Lawry, J. (2000). An Alternative Interpretation of Linguistic Variables and Computing with Words. In IPMU 2000.

An Alternative Interpretation of Linguistic Variables and Computing with Words

Jonathan Lawry

A.I. Group,
Department of Engineering Mathematics,
University of Bristol, Bristol BS8 1TR,
United Kingdom
j.lawry@bristol.ac.uk

Abstract: An alternative theory of linguistic variables is introduced based on voting model semantics. This theory is then applied to computing with words whereby a calculus is introduced for inference from linguistic facts and rules.

Keywords: Computing with words, linguistic variable, linguistic syllogism, linguistic fact, linguistic rule

1. Introduction

The use of fuzzy sets is clearly central to computing with words as they provide a means of modelling the vagueness underlying most natural language terms. Their introduction should not, however, result in large-scale increases in computational complexity. Unfortunately such increases do seem to result from the use of the extension principle in computing with words [14]. In the sequel we shall introduce an alternative approach to computing with words based on mass assignment theory [2] that avoids complexity problems of the above type. The theory is based on the use of a restricted notion of linguistic variable together with ideas taken from the mass assignment theory of the probability of fuzzy events [3]. To proceed, however, it is first necessary to give a brief introduction to mass assignment theory.

2 Basic Mass Assignment Theory

A mass assignment for a fuzzy concept, introduced by Baldwin [2], can be interpreted as the probability distribution over possible definitions of the concept. These varying definitions might be provided by a population of voters where each is asked to give his or her crisp definition of the concept.

Definition 2.1 (Mass Assignment)
Let $\boldsymbol{f}$ be a fuzzy subset of a finite universe Ω such that the range of the membership function of $\boldsymbol{f}$, χ_f, is $\{y_1,\ldots,y_n\}$ where $y_i > y_{i+1} > 0$. Then the mass assignment of $\boldsymbol{f}$, denoted $\boldsymbol{m_f}$, is a probability distribution on 2^{Ω} satisfying $\boldsymbol{m_f}(F_i) = y_i - y_{i+1}$ where $F_i = \{x \in \Omega | \chi_f(x) \geq y_i\}$ for $i = 1,\ldots,n$. $\{F_i\}_{i=1}^n$ are referred to as the focal elements(sets) of $\boldsymbol{m_f}$.

The notion of mass assignment suggests a means of conditioning a variable relative to a fuzzy constraint. That is given a variable X and the constraint X is f we can determine a conditional probability distribution on X referred to as the *least prejudiced distribution of* f (denoted lp_f) [2],[3]. The least prejudiced distribution is a special case of the pignistic probability distribution in Smets Transferable Belief Model [11].

Definition 2.2 (Least Prejudiced Distribution)
For f a fuzzy subset of a finite universe Ω such that f is normalised then the least prejudiced distribution of f is a probability distribution on Ω given by

$$\forall x \in \Omega \; lp_f(x) = \sum_{F_i : x \in F_i} \frac{\boldsymbol{m}_f(F_i)}{|F_i|}$$

where $\boldsymbol{m}_f$ is the mass assignment of f and $\{F_i\}_i$ is the corresponding set of focal elements.

The notion of least prejudiced distribution provides a mechanism by which we can, in a sense, convert a fuzzy set into a probability distribution. That is, in the absence of any prior knowledge, we might, on being told f, naturally infer the distribution lp_f. If, however, fuzzy sets are to serve as descriptions of probability distributions the converse must also hold. In other words, given a probability distribution we require it to hold that there is a unique fuzzy set conditioning on which yields this distribution.

Theorem 2.3
Let $\boldsymbol{Pr}$ be a probability distribution on a finite universe Ω taking as a range of values $\{p_1,\ldots,p_n\}$ where $0 \leq p_{i+1} < p_i \leq 1$ and $\sum_{i=1}^{n} p_i = 1$. Then $\boldsymbol{Pr}$ is the

least prejudiced distribution of a fuzzy set f if and only if f has mass assignment given by $\boldsymbol{m}_f(F_i) = y_i - y_{i+1}$ for $i = 1,\ldots,n$ where $F_i = \{x \in \Omega | \boldsymbol{Pr}(x) \geq p_i\}$ and

$$y_i = |F_i| p_i + \sum_{j=i+1}^{n} \left(|F_j| - |F_{j+1}|\right) p_j$$

Proof (see [5])

Curiously this transformation is identical to the bijective transformation method proposed by Dubois and Prade [6] although the motivation here is somewhat different.

3 A Mass Assignment Theory of Linguistic Variables

The concept of a linguistic variable was first introduced by Zadeh (see [12]) as a model of how words or labels can represent vague concepts in natural language. More formally:

Definition 3.1 (Linguistic Variable)

A linguistic variable is a quadruple $\langle L, T(L), \Omega, M \rangle$[1] in which L is the name of the variable, $T(L)$ is a finite term set of labels or words (i.e. the linguistic values), Ω is a universe of discourse and M is a semantic rule.

