Similarity, Approximations and Vagueness*

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

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

Abstract. The relation of similarity is essential in understanding and developing frameworks for reasoning with vague and approximate concepts. There is a wide spectrum of choice as to what properties we associate with similarity and such choices determine the nature of vague and approximate concepts defined in terms of these relations. Additionally, robotic systems naturally have to deal with vague and approximate concepts due to the limitations in reasoning and sensor capabilities. Halpern [1] introduces the use of subjective and objective states in a modal logic formalizing vagueness and distinctions in transitivity when an agent reasons in the context of sensory and other limitations. He also relates these ideas to a solution to the Sorities and other paradoxes. In this paper, we generalize and apply the idea of similarity and tolerance spaces [2,3,4,5], a means of constructing approximate and vague concepts from such spaces and an explicit way to distinguish between an agent's objective and subjective states. We also show how some of the intuitions from Halpern can be used with similarity spaces to formalize the above-mentioned Sorities and other paradoxes.

1 Introduction and Preliminaries

1.1 Introduction

In a recent paper, Halpern [1] points out the tight correlation between similarity notions on individuals and their relation to vague predicates. He also considers a distinction between the subjective and objective realities of agent systems and how standard properties of similarity such as transitivity do not necessarily make sense when taking into account epistemic and subjective states of agent systems. Objectively, viewing similarity as an equivalence relation may make sense, but when taking into account capabilities of agents to discern, or their subjective psychological states, it may not make sense to view similarity as a transitive relation. One can also find other examples where intransitivity may hold at the objective level, but not at the subjective level.

When viewing similarity and vagueness in this respect, it turns out that a number of interesting reasoning paradoxes such as the *Sorities Paradox*, can be explained in a matter both satisfactory in the formal sense and also in the pragmatic sense, where agents would have to represent and reason about such concepts as heaps. In attacking these problems, [1] proposes a modal logic which semantically represents both the subjective

* Supported in part by a WITAS UAV project grant under the Wallenberg Foundation, Sweden and an NFFP03 grant (COMPAS).

D. Ślęzak et al. (Eds.): RSFDGrC 2005, LNAI 3641, pp. 541–550, 2005.

and objective states accessible by an agent and also allows for the ability to distinguish between *perception reports* and what an agent may definitely know about its objective state. This is done by introducing two modal operators.

Rather than going the modal route, we introduce a general method for modeling similarity relations, approximate sets, and vague predicates. We show how this formal framework can be used to model scenarios associated with an agent, its objective and subjective realities, similarity relations contributing to the definition of vague or approximate predicates, and the sensory limitations essentially defining what it formally means for an agent to have a subjective view of reality as observed through its sensory filters. The basis for this representational capability are *similarity spaces*, *neighborhoods of individuals* derived from such spaces and *approximate or vague predicates* defined using such neighborhoods. We also show that when restricted to finite domains in a relational database framework, inference associated with the approach is tractable.

Before providing the formal framework, we describe an intuitive scenario from [1]. Relative predicates associated with the sensing modalities are often difficult to represent and define due to the subjective nature of the concepts involved. For example, given samples of beverages and the task of stating which one is sweeter than which, it is difficult to characterize a comparative component of the definition while keeping it consistent with the objective sensor data from which it is grounded.

On the one hand, similarity of sweetness is transitive relative to the number of grains of sugar in beverages, but at a more subjective level, transitivity breaks down. We often experience such comparative situations where beverage A's sweetness is indistinguishable from beverage B, and B's sweetness is indistinguishable from beverage C, but A's sweetness is in fact quite distinguishable from C. Distinguishability at this level is qualitatively different from that at the granular level where a beverage with *n* grains of sugar is indistinguishable from that with *n+1* grains of sugar and so on and so forth.

It is obvious to see how the Sorities Paradox is related to this issue. At an objective level, heaps are simply piles of sand with a certain number of grains in them. At the subjective level they are based on subjective perception reports which do not necessarily reflect transitive nature of the sensor data, but should still remain consistent with it.

