Label Semantics as a Framework for Granular Modelling

Jonathan Lawry

Artificial Intelligence Group, Department of Engineering Mathematics
University of Bristol, BS8 1TR, United Kingdom
J.Lawry@bris.ac.uk

Summary. An alternative perspective on granular modelling is introduced where an information granule characterises the relationship between a label expression and elements in an underlying perceptual space. Label semantics is proposed as a framework for representing information granules of this kind. Mass relations and linguistic decision trees are then introduced as two types of granular models in label semantics. Finally, its shown how linguistic decision trees can be combined within an attribute hierarchy to model complex multi-level composite mappings.

1 Introduction to Granular Modelling

Fundamental to human communication is the ability to effectively describe the continuous domain of sensory perception in terms of a finite set of description labels. It is this process of *granular modelling* which permits us to process and transmit information efficiently at a suitable level of detail, to express similarity and difference between perceptual experiences and to generalize from current knowledge to new situations. Furthermore, it allows us to express information and knowledge in a way that is robust to small variations, noise and sensory aggregations in a complex multi-dimensional and evolving perceptual environment. Given these advantages, the formalization of granular models within a mathematical theory can allow for the effective modelling of complex multi-dimensional systems in such a way as to be understandable to practitioners who are not necessarily experts in formal mathematics.

The use of labels as a means of discretizing information plays a central role in granular modelling. Indeed one possible definition of an *information granule* could be in terms of the mapping between labels and domain values as follows:

> An information granule is a characterisation of the relationship between a discrete label or expression and elements of the underlying (often continuous) perceptual domain which it describes.

From this perspective crisp sets, fuzzy sets [19], random sets [10] and rough sets [11] can all correspond to information granules in that they can be used to characterise just such a relationship between a label and the elements of the

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 87–102, 2008.
springerlink.com

underlying domain. A typical form of information granule is as the extension of the concept symbolically represented by a given label. For a label L the extension of L identifies the set of domain elements to which L can be truthfully or appropriately applied. Fuzzy sets, random sets and rough sets are then mechanisms according to which gradedness, uncertainty and imprecision can respectively be introduced into the definition of concept extensions.

The above definition of information granule should be contrasted with that of Zadeh [20] who explains granularity in terms of (possibly fuzzy) clusters of points as follows:

> A granule is a clump of objects (points) which are drawn together by indistinguishability, similarity, proximity and functionality.

However, while different there are a number of clear connections between the two notions of information granule. Gärdenfors [3] introduces *conceptual spaces* as metric spaces of sensory inputs in which the extensions of concepts correspond to convex regions. Certainly from this perspective elements within the extension of a concept are indeed likely to be linked in terms of their similarity and proximity to one another. Also the functionality of an object can directly inference the way that it is labelled or classified. For example. the labelling of parts of the face as nose, mouth, ear etc is, as noted by Zadeh [20], significantly dependant on their respective functions.

Label semantics [5], [6] is a representation framework to encode the conventions for allocating labels and compound expressions generated from labels, as descriptions of elements from the underlying domain. As such it provides a useful tool for granular modelling when formulated as above with an emphasis on the association of points and labels. The notion of vagueness is also closely related to that of information granularity in that for most examples of information processing in natural language the information granules are not precisely defined. Indeed this semantic imprecision can often result in more flexible and robust granular models. Label semantics is based on an epistemic theory of vagueness [18] according to which the individual agents involved in communication believe in the existence of language conventions shared across the population of communicators but are (typically) uncertain as to which of the available labels can be appropriately used to describe any given instance.

2 Underlying Philosophy of Vagueness

In contrast to fuzzy set theory [19], *label semantics* encodes the meaning of linguistic labels according to how they are used by a population of communicating agents to convey information. From this perspective, the focus is on the decision making process an intelligent agent must go through in order to identify which labels or expressions can actually be used to describe an object or value. In other words, in order to make an assertion describing an object in terms of some set of linguistic labels, an agent must first identify which of these labels are appropriate or assertible in this context. Given the way that individuals learn language through an ongoing process of interaction with the other communicating agents and with the environment, then we can expect there to be considerable uncertainty associated with any decisions of this kind. Furthermore, there is a subtle assumption central to the label semantic model, that such decisions regarding appropriateness or assertibility are meaningful. For instance, the fuzzy logic view is that vague descriptions like 'John is *tall*' are generally only partially true and hence it is not meaningful to consider which of a set of given labels can truthfully be used to described John's height. However, we contest that the efficacy of natural language as a means of conveying information between members of a population lies in shared conventions governing the appropriate use of words which are, at least loosely, adhere to by individuals within the population.

In our everyday use of language we are continually faced with decisions about the best way to describe objects and instances in order to convey the information we intend. For example, suppose you are witness to a robbery, how should you describe the robber so that police on patrol in the streets will have the best chance of spotting him? You will have certain labels that can be applied, for example *tall, short, medium, fat, thin, blonde, etc*, some of which you may view as inappropriate for the robber, others perhaps you think are definitely appropriate while for some labels you are uncertain whether they are appropriate or not. On the other hand, perhaps you have some ordered preferences between labels so that *tall* is more appropriate than *medium* which is in turn more appropriate than *short*. Your choice of words to describe the robber should surely then be based on these judgements about the appropriateness of labels. Yet where does this knowledge come from and more fundamentally what does it actually mean to say that a label is or is not appropriate? Label semantics proposes an interpretation of vague description labels based on a particular notion of appropriateness and suggests a measure of subjective uncertainty resulting from an agent's partial knowledge about what labels are appropriate to assert. Furthermore, it is suggested that the vagueness of these description labels lies fundamentally in the uncertainty about if and when they are appropriate as governed by the rules and conventions of language use. The underlying assumption here is that some things can be correctly asserted while others cannot. Exactly where the dividing line lies between those labels that are and those that are not appropriate to use may be uncertain, but the assumption that such a division exists would be a natural precursor to any decision making process of the kind just described.