The semantic rule M then is defined as a function that associates a normalised fuzzy subset of Ω with each word in $T(L)$. In other words the fuzzy set $M(w)$ can be viewed as encoding the meaning of w so that for $x \in \Omega$ the membership value $\chi_{M(w)}(x)$ quantifies the suitability or applicability of the word w as a label for the value x. We can regard the semantic function M as being determined by a group voting model (see [2], [8] and [9]) across a population of voters as follows. Each voter is asked to provide the subset of words from the finite set $T(L)$ which are appropriate as labels for the value x. The membership value $\chi_{M(w)}(x)$ is then taken to be the proportion of voters who include w in their set of labels.

Example 3.2

Consider the set of words {*small(s)* , *medium(m)*, *large(l)*} as labels of a linguistic variable *SIZE* describing values in U=[0,100]. Given a set of 10 voters a possible voting pattern for the value 25 is

1	2	3	4	5	6	7	8	9	10
s *m*	*s* *m*	*s* *m*	*s* *m*	*s* *m*	*s*	*s*	*s*	*s*	s

This gives $\chi_{M(small)}(25) = 1$ and $\chi_{M(medium)}(25) = 0.5$. Now this voting pattern can be represented by a mass assignment {*small*, *medium*}:0.5,{*small*}:0.5 on the power set of {*small*, *medium*, *large*}. This in turn represents a fuzzy set on the set of words, namely *small*/1+*medium*/0.5. This fuzzy set can be viewed as a linguistic description of the value 25 in terms of the words *small*, *medium* and *large* and is denoted by $des_{SIZE}(25)$. Notice that the linguistic description of 25 can be expressed in terms of the semantic function M in the following manner:

$$small / \chi_{M(small)}(25) + medium / \chi_{M(medium)}(25)$$

Hence, in practice we need only to define the fuzzy sets M(*small*), M(*medium*) and M(*large*) from which we can determine any linguistic description. The idea of using fuzzy sets on words to describe values was first proposed by Baldwin in [2]. More formally, a linguistic description of a value is defined by:

Definition 3.3 (Linguistic Description of a Value)

Let $x \in \Omega$ then the linguistic description of x relative to the linguistic variable L is the fuzzy subset of $T(L)$

$$des_L(x) = \sum_{w \in T(L)} w / \chi_{\boldsymbol{M}(w)}(x)$$

In cases where the linguistic variable is fixed we drop the subscript L and write $des(x)$. This notion can be extended to the case where the value given is a fuzzy subset of Ω in which case the appropriate linguistic description is defined as follows.

Definition 3.4 (Linguistic Description of a Fuzzy Set)

Let $f \subseteq_f \Omega$ then the linguistic description of f relative to the linguistic variable L is a fuzzy subset of $T(L)$, $des_L(f)$ satisfying

$$\forall w \in T(L)\ lp_{des_L(f)}(w) = \sum_{F_i} \frac{\boldsymbol{m}_f(F_i)}{|F_i|} \sum_{x \in F_i} lp_{des_L(x)}(w)$$

if Ω is finite and $\int_0^1 \int_{f_y} \frac{lp_{des_L(x)}(w)}{\lambda(f_y)}\, dx\, dy$ in the continuous case.

[1] In [13] Zadeh originally defined a linguistic variable as a quintuple by including a syntactic rule according to which new terms (i.e. linguistic values) could be formed by applying hedges to existing words. This, however, allows for the term set to be infinite.

Given these constraints $des_L(f)$ can be determined according to the transformation described in theorem 2.3. The intuition underlying this definition is as follows. For each focal set F_i (or α-cut f_y in the continuous case) we average the probability of w being selected to label values in F_i. This is then averaged across the focal sets to give the overall probability of w. In general we take the expression linguistic description to mean a fuzzy subset of the term set of a linguistic variable.

4 Reasoning with Linguistic Syllogisms

Syllogisms are a well known classical inference schema of the form

$$\frac{\text{All } X \text{ are } Y \qquad \text{All } Y \text{ are } Z}{\text{All } X \text{ are } Z}$$

where X, Y and Z are properties. This schema was extended in [13] and [7] to allow the use of linguistic quantifiers such as *most, few* and *several*. Here we adopt a different approach to Zadeh and interpret linguistic quantifiers as words describing probability values determined according to mass assignment theory [3]. More specifically, let the set of words describing probability values be $T_{\boldsymbol{Pr}} = \{Q_1, \cdots, Q_n\}$ with corresponding semantic function $M_{\boldsymbol{Pr}}$. Then for any variable α representing the value of the probability of some specific event or conditional event define the linguistic variable describing α as $\langle L_\alpha, T_{\boldsymbol{Pr}}, [0,1], M_{\boldsymbol{Pr}} \rangle$. As is noted in [7] and [4] the fact that probabilities are not truth functional means that a schema for syllogisms with linguistic quantifiers cannot be translated directly from the classical schema. For instance, probability theory does not allow us to make any inference regarding the shoe size of Swedes from the facts '*most* Swedes are *tall*' and '*most tall* people have *large* shoe sizes'. In view of this we define a linguistic syllogism schema along similar lines to [7] as follows:

