Tolerance Spaces and Approximative Representational Structures*

Patrick Doherty[1], Witold Łukaszewicz[1,2], and Andrzej Szałas[1,2]

[1] Department of Computer Science, University of Linköping, Sweden,
patdo@ida.liu.se
[2] The College of Economics and Computer Science, Olsztyn, Poland
{witlu,andsz}@ida.liu.se

Abstract. In traditional approaches to knowledge representation, notions such as tolerance measures on data, distance between objects or individuals, and similarity measures between primitive and complex data structures are rarely considered. There is often a need to use tolerance and similarity measures in processes of data and knowledge abstraction because many complex systems which have knowledge representation components such as robots or software agents receive and process data which is incomplete, noisy, approximative and uncertain. This paper presents a framework for recursively constructing arbitrarily complex knowledge structures which may be compared for similarity, distance and approximativeness. It integrates nicely with more traditional knowledge representation techniques and attempts to bridge a gap between approximate and crisp knowledge representation. It can be viewed in part as a generalization of approximate reasoning techniques used in rough set theory. The strategy that will be used is to define tolerance and distance measures on the value sets associated with attributes or primitive data domains associated with particular applications. These tolerance and distance measures will be induced through the different levels of data and knowledge abstraction in complex representational structures. Once the tolerance and similarity measures are in place, an important structuring generalization can be made where the idea of a *tolerance space* is introduced. Use of these ideas is exemplified using two application domains related to sensor modeling and communication between agents.

1 Introduction

In traditional approaches to knowledge representation, notions such as tolerance measures on data, distance between objects or individuals, and similarity measures between primitive and complex data structures such as properties and relations, elementary and complex descriptors, decision rules, information systems, and relational databases, are rarely considered. This is unfortunate because

* Supported in part by the WITAS project grant under the Wallenberg Foundation, Sweden and KBN grant 8 T11C 009 19.

many complex systems which have knowledge representation components such as robots or software agents receive and process data which is incomplete, noisy, approximative and uncertain. There is often a need to use tolerance and similarity measures in processes of data and knowledge abstraction and in communication between agents.

This is a particular problem in the area of cognitive robotics where data input by sensors has to be fused, filtered and integrated with more traditional qualitative knowledge structures. A great many levels of knowledge abstraction and data reduction must be used as one tries to integrate newly acquired raw data with existing data which has previously been abstracted and represented explicitly in the form of more qualitative data and knowledge structures. It is also a problem for software agents on the world wide web where knowledge structures are continually required to be compared and merged and agents are obligated to communicate with each other using similar, but unidentical ontologies or vocabularies.

This paper presents a framework for recursively constructing arbitrarily complex knowledge structures which may be compared for similarity, distance and approximativeness. It integrates nicely with more traditional knowledge representation techniques and attempts to bridge a gap between approximate and crisp knowledge representation [2]. It can be viewed in part as a generalization of approximate reasoning techniques used in rough set theory [5] where an approximate relation is represented as having both an upper and lower approximation represented as classical sets and an individual in a domain of discourse has additional structure in terms of attribute/value pairs. It also has connections to recent work by Gärdenfors with conceptual spaces [4].

Ontologically, the world is viewed as consisting of individual elements with associated sets of attribute/value pairs. Each attribute has a value set and tolerance relations will be associated with each value set inducing a neighborhood relation. Arbitrarily complex data structures and representational systems are constructed recursively from the primitive notions of individual, attribute and value. Consequently, notions of tolerance and similarity can be induced through these structures via the tolerance and similarity measures placed on primitive data or value sets. For example, a set of values for each attribute associated with an individual may be viewed as a tuple. A set of one-tuples is a property, a set of k-tuples is a k-relation, sets of relations are associated with relational structures. In regard to relations, upper and lower approximations to these can be derived through use of the individual tolerance relations.

The representational structures constructed in this manner are viewed as information granules and have a great deal of representational fluidity. They can be combined, compared for tolerance and similarity, reasoned about approximatively, and often represented using traditional database techniques. The latter is especially important for integration with legacy knowledge structures and logical inferencing techniques. Use of the framework will be exemplified using two fundamentally important potential applications: sensor to symbolic data conversions and communication between software agents. The techniques are currently being applied to real-world applications in both the cognitive robotics and software agents application domains.

Let us begin with the notion of tolerance and tolerance measures. Webster's dictionary defines tolerance as "the amount of variation allowed from a standard, accuracy, etc."

For example, suppose a system receives data about an attribute a from two sources, where source one asserts that that $a = 1.04$ and source two asserts that $a = 0.98$. Depending on the context, the system might want to consider the values 1.04 and 0.98 as the same relative to some tolerance measure since their distance is only 0.06. In another application this difference may have serious repercussions on system safety, so it is important to make sure that tolerance measures are contextual and can be tuned either automatically or manually relative to the application and context at hand.