The above argument is very close to the epistemic view of vagueness as expounded by Timothy Williamson [18]. Williamson assumes that for the extensions of a vague concept there is a precise but unknown dividing boundary between it and the extension of the negation of that concept. However, while there are marked similarities between the epistemic theory and the label semantics view, there are also some subtle differences. For instance, the epistemic view would seem to assume the existence of some objectively correct, but unknown, definition of a vague concept. Instead of this we argue that individuals when faced with decision problems regarding assertions find it useful as part of a decision making strategy to assume that there is a clear dividing line between those labels which are and those which are not appropriate to describe a given instance. We refer to this strategic assumption across a population of communicating agents as the *epistemic stance*, a concise statement of which is as follows:

> *Each individual agent in the population assumes the existence of a set of labelling conventions, valid across the whole population, governing what linguistic labels and expressions can be appropriately used to describe particular instances.*

In practice these rules and conventions underlying the appropriate use of labels would not be imposed by some outside authority. In fact, they may not exist at all in a formal sense. Rather they are represented as a distributed body of knowledge concerning the assertability of predicates in various cases, shared across a population of agents, and emerging as the result of interactions and communications between individual agents all adopting the epistemic stance. The idea is that the learning processes of individual agents, all sharing the fundamental aim of understanding how words can be appropriately used to communicate information, will eventually converge to some degree on a set of shared conventions. The very process of convergence then to some extent vindicates the epistemic stance from the perspective of individual agents. Of course, this is not to suggest complete or even extensive agreement between individuals as to these appropriateness conventions. However, the overlap between agents should be sufficient to ensure the effective transfer of useful information.

A further distinction between our view of appropriateness and the epistemic view of Williamson can be found in the local, or instance-based, nature of appropriateness judgements. Arguments in favour of the epistemic view concern the existence of a precise boundary between the extension of a concept and that of its negation. The appropriateness of labels on the other hand is judged with reference to a particular instance. From this perspective it is unlikely that agents would generate an explicit representation of the extension of a vague concept. Instead their knowledge would be based on previous experience of assertions about similar instances from a range of other agents and a subsequent process of interpolation between these examples. In most cases decision problems about assertions would then typically concern a particular instance, so that the problem of identifying concept boundaries would not be directly considered.

The epistemic stance allows agents to meaningfully apply epistemic models of uncertainty to quantify their subjective belief in whether certain labels are appropriate. In the sequel we will introduce two related probabilistic measures of an agent's uncertainty concerning the appropriateness of vague expressions and explore the resulting calculus.

3 Label Semantics

Label semantics proposes two fundamental and inter-related measures of the appropriateness of labels as descriptions of an object or value. We begin by assuming that for all agents there is a fixed shared vocabulary in the form of a finite set of basic labels LA for describing elements from the underlying universe Ω. These are building blocks for more complex compound expressions which can then also be used as descriptors as follows. A countably infinite set of expressions LE can be generated through recursive applications of logical connectives to the basic labels in LA. So for example, if Ω is the set of all possible rgb values and LA is the set of basic colour labels such as *red*, *yellow* ,*green*, *orange* etc then LE contains those compound expressions such as *red & yellow*, *not blue nor orange* etc. The measure of appropriateness of an expression $\theta \in LE$ as a description of instance x is denoted by $\mu_\theta(x)$ and quantifies the agent's subjective probability that θ can be appropriately used to describe x. From an alternative perspective, when faced with describing instance x, an agent may consider each label in LA and attempt to identify the subset of labels that are appropriate to use. This is a totally meaningful endeavour for agents who adopt the epistemic stance. Let this complete set of appropriate labels for x be denote by $\mathcal{D}_x$. In the face of their uncertainty regarding labelling conventions agents will also be uncertain as to the composition of $\mathcal{D}_x$, and we represent this uncertainty with a probability mass function $m_x : 2^{LA} \rightarrow [0, 1]$ defined on subsets of labels. Hence, for the subset of labels $\{red, orange, yellow\}$ and rgb value x, $m_x(\{red, orange, yellow\})$ denotes the subjective probability that $\mathcal{D}_x = \{red, orange, yellow\}$, or in other words that $\{red, orange, yellow\}$ is the complete set of basic colour labels with which it is appropriate to describe x. We now provide formal definitions for the set of expressions LE and for mass functions m_x, following which we will propose a link between the two measures $\mu_\theta(x)$ and m_x for expression $\theta \in LE$.

Definition 1. *Label Expressions*
The set of label expressions LE generated from LA, is defined recursively as follows:

- *If* $L \in LA$ *then* $L \in LE$
- *If* $\theta, \varphi \in LE$ *then* $\neg\theta, \theta \wedge \varphi, \theta \vee \varphi \in LE$.

Definition 2. *Mass Function on Labels*
$\forall x \in \Omega$ *a mass function on labels is a function* $m_x : 2^{LA} \rightarrow [0, 1]$ *such that* $\sum_{S \subseteq LA} m_x(S) = 1$.

Note that there is no requirement for the mass associated with the empty set to be zero. Instead, $m_x(\emptyset)$ quantifies the agent's belief that none of the labels are appropriate to describe x. We might observe that this phenomena occurs frequently in natural language, especially when labelling perceptions generated along some continuum. For example, we occasionally encounter colours for which none of our available colour descriptors seem appropriate. Hence, the value $m_x(\emptyset)$ is an indicator of the describability of x in terms of the labels LA.

