $T_\aleph$, is syntactically mapped from (H, D, A) by a syntactic grammar function for $\aleph$, i.e. $G_\aleph$, which is defined as follows

$$G_\aleph : (H, D, A) \mapsto T_\aleph. \tag{1}$$

$\mapsto$ denotes "be linguistically mapped to". H, D, and A are increasingly vectorized or ordered, and are defined in Definitions 4–7 as blow respectively.

Definition 4 *((A)).* An atomic linguistic term a is the base linguistic element used to describe the statement, or to measure the objects in the initial sense. A vectorized atomic linguistic variable A is the vector of atomic linguistic terms, i.e. $a \in A$, and has the form: $A = [a_j]_{j=1}^n = [a_1, \cdots, a_n]$ such that $a_1 \le \cdots \le a_n$ (or denoted by $\le_{j=1}^n(a_j)$). n is an odd number and $n \ge 3$. For example, A= [Poor, Weak, Fair, Good, Excellent] where Poor ≤ Weak ≤ Fair ≤ Good ≤ Excellent.

Definition 5 *((H)).* A hedge linguistic term h is used to hedge or adjust the linguistic quantity in a. A vectorized hedge linguistic variable H is of the form: $H = [h_i]_{i=1}^\eta$ such that $\le_{i=1}^\eta(h_i)$ and $h \in H$. For example, = [Little, Quite, Much] such that Little ≤ Quite ≤ Much.

Definition 6 *((D)).* A directional term d is used to define a direction for linguistic quantity hedge h. A vectorised directional linguistic variable D usually consists of a vector of three ordinal directional terms, i.e. $D = [d_i]_{i=1}^3 = [d^-, d^\theta, d^+]$ such that $\le_{i=1}^3(d_i)$ (or $d^- \le d^\theta \le d^+$). d^θ is a static direction. d^+ modifies h as the positive linguistic quantity modifier whilst d^- is the negative linguistic quantity modifier since h is the pure linguistic quantity in nature. For example, D = [Below, Absolutely, Above] such that Below ≤ Absolutely ≤ Above.

h and d can form a directional hedge linguistic term $\vec{h}$ defined as below.

Definition 7 *(($\to$ H)).* Let $\vec{H} = [\vec{h}_i]_{i=1}^m$ such that $\le_{i=1}^m(\vec{h}_i)$. To extend its properties, then

$$\vec{h}^- = h \oplus d^-, \quad \text{where} \quad \vec{h}^- \in \vec{H}^- = [\vec{h}_i]_{i=1}^\eta, \ \eta = |\vec{H}^-| \tag{2}$$

$$\vec{h}^\theta = h \oplus d^\theta \quad \text{where} \quad \vec{h}^\theta = \vec{h}_{\eta+1} \tag{3}$$

$$\vec{h}^+ = h \oplus d^+ \quad \text{where} \quad \vec{h}^+ \in \vec{H}^+ = [\vec{h}_i]_{i=\eta+2}^m \tag{4}$$

$\vec{h}^-$ is a negative hedge linguistic term, $\vec{h}^\theta$ is a static hedge linguistic term, and $\vec{h}^+$ is a positive hedge linguistic term. $|\cdot|$ returns the cardinal number of a set. $\oplus$ is a linguistic combination operator merging two linguistic terms ("+" is the addition between two numbers). A syntactic function for $\vec{H}$, $G_{\vec{H}} : (H, D) \mapsto \vec{H}$, is derived as below.

Proposition 1. $(\vec{H} = G_{\vec{H}}(H, D))$. If $H = [h_i]_{i=1}^\eta$ and $D = [d^-, d^\theta, d^+]$, then

$$\vec{H} = G_{\vec{H}}(H, D) = [\vec{h}_i]_{i=1}^m = [(h_\eta \oplus d^-), \ \ldots, \ (h_1 \oplus d^-), \ d^\theta, \ (h_1 \oplus d^+), \ \ldots, \ (h_\eta \oplus d^+)] \tag{6}$$

For example, $\vec{H}$ = ["Much Below", "Quite Below", "Little Below", "Absolutely", "Little Above", "Quite Above", "Much Above"]. When $\vec{h}$ and a are structured as a new term, the definition of a compound linguistic term is given as below.