The strategy that will be used is to define tolerance and distance measures on the value sets associated with attributes or primitive data domains associated with particular applications. These tolerance and distance measures will be induced through the different levels of data and knowledge abstraction in complex representational structures. The representational structures will in some sense inherit the tolerance measures from the primitive data domains and value sets used in these structures at lower levels of abstraction and taken into account when comparing for similarity or reasoning. By defining parameterized measures of tolerance via distance measurements on values sets and primitive domains, one can cluster sets of values into tolerance neighborhoods and view the clusters as individual elements. Similarly, individuals whose identities are dependent on sets of attribute/value pairs can also be clustered into tolerance neighborhoods and viewed as indiscernible entities to a particular degree of tolerance when used in other data structures.

The basic primitive in the ideas presented is that of a tolerance function. Let's begin with a value set V and two elements $x, y \in V$. A tolerance function τ provides us with a distance measure between x and y normalized to the real interval $[0, 1]$ where the higher the value, the closer in tolerance the two elements are. Given a parameter $p \in [0, 1]$, a tolerance relation τ^p is then introduced among individuals with a threshold p which tunes the tolerance to be within a certain degree. If $\tau(x, y) \geq p$ then the pair $\langle x, y \rangle$ is in the relation τ^p. Both the tolerance function and the parameter p must be provided by a knowledge engineer or must be machine learned. One can continually refine these values.

Once this is done for individual value sets or primitive data domains, it can be generalized to tuples of values and tolerance can be measured between two tuples $\langle x_1, \ldots, x_k \rangle$ and $\langle y_1, \ldots, y_k \rangle$ using pairwise comparison of associated tolerance relations.

Given a value set V with associated tolerance measures, we can then take subsets $V_1, V_2 \subseteq V$ and induce tolerance measures and neighborhood functions on the subsets. Likewise, given a set T of k-tuples with associated tolerance measures, we can then take subsets $T_1, T_2 \subseteq T$ and induce tolerance measures and neighborhood functions on the subsets. Subsets of V can be viewed as properties or concepts and subsets of T can be viewed as k-argument relations.

These ideas can be generalized further to sets of sets and sets of sets of tuples, where the tolerance and similarity measures between these structures is induced from the primitive tolerance measures in the base value sets. Once the tolerance and similarity measures are in place, an important structuring generalization can be made where the idea of a *tolerance space* is introduced.

Given a universe U of objects in a tolerance space with the associated tolerance measures, we can provide a generalization of the notions of upper and lower approximations on sets used in rough set theory to subsets of U. The lower and upper approximations will again be induced from the particular tolerance measures provided by the tolerance space in question. Rather than using equivalence classes of individuals constructed from subsets of attributes as in rough set theory, one would work instead with neighborhoods generated from neighborhood functions of individuals.

There is an interesting connection between the idea of tolerance spaces proposed in this chapter and the work of Gärdenfors with conceptual spaces (see, e.g., [4]). Conceptual spaces are built up using multi-dimensional spaces of quality dimensions (attributes) and providing geometric constraints between these dimensions in order to model distance measures and similarity. However, we use the notion of semi-distances rather than of distances. Tolerance spaces contribute to a generalization of conceptual spaces in the sense that concepts can be generalized to approximate concepts based on tolerance measures and the geometric constraints used are less rigid than with conceptual spaces. In order to place tolerance spaces in the proper context with conceptual spaces, we define a simple version of conceptual spaces and show how tolerance spaces may be integrated in this framework.

In the remainder of the paper, the basic framework will be presented and then exemplified using two applications.

2 Conceptual Spaces

A *semi-metric space* is a pair $\langle A, \delta \rangle$, where A is a set and δ is a function

$$\delta : A \times A \longrightarrow \mathcal{R}$$

which, for all $x, y \in A$, satisfies:

$$\delta(x,y) \geq 0,\ \delta(x,x) = 0 \text{ and } \delta(x,y) = \delta(y,x).$$

Any function δ satisfying the above properties is called a *semi-metric* for A and $\delta(x,y)$ is called the *semi-distance* between x and y.

Definition 2.1. *Let U be a finite nonempty set of objects. By a* quality dimension over U *we understand any semi-metric space $\langle U, \delta \rangle$. By a* conceptual space over U *we mean any pair $\langle U, Q \rangle$, where Q is a finite set of quality dimensions over U.* ■

Quality dimensions usually correspond to attributes of objects together with a semi-distance defined on the attributes value domains. For example, if one measures colors of objects, quality dimensions can correspond to hue, chromaticity and brightness. The concept "fruit" may have dimensions corresponding to weight, taste, color, etc.

Usually, with any quality dimension one associates a relational structure representing a domain of values corresponding to the quality dimension, together with functions and relations allowing one to calculate (semi-)distances.

For instance, with the quality dimension "weight" one can associate a relational structure defining arithmetic on the real numbers.

3 Tolerance and Inclusion Functions

We begin by defining a tolerance function on individuals. From this a parameterized tolerance relation follows naturally.