Now depending on labelling conventions there may be certain combinations of labels which cannot all be appropriate to describe any object. For example, *small* and *large* cannot both be appropriate. This restricts the possible values of $\mathcal{D}_x$ to the following set of focal elements:

Definition 3. *Set of Focal Elements*
Given labels LA together with associated mass assignment $m_x : \forall x \in \Omega$, the set of focal elements for LA is given by $\mathcal{F} = \{S \subseteq LA : \exists x \in \Omega,\ m_x(S) > 0\}$.

The link between the mass function m_x and the appropriateness measures $\mu_\theta(x)$ is motivated by the intuition that the assertion 'x is θ' directly provides information dependent on θ, as to what are the possible values for $\mathcal{D}_x$. For example, the assertion 'x is *blue*' would mean that *blue* is an appropriate label for x, from which we can infer that $blue \in \mathcal{D}_x$. Similarly, the assertion 'x is *green and not blue*' would mean that *green* is an appropriate label for x while *blue* is not, so that we can infer $green \in \mathcal{D}_x$ and $blue \notin \mathcal{D}_x$. Another way of expressing this information is to say that $\mathcal{D}_x$ must be a member of the set of sets of labels which contain *green* but do not contain *blue* i.e. $\mathcal{D}_x \in \{S \subseteq LA : green \in S,\ blue \notin S\}$. More generally, we can define a functional mapping λ from LE into $2^{2^{LA}}$ (i.e. the set containing all possible sets of label sets) for which the assertion 'x is θ' enables us to infer that $\mathcal{D}_x \in \lambda(\theta)$. This mapping is defined recursively as follows:

Definition 4. *λ-mapping*
$\lambda : LE \to 2^{\mathcal{F}}$ is defined recursively as follows: $\forall \theta,\ \varphi \in LE$

- $\forall L \in LA\ \lambda(L) = \{S \in \mathcal{F} : L \in S\}$
- $\lambda(\theta \wedge \varphi) = \lambda(\theta) \cap \lambda(\varphi)$
- $\lambda(\theta \vee \varphi) = \lambda(\theta) \cup \lambda(\varphi)$
- $\lambda(\neg\theta) = \lambda(\theta)^c$.

The λ-mapping then provides us with a means of evaluating the appropriateness measure of an expression θ directly from m_x, as corresponding to the subjective probability that $\mathcal{D}_x \in \lambda(\theta)$ so that:

Definition 5. *Appropriateness Measures*
For any expression $\theta \in LE$ and $x \in \Omega$, the appropriateness measure $\mu_\theta(x)$ can be determined from the mass function m_x according to:

$$\forall \theta \in LE\ \mu_\theta(x) = \sum_{S \in \lambda(\theta)} m_x(S).$$

From this relationship the following list of general properties hold for expressions θ and φ in LE [5]:

Theorem 1. *Lawry [5],[6]*

- *If $\theta \models \varphi$ then $\forall x \in \Omega\ \mu_\theta(x) \leq \mu_\varphi(x)$*
- *If $\theta \equiv \varphi$ then $\forall x \in \Omega\ \mu_\theta(x) = \mu_\varphi(x)$*
- *If θ is a tautology then $\forall x \in \Omega\ \mu_\theta(x) = 1$*
- *If θ is a contradiction then $\forall x \in \Omega\ \mu_\theta(x) = 0$*
- *$\forall x \in \Omega\ \mu_{\neg\theta}(x) = 1 - \mu_\theta(x)$.*

Notice, here that the laws of excluded middle and non-contradiction are preserved since for any expression θ, $\lambda(\theta \vee \neg\theta) = \lambda(\theta) \cup \lambda(\theta)^c = 2^{2^{LA}}$ and

$\lambda(\theta \wedge \neg\theta) = \lambda(\theta) \cap \lambda(\theta)^c = \emptyset$. Also, the idempotent condition holds since $\lambda(\theta \wedge \theta) = \lambda(\theta) \cap \lambda(\theta) = \lambda(\theta)$.

The λ-mapping provides us with a clear formal representation for linguistic constraints, where the imprecise constraint 'x is θ' on x is interpreted as the precise constraint $\mathcal{D}_x \in \lambda(\theta)$ on $\mathcal{D}_x$.

3.1 Ordering Labels

As discussed above an agent's estimation of both m_x and $\mu_\theta(x)$ should depend on their experience of language use involving examples similar to x. Clearly the form of this knowledge is likely to be both varied and complex. However, one natural type of assessment for an agent to make would be to order or rank label in terms of their estimated appropriateness for x. This order information could then be combined with estimates of appropriateness measure values for the basic labels (i.e. elements of LA) in order to provide estimates of values for compound expressions (i.e. elements of LE). Hence we assume that:

An agent's knowledge of label appropriateness for an instance x, can be represented by an ordering on the basic labels LA and an allocation of uncertainty values to the labels consistent with this ordering.

Effectively we are assuming that through a process of extrapolation from experience agents are, for a given instance, able to (at least partially) rank labels in terms of their appropriateness and then, consistent with this ranking, to estimate a subjective probability that each label is appropriate. On the basis of both the ordering and probability assignment to basic labels the agent should then be able to evaluate the appropriateness measure of more complex compound expressions. The ranking of available labels would seem to be an intuitive first step for an agent to take when faced with the decision problem about what to assert. Also, the direct allocation of probabilities to a range of complex compound expressions so that the values are internally consistent is a fundamentally difficult task. Hence, restricting such evaluations to only the basic labels would have significant practical advantages in terms of computational complexity.

Definition 6. *(Ordering on Labels)*
For $x \in \Omega$ let $\preceq_x$ be an ordering on LA such that for $L, L' \in LA$, $L' \preceq_x L$ means that L is at least as appropriate as a label for x as L'.