$$\frac{(Q_i)\ X \text{ are } Y, \qquad (Q_j)\ (X \text{ and } Y) \text{ are } Z, (Q_k)\ (X \text{ and } \neg Y) \text{ are } Z}{(des)\ X \text{ are } Z}$$

where $Q_i, Q_j, Q_k \in T_{\boldsymbol{Pr}}$ and $des \subseteq_f T_{\boldsymbol{Pr}}$. This can be translated into linguistic variable form in the following manner. Let $\alpha = \boldsymbol{Pr}(Y|X)$, $\beta = \boldsymbol{Pr}(Z|X,Y)$, $\gamma = \boldsymbol{Pr}(Z|X,\neg Y)$ and $\theta = \boldsymbol{Pr}(Z|X)$ then the linguistic syllogism schema is equivalent to

$$\frac{L_\alpha = Q_i,\ L_\beta = Q_j,\ L_\gamma = Q_k}{L_\theta \text{ is } des}$$

Now according to the theorem of total probability

$$\boldsymbol{Pr}(Z|X) = \boldsymbol{Pr}(Z|X,Y)\,\boldsymbol{Pr}(Y|X) + \boldsymbol{Pr}(Z|X,\neg Y)\,\boldsymbol{Pr}(\neg Y|X)$$

So we have that $\theta = \beta\alpha + \gamma(1-\alpha)$. Notice that if $\boldsymbol{Pr}(Z|X,Y) = \boldsymbol{Pr}(Z|Y)$ and $\boldsymbol{Pr}(Z|X,\neg Y)$ is unknown then the classical form of the schema is retained. Essentially, in this case, we are assuming that Z is *conditionally independent* [10] of X given Y. This assumption is likely to be most justifiable when the populations sizes of objects satisfying X and Y are similar.

Now given the above constraints posterior distributions of α, β and γ can be determined. For $Q \in T_{\boldsymbol{Pr}}$ the posterior density given Q is, from Bayes theorem

$$p(\alpha|Q) = \left[\boldsymbol{Pr}(Q|\alpha)p(\alpha)\right] \Big/ \boldsymbol{Pr}(Q)$$

Now according to the voting model semantics $\boldsymbol{Pr}(Q|\alpha)$ is interpreted as being the probability that a randomly selected voter will pick the label Q when presented with value α. From section 3 we recall that this is given by $lp_{des_{L_\alpha}(\alpha)}(Q)$. Hence, assuming that, *a priori,* all probability values are equally likely (i.e. $p(\alpha)$ is the uniform distribution) then we have

$$p(\alpha|Q) = lp_{des_{L_\alpha}(\alpha)}(Q) \Big/ \int_0^1 lp_{des_{L_\alpha}(\alpha)}(Q)\, d\alpha$$

Furthermore, since we may assume that the variables α, β and γ are independent it follows that the expected value of θ is given by

$$E(\theta) = E(\beta)E(\alpha) + E(\gamma)(1 - E(\alpha))$$

where $E(\alpha) = \int_0^1 \alpha\, p(\alpha|Q_i)\, d\alpha$, $E(\beta) = \int_0^1 \beta\, p(\beta|Q_j)$ and $E(\gamma) = \int_0^1 \gamma\, p(\gamma|Q_k)\, d\gamma$. Hence, we take as our estimate for θ, $\hat{\theta} = E(\beta)E(\alpha) + E(\gamma)(1 - E(\alpha))$. From this we can determine a linguistic description so that $des = \left(des_{L_\theta}(\hat{\theta})\right)$.

Example 4.1

Consider the following linguistic syllogism.

$$\frac{most\ X \text{ are } Y \qquad most\ (X \text{ and } Y) \text{ are } Z,\ several\ (X \text{ and } \neg Y) \text{ are } Z}{des\ X \text{ are } Z}$$

where $T_{\boldsymbol{Pr}} = \{few, several, most\}$ and $M_{\boldsymbol{Pr}}$ is such that $M_{\boldsymbol{Pr}}(few) = [0:1\ 0.25:1\ 0.5:0]$, $M_{\boldsymbol{Pr}}(several) = [0:0\ 0.25:1\ 0.75:1\ 1:0]$ and $M_{\boldsymbol{Pr}}(most) = [0.5:0\ 0.75:1\ 1:1]$.

For the sake of simplicity we shall take T_{Pr} to be defined in this way for all subsequent examples.
Now

$$lp_{des(\alpha)}(most) = 2\alpha - 1 \text{ for } 0.5 \le \alpha \le 1$$
$$= 0 \text{ otherwise}$$

and hence

$$p(\alpha|most) = lp_{des_{L_\alpha}(\alpha)}(most) \Big/ \int_0^1 lp_{des_{L_\alpha}(\alpha)}(most)\, d\alpha$$
$$= 8\alpha - 4 \text{ for } 0.5 \le \alpha \le 1$$
$$= 0 \text{ otherwise}$$

From this we obtain that

$$E(\alpha) = E(\beta) = \int_{0.5}^{1} (8x - 4)x\, dx = 0.833333$$

Similarly, we find that $E(\gamma) = 0.5$. Therefore, $\hat{\theta} = 0.77777$ and $des = des_{L_{\hat{\theta}}}(0.77777)$=*several*/0.8889 +*most*/1. In other words, from a voting model perspective, around 89% of voters believe that either *several* or *most* are suitable as labels for the probability of Z given X and around 11% think that only *most* is appropriate.