Definition 3.1. *By a* tolerance function *on a set U we mean any function $\tau : U \times U \longrightarrow [0,1]$ such that for all $x, y \in U$,*

$$\tau(x,x) = 1 \quad \textit{and} \quad \tau(x,y) = \tau(y,x). \qquad \blacksquare$$

Given a conceptual space $\langle U, Q\rangle$ and a quality dimension $\langle U, \delta\rangle \in Q$, a *tolerance function τ, based on the quality dimension* can be defined as follows:

$$\tau(u,u') \stackrel{\text{def}}{=} 1 - \frac{\delta(u,u')}{\max\{\delta(x,y) : x, y \in U\}}. \tag{1}$$

Of course, the same approach could be used for an attribute a and its value set V_a in a complex knowledge structure, provided δ is given, without appeal to conceptual spaces.

Definition 3.2. *For $p \in [0,1]$ by a* tolerance relation to a degree at least p based on τ, *we mean the relation τ^p given by*

$$\tau^p \stackrel{\text{def}}{=} \{\langle x, y\rangle \mid \tau(x,y) \geq p\}.$$

The relation τ^p is also called the parameterized tolerance relation. ■

In the rest of the paper, $\tau^p(x,y)$ is used to denote the characteristic function for the relation τ^p.

Intuitively, $\tau(x,y)$ provides a degree of similarity between x and y, whereas $\tau^p(x,y)$ states that the degree of similarity between x and y is at least p. In what follows we limit ourselves to tolerance relations where it is assumed that the parameter p has been provided and is tuned to fit particular applications.

Often one considers objects to be similar if a given distance between them is not greater than a given threshold, say d. Given a quality dimension $\langle U, \delta \rangle$ and a threshold $d \geq 0$, one can define the parameter p from Definition 3.2 to be

$$p \stackrel{\text{def}}{=} 1 - \frac{d}{\max\{\delta(x,y) : x, y \in U\}}. \tag{2}$$

A parameterized tolerance relation is used to construct tolerance neighborhoods for individuals.

Definition 3.3. *By a* neighborhood function wrt τ^p *we mean a function given by*

$$n^{\tau^p}(u) \stackrel{\text{def}}{=} \{u' \in U \mid \tau^p(u, u') \text{ holds}\}.$$

By a neighborhood of u wrt τ^p *we mean the value* $n^{\tau^p}(u)$. ■

4 Tolerance Spaces

The concept of tolerance spaces plays a fundamental rôle in our approach.

Definition 4.1. *A* tolerance space *is defined as the tuple* $TS = \langle U, \tau, p \rangle$*, which consists of*

- *a nonempty set* U*, called the* domain *of* TS*;*
- *a* tolerance function τ
- *a* tolerance parameter $p \in [0, 1]$.

The parameterized tolerance relation τ^p *is defined as in Definition 3.2.* ■

Given a universe U of individuals, a set of attributes A and a set $X \subseteq U$, one often considers the lower and upper approximation of X as defined in terms of a partitioning of the universe U in indiscernibility classes relative to a subset of the attributes A. Given a tolerance space $TS = \langle U, \tau, p \rangle$, rather than considering an individual's indiscernibility class as a basis for defining the lower and upper approximation of $X \subseteq U$, we can instead use the neighborhood of an individual induced by the tolerance function/parameter pair(s) provided by the tolerance space. In addition, we can tune our definition of upper approximation via a parameter q which determines how much of a neighborhood must be part of X in order for it to be included in the upper approximation.[3]

Below, for any set X, by $|X|$ we mean the cardinality of X.

[3] A different approach, based on a notion of approximation spaces, object neighborhoods and rough inclusion, has been introduced in [6].

Definition 4.2. *Let* $U_1, U_2 \subseteq U$*. By the* standard inclusion function *we mean the function given by*

$$\mu(U_1, U_2) \stackrel{\text{def}}{=} \begin{cases} \dfrac{|U_1 \cap U_2|}{|U_1|} & \text{if } U_1 \neq \emptyset \\ 1 & \text{otherwise.} \end{cases}$$

Let $TS = \langle U, \tau, p \rangle$ *be a tolerance space and* $X \subseteq U$*. The* lower and upper approximations of X wrt TS to a degree $q \in [0,1]$, $X^q_{TS^+}$ *and* $X^q_{TS^\oplus}$*, are defined by*

$$X^q_{TS^+} = \{u \in U : \mu(n^{\tau^p}(u), X) = 1\}, X^q_{TS^\oplus} = \{u \in U : \mu(n^{\tau^p}(u), X) > q\}.$$

The approximations $X^0_{TS^+}, X^0_{TS^\oplus}$ *are called* the lower and upper approximations of X wrt TS *and are often denoted by* $X_{TS^+}, X_{TS^\oplus}$*, respectively.* ■

5 Defining Tolerance on Complex Representational Structures

In this section we show how to induce a tolerance relation on complex structures on the basis of a tolerance relation defined on domain elements.