The identification by an agent of an ordering on labels $\preceq_x$ for a particular $x \in \Omega$ (as in definition 6), restricts the possible label sets which they can then consistently allocate to $\mathcal{D}_x$. For instance, $L' \preceq_x L$ then this implies that if $L' \in \mathcal{D}_x$ then so is $L \in \mathcal{D}_x$, since L is as least as appropriate a description for x as L'. Hence, given $\preceq_x$ for which $L' \preceq_x L$ it must hold that $m_x(S) = 0$ for all $S \subseteq LA$ where $L' \in S$ and $L \notin S$. Trivially, from definition 5 this also means that $\mu_{L'}(x) \leq \mu_L(x)$. Given these observations an important question is whether the information provided by ordering $\preceq_x$ together with a set of appropriateness values $\mu_L(x) : L \in LA$ for the basic labels, consistent with $\preceq_x$, is sufficiently to specify a unique mass function m_x? Notice that in the label semantics framework the identification of a unique mass function m_x in this way immediately enables the agent to apply definition 5 in order to evaluate the appropriateness $\mu_\theta(x)$ of any compound expression θ from the appropriateness measure values for the basis labels. In fact, in the case that $\preceq_x$ is a total (linear) ordering it is not difficult to see that such a unique mapping does indeed exist between the mass function and the appropriateness measures of basic labels. To see this suppose that we index the labels in LA so that $L_n \preceq_x L_{n-1} \preceq_x \ldots \preceq_x L_1$ with corresponding appropriateness measures $\mu_{L_n}(x) = a_n \leq \mu_{L_{n-1}}(x) = a_{n-1} \leq \ldots \leq \mu_{L_1}(x) = a_1$. Now from the above discussion we have that in this case the only possible values for $\mathcal{D}_x$ are from the nested sequence of sets $\emptyset, \{L_1\}, \{L_1, L_2\}, \ldots, \{L_1, \ldots, L_i\}, \ldots, \{L_1, \ldots, L_n\}$. This together with the constraints imposed by definition 5 that for each label $a_i = \mu_{L_i}(x) = \sum_{S:L_i \in S} m_x(S)$ results in the following unique mass function:

$$m_x := \{L_1, \ldots, L_n\} : a_n, \ldots, \{L_1, \ldots, L_i\} : a_i - a_{i+1},$$
$$\ldots, \{L_1\} : a_1 - a_2, \emptyset : 1 - a_1$$

Hence, for $\preceq_x$ a total ordering we see that $\mu_\theta(x)$ can be determined as a function of the appropriateness measure values $\mu_L(x) : L \in LA$ on the basic labels. For an expression $\theta \in LE$, this function is a composition of the above mapping, in order to determine a unique mass function, and the consequent summing of mass function values across $\lambda(\theta)$, as given in definition 5, to evaluate $\mu_\theta(x)$. Although functional in this case, the calculus for appropriateness measures cannot be truth-functional in the sense of fuzzy logic since appropriateness measures satisfy all the classical Boolean laws and a well known result due to Dubois and Prade [2] shows that no truth-functional calculus can in general preserve all such laws. For a more detailed discussion of the difference between functionality and truth-functionality see Lawry [6]. The following theorem shows that in the case where $\preceq_x$ is a total ordering the max and min combination rules can be applied in certain restricted cases:

Theorem 2. *[5, 16]*
Let $LE^{\wedge,\vee} \subseteq LE$ denote those expressions generated recursively from LA using only the connectives $\wedge$ and $\vee$. If the appropriateness measures on basic labels are consistent with a total ordering $\preceq_x$ on LA then $\forall \theta, \varphi \in LE^{\wedge,\vee}$ it holds that:

$$\mu_{\theta\wedge\varphi}(x) = \min(\mu_\theta(x), \mu_\varphi(x)), \ \mu_{\theta\vee\varphi}(x) = \max(\mu_\theta(x), \mu_\varphi(x)).$$

In the case that $\preceq_x$ is only a partial ordering on LA then in general this does not provide the agent with sufficient information to determine a unique mass function from the appropriateness measure values on the basic labels. Instead, further information is required for the agent to evaluate a mass function and consequently the appropriateness of compound label expressions. In Lawry [7] it is proposed that this additional information takes the form of conditional independence constraints imposed by a Bayesian network generated by $\preceq_x$. These additional assumptions are then sufficient to determine m_x uniquely. Details of this approach, however, are beyond the scope of this current paper. Instead in the examples presented in the sequel we will assume that the ordering $\preceq_x$ is total.

4 Granular Models in Label Semantics

In label semantics information granules correspond to appropriateness measure for a fixed expression i.e. an information granule is a function $\mu_\theta : \Omega \rightarrow [0,1]$ for some label expression $\theta \in LE$. In fact, within the scope of this definition we can also use mass functions to represent information granules. Specifically, for a fixed focal set $F \subseteq LA$ an information granule may also be represented by the function corresponding to the values of $m_x(F)$ as x varies across Ω. To see this notice that the value $m_x(F)$ can also be represented by the appropriateness measure $\mu_{\alpha_F}(x)$ where $\alpha_F = (\bigwedge_{L \in F} L) \wedge (\bigwedge_{L \notin F} \neg L)$ is the label expression stating that all and only the labels in F are appropriate. Hence, for a focal set $F \subseteq LA$ the corresponding information granule is the function $\mu_{\alpha_F} : \Omega \rightarrow [0,1]$. For example, in figure 1 information granules are defined in terms of the appropriateness measures for labels *low*, *medium* and *high*, represented by trapezoidal functions of $x \in \Omega = [0,30]$. Assuming a total ordering $\preceq_x$ on labels for all $x \in \Omega$ results in mass functions m_x for the focal sets $\mathcal{F} = \{\{l\},\{l,m\},\{m\},\{m,h\},\{h\}\}$, shown as triangular functions in figure 2. These triangular functions then correspond to the information granules generated by the focal sets in $\mathcal{F}$. The direct use of focal sets as information granules in granular models can in some cases allow for more straightforward information processing. In particular, note that the mass function m_x defines a probability distribution on $\mathcal{D}_x$ which can in turn make it relatively straightforward to evaluate probability values from a granular model based on such functions.