The literature on similarity is vast and it is often the case that different properties of the associated relation are played off against the other, such as transitivity versus intransitivity, symmetry versus anti-symmetry, etc. One can relax the requirement of a tradeoff in many ways. Some such relaxations are introduced in Section 1.3 and used throughout the paper. In summary, our main goal is to introduce a general framework for representing similarity structures, which permits the definition of vague sets/relations in a meaningful and intuitive way. This will be partly verified by modeling some of the interesting scenarios presented in [1]. The starting point for the approach we propose was initiated by [6], but substantially generalized and applied in [2,3,4,5].

1.2 Paper Structure

In the remainder of this section, we provide some preliminary definitions. In Section 2, we consider the objective and subjective levels of an agent system interfacing with sensors to an external environment. We then relate these levels to vague concepts. In Section 3, we introduce similarity spaces which are the formal vehicle for making distinctions between objective and subjective perceptual descriptions and formalizing vagueness using approximate sets constructed from similarity-based neighborhoods. In Section 5, we formalize a number of examples including those already mentioned. In Section 6, we state some results on the complexity of the approach. In Section 7, we refer to some of the related literature and conclude the paper.

1.3 Preliminaries

Below we assume that $[0, 1]$ is the closed interval of all real numbers between 0 and 1, ordered by the standard ordering on reals $\leq$. We shall also use value Υ, meaning "unknown", which is not in $[0, 1]$ and is incomparable wrt $\leq$ with any number of $[0, 1]$. Let U be a set, $\sigma : U \times U \longrightarrow [0, 1] \cup \{\Upsilon\}$ be a binary function on U and $p \in [0, 1]$ be a given real number. Then, σ is called

- *p-serial* iff for any $x \in U$ there is $y \in U$ such that $\sigma(x, y) \geq p$
- *p-reflexive* iff for any $x \in U$, $\sigma(x, x) \geq p$ (note: p-reflexivity implies p-seriality)
- *p-symmetric* iff for any $x, y \in U$, $\sigma(x, y) \geq p$ implies $\sigma(y, x) \geq p$
- *p-transitive* iff for any $x, y, z \in U$, $\sigma(x, y) \geq p$ and $\sigma(y, z) \geq p$ implies $\sigma(x, z) \geq p$

One can also relax transitivity, as is often done in the fuzzy set area (cf. [7]).

2 Objectiveness, Subjectiveness and Vagueness

It is often the case that an intelligent system interfaces to external environment through the use of real sensors as in the robotics domain or through virtual sensors as in the software agent domain. Already, at this sensor interface level, there is a gap between what the world is actually like and what the robot or software agent is capable of perceiving given a particular sensor suite. For example, a red car may often be perceived by a robot to be brown in color due to special lighting conditions. There is an additional gap between raw sensor data and additional qualitative structures derived via the raw data and additional data fusion and knowledge construction processes. For example, a vehicle which is objectively on a road may be perceived by sensors to be both on and off the road due to sensor noise and inaccuracies, but at a qualitative level, a normative decision has been made to view the vehicle as being completely on the road.

In order to make these distinctions clear, we assume the existence of an *objective reality* independent of any agent's particular perceptive capabilities and the existence of a *subjective reality* specific to an agent. Each agent may or may not have different subjective realities and one agent may in fact have several subjective realities based on its particular configuration and context. Assuming the distinction between objective and subjective realities of an agent, we can refer to an agent's objective state in addition to its subjective states. This distinction is central to Halpern's approach [1] and we will show how our framework can clearly model this distinction in a highly flexible manner. We also use the term *subjective perception* to refer to perceptual activity which results in the generation of perception reports associated with the subjective state(s) of an agent.

In addition to perception reports regarding properties and relations between objects, perception reports about objects themselves and their similarity or lack thereof is equally important as input to reasoning processes. Subjective perception often cannot distinguish objects which are different at the objective level. In some situations this leads to *borderline cases*, where the observer cannot classify objects relative to a given concept. For example, we may not be able to state unequivocally that a vehicle is too close to another or that it is moving too fast relative to a specified speed limit.