It is also possible that there is uncertainty regarding the labels for $\boldsymbol{Pr}(Y|X)$, $\boldsymbol{Pr}(Z|X,Y)$ and $\boldsymbol{Pr}(Z|X,\neg Y)$. For instance, consider the natural language constraints regarding a number of batches of a certain type of component produced by a rather poor factory.

Several or quite possibly *most* of the components checked were faulty

Few although in some cases *several* of the components were checked

Query: What is the overall proportion of components that are faulty?

In order to model such situations we introduce the notion of a general linguistic syllogism characterised by the following schema.

$$\frac{(des_1)X \text{ are } Y \qquad (des_2)(X \text{ and } Y) \text{ are } Z,\ (des_3)(X \text{ and } \neg Y) \text{ are } Z}{(des)X \text{ are } Z}$$

where des_1, des_2, des_3 and des are linguistic descriptions over T_{Pr}.

The corresponding linguistic variable interpretation is then given by

$$\frac{L_\alpha \text{ is } des_1,\ L_\beta \text{ is } des_2,\ L_\gamma \text{ is } des_3}{L_\theta \text{ is } des}$$

Now as for standard linguistic syllogisms these constraints determine posterior distributions on α, β and γ. In this case, however, we are conditioning on linguistic descriptions rather than words. We define

$$p(\alpha|des) = \sum_{D_j} p(\alpha|D_j)\boldsymbol{m}_{des}(D_j)$$

where $\{D_j\}_j$ are the focal sets of $\boldsymbol{m}_{des}$ and where from Bayes theorem

$$p(\alpha|D_j) = \sum_{Q \in D_j} lp_{des_{L_\alpha}(\alpha)}(Q) \Big/ \int_0^1 \sum_{Q \in D_j} lp_{des_{L_\alpha}(\alpha)}(Q)\, d\alpha$$

Expected values for α, β and γ can be determined from the posterior distributions and an estimate for θ determined as before.

Example 4.2
Consider the following general linguistic syllogism

$$\frac{(most/1 + several/0.8)\ X \text{ are } Y \qquad (few/0.2 + several/1)\ (X \text{ and } Y) \text{ are } Z \qquad (few/1 + several/0.5)\ (X \text{ and } \neg Y) \text{ are } Z}{(des)\ X \text{ are } Z}$$

Now $\boldsymbol{m}_{most/1+several/0.8} = \{most, several\} : 0.8$, $\{most\} : 0.2$ so that

$$p(\alpha|most/1 + several/0.8) = p(\alpha|\{most, several\})0.8 + p(\alpha|most)0.2$$

where $p(\alpha|\{most, several\}) =$

$$\frac{lp_{des_{L_\alpha}}(most) + lp_{des_{L_\alpha}}(several)}{\int_0^1 lp_{des_{L_\alpha}}(most) + lp_{des_{L_\alpha}}(several)\, d\alpha} =$$

$$1 - lp_{des_{L_\alpha}(\alpha)}(few) \Big/ 1 - \int_0^{0.5} lp_{des_{L_\alpha}(\alpha)}(few)$$

where

$$lp_{des_{L_\alpha}(\alpha)}(few) = 1 - 2\alpha \text{ for } 0 \le \alpha \le 0.5$$
$$= 0 \text{ otherwise}$$

and

$$p(\alpha|most) = lp_{des_{L_\alpha}}(most) \Big/ \int_0^1 lp_{des_{L_a}}(most)\, d\alpha$$

where

$$lp_{des_{L_\alpha}(\alpha)}(most) = 2\alpha - 1 \text{ for } 0.5 \le \alpha \le 1$$

$$= 0 \text{ otherwise}$$

Hence, $p(\alpha|most/1 + several) =$

$$0.8\left(\frac{8}{3}\alpha\right) = 2.13333\alpha \text{ for } 0 \le \alpha \le 0.5$$

$$0.8\left(\frac{4}{3}\right) + 0.3(8\alpha - 4) = 1.6\alpha + 0.266667 \text{ otherwise}$$

Similarly we find that

$$p(\beta \mid few/0.2 + several/1)$$

$$= 3.2\beta + 0.266667 \text{ for } 0 \le \beta \le 0.5$$

$$= -3.73333\beta + 3.73333 \text{ otherwise}$$

and

$$p(\gamma|few/1 + several/0.5)$$

$$= 2.66667 - 4\gamma \text{ for } 0 \le \gamma \le 0.5$$

$$= -1.33333\gamma + 1.33333 \text{ otherwise}$$

From these distributions we obtain that $E(\alpha) = 0.655556$, $E(\beta) = 0.477778$ and $E(\gamma) = 0.277778$ so that $\theta = E(\alpha)E(\beta) + (1 - E(\alpha))E(\gamma)$
$= 0.655556(0.477778) + (1 - 0.655556)0.277778$
$= 0.40889$ and $des = des_{L_\theta}(0.40889) =$
$few/0.364 + several/1$

The linguistic syllogism schema only provides a model for a rather restricted form of computing with words. In the following sections we shall attempt to provide more general models of reasoning with linguistic variables. The framework will be based on ideas taken from the Fril programming language [1].