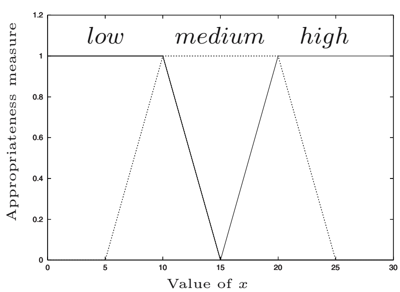

Fig. 1. Appropriateness measure values for labels *low*, *medium*, *high* viewed as a function of x as x varies across $\Omega = [0,30]$

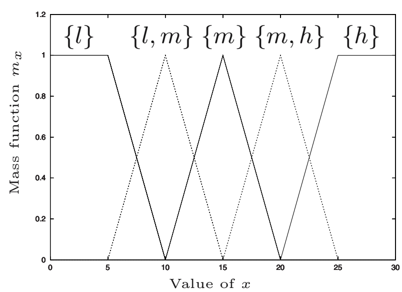

Fig. 2. Mass function values for the sets $\{low\}$, $\{low, medium\}$, $\{medium\}$, $\{medium, high\}$, $\{high\}$ viewed as a function of x as x varies across $\Omega = [0,30]$

Consider the following formalization of a simple modelling problem: Given attributes $x_1, \ldots, x_{k+1}$ with universes $\Omega_1, \ldots, \Omega_{k+1}$ suppose that x_{k+1} is dependent on $x_1, \ldots, x_k$ according to some functional mapping $g : \Omega_1 \times \ldots \times \Omega_k \to \Omega_{k+1}$ (i.e. $x_{k+1} = g(x_1, \ldots, x_k)$). In the case that Ω_{k+1} is finite then this is referred to as a classification problem whereas if Ω_{k+1} is an infinite subset of $\mathbb{R}$ (typically a closed interval) then it is referred to as a prediction or regression problem. For a learning problem, information regarding this function is then provided by a training database containing vectors of input values together with their associated output. Let this database be denoted by $DB = \{\langle x_1(i), \ldots, x_k(i), x_{k+1}(i)\rangle : i = 1, \ldots, N\}$. For a more general modelling problem information on g may take a variety of forms including qualitative information elicited from domain experts.

Label semantics can be used to represent linguistic rule based models which provide an approximation $\hat{g}$ of the underlying function mapping g. Here we consider two such models; mass relations and linguistic decision trees. For both these approaches we use appropriateness measures to define a set of labels describing each attribute $LA_j : j = 1, \ldots, k+1$ with associated label expressions $LE_j : j = 1, \ldots, k+1$ and focal sets $\mathcal{F}_j : j = 1, \ldots, k+1$. We will also describe how these models can be used within a hierarchical structure to provide a decomposed model for high-dimensional mappings.

4.1 Mass Relational Models

If we consider the problem of describing an object or instance on the basis of k attributes $x_1, \ldots, x_k$ then we need to jointly quantify the appropriateness of labels in each of the associated labels sets $LA_j : j = 1, \ldots, k$ to describe each attribute. In other words, we need to define a joint mass function on $\mathcal{D}_{x_1} \times \ldots \times \mathcal{D}_{x_k}$ mapping from $\mathcal{F}_1 \times \ldots \times \mathcal{F}_k$ into $[0, 1]$. We refer to such joint mass functions on label sets as a mass relations. Mass relations can be used to represent a granular model of the function g. Typically, this is achieved by defining a mass relation between input focal sets conditional on each of the output focal sets in $\mathcal{F}_{k+1}$. Together these can then be used to infer a mass functions on output focal sets given a vector of input attribute values.

Definition 7. *Mass Relations*
A mass relation is a conditional function $m : \mathcal{F}_1 \times \ldots \times \mathcal{F}_k \to [0, 1]$ *such that for* $F_i \in \mathcal{F}_i : i = 1, \ldots, k+1$, $m(F_1, \ldots, F_k | F_{k+1})$ *is the conditional joint mass function value of the input focal sets* $F_1, \ldots, F_k$ *given output focal set* F_{k+1}. *This can be evaluated from a database* DB *according to:*

$$m(F_1, \ldots, F_k | F_{k+1}) = \frac{\sum_{i \in DB} \prod_{j=1}^{k+1} m_{x_j(i)}(F_j)}{\sum_{i \in DB} m_{x_{k+1}(i)}(F_{k+1})}$$

A set of mass relations conditional on each of the output focal sets in $\mathcal{F}_{k+1}$ *generates a set of weighted rules of the form:*

$$(\mathcal{D}_{x_1} = F_1) \wedge \ldots \wedge (\mathcal{D}_{x_k} = F_k) \rightarrow (\mathcal{D}_{x_{k+1}} = F_{k+1}) : w \ \textit{where}$$
$$w = m(F_{k+1}|F_1, \ldots, F_k) = \frac{m(F_1, \ldots, F_k|F_{k+1})m(F_{k+1})}{\sum_{F_{k+1}} m(F_1, \ldots, F_k|F_{k+1})m(F_{k+1})}$$
$$\textit{and } m(F_{k+1}) = \frac{1}{N} \sum_{i \in DB} m_{x_{k+1}(i)}(F_{k+1}).$$