According to the literature (see, e.g., [8]), a concept is vague when it has borderline cases, i.e., some objects cannot be classified to the concept or to its complement with certainty. In this paper vagueness is modelled by introducing similarity-based approximations of concepts. More specifically, the lower approximation of a concept consists of objects that are known to belong to the concept and the upper approximation of the concept consists of objects that might belong to the concept.

Observe that even the properties of similarity notion might be substantially different at the objective and subjective level, as illustrated by the following examples.

Example 2.1. Consider a robot equipped with a camera. Assume that the camera's field of view does not allow the robot to fully observe itself, which is a very strong perceptual limitation. In this case the similarity relation on the objective level is reflexive, while on the robot's subjective level it does not have to be reflexive, since the robot cannot observe itself (but might be p-reflexive and/or p-serial, for some p). □

Example 2.2. Assume that in a given application one considers a similarity relation, $\sim$, between children and parents. On the objective level, it is defined to satisfy $x \sim y$ iff $[Child(x, y) \wedge Sex(x) = Sex(y)]$. Then $\sim$ is not symmetric.[1] Now, suppose that on the subjective level one cannot recognize whether $Child(x, y)$ holds. In this case similarity is defined as $x \sim_s y$ iff $[\Upsilon \wedge Sex(x) = Sex(y)]$, which is symmetric. □

Example 2.3. Consider the similarity between persons in the set $\{P_1, P_2, P_3\}$. This relation, on the objective level, does not have to be transitive, since similarities between persons P_1 and P_2 as well as between P_2 and P_3 do not have to imply the similarity between P_1 and P_3. On the other hand, subjectively, a robot might not be able to distinguish between P_1, P_2 and P_3, which makes the similarity relation transitive. □

3 Similarity Spaces

Similarity spaces are used as the formal mechanism for representing the indistinguishability of individuals in a specific domain of discourse. Similarity spaces are quite versatile in use. They are also used as a basis for defining approximate sets and vague predicates in addition to modeling the sensory limitations of agents and provide a formal basis for constructing and analyzing subjective state.

Similarity spaces [2] generalize tolerance spaces as defined in [3]. Comparing the current approach to the approaches of [2,3], we assume that the similarity function can return the value Υ, since the similarity between some objects might be unknown. Also, as argued in [2], and advocated in Examples 2.1 and 2.2, we also relax the requirements that similarity has to be symmetric or reflexive. However, in order to make approximations intuitively meaningful, we will require the seriality of similarity spaces.

[1] In fact, one usually compares children to parents, not vice versa and it might be desirable that computer reflects this human behavior.

Definition 3.1. *By a* similarity function *on a set U we mean any function $\sigma : U \times U \longrightarrow [0,1] \cup \{\Upsilon\}$. A similarity function σ is called a* total similarity function *if, for any $x, y \in U$, $\sigma(x,y) \in [0,1]$. For $p \in [0,1]$, by a* similarity relation to a degree at least *p*, based on σ, *we mean the relation $\sigma^p = \{\langle x, y\rangle \mid \sigma(x,y) \geq p\}$. Such defined σ^p is also simply called the* similarity relation. □

A similarity relation is used to construct similarity neighborhoods for individuals.

Definition 3.2. *By a* neighborhood *of u wrt σ^p we mean the pair of sets $n^{\sigma^p}(u) = \langle n_+^{\sigma^p}(u), n_\oplus^{\sigma^p}(u)\rangle$, where $n_+^{\sigma^p}(u) = \{u' \in U \mid \sigma^p(u,u')$ holds$\}$ is called the* lower approximation of the neighborhood*, and $n_\oplus^{\sigma^p}(u) = n_+^{\sigma^p}(u) \cup \{y \mid \sigma(u,y) = \Upsilon\}$ is called the* upper approximation of the neighborhood. □

The lower approximation $n_+^{\sigma^p}(u)$ consists of elements which, in the context of available knowledge, are surely similar enough to u, while the upper approximation $n_\oplus^{\sigma^p}(u)$ consists additionally of elements that might be similar to u due to the unknown status of the similarity function. Note that in the case when σ is a total similarity function, we have that $n_+^{\sigma^p}(u) = n_\oplus^{\sigma^p}(u)$, thus the neighborhood can be considered as a single crisp set rather than pair of approximations.