5 Inference from Linguistic Facts.

A linguistic fact will be an instantiation of a linguistic variable qualified by a linguistic description acting as a quantifier. More specifically, let X be a variable with universe Ω and L_X is a linguistic variable describing X such that $T(L_X) = \{w_1, \cdots, w_m\}$ then a linguistic fact describing X has the form

$$((L_X = w_i)) : (des_i)$$

where $des_i \subseteq_f T_{\mathbf{Pr}}$. Now given a knowledge base of such constraints $((L_X = w_i)) : (des_i)$ for $i = 1, \cdots, m$ and letting $\alpha_i = \mathbf{Pr}(L_X = w_i)$ we can express this in terms of constraints on linguistic variables as follows:

$$L_{\alpha_i} \text{ is } des_i \text{ for } i = 1, \cdots, m \text{ and } \sum_{i=1}^{m} \alpha_i = 1$$

Let this set of constraints be denoted by K. Now conditional distributions $p(\alpha_i|des_i)$ can be determined as before and assuming independence up to the constraint $\sum_{i=1}^{n} \alpha_i = 1$ a joint posterior distribution on $\alpha_1, \cdots, \alpha_m$ is given by $\forall \bar{\alpha} \in V(m)$

$$p(\bar{\alpha}|K) = \prod_{i=1}^{m} p(\alpha_i|des_i) \Big/ \int_{V(m)} \prod_{i=1}^{m} p(\alpha_i|des_i)\, dV(m)$$

where $V(m) = \left\{\bar{\alpha} \in [0,1]^m \,\middle|\, \sum_{i=1}^{m} \alpha_i = 1\right\}$.

From this joint distribution posterior marginal distributions can be determined for $\alpha_1, \cdots, \alpha_m$ such that

$$p_i(\alpha|K) = \int_{\{\bar{\alpha} \in V(m) |\, \alpha_i = \alpha\}} p(\bar{\alpha}|K)\, dV(m)$$

Hence, we can obtain the estimate

$$\hat{\alpha}_i = \int_0^1 \alpha\, p_i(\alpha|K)\, d\alpha$$

Converting the probability distribution $w_i : \hat{\alpha}_i$ for $i = 1, \cdots, m$ into a fuzzy set according to theorem 2.3 we can infer a linguistic description of X. Alternatively, a numerical estimate of the value of X can be determined by $\hat{X} = \sum_{i=1}^{m} E(X|w_i)\, \hat{\alpha}_i$ where $E(X|w_i) = \int_\Omega X\, p(X|w_i)\, dX$ and where

$$p(X|w_i) = lp_{des_{L_X}(X)}(w_i) \Big/ \int_\Omega lp_{des_{L_X}(X)}(w_i)\, dX$$

Example 5.1

Consider the following statement with associated query adapted from an example given by Baldwin [4]:

The examination marks of a good project student will be *excellent* in *most* subjects *good* in *several* subjects and *poor* in only a *few* subjects.

What is the average exam mark of a good project student?

Let *MARK* be the percentage mark of a good project student so that $\Omega = [0,100]$ and let L_{MARK} be the linguistic variable describing *MARK* with $T(L_{MARK}) =$ {*poor*, *mediocre*, *good*, *excellent*}. The associated semantic function is defined by $M_{MARK}(poor) = [0:1\ 45:1\ 50:0]$, $M_{MARK}(mediocre) = [40:0\ 45:1\ 55:1\ 60:0]$, $M_{MARK}(good) =$

$[50:0\ 55:1\ 65:1\ 70:0]$ and $M_{MARK}(excellent) = [60:0\ 65:1\ 100:1]$. The problem can be expressed by the following linguistic facts:

$((L_{MARK} = poor)):(few)$
$((L_{MARK} = mediocre)):(few/1 + several/1 + most/1)$
$((L_{MARK} = good)):(several)$
$((L_{MARK} = excellent)):(most)$

Let $\alpha_1 = \boldsymbol{Pr}(poor)$, $\alpha_2 = \boldsymbol{Pr}(mediocre)$, $\alpha_3 = \boldsymbol{Pr}(good)$ and $\alpha_4 = \boldsymbol{Pr}(excellent)$ then we have the following constraints denoted by K:

$L_{\alpha_1} = few$, L_{α_2} *is* $few/1 + several/1 + most/1$

$L_{\alpha_3} = several$, $L_{\alpha_4} = most$

and $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$

Given this knowledge we obtain a joint posterior distribution.

$$\forall\langle\alpha_1,\cdots,\alpha_4\rangle \in V(4)\ p(\alpha_1,\alpha_2,\alpha_3,\alpha_4|K) = \frac{p(\alpha_1|few)p(\alpha_3|several)p(\alpha_4|most)}{\int_{V(4)} p(\alpha_1|few)p(\alpha_3|several)p(\alpha_4|most)dV(4)}$$