Given a vector of input values $\mathbf{x} = \langle x_1, \ldots, x_k \rangle$ we can use Jeffrey's rule [4] to determine a mass function on the output focal sets $\mathcal{F}_{k+1}$ from a mass relation between $\mathcal{F}_1 \times \ldots \times \mathcal{F}_k$ and $\mathcal{F}_{k+1}$, as follows:

$$\forall F_{k+1} \in \mathcal{F}_{k+1} \ m(F_{k+1}|\mathbf{x}) =$$
$$\sum_{F_1 \in \mathcal{F}_1} \cdots \sum_{F_k \in \mathcal{F}_k} m(F_{k+1}|F_1, \ldots, F_k) m_{\mathbf{x}}(F_1, \ldots, F_k) \text{ where}$$
$$m_{\mathbf{x}}(F_1, \ldots, F_k) = \prod_{i=1}^{k} m_{x_i}(F_i)$$

In practice it can be computationally expensive to calculate the mass relation exactly and typically we need to use some form of approximation. One of the simplest is to assume conditional independence between $\mathcal{D}_{x_1}, \ldots \mathcal{D}_{x_k}$ given the values of $\mathcal{D}_{x_{k+1}}$. In this case:

$$m(F_1, \ldots, F_k|F_{k+1}) = \prod_{i=1}^{k} m(F_i|F_{k+1})$$

An extension to this approach involves searching for dependency groupings amongst the attributes and assume conditional independence (given F_{k+1}) between these groups (see [14] for details).

Recent applications of mass relations have focussed on the area of flood prediction where they have been used to model river flow [13] and also tidal surges up to the Thames barrier in London [15].

4.2 Linguistic Decision Trees

A linguistic decision tree is a decision tree with attributes as nodes and linguistic descriptions of attributes as branches. Also associated with each branch, there is a mass function over the output focal sets.

Definition 8. *Linguistic Decision Trees (LDT)*
A linguistic decision tree is a decision tree where the nodes are attributes from $x_1, \ldots, x_k$ and the edges are label expressions describing each attribute. More formally, supposing that the j'th node at depth d is the attribute x_{j_d} then there is a set of label expressions $\mathcal{L}_{j,d} \subseteq LE_i$ forming the edges from x_{j_d} such that $\lambda(\bigvee_{\theta \in \mathcal{L}_{j,d}} \theta) \supseteq \mathcal{F}_{j_d}$ and $\forall \theta, \varphi \in \mathcal{L}_{j,d}\ \lambda(\theta \wedge \varphi) \cap \mathcal{F}_{j_d} = \emptyset$. Also a branch B from a LDT consists of a sequence of expressions $\varphi_1, \ldots, \varphi_m$ where $\varphi_d \in \mathcal{L}_{j,d}$ for some

$j \in \mathbb{N}$ for $d = 1, \ldots, m$, augmented by a conditional mass value $m(F_{k+1}|B)$ for every output focal set $F_{k+1} \in \mathcal{F}_{k+1}$. Hence, every branch B encodes a set of weighted linguistic rules of the form:

$$(x_{j_1} \textit{ is } \varphi_1) \wedge \ldots \wedge (x_{j_m} \textit{ is } \varphi_m) \rightarrow (\mathcal{D}_{x_{k+1}} = F_{k+1}) : m(F_{k+1}|B)$$

where x_{j_d} is a depth d attribute node.

Also the mass assignment value $m(F_{k+1}|B)$ can be determined from DB according to:

$$m(F_{k+1}|B) = \frac{\sum_{i \in DB} m_{x_{k+1}(i)}(F_{k+1}) \prod_{d=1}^{m} \mu_{\varphi_d}(x_{j_d}(i))}{\sum_{i \in DB} \prod_{d=1}^{m} \mu_{\varphi_d}(x_{j_d}(i))}$$

Notice that a branch of a linguistic decision tree can be rewritten using the λ-function so that it refers only to constraints on $\mathcal{D}_{x_i} : i = 1, \ldots, k$. This means that the rules generated by LDT branches are a more general form of the rules generated by mass relations. For example, the branch rule

$$(x_{j_1} \text{ is } \varphi_1) \wedge \ldots \wedge (x_{j_m} \text{ is } \varphi_m) \rightarrow (\mathcal{D}_{x_{k+1}} = F_{k+1}) : m(F_{k+1}|B)$$

can be rewritten as

$$(\mathcal{D}_{x_{j_1}} \in \lambda(\varphi_1)) \wedge \ldots \wedge (\mathcal{D}_{x_{j_m}} \in \lambda(\varphi_m)) \rightarrow (\mathcal{D}_{x_{k+1}} = F_{k+1}) : m(F_{k+1}|B)$$

Given a vector of input attribute values $\mathbf{x} = \langle x_1, \ldots, x_k \rangle$ we can use a LDT to determine a mass function on output focal sets as follows: Suppose the LDT has branches $B_1, \ldots, B_t$ each with an associated mass function $m(\bullet|B_j) : j = 1, \ldots, t$