Definition 3.3. *A* similarity space *is defined as tuple $\Sigma = \langle U, \sigma, p\rangle$, consisting of*

- *a nonempty set U, called the* domain *of Σ*
- *a* similarity function σ
- *a* similarity threshold $p \in [0,1]$.

If σ is a total similarity function, then Σ is called total*. If σ is p-serial (p-reflexive, p-symmetric, p-transitive) then Σ is called* serial *(*reflexive, symmetric, transitive*).* □

Tolerance spaces, as defined in [3], are total reflexive and symmetric similarity spaces (cf. [9]). Since reflexivity implies seriality, tolerance spaces are serial similarity spaces.

4 Approximations and Vagueness

Let us define the notions of approximation and vagueness as understood in this paper.

Definition 4.1. *Let $\Sigma = \langle U, \sigma, p\rangle$ be a serial similarity space and let $S \subseteq U$. The* lower and upper approximation of S wrt Σ*, denoted respectively by $S_{\Sigma+}$ and $S_{\Sigma\oplus}$, are defined by $S_{\Sigma+} \stackrel{\text{def}}{=} \{u \in U : n_+^{\sigma^p}(u) \subseteq S\}$ and $S_{\Sigma\oplus} \stackrel{\text{def}}{=} \{u \in U : n_\oplus^{\sigma^p}(u) \cap S \neq \emptyset\}$.* □

By $S_{\Sigma-}$ and $S_{\Sigma\ominus}$ we denote the complement of $S_\oplus$ and of S_+, respectively. The *boundary region* of S, denoted by $S_{\Sigma\pm}$, is defined as $(S_{\Sigma\oplus} - S_{\Sigma+})$.

Intuition behind Definition 4.1 is depicted in Fig.1. The element marked by $\triangledown$ is in the lower approximation of S – its whole lower approximation neighborhood is included in S. The element marked by $\square$ is in the boundary region – its upper approximation neighborhood contains elements which are in S and elements outside S. Finally, the element marked by $\triangle$ is outside of the upper approximation – its whole upper approximation neighborhood is outside S. Given a particular similarity space, one can be sure that $\triangledown$ is in S, $\triangle$ is outside S. The membership of $\square$ in S cannot be determined.

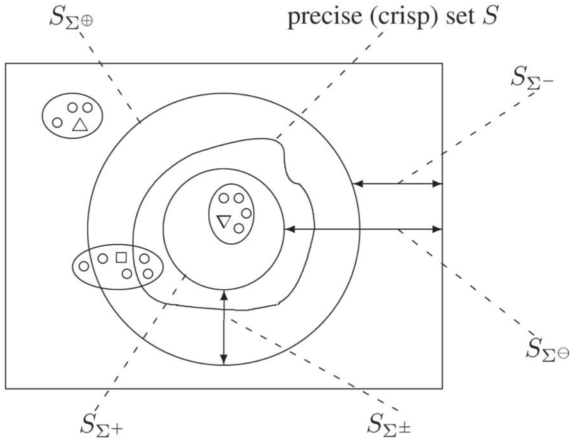

Fig. 1. Approximations of set S wrt a similarity space Σ

We observe[2] that the following proposition holds.

Proposition 4.2. *For any serial $\Sigma = \langle U, \sigma, p \rangle$ and $S \subseteq U$, we have $S_{\Sigma+} \subseteq S_{\Sigma\oplus}$.* □

If Σ is not serial, then, in general, the above inclusion does not hold. Without this property, the intuitive idea of an approximate set being bound set theoretically from below and above would not hold.

Definition 4.3. *Given a similarity space $\Sigma = \langle U, \sigma, p \rangle$, by a* vague set over Σ *or simply a* vague set[3] *(when Σ is known from context), we shall understand a pair $\langle S_{\Sigma+}, S_{\Sigma\oplus} \rangle$, where $S \subseteq U$. $\langle S_{\Sigma+}, S_{\Sigma\oplus} \rangle$ is called a* crisp set *over Σ (or simply a* crisp set*, when Σ is known), when $S_{\Sigma+} = S_{\Sigma\oplus}$. Then we write $S_{\Sigma+}$ rather than the whole pair.* □

Observe that in the definition above, boundary regions model borderline cases. Moreover, the parameter p of Σ may vary when perceptual capabilities of an observer change. Consequently, for a given concept its boundary region is not definitely fixed but contextual. Note also that relations are sets of tuples and can be approximated given a similarity space on tuples of the corresponding type.