Now

$$\int_{V(4)} p(\alpha_1|few)p(\alpha_3|several)p(\alpha_4|most)dV(4) =$$

$$\int_0^{0.5} p(\alpha_1|few) \int_0^{0.5-\alpha_1} p(\alpha_3|several) \int_{0.5}^{1-\alpha_1-\alpha_3} p(\alpha_4|most)\, d\alpha_4 d\alpha_3 d\alpha_1$$

$$= \int_0^{0.5} (4-8\alpha_1) \int_0^{0.5-\alpha_1} 4\alpha_3 \int_{0.5}^{1-\alpha_1-\alpha_3} (8\alpha_4 - 4)\, d\alpha_4 d\alpha_3 d\alpha_1 = 1/36$$

Hence, $\forall\langle\alpha_1,\cdots,\alpha_4\rangle \in V(4)\ p(\alpha_1,\alpha_2,\alpha_3,\alpha_4|K) =$

$36(4-8\alpha_1)(4\alpha_3)(8\alpha_4 - 4)$ for $0 \le \alpha_1 \le 0.5$

$0 \le \alpha_3 \le 0.5$ and $0.5 \le \alpha_4 \le 1$

$= 0$ otherwise

From this we obtain the following marginal distributions

$p_1(\alpha|K) =$

$-384\alpha^5 + 480\alpha^2 - 120\alpha + 960\alpha^4 - 960\alpha^3 + 12$

for $0 \le \alpha \le 0.5$

$= 0$ otherwise

$p_2(\alpha|K) = -76.865\alpha^5 + 192\alpha^2 - 192\alpha^3 + 72\alpha + 9.6$

for $0 \le \alpha \le 0.5$

$= 0$ otherwise

$p_3(\alpha|K) = -384\alpha^5 - 384\alpha^2 + 576\alpha^3 + 72\alpha$

for $0 \le \alpha \le 0.5$

$= 0$ otherwise

$p_4(\alpha|K) =$

$-384\alpha^5 - 768\alpha^2 + 960\alpha^4 - 384\alpha^3 + 768\alpha - 192$

for $0.5 \le \alpha \le 1$

$= 0$ otherwise

Taking the expected values of these distribution gives us; $\hat{\alpha}_1 = 0.0714286$, $\hat{\alpha}_2 = 0.0857143$, $\hat{\alpha}_3 = 0.171429$ and $\hat{\alpha}_4 = 0.671429$. Converting the resulting distribution into a fuzzy set yields the following linguistic description of the average mark of a good project student.

$excellent/1 + good/0.5 + mediocre/0.329 + poor/0.286$

If a numerical value is required we can use Bayes theorem to determine conditional distributions on *MARK*, take expected values and finally evaluate a linear combination of these expected values relative to the above distribution. In this case we obtain that $E(MARK|poor) = 22.5926$, $E(MARK|mediocre) = 50$, $E(MARK|good) = 60$ and $E(MARK|excellent) = 82.38$. Hence, the estimated value for *MARK* is given by

$0.0714286(22.5926) + 0.0857143(50) + 0.171429(60)$
$+0.671429(82.38) = 71.498\%$

6 Linguistic General Rules

In addition to the ability to reason from linguistic facts it is also desirable to be able to make inferences from knowledge bases consisting of linguistic rules and facts. Here we introduce a type of linguistic rule based on the Fril general rule [1] and [2] with the following form:

$$\big((L_X = w_i)$$
$$\big((C_1)\cdots(C_r)\big) : \big((des_{i,1})\cdots(des_{i,r})\big)$$

where $\{C_j\}_{j=1}^r$ are a set of mutually exclusive and exhaustive constraints, $w_i \in T(L_X)$ and $\boldsymbol{des}_{i,j} \subseteq_f T_{\boldsymbol{Pr}}$. If $\alpha_{i,j} = \boldsymbol{Pr}(L_X = w_i|C_j)$ for $j = 1,\cdots,r$ then a knowledge base of such rules where $i = 1,\cdots,m$ can be interpreted as generating the following constraints:

$L_{\alpha_{i,j}}$ is $des_{i,j}$ for $i = 1,\cdots,m$ and $j = 1,\cdots,r$

$$\sum_{i=1}^{m} \alpha_{i,j} = 1 \text{ for } j = 1,\cdots,r$$

Now for a fixed j the constraints, denoted K_j, have the same form as those generated by a knowledge base of linguistic facts and hence we can determine a posterior distribution on $V(m)$ of the form;

$$p\left(\alpha_{1,j},\cdots,\alpha_{m,j}\middle|K_j\right)=$$
$$\prod_{i=1}^{m} p\left(\alpha_{i,j}\middle|des_{i,j}\right)\Big/\int_{V(m)}\prod_{i=1}^{m} p\left(\alpha_{i,j}\middle|des_{i,j}\right)dV(m)$$