Consider a tolerance space $TS = \langle U, \tau, p \rangle$. First, we would like to extend the tolerance and neighborhood functions induced by TS to deal with subsets of U. We shall need a notion of generalized inclusion function ν^{τ^p} which will be used as a basis for measuring similarity between complex information structures.

One of the important motivations behind the definition provided is that we require a generalized inclusion function to coincide with the standard inclusion function in the case of a trivial tolerance space (identifying equal elements and distinguishing elements that are not equal).[4]

Definition 5.1. *Let* U *be a set and* $U_1, U_2 \subseteq U$*. By the* generalized inclusion function induced by τ^p *we mean the function given by*

$$\nu^{\tau^p}(U_1, U_2) \stackrel{\text{def}}{=} \begin{cases} \dfrac{|\{u_1 \in U_1 : \exists u_2 \in U_2 [u_1 \in n^{\tau^p}(u_2)]\}|}{|U_1|} & \text{if } U_1 \neq \emptyset \\ 1 & \text{otherwise.} \end{cases}$$

For $q \in [0,1]$*, we say that* U_1 is included in U_2 to a degree at least q wrt ν^{τ^p} *iff* $\nu^{\tau^p}(U_1, U_2) \geq q$*.*

In the case of tuples[5] $U_1 = \langle u_1, \ldots, u_n \rangle$ *and* $U_2 = \langle u'_1, \ldots, u'_n \rangle$*, by the* generalized inclusion function over tuples, induced by τ^p *we mean the function given by*

$$\nu_o^{\tau^p}(U_1, U_2) \stackrel{\text{def}}{=} \begin{cases} \dfrac{|\{u_i : 1 \leq i \leq n \text{ and } u_i \in n^{\tau^p}(u'_i)\}|}{|U_1|} & \text{if } n \neq 0 \\ 1 & \text{otherwise.} \end{cases}$$ ■

[4] We also require such "continuity" in other definitions. Namely, the trivial tolerance space should always lead to standard notions that are accepted when tolerance is not considered.

[5] I.e., ordered sets of the same cardinality.

In the sequel we write $\nu_{TS}^{\tau p}$ and $n_{TS}^{\tau p}$, respectively, to denote $\nu^{\tau p}$ and n^{τ^p}, where τ^p is a tolerance relation induced from a tolerance space TS.[6]

Definition 5.2. *Let $TS = \langle U, \tau, p\rangle$ be a tolerance space. By a* power tolerance space induced by TS *we mean $T^{TS} = \langle U^{TS}, \tau^{TS}, s\rangle$, where*

- $U^{TS} \stackrel{\text{def}}{=} 2^U$*, is the set of all subsets of* U
- *for* $U_1, U_2 \in U^{TS}$, $\tau^{TS}(U_1, U_2) \stackrel{\text{def}}{=} \min\left\{\nu^{\tau^p}(U_1, U_2), \nu^{\tau^p}(U_2, U_1)\right\}$
- $s \in [0, 1]$ *is a tolerance parameter.* ∎

We define tolerance and neighborhood functions on tuples of elements in a similar manner.

Definition 5.3. *Let $TS = \langle U, \tau, p\rangle$ be a tolerance space. By a k-*tuple tolerance space induced by TS *we mean $T^{TS^k} = \langle U^{TS^k}, \tau^{TS^k}, s\rangle$, where*

- $U^{TS^k} \stackrel{\text{def}}{=} \underbrace{U \times \ldots \times U}_{k-times}$*, is the set of all k-tuples of* U
- *for* $U_1, U_2 \in U^{TS^k}$, $\tau^{TS^k}(U_1, U_2) \stackrel{\text{def}}{=} \nu_o^{\tau^p}(U_1, U_2) = \nu_o^{\tau^p}(U_2, U_1)$,[7]
- $s \in [0, 1]$ *is a tolerance parameter*[8]. ∎

Let us summarize the methodology we propose:

- we start with a quantitative representation of the similarity of considered concepts given by semi-distance or tolerance functions (see Definitions 2.1 and 3.1)
- the definition of tolerance spaces (Definition 4.1) and neighborhoods (Definition 3.3) allows us to transform the quantitative representation of the similarity into a qualitative representation of the concepts. Such a transformation can also be applied to complex representational structures using Definitions 5.2 and 5.3. Tolerance parameters allow us to tune the similarities to fit particular application domains
- the approximations provided in Definition 4.2 allow us to isolate objects that surely satisfy a given property and that might satisfy the property. In consequence, we also obtain a characterization of objects that surely do not satisfy the property
- finally one can apply various deduction mechanisms to reason about the considered concepts (see, e.g., [2]).

---

[6] We often drop the superscripts and subscripts when the tolerance spaces and relations are known from context.

[7] The equality between $\nu_o^{\tau^p}(U_1, U_2)$ and $\nu_o^{\tau^p}(U_2, U_1)$ follows from the symmetry of τ.