We assume that the objective level is specified by means of crisp or vague relations, e.g., stored in a relational or deductive database, or defined by means of formulas of a given logic with underlying relational semantics. We also assume that any serial similarity space Σ reflects perceptual limitations of an observer's subjective level:

Definition 4.4. *Assume $\Sigma = \langle U, \sigma, p \rangle$ is a serial similarity space and let $Z = \langle X, Y \rangle$ be a vague set over U. Then a* perception report of Z wrt $\Sigma_1 = \langle U_1, \sigma_1, p_1 \rangle$ *is defined as the vague set $\langle X_{\Sigma_1^+}, Y_{\Sigma_1^\oplus} \rangle$.* □

[2] E.g., by a slight generalization of the corresponding argument given in [2].

[3] We sometimes use the term *approximate set* as a synonym for "vague set". We also deal with *vague predicates* to mean that their extensions are vague sets (of tuples of respective arity).

Definition 4.4 allows one to model perception reports pertaining to both the objective and subjective levels of different agents. For example, in agent communication, a receiving agent may view the sending agent's knowledge as objective and then apply a perceptual filter in terms of its current perceptual limitations. The resulting interpretation of knowledge is that perceived subjectively by the receiver and it is different from that perceived "objectively" by the sender. For example, the vague set Z can be provided by an agent with some perceptual limitations modelled by Σ. Receiver approximates Z using its own filter Σ_1. These ideas are developed, in the context of tolerance spaces, in [4,5], and can easily be generalized onto arbitrary serial similarity spaces.

We also have the following theorem about the unfalsifiability of perception in the case of serial similarity spaces. It essentially states that once an element is surely perceived to be in a set under observation, it cannot be further classified not to be in the set, even when the similarity threshold is arbitrarily changed (without violating seriality). Similarly, once it is surely perceived to be outside the set, it cannot be further classified to be in the set no matter what the similarity threshold is (again, retaining seriality).

Theorem 4.5. *Let $\Sigma = \langle U, \sigma, p\rangle$ and $\Sigma_1 = \langle U, \sigma, q\rangle$ be any serial similarity spaces.*[4] *Then, for any vague set $Z = \langle X, Y\rangle$ over U, we have:*

1. *if $s \in X_{\Sigma^+}$ then $s \in X_{\Sigma_1^\oplus}$*
2. *if $s \in Y_{\Sigma^-}$ then $s \in Y_{\Sigma_1^\ominus}$*

Proof. We prove the first part. The second is symmetric. Consider the following cases:

$q \le p$: then $n_\oplus^{\sigma^p}(s) \subseteq n_\oplus^{\sigma^q}(s)$. If $s \in X_{\Sigma^+}$ then, by seriality of Σ, $s \in X_{\Sigma^\oplus}$. Thus $n_\oplus^{\sigma^p}(s) \cap X \neq \emptyset$, i.e., $n_\oplus^{\sigma^q}(s) \cap X \neq \emptyset$, i.e., $s \in X_{\Sigma_1^\oplus}$

$q > p$: then $n_+^{\sigma^q}(s) \subseteq n_+^{\sigma^p}(s)$. Thus, if $s \in X_{\Sigma^+}$ then $n_+^{\sigma^q}(s) \subseteq n_+^{\sigma^p}(s) \subseteq X$. By seriality of Σ_1, $n_+^{\sigma^q}(s) \neq \emptyset$. Thus $n_+^{\sigma^q}(s) \cap X \neq \emptyset$. By definition, $n_+^{\sigma^q}(s) \subseteq n_\oplus^{\sigma^q}(s)$, hence $n_\oplus^{\sigma^q}(s) \cap X \neq \emptyset$, i.e., $s \in X_{\Sigma_1^\oplus}$. □

In the non-serial case the above theorem does not hold, as is shown in Example 5.1.