Marginal distributions $p_i\left(\alpha_{i,j}\middle|K_j\right)$ can then be determined as before and expected values obtained in order to obtain estimates $\hat{\alpha}_{i,j}=E\left(\alpha_{i,j}\middle|K_j\right)$. Furthermore, according to Jeffrey's rule we have that

$$\boldsymbol{Pr}\left(L_X=w_i\right)=\sum_{j=1}^{r}\boldsymbol{Pr}\left(L_X=w_i|C_j\right)\boldsymbol{Pr}\left(C_j\right)$$

Let $\alpha_i=\boldsymbol{Pr}\left(L_X=w_i\right)$ for $i=1,\cdots,m$ and $\beta_j=\boldsymbol{Pr}\left(C_j\right)$ for $j=1,\cdots,r$. Now the values of β_j might be specified directly by means of standard Fril rules $\left(\left(C_j\right)\right):\beta_j$ for $j=1,\cdots,r$ in which case $\hat{\alpha}_i=\sum_{j=1}^{r}\hat{\alpha}_{i,j}\beta_j$. Alternatively, we may have a set of linguistic facts $\left(\left(C_j\right)\right):\left(des_j'\right)$ where $des_j\subseteq_f T_{\mathbf{Pr}}$ for $j=1,\cdots,r$. In this case we take $\hat{\beta}_j$ to be the expected value of the distribution $p\left(\beta_j\middle|des_j'\right)$ so that $\hat{\alpha}_i=\sum_{j=1}^{r}\hat{\alpha}_{i,j}\hat{\beta}_j$. As before the distribution generated on $T\left(L_X\right)$ can be transformed into a fuzzy set to obtain a linguistic description of X or a numerical estimate can be obtained according to $\hat{X}=\sum_{i=1}^{m}\hat{\alpha}_i E\left(x|w_i\right)$. Notice that the linguistic facts regarding X may be inferred form a direct fuzzy constraint on X. For instance, the fuzzy constraint X is f generates the linguistic facts;

$$\left(\left(L_X=w_i\right)\right):lp_{des_{L_X}(f)}\left(w_i\right)\text{ for }i=1,\cdots,m$$

where $des_{L_X}(f)$ is the linguistic description of f determined according to definition 3.4

Example 6.1
Consider the following rule base and associated query relating to the income of project managers.

Most senior project managers have *good* salaries and at least *several* have *very good* salaries.

Most experienced project managers have *moderate* salaries and *several* have *good* salaries. However, *few* have *very good* salaries.

Few junior project managers have *low* salaries.

Query: A certain project manager has been employed for *about fifteen* years. What will be his expected salary ?

Let *SAL* be the project manager's salary with $T\left(L_{SAL}\right)=\{low,\ moderate,\ good,\ very\ good\}$. Also, let *YRS* be the project manager's number of years of experience defined on the universe [0,40]. $T\left(L_{YRS}\right)=\{junior,\ experienced,\ senior\}$ where $M(junior)=$ [0:1 5:1 10:0], $M(experienced)$ = [0:0 5:1 15:1 20:0] and $M(senior)$=[10:0 15:1 40:1]. The knowledge base can be expressed in terms of linguistic general rules by:

((L_{SAL}=*low*)
(L_{YRS} = *junior*)
(L_{YRS} = *experienced*)
(L_{YRS} = *senior*)
: ((*few*)({*few, several, most*})({*few, several, most*}))

((L_{SAL}=*moderate*)
(L_{YRS} = *junior*)
(L_{YRS} = *experienced*)
(L_{YRS} = *senior*)
: (({*few, several, most*})(*most*)({*few, several, most*}))

((L_{SAL}=*good*)
(L_{YRS} = *junior*)
(L_{YRS} = *experienced*)
(L_{YRS} = *senior*)
: (({*few, several, most*})(*several*)(*most*))

((L_{SAL}=*good*)
(L_{YRS} = *junior*)
(L_{YRS} = *experienced*)
(L_{YRS} = *senior*)
: (({*few, several, most*})(*few*)({*several, most*}))

From these rules we generate the following three sets of linguistic constraints from which we use the methods described above to estimate the relevant probability values. For example,

$\underline{K_1}$ For $\alpha_{1,1}=\boldsymbol{Pr}\left(low|junior\right)$,
$\alpha_{2,1}=\boldsymbol{Pr}\left(moderate|junior\right)$, $\alpha_{3,1}=\boldsymbol{Pr}\left(good|junior\right)$
and $\alpha_{4,1}=\boldsymbol{Pr}\left(very\ good|junior\right)$
$L_{\alpha_{1,1}}=few$, $L_{\alpha_{2,1}}\in\{few, several, most\}$
$L_{\alpha_{3,1}}\in\{few, several, most\}$, $L_{\alpha_{4,1}}\in\{few, several, most\}$
$\alpha_{1,1}+\alpha_{2,1}+\alpha_{3,1}+\alpha_{4,1}=1$

From which we obtain $\hat{\alpha}_{1,1} = 0.135294$ and $\hat{\alpha}_{2,1} = \hat{\alpha}_{3,1} = \hat{\alpha}_{4,1} = 0.288235$