[8] The tolerance paramter s specified in definitions 5.2 and 5.3 is not used in this paper.

Object	Length (cm)	Wingspan (cm)	Weight (g)	Color
blue jay	28	41	85	blue and grey
gray jay	29	46	70	grey with white and black
rusty blackbird	23	36	60	black
brewer's blackbird	23	39	63	black
european starling	22	41	82	black to brown

Table 1. Description of birds.

6 An Example

In the following example we will use the data in Table 1 to exemplify the definition and use of tolerance spaces, where objects (birds) are characterized by attributes Length, Wingspan, Weight and Color.

For simplicity of presentation, below we use a separate domain for each of the attributes. Of course, the domains can simply be encoded by a single domain.

Let $\delta(x, y) \stackrel{\text{def}}{=} abs(x - y)$ be a distance function, where $abs(z)$ stands for the absolute value of z.

We first define a tolerance space for the integer value domain V_L of the attributes Length and Wingspan. We use a threshold of 5cm. The corresponding tolerance space $TS_L = \langle V_L, \tau_L, p_L \rangle$ is defined by:

$$V_L = \{x : 20 \leq x \leq 50\},\ \tau_L(x, y) = 1 - \frac{\delta(x, y)}{\delta(20, 50)},\ p_L = 1 - \frac{5}{\delta(20, 50)}.$$

Now $n^{\tau_L{}^{p_L}}(x) = \{y \in V_L : abs(x - y) \leq 5\} = \{y \in V_L : \tau_L^{P_L}(x, y)\}$.

Similarly one can define a tolerance space for the integer value domain V_W of the attribute Weight. We use a threshold of 10g. The tolerance space $TS_W = \langle V_W, \tau_W, p_W \rangle$ is defined by:

$$V_W = \{x : 60 \leq x \leq 90\},\ \tau_W(x, y) = 1 - \frac{\delta(x, y)}{\delta(60, 90)},\ p_W = 1 - \frac{10}{\delta(60, 90)}.$$

Now $n^{\tau_W{}^{p_W}}(x) = \{y \in V_L : abs(x - y) \leq 10\} = \{y \in V_L : \tau_W^{P_W}(x, y)\}$.

We define a tolerance space for the symbol value domain V_C of the attribute Color to be $TS_C = \langle V_C, \tau_C, p_C \rangle$, where:

- V_C consists of colors listed in the column of Table 1 labelled by Color
- for any color c, $\tau_C(c, c) = 1$. We also assume that

$$\tau_C(\textsf{black}, \textsf{black to brown}) = \tau_C(\textsf{black to brown}, \textsf{black}) = 0.9$$

- $p_C = 0.85$.

Assuming such tolerance spaces, we can conclude that rusty blackbird is similar to brewer's blackbird, since the first one is characterized by attributes $\langle 23, 36, 60, \textsf{black} \rangle$ and the second by attributes $\langle 23, 39, 63, \textsf{black} \rangle$. By Definition 5.3,

$$\tau^{TS^4}(\langle 23, 36, 60, \textsf{black} \rangle, \langle 23, 39, 63, \textsf{black} \rangle) = 4/4 = 1,$$

since $23 \in n^{\tau_L{}^{pL}}(23)$, $36 \in n^{\tau_L{}^{pL}}(39)$, $60 \in n^{\tau_L{}^{pL}}(63)$ and $\mathsf{black} \in n^{\tau_C{}^{pC}}(\mathsf{black})$.

For blue jay and grey jay we can conclude that:

$$\tau^{TS^4}(\langle 28, 41, 85, \mathsf{blue\ and\ grey}\rangle, \langle 29, 46, 70, \mathsf{grey\ with\ white\ and\ black}\rangle) = 2/4,$$

since $28 \in n^{\tau_L{}^{pL}}(29)$, $41 \in n^{\tau_L{}^{pL}}(46)$, $85 \notin n^{\tau_L{}^{pL}}(70)$ and
$\mathsf{blue\ and\ grey} \notin n^{\tau_C{}^{pC}}(\mathsf{grey\ with\ white\ and\ black})$.

For rusty blackbird and european starling we can conclude that:

$$\tau^{TS^4}(\langle 23, 36, 60, \mathsf{black}\rangle, \langle 22, 41, 82, \mathsf{black\ to\ brown}\rangle) = 3/4,$$

since $23 \in n^{\tau_L{}^{pL}}(22)$, $36 \in n^{\tau_L{}^{pL}}(41)$, $60 \notin n^{\tau_L{}^{pL}}(82)$ and
$\mathsf{black} \in n^{\tau_C{}^{pC}}(\mathsf{black\ to\ brown})$.

One can further define tolerance spaces on collections of birds, using Definition 5.2, relations defined on birds, etc.

7 Applications

7.1 Sensor Models and Tolerance Spaces

In this section, we provide a simple sensor model[9] and one method for modeling uncertainty in sensor data which integrates well with tolerance spaces. We also discuss the construction of virtual sensors from combinations of actual and other virtual sensors.