5 Examples

The first example provides a counter-example to Theorem 4.5 in the case of non-seriality of the underlying similarity spaces.

Example 5.1. Let us go back to Example 2.1. Assume that there are two objects: Ob and the robot Ro, and that the robot determines similarity $\sigma(Ob, Ob) = 1.0$, $\sigma(Ob, Ro) = \sigma(Ro, Ob) = 0.8$ and $\sigma(Ro, Ro) = 0.6$. Consider similarity spaces $\Sigma = \langle\{Ob, Ro\}, \sigma, 0.8\rangle$ and $\Sigma_1 = \langle\{Ob, Ro\}, \sigma, 1.0\rangle$. Σ is serial while Σ_1 is not. We have: $\{Ob\}_{\Sigma^+} = \{Ro\}$ and $\{Ob\}_{\Sigma_1^\oplus} = \{Ob\}$. Thus Ro is in $\{Ob\}_{\Sigma^+}$ but not in $\{Ob\}_{\Sigma_1^\oplus}$ which, in the non-serial case, falsifies Theorem 4.5 with $X = \{Ob\}$. □

Consider the heap example and the Sorites Paradox, widely discussed, also in [1].

---

[4] Thus, it is sufficient to require that σ is $\big(\max\{p, q\}\big)$-serial.

Example 5.2. Let *Heap*(n) be a predicate, denoting that n grains of sand make a heap. Assume that an agent is asked to recognize heaps, provided that the objective level definition of *Heap* is *Heap*$(n) = n \geq 100$. Assume further that the agent's subjective reality is modeled in terms of a perception filter given by the similarity space $\langle N, \sigma, 0.8\rangle$, where N is the set of natural numbers and

$$\sigma(k, m) = 1 - \frac{| k - m |}{\max\{1, k, m\}}\,.$$ [5]

Since σ is total, we have that $n_{+}^{\sigma^{0.8}}(i) = n_{\oplus}^{\sigma^{0.8}}(i) = \{j \mid \sigma(i, j) \geq 0.8\}$. We now calculate approximations of $[100, +\infty]$ which is the set of natural numbers satisfying the predicate *Heap*. Using Definition 4.1, we get $Heap_{\Sigma+} = [125, +\infty)$ and $Heap_{\Sigma\oplus} = [80, +\infty)$. Consequently, we can put $Heap_{\Sigma-} = N - Heap_{\Sigma\oplus} = [0, 79]$ and $Heap_{\Sigma\pm} = Heap_{\Sigma\oplus} - Heap_{\Sigma+} = [80, 125)$. $Heap_{\Sigma+}(n)$ implies $Heap_{\Sigma+}(n+1)$, but the converse implication is satisfied only when n is greater than or equal to a certain natural number, in our case, $n = 125$. Similarly, $Heap_{\Sigma\oplus}(n+1)$ implies $Heap_{\Sigma\oplus}(n)$ for $n \geq 80$.

The above induction is related to the Sorites Paradox. We share the opinion that the paradox results from mixing the objective and subjective levels. Quantitative measurements on the objective level are often not reflected by a qualitative change on the subjective level. One cannot expect a relatively small change, that cannot be observed due to perceptual limitations, to cause qualitative change at the agent's subjective level. "Grains of sand" are too small to register a qualitative change from non-heap to heap.

To be more specific, consider similarity $\sigma(1000, 1001) = 1 - 1/1001 = 0.999$ between 999 and 1000 grains of sand. Given particular perceptual limitations, say $p = 0.99$, we have that 999 and 1000 grains are indiscriminately different. One grain is too small "particle" to be recognized in this context and to make a qualitative change. □

In Example 5.2 we computed the subjective perception of *Heap* wrt the considered similarity space. One can additionally solve two other related tasks, where the subjective perception of *Heap* is understood according to Definition 4.4:

- given a subjective perception of *Heap* and a similarity space Σ, determine the objective level definition of *Heap*
- given an objective level definition of *Heap*, a subjective perception of *Heap* and a similarity space σ, determine the similarity threshold p of Σ,