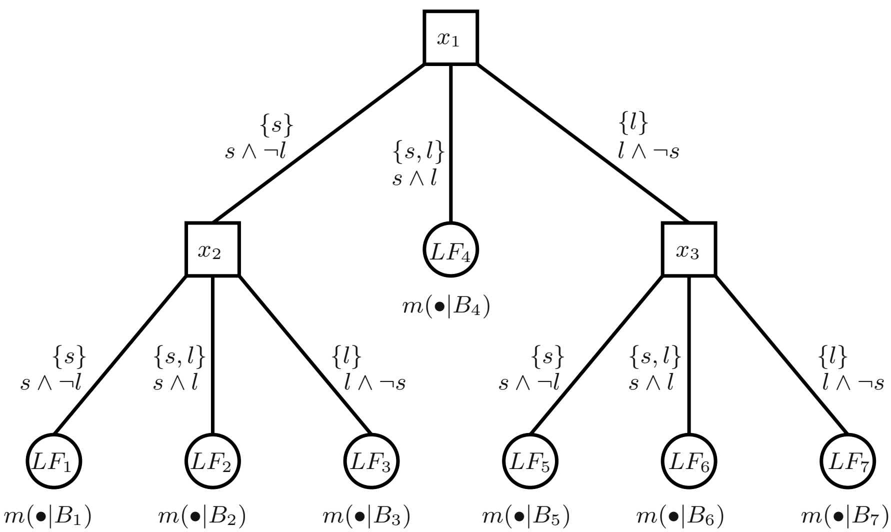

Fig. 3. An example of a linguistic decision tree

on $\mathcal{F}_{k+1}$. By applying Jeffrey's rule we obtain an aggregated mass function on $\mathcal{F}_{k+1}$ for a given input vector $\mathbf{x}$ according to:

$$m(F_{k+1}|\mathbf{x}) = \sum_{j=1}^{t} m(F_{k+1}|B_j)P(B_j|\mathbf{x}) \text{ where if}$$

$$B_j = (x_{j_1} \text{ is } \varphi_1) \wedge \ldots \wedge (x_{j_m} \text{ is } \varphi_m) \text{ then } P(B_j|\mathbf{x}) = \prod_{d=1}^{m} \mu_{\varphi_d}(x_{j_d})$$

Notice that from a computational viewpoint the output mass function above is determined from the LDT as a function of the input masses $m_{x_1}, \ldots, m_{x_k}$ only. From this perspective a LDT can be viewed as a function mapping from mass functions on the input attribute labels to a mass function on the output labels.

$$m(\bullet|\mathbf{x}) = LDT(m_{x_1}, \ldots, m_{x_k})$$

Figure 3 shows an example of a simple linguistic decision tree involving only three input attributes x_1, x_2 and x_3. Each of these is described by the same set of two overlapping labels $LA = \{small, large\}$, so that we have focal sets $\{\{s\}, \{s, l\}, \{l\}\}$. Each branch in figure 3 is labelled with its associated linguistic expressions together with their corresponding representation in terms of focal sets. In this case all the linguistic expressions involved are atomic in nature so that their λ-mappings contain only one focal set. For example, the branch B_2 ending in the leaf node LF_2 encodes the follow rules: For all $F_{k+1} \in \mathcal{F}_{k+1}$,

$$(x_1 \text{ is } s \wedge \neg l) \wedge (x_2 \text{ is } s \wedge l) \rightarrow (\mathcal{D}_{x_{k+1}} = F_{k+1}) : m(F_{k+1}|B_2)$$

Now $\lambda(s \wedge \neg l) = \{\{s\}\}$ and $\lambda(s \wedge l) = \{\{s, l\}\}$ and hence the above rule has the following focal set representation:

$$(\mathcal{D}_{x_1} = \{s\}) \wedge (\mathcal{D}_{x_2} = \{s, l\}) \rightarrow (\mathcal{D}_{x_{k+1}} = F_{k+1}) : m(F_{k+1}|B_2)$$

Linguistic decision trees can be learnt from data using the LID3 algorithm [12]. This is an extension of ID3 to allow for the type of calculations on mass functions required for a LDT. Recent application of LID3 include classification of radar images [9] and online path planning [17].

4.3 Linguistic Attribute Hierarchies

In many cases the function g is complex and it is difficult to define $\hat{g}$ as a direct mapping between $x_1, \ldots, x_k$ and x_{k+1}. Attribute hierarchies [1] are a well known approach to this problem and involve breaking down the function g into a hierarchy of sub-functions each representing a new intermediate attribute. A bottom-up description of this process is as follows: The set of original attributes $\{x_1, \ldots, x_k\}$ are partitioned into attribute subsets $S_1, \ldots, S_m$ and new attributes $z_1, \ldots z_m$ are defined as functions of each partition set respectively, so that $z_i = G_i(S_i)$ for $i = 1, \ldots, m$. The function g is then defined as a new function F of

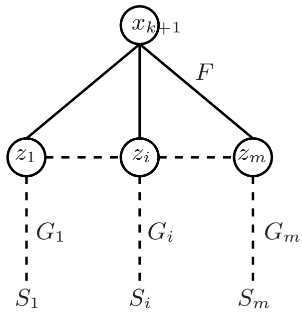

Fig. 4. Attribute hierarchy showing partition of attributes

the new attributes $z_1, \ldots z_m$, so that $x_{k+1} = g(x_1, \ldots, x_k) = F(z_1, \ldots, z_m) = F(G_1(S_1), \ldots, F(G_m(S_m))$. The same process can then be repeated recursively for each partition set S_i, to generate a new layer of new variables as required.

The identification of attribute hierarchies and their associated functional mappings is often a highly subjective process involving significant uncertainty and imprecision. Hence, the relationship between certain levels in the hierarchy, can best be described in terms of linguistic rules and relations. This can allow for judgements and rankings to be made at a level of granularity appropriate to the level of precision at which the functional mappings can be realistically defined. In linguistic attribute hierarchies the functional mappings between parent and child attribute nodes in the attribute hierarchy are defined in terms of weighted linguistic rules (typically linguistic decision trees) which explicitly model both the uncertainty and vagueness which often characterises our knowledge of such aggregation functions.