A sensor is used to measure one or more physical attributes in an environment E. The value sets associated with a physical attribute might be the real numbers, as in the case of measurement of the temperature or velocity of an object; Boolean values, as in the measurement of the presence or absence of an object such as a red car; integer values, as in the case of measurement of the number of vehicles in a particular intersection; or scalar values, such as the specific color of a vehicle. An environment E can be viewed as an abstract entity containing a collection of physical attributes that are measurable. Vectors or n-dimensional arrays of attribute/value pairs could be used to represent a particular environment. One may want to add a temporal argument to E, so the current state of the environment is dynamic and changes with time.

We denote a sensor S_i as a function of the environment E and time point t, $S_i(E, t)$. S_i is a function which returns a pair of functions,

$$S_i(E, t) = \{V_i(t), \epsilon_i(t)\}.$$

Depending on the type of sensor being modeled, $V_i(t)$ will be a function that returns the values of the physical attributes associated with the sensor. V_i might

[9] This model is based on a generalization of that in [1].

return a single value, as in the case of a single temperature sensor, or a vector or array of values for more complex sensors.

For any physical attribute measured, explicit accuracy bounds will be supplied in the form of $\epsilon_i(t)$. The temporal argument is supplied since the accuracy of a sensor may vary with time. As in the case of V_i, ϵ_i might return a single accuracy bound or a vector or array of accuracy bounds.

For example, suppose S_{temp} is a sensor measuring the temperature of a PC104 box on an unmanned aerial vehicle. Let a_{temp} be the physical attribute associated with temperature in the environment, where the actual temperature is $E(t)(a_{temp})$ and the value returned by the sensor is $V_i(t)(a_{temp})$. The following constraint holds:

$$E(t)(a_{temp}) \in [V_i(t)(a_{temp}) - \epsilon_i(t), V_i(t)(a_{temp}) + \epsilon_i(t)].$$

By using tolerance spaces, accuracy bounds for a physical attribute can be represented equivalently as tolerance relations to degree p on the value set for the attribute. In this manner, we can use neighborhood functions to reason about the tolerance or accuracy neighborhoods around individual sensor readings and combine these into neighborhoods for more complex virtual sensors.

In the following, we will drop the temporal argument for ϵ and assume the accuracy bounds for attributes do not change with time. Let $TS_{S_{ik}} = \langle V_{S_{ik}}, \tau_{s_{ik}}, p_{s_{ik}} \rangle$ be a tolerance space for the kth physical attribute, a_{i_k} associated with the sensor S_i, where,

- $V_{S_{ik}} = \{x \mid lb \leq x \leq ub, x \in D\}$, where D is a value domain such as the reals or integers. It is assumed that the legal values for a physical attribute have a lower and upper bound, lb, ub. We associate a distance measurement $\delta(x) = \mid x - y \mid$ with the value set $V_{S_{ik}}$, which includes all the values that can be read from the sensor S_i.
- Both the tolerance function $\tau_{s_{ik}}$, and the tolerance parameter $p_{s_{ik}}$ are defined as follows,

$$\tau_{s_{ik}}(x, y) = 1 - \frac{\delta(x, y)}{\delta(lb, ub)}, \quad p_{s_{ik}} = 1 - \frac{\epsilon_i}{\delta(lb, ub)}.$$

The neighborhood function can be used to compute the possible actual values of a physical attribute in the environment, given a sensor reading, under the assumption that the accuracy bounds have been generated correctly for a particular sensor and the sensor remains calibrated. For example, if $V_i(a_{temp})$ is the current value measured by the sensor S_i then we would know that $E(a_{temp}) \in n^{p_{s_{ik}}}(V_i(a_{temp}))$. So, the tolerance neighborhood around a sensor reading always contains the actual value of the physical attribute in the environment E and it would be correct to reason with the neighborhoods of sensor values, rather than the sensor value itself.

We can then use these physical attributes and their associated tolerance spaces to construct more complex attributes and knowledge structures in terms of these.

These new attributes and knowledge structures would inherit the accuracy (inaccuracy) of the primitive sensor data used in their construction.

7.2 Mutual Understanding between Tolerance Agents

Consider a multi-agent application in a complex environment such as the world wide web for software agents, or a natural disaster in an urban area for physical robots. Each agent will generally have its own view of its environment due to a number of factors such as the use of different sensor suites, knowledge structures, reasoning processes, etc. Agents may also have different understandings of the underlying concepts which are used in their respective representational structures and will measure objects and phenomena with different accuracy. How then can agents with different knowledge structures and perceptive accuracies understand each other and effect meaningful communication and how can this be modeled? In this section, both tolerance spaces and upper and lower approximations on agent concepts and relations are used to define a means for agents to communicate when different sensor capabilities and different levels of accuracy in knowledge structures are assumed.

We begin with a broad definition of a *tolerance agent*.