Example 5.3. Assume that we compare the sweetness of coffee cups by comparing the number of sugar grains in each cup. Suppose further that objectively two cups containing k and m grains of sugar, respectively, are of the same sweetness, denoted by $k \sim n$, when for some natural number i, both k and m are in the same interval $[i*10, i*10+9]$. Clearly, $\sim$ is transitive. Suppose, a robot is given the task to measure sweetness, but has some subjective limitations pertaining to its sensors. Limitations are represented as the similarity space $\Sigma = \langle\{0, 1, 2, \ldots\}, \sigma, 0.6\rangle$, where σ is defined as in Example 5.2. Now, $\sigma(2, 3) = 2/3 \geq 0.6$, $\sigma(3, 5) = 0.6 \geq 0.6$ and $\sigma(2, 5) = 0.4 \not\geq 0.6$, i.e., $2, 3$ and $3, 5$ are of the same sweetness wrt Σ, whereas $2, 5$ are not. Thus, on the subjective level, transitivity does not apply. □

[5] This similarity function is just an example. However, it reflects the intuition that heaps, say, of 1000 and 1001 grains are more similar than heaps of, e.g., 100 and 101 grains.

Example 5.4. Let $Fast(c)$ be a predicate denoting that a car c's speed is very high. We assume here that each car is characterized by its speed, measured by a radar.[6] The speed of a car c is given by a function $S(c)$, whose value is in $[0, 200] \cup \Upsilon$. Assume that an agent is asked to identify fast cars, provided that the objective level definition of $Fast$, given particular road conditions, is $Fast(c) \equiv [S(c) \geq 80]$.

Suppose that the agent's subjective perceptual capability is modelled by a similarity space $\Sigma = \langle C, \sigma, 0.9 \rangle$, where C is the set of considered cars and

$$\sigma(c_1, c_2) \stackrel{\text{def}}{=} \begin{cases} 1 \; - \mid S(c_1) - S(c_2) \mid /200 & \text{when } S(c_1), S(c_2) \neq \Upsilon \\ 1 \quad \text{when } S(c_1) = S(c_2) = \Upsilon \\ \Upsilon \quad \text{otherwise.} \end{cases}$$

We have $n_{+}^{\sigma^{0.9}}(c) = \{c' \mid \sigma(S(c), S(c')) \geq 0.9\}$. In the same way we have $n_{\oplus}^{\sigma^{0.9}}(c) = n_{+}^{\sigma^{0.9}}(c) \cup \{c' \mid \sigma(S(c), S(c')) = \Upsilon\}$. We now compute approximations of the set $FC = \{c \in C \mid S(c) \geq 80\}$. Using Definition 4.1, we finally obtain that $FC_{\Sigma+} = \{c \in C \mid S(c) \geq 100\}$ and $FC_{\Sigma\oplus} = \{c \in C \mid S(c) \geq 60 \text{ or } S(c) = \Upsilon\}$. □

6 Complexity of the Approach

The approach we propose is tractable in the case of finite domains, as shown below.

Definition 6.1. *A similarity space $\Sigma = \langle U, \sigma, p \rangle$ is* tractable*, if U is finite and, for all $a, b \in U$, $\sigma(a, b)$ can be computed in deterministic polynomial time in the size of U.* □

Assume the considered relations are stored in relational or deductive databases tractable intensional part (e.g., expressed by the classical first-order rules or fixpoint calculus – see, e.g., [10]). Under such assumptions, we have the following proposition.

Proposition 6.2. *Let $\Sigma = \langle U, \sigma, p \rangle$ be a tractable similarity space. Then, for any set $S \subseteq U$, approximations $S_{\Sigma+}$ and $S_{\Sigma\oplus}$ are computable in deterministic polynomial time in the size of U.* □

In consequence, any query referring to $R_{\Sigma+}$ and $R_{\Sigma\oplus}$, where R is a relation on U, are computable in deterministic polynomial time in the size of U.