In linguistic attribute hierarchies the functional relationship between child and parent nodes are not defined precisely. Instead the labels for a parent attribute are defined in terms of the labels describing the attributes corresponding to its child nodes, by means of a linguistic decision tree. To illustrate this idea consider the following simple linguistic attribute hierarchy as shown in figure 5. Here we have 4 input attributes $x_1, \ldots, x_4$ and output attribute x_5, these being described by label sets $LA_1, \ldots, LA_5$ with focal sets $\mathcal{F}_1, \ldots, \mathcal{F}_5$ respectively. The labels for x_5 are defined in terms of the labels for two intermediate level attributes z_1 and z_2 by a linguistic decision tree LDT_1. Let LA_{z_1}, LA_{z_2} and $\mathcal{F}_{z_1}$, $\mathcal{F}_{z_2}$ be the labels and focal sets for z_1 and z_2 respectively. Furthermore, the labels for z_1 are defined in terms of those for x_1 and x_2 according to linguistic decision tree LDT_2, and the labels for z_2 are defined in terms of those for x_3 and x_4 according to linguistic decision tree LDT_3. Information is then propagated up through the hierarchy as mass functions on the relevant focal sets. Specifically, LDT_2 combines mass functions on $\mathcal{F}_1$ and $\mathcal{F}_2$ in order to generate a mass function on $\mathcal{F}_{z_1}$. Similarly mass functions on $\mathcal{F}_3$ and $\mathcal{F}_4$ are combined using LDT_3 to generate a mass function on $\mathcal{F}_{z_2}$. These two mass functions on $\mathcal{F}_{z_1}$ and $\mathcal{F}_{z_2}$ respectively are then combined according to LDT_1 in order to obtain a mass assignment on the output

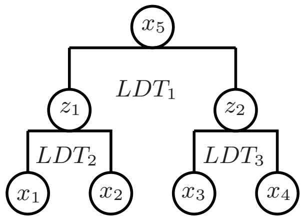

Fig. 5. Example of a simple linguistic attribute hierarchy

focal sets $\mathcal{F}_5$ conditional on the inputs. At the mass function level the complete mapping is:

$$m(\bullet|x_1,\ldots,x_4) = LDT_1(LDT_2(m_{x_1}, m_{x_2}), LDT_3(m_{x_3}, m_{x_4}))$$

5 Conclusions

An alternative perspective on granular modelling has been introduced where information granules encode the relationship between description labels and the underlying perceptual domain. Label semantics has been introduced as a framework for modelling linguistic vagueness and granularity. In this framework information granules are appropriateness measures and mass functions which quantify the appropriateness of label expressions to describe elements from the underlying universe. Two types of granular models have been described; mass relations and linguistic decision trees. These encode the relationship between labels on input values and those on the output values in an imprecisely defined functional mapping. In addition, we have shown how linguistic decision trees can be used as part of attribute hierarchies to combine information in complex multi-level composite mappings.

References

1. Bohanec, M., Zupan, B.: A Function-Decomposition Method for Development of Hierarchical Multi-Attribute Decision Models. Decision Support Systems 36, 215–223 (2004)
2. Dubois, D., Prade, H.: An Introduction to Possibility and Fuzzy Logics. In: Smets, P., et al. (eds.) Non-Standard Logics for Automated Reasoning, pp. 742–755. Academic Press, London (1988)
3. Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought. MIT Press, Cambridge (2000)
4. Jeffrey, R.C.: The Logic of Decision. Gordon and Breach, New York (1965)
5. Lawry, J.: A framework for linguistic modelling. Artificial Intelligence 155, 1–39 (2004)
6. Lawry, J.: Modelling and Reasoning with Vague Concepts. Springer, Heidelberg (2006)

7. Lawry, J.: Appropriateness measures: An uncertainty model for vague concepts, Synthese (to appear, 2007)
8. Lawry, J., He, H.: Linguistic Attribute Hierarchies for Multiple-Attribute Decision Making. In: Proceedings 2007 IEEE International Conference on Fuzzy Systems FUZZ-IEEE 2007 (2007)
9. McCulloch, D.R., et al.: Classification of Weather Radar Images using Linguistic Decision Trees with Conditional Labelling. In: Proceedings 2007 IEEE International Conference on Fuzzy Systems FUZZ-IEEE 2007 (2007)
10. Nguyen, H.T.: On Modeling of linguistic Information using Random Sets. Information Science 34, 265–274 (1984)
11. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishing, Dordrecht (1991)
12. Qin, Z., Lawry, J.: Decision Tree Learning with Fuzzy Labels. Information Sciences 172, 91–129 (2005)
13. Randon, N., et al.: River Flow Modelling Based on Fuzzy Labels. In: Information Processing and Management of Uncertainty IPMU 2004 (July 2004)
14. Randon, N., Lawry, J.: Classification and Query Evaluation using Modelling with Words. Information Sciences 176, 438–464 (2006)
15. Randon, N., et al.: Fuzzy Bayesian Modelling of Sea-Level along the East Coast of Britain. IEEE Transaction on Fuzzy Systems (to appear, 2007)
16. Tang, Y., Zheng, J.: Linguistic Modelling based on Semantic Similarity Relation amongst Linguistic Labels. Fuzzy Sets and Systems 157, 1662–1673 (2006)
17. Turnbull, O., et al.: Fuzzy Decision Tree Cloning of Flight Trajectory Optimisation for Rapid Path Planning. In: Proceedings 45 IEEE Conference on Decision and Control (2006)
18. Williamson, T.: Vagueness, Routledge (1994)
19. Zadeh, L.A.: Fuzzy Sets. Information and Control 8(3), 338–353 (1965)
20. Zadeh, L.A.: Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems 90, 111–127 (1997)