Definition 7.1. *By a* tolerance agent *we shall understand any pair* $\langle Ag, TS \rangle$*, where* Ag *is an agent and* TS *is a tolerance space.* ∎

The assumption is that the Ag part of an agent consists of common functionalities normally associated with agents such as planners, reactive and other methods, knowledge bases or structures, etc. The knowledge bases or structures are also assumed to have a relational component consisting of approximate relations which are derived and viewed through the agents limited sensor capabilities. When the agent introspects and queries its own knowledge base these limited perceptive capabilities should be reflected in any answer to a query.

The following definition will be used to provide a tolerance limited semantics for queries in the context of a particular tolerance space.

Definition 7.2. *Let* $TS = \langle U, \tau, p \rangle$ *be a tolerance space. Consider a pair of sets,* $Z = \langle X, Y \rangle$ *, such that* $X \subseteq Y$.[10] *By a* lower *and* upper approximation of Z wrt TS *we mean*

$$Z^{\tau^p}_{TS^+} \stackrel{\text{def}}{=} \{u \in U : n^{\tau^p}_{TS}(u) \subseteq X\}$$
$$Z^{\tau^p}_{TS^\oplus} \stackrel{\text{def}}{=} \{u \in U : n^{\tau^p}_{TS}(u) \cap Y \neq \emptyset\}.$$

$Z^{\tau^p}_{TS^-}$ *is defined as* $-Z^{\tau^p}_{TS^\oplus}$. ∎

To keep the exposition concise, simple queries, such as $R(\mathsf{a})$ will be used, where R is a relation symbol and a is a constant symbol. Due to its limited perceptive

[10] Intuitively, X and Y correspond to a lower and upper approximation of a set.

capabilities, one can assume that the agent may not recognize the difference between a and other objects in the neighborhood of a. Thus, the agent can be sure that $R(\mathsf{a})$ holds only if all elements in the neighborhood of a satisfy R. The agent also can not exclude the possibility that $R(\mathsf{a})$ holds if there is at least one element in the neighborhood of a satisfying R. Consequently, it is clear that R can be viewed as a set such that:

– its lower approximation only contains elements that, together with all elements in their neighborhood, satisfy R
– its upper approximation contains elements for which there is at least one element in their neighborhood that satisfies R.

Moreover, the set itself is given only via its approximations.

The following example illustrates this approach.

Example 7.3. Let *TA* be a tolerance agent with the following domain of discourse: Mary, lR, mR, dR, where the latter three elements denote "light red", "medium red" and "dark red", respectively. *TA*'s knowledge base contains the following three facts:

$$Likes(\mathsf{Mary}, \mathsf{lR}),\ Likes(\mathsf{Mary}, \mathsf{mR}),\ \ \neg Likes(\mathsf{Mary}, \mathsf{dR}),$$

Assume further that the single tolerance relation associated with the tolerance space of *TA* identifies lR with mR and mR with dR.

Suppose agent *TA* is given the task of verifying whether Mary likes a color it directly senses as being lR. Based on the agent's tolerance relation, its sensors are not capable of recognizing the difference between lR and mR. However, Mary likes both colors, so *TA* can be sure that she likes the color sensed by TA with certainty.

If *TA* directly sensed the color as mR then it could not be sure whether Mary likes this color or not, since it does not perceive any difference between mR and dR. The sensed color might actually be dR which Mary does not like. On the other hand, *TA* could not exclude the alternative that Mary likes this color, as it could equally well be mR.

In summary, lR is in the lower approximation of the (unary) relation $likes(\mathsf{Mary}, x)$ and mR and dR are in the upper approximation of the relation. The agent *TA* would use these approximations of the relation together with its knowledge and associated tolerance space when answering questions about Mary's likes or dislikes. ■

These intuitions are formalized in the following definition.

Definition 7.4. *Let* TA $= \langle Ag, TS \rangle$ *be a tolerance agent. Then the semantics of a relation R wrt* TA *is given by:*

$$R_{TA^+} \stackrel{\text{def}}{=} R_{TS^+},\ R_{TA^\oplus} \stackrel{\text{def}}{=} R_{TS^\oplus} \text{ and } R_{TA^-} \stackrel{\text{def}}{=} R_{TS^-},$$

where R_{TS^+}, $R_{TS^\oplus}$ and R_{TA^-} are as defined in Definition 7.2. ■

Remark 7.5. It is important to note that Definition 7.4 refers to an arbitrary relation. Since any first-order or fixpoint query to a RDB returns a relation as its result, the definition also provides us with the semantics of queries asked to and answered by tolerance agents. ■

Example 7.6. Consider the tolerance agent TA again, with the same tolerance space and facts given in Example 7.3. The answer returned by agent TA to the sample query, $Likes(\mathsf{Mary}, x)$, will be computed using Definition 7.4.