7 Relation to Other Approaches and Conclusion

The use of similarity spaces is a generalization of Pawlak's [6] pioneering work with rough sets where indiscernibility among individuals is modeled in terms of equivalence classes on feature/value pairs. In this paper, we also extend the results in [2,3,4,5]. The incentive for this generalization is due to the novel manner in which Halpern [1] approaches problems of intransitivity and vagueness. Fuzzy sets [11,12,13] provide another means of modeling vagueness and [14,15] provide insights into how one can formally translate between fuzzy and rough sets. These techniques can also be applied to our generalizations. There is also some relevant work outside the soft computing genre which attempts to provide methods and techniques for reasoning with approximate relations of which ([16,17,18,19,20]) are representative.

[6] Note that a radar may not be able to measure the speed of some cars.

References

1. Halpern, J.: Intransitivity and vagueness. In Dubois, D., Welty, C., Williams, M.A., eds.: Proc. of 9th Int. Conf. KR'2004, AAAI Press (2004) 121–129
2. Doherty, P., Szałas, A.: On the correspondence between approximations and similarity. In Tsumoto, S., Slowinski, R., Komorowski, J., Grzymala-Busse, J., eds.: Proc. of 4th Int. Conf. RSCTC'2004. Volume 3066 of LNAI., Springer-Verlag (2004) 143–152
3. Doherty, P., Łukaszewicz, W., Szałas, A.: Tolerance spaces and approximative representational structures. In Günter, A., Kruse, R., Neumann, B., eds.: Proc. 26th German Conf. on AI. Volume 2821 of LNAI., Springer-Verlag (2003) 475–489
4. Doherty, P., Łukaszewicz, W., Szałas, A.: Approximate databases and query techniques for agents with heterogenous perceptual capabilities. In: Proc. of the 7th Int. Conf. on Information Fusion, FUSION'2004. (2004) 175–182
5. Doherty, P., Łukaszewicz, W., Szałas, A.: Approximative query techniques for agents with heterogeneous ontologies and perceptive capabilities. In Dubois, D., Welty, C., Williams, M.A., eds.: Proc. of the 9th Int. Conf. KR'2004, AAAI Press (2004) 459–468
6. Pawlak, Z.: Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
7. Klir, G., Folger, T.: Fuzzy Sets, Uncertainty, and Information. Prentice Hall (1988)
8. Williamson, T.: Vagueness. Routledge (1994)
9. Skowron, A., Stepaniuk, J.: Tolerance approximation spaces. Fundamenta Informaticae **27** (1996) 245–253
10. Abiteboul, S., Hull, R., Vianu, V.: Foundations of Databases. Addison-Wesley (1996)
11. Zadeh, L.: Fuzzy sets. Information and Control **8** (1965) 333–353
12. Zadeh, L.: A new direction in AI: Toward a computational theory of perceptions. AI Magazine **22** (2001) 73–84
13. Murai, T., Kanemitsu, H., Shimbo, M.: Fuzzy sets and binary-proximity-based rough sets. Information Sciences **104** (1998) 49–80
14. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. International Journal of General Systems **17** (1990) 191–209
15. Dubois, D., Prade, H.: Putting rough sets and fuzzy sets together. In Slowinski, R., ed.: Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory, Kluwer (1992) 204–232
16. McCarthy, J.: Approximate objects and approximate theories. In Cohn, A., Giunchiglia, F., Selman, B., eds.: Proc. 7th Int. Conf. KR'2000, San Francisco, Ca., Morgan Kaufmann Pub., Inc. (2000) 519–526
17. Kautz, H., Selman, B.: Knowledge compilation and theory approximation. Journal of the ACM **43** (1996) 193–224
18. Lin, F.: On strongest necessary and weakest sufficient conditions. In Cohn, A., Giunchiglia, F., Selman, B., eds.: Proc. 7th Int. Conf. KR'2000, Morgan Kaufmann (2000) 167–175
19. Cadoli, M.: Tractable Reasoning in Artificial Intelligence. Volume 941 of LNAI. Springer-Verlag, Berlin Heidelberg (1995)
20. Doherty, P., Łukaszewicz, W., Szałas, A.: Computing strongest necessary and weakest sufficient conditions of first-order formulas. Proc. IJCAI'2001 (2001) 145 – 151