Similarly we find from K_2 that $\hat{\alpha}_{1,2} = 0.0857134$, $\hat{\alpha}_{2,2} = 0.671429$, $\hat{\alpha}_{3,2} = 0.171429$, $\hat{\alpha}_{4,2} = 0.0714286$ and from K_3 that $\hat{\alpha}_{1,3} = \hat{\alpha}_{2,3} = 0.0837997$, $\hat{\alpha}_{3,3} = 0.665734, \hat{\alpha}_{4,3} = 0.1666667$

Now let *about fifteen (abft)* be a fuzzy subset on [0,40] such that *about fifteen* = [12:0 15:1 18:0] then according to definition 3.4 we have that:

$$lp_{des_{L_{YRS}}(abft)}(exp.) = \int_0^1 \int_{abft_y} \frac{lp_{des_{L_{YRS}}(x)}(exp.)}{\lambda(abft_y)}\, dxdy$$

$$= \int_0^1 \int_{3y+12}^{18-3y} \frac{lp_{des_{L_{YRS}}(x)}(exp.)}{6(1-y)}\, dxdy$$

$$= \frac{1}{6}\int_0^1 \frac{1}{1-y} \int_{3y+12}^{18-3y} \frac{20-x}{10}\, dxdy = 0.5$$

Similarly we find that $lp_{des_{L_{YRS}}(abft)}(junior) = 0$ and $lp_{des_{L_{YRS}}(abft)}(senior) = 0.5$ giving us the linguistic facts

$((L_{YRS} = junior)):0$
$((L_{YRS} = experienced)):0.5$
$((L_{YRS} = senior)):0.5$

Hence, we infer that $\boldsymbol{Pr}(low) = 0.5\hat{\alpha}_{1,2} + 0.5\hat{\alpha}_{1,3}$
$= 0.5(0.0857143 + 0.0837997) = 0.0847565$

In the same way we obtain that $\boldsymbol{Pr}(moderate) = 0.3776$, $\boldsymbol{Pr}(good) = 0.4186$ and $\boldsymbol{Pr}(very\ good) = 0.11905$. Transforming this distribution into a fuzzy set gives the following linguistic description of the expected salary of the project manager:

$$low/0.339 + moderate/0.959 + good/1 + very\ good/0.442$$

7 Conclusion

We have outlined an approach to computing with words based on a mass assignment theory of linguistic variables. The methods described have a clear interpretation in terms voting model semantics and avoid many of the complexity problems associated with other approaches.

Acknowledgements

Thanks to Prof. Senen Barro and Dr. Alberto Bugarin for their many insightful comments.

References

[1] J.F. Baldwin, T.P. Martin, B.W. Pilsworth (1988), *FRIL Manual (Version 4.0)*, FRIL Systems Ltd, Bristol Business Centre, Maggs House, Queens Road, Bristol BS8 1QX, UK.

[2] J.F. Baldwin, T.P. Martin, B.W. Pilsworth (1995), *FRIL -Fuzzy and Evidential Reasoning in A.I*, Research Studies Press, John Wiley.

[3] J.F. Baldwin, J. Lawry, T.P. Martin (1996), "A Mass Assignment Theory of the Probability of Fuzzy Events", *Fuzzy Sets and Systems* 83, pp353-367

[4] J.F. Baldwin (1997), 'Computing with Words and Fuzzy Data Browser', *Proceedings of Unicom Conference on Fuzzy Logic: Applications and Future Directions*.

[5] J.F. Baldwin, J. Lawry, T.P. Martin (1998), 'The Application of Generalised Fuzzy Rules to Machine Learning and Automated Knowledge Discovery', *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* Vol. 6, No. 5, pp459-487

[6] D. Dubois, H. Prade (1983), 'Unfair Coins and Necessity Measures: Towards a Possibilistic Interpretation of Histograms', *Fuzzy Sets and Systems* 10, pp15-20.

[7] D. Dubois, H. Prade (1988), 'On Fuzzy Syllogisms', *Computational Intelligence*, Vol.4, pp171-179

[8] B.R. Gaines (1978), 'Fuzzy and Probability Uncertainty Logics', *Journal of Information and Control* 38, pp154-169

[9] J. Lawry (1998), 'A Voting Mechanism for Fuzzy Logic' *The International Journal of Approximate Reasoning* 19, pp315-333.

[10] J. Pearl (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann.

[11] P. Smets (1994), 'The transferable belief model', *Artificial Intelligence,* 66, pp191-234

[12] L.A. Zadeh, 'The concept of a linguistic variable and its applications to approximate reasoning', *Part I: Information Sciences 8* (1975) pp199-249; *Part II: Information Sciences 8* (1975) pp301-357; *Part III: Information Sciences* 9 (1976) pp 43-80

[13] L.A. Zadeh (1983), 'A Computational Approach to Fuzzy Quantifiers in Natural Language', *Comp. & Mechs. with Applications* Vol. .9, No.1, pp149-184

[14] L.A. Zadeh (1996), *Fuzzy Logic =Computing with Words*, IEEE Transactions on Fuzzy Systems, Vol. 4, No. 2, pp103-111