According to Example 7.3,

$$Likes = \{\langle \mathsf{Mary}, \mathsf{lR}\rangle, \langle \mathsf{Mary}, \mathsf{mR}\rangle\}.$$

Consequently, $Likes$ is approximated by TA as follows:

$$\begin{aligned} Likes_{TA^+} &\stackrel{\text{def}}{=} Likes_{TS^+} = \{u \mid u \in \{\langle \mathsf{Mary}, \mathsf{lR}\rangle, \langle \mathsf{Mary}, \mathsf{mR}\rangle, \langle \mathsf{Mary}, \mathsf{dR}\rangle\} \\ &\quad \text{and } n_{TS}(u) \subseteq Likes\} = \{\langle \mathsf{Mary}, \mathsf{lR}\rangle\} \\ Likes_{TA^\oplus} &\stackrel{\text{def}}{=} Likes_{TS^\oplus} = \{u \mid u \in \{\langle \mathsf{Mary}, \mathsf{lR}\rangle, \langle \mathsf{Mary}, \mathsf{mR}\rangle, \langle \mathsf{Mary}, \mathsf{dR}\rangle\} \\ &\quad \text{and } n_{TS}(u) \cap Likes \neq \emptyset\} = \{\langle \mathsf{Mary}, \mathsf{lR}\rangle, \langle \mathsf{Mary}, \mathsf{mR}\rangle, \langle \mathsf{Mary}, \mathsf{dR}\rangle\}. \end{aligned}$$

Thus,the following facts hold:

$$\begin{aligned} &Likes(\mathsf{Mary}, \mathsf{lR})_{TA^+},\ Likes(\mathsf{Mary}, \mathsf{lR})_{TA^\oplus}, \\ &Likes(\mathsf{Mary}, \mathsf{mR})_{TA^\oplus},\ Likes(\mathsf{Mary}, \mathsf{dR})_{TA^\oplus}. \end{aligned}$$

These results reflect the intuitions described in Example 7.3. ■

Given that two tolerance agents have different tolerance spaces it becomes necessary to define the meaning of queries and answers relative to the two tolerance agents. As advocated before, a tolerance agent, when asked about a relation, answers by using the approximations of the relation wrt its tolerance space. On the other hand, the agent that asked the query has to understand the answer provided by the other agent wrt to its own tolerance space. The dialog between agents, say TA_1 (query agent) and TA_2 (answer agent), conforms then to the following schema:

1. TA_1 asks a query Q to TA_2
2. TA_2 computes the answer approximating it according to its tolerance space and returns as an answer the approximations $QA = \langle Q_{TA_2^+}, Q_{TA_2^\oplus}\rangle$
3. TA_1 receives QA as input and approximates it according to its own tolerance space. The resulting approximations provide the answer to the query, as understood by TA_1.

In order for the schema to work properly, it has to be assumed that the two agents operate with a common vocabulary when communicating. This does not imply that the agents need to have the same vocabulary, simply that there is some overlap.

The definition describing this interaction now follows.

Definition 7.7. *Let TA_1, TA_2 be tolerance agents and let Q be a query, expressed in a logic, which is asked by TA_1 and answered by TA_2. Then the* meaning of the query *is given by the following approximations:*

$$\langle \langle Q_{TA_2^+}, Q_{TA_2^\oplus} \rangle_{TA_1^+}, \langle Q_{TA_2^+}, Q_{TA_2^\oplus} \rangle_{TA_1^\oplus} \rangle. \tag{3}$$

■

The notion of mutual understanding used by communicating agents of this type is developed in full in [3].

8 Summary

This paper presents a framework for recursively constructing arbitrarily complex knowledge structures which may be compared for similarity, distance and approximativeness. The techniques used attempt to bridge a gap between quantitative representations of data in terms of attribute/value pairs and their use in qualitative knowledge representations at different levels of abstraction. The qualitative representations inherit the approximativeness of their component structures through the use of neighborhoods of objects and upper and lower approximations induced through their use. Tolerance spaces provide a structured means of constructing complex representational structures. These ideas have been exemplified by using the techniques for sensor modeling and signal to symbol conversions and for representing approximate queries between agents with heterogenous perceptive capabilities.

References

1. R. R. Brooks and S.S. Iyengar. *Multi-Sensor Fusion*. Prentice-Hall, 1998.
2. P. Doherty, W. Łukaszewicz, A. Skowron, and A. Szałas. Combining rough and crisp knowledge in deductive databases. In *Rough-Neuro Computing: Techniques for Computing with Words*. 2003.
3. P. Doherty, W. Łukaszewicz, and A Szałas. On mutual understanding among communicating agents. In B. Dunin-Keplicz and R. Verbrugge, editors, *Proceedings of Workshop on Formal Approaches to Multi-agent Systems (FAMAS'03)*, pages 83–97, 2003.
4. P. Gärdenfors. *Conceptual spaces: the geometry of thought*. MIT Press, Cambridge, Mass., 2000.
5. Z. Pawlak. *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht, 1991.
6. A. Skowron and J. Stepaniuk. Tolerance approximation spaces. *Fundamenta Informaticae*, 27:245–253, 1996.