Similarity Relations and Independence Concepts

Frank Klawonn[†] and Rudolf Kruse[‡]

[†] Department of Computer Science, University of Applied Sciences, Wolfenbüttel, Germany

[‡] Department of Computer Science, Otto-von-Guericke University, Magdeburg, Germany

Abstract This paper addresses the definition of independence concepts in the context of similarity relations. After motivating the need for independence concepts basic ideas from similarity relations and their connections to fuzzy systems are reviewed. Three different independence notions are discussed and investigated in the framework of similarity relations. The results show that there are significant differences for independence concepts in a probabilistic setting and in the framework of similarity relations.

1 Introduction

Similarity is a very fundamental concept used in approximate and cased-based reasoning. There are many different ways to model similarity. In this paper we mainly focus on similarity as a dual concept to the notion of distance. When dealing with a real-valued attribute a distance function is an elementary notion, easy to define and to comprehend. As long as only a single variable is considered, the distance must take a context dependent scaling into account. However, when attribute vectors are used, it is usually not sufficient to aggregate the distances of the single attributes in an independent fashion. The overall distance or similarity of two elements that are described by the same vector of attributes, but with different values, crucially depends on the interaction and dependencies between the attributes. Within probability theory the notion of independence is a well defined and experienced concept. In other fields, related to but different from probability theory, like possibility theory (Bouchon-Meunier et al., 2004; De Cooman, 1997) or belief functions (Yaghlane et al., 2002) the definition of independence becomes difficult.

In this paper, we discuss the notion of independence in the context of distance-based similarity measures. To better illustrate the underlying questions and consequences, we use an interpretation of fuzzy systems based on similarity relations. We show that the concept of independence for similarity relations is crucial, but it is not at all obvious, how to define it. Certain definitions lead also to unusual properties like asymmetric independence.

The paper is organized as follows. Section 2 motivates the use and importance of independence considerations within modelling imperfect knowledge, especially in combination with data available for training or tuning model parameters. Section 3 briefly reviews basic ideas from similarity relations and explains their connection to fuzzy systems. General considerations about modelling independence under different aspects are discussed in section 4. The application of the considerations to similarity relations is investigated in section 5, before we come to the final conclusions in section 6.

2 Modelling Imperfect Knowledge Enhanced by Data

Classical two-valued logic and standard deduction system are designed to model crisp facts and perfect knowledge. Although this is suitable for certain applications, in knowledge-based systems it is very often desirable to include imperfect knowledge. It would take too much space to discuss all facets of imperfect and uncertain knowledge. A main characteristic is that numbers or weights are assigned to propositions, events or statements. Of course, the meaning of the numbers is crucial and determines how to operate with the imperfect knowledge and the assigned weights or numbers. Probability theory provides the most popular model. It provides only an abstract framework that leaves space for an interpretation. The frequentistic view of probabilities is probably the most common one. The numbers or weights – in this context they are called probabilities – represent relative frequencies of events in experiments that are assumed to be be repeatable "arbitrarily often" in an "independent" manner. Although this seems to be appealing and intuitive, it has certain problems and limitations. Other interpretations in terms of subjective probabilities within a framework of rational betting behaviours (see for instance (O'Hagan and Forster, 2004)) or in a game-theoretic setting (Shafer, 2006) put a stronger emphasis on the evaluation of knowledge and experience that does not have to be based on observations in terms of counting relative frequencies.

Other examples are belief functions within Dempster-Shafer theory (Shafer, 1976) or within the transferable belief model (Smets and Kennes, 1994), possibility theory (Dubois and Prade, 2001) or preferences. In all these models, the interpretation of the weights or numbers determines how to operate with them. Fuzzy systems are an example where the the interpretation of the weights – in this case degrees of membership – is not straightforward in most cases and sometimes the choice of operations looks very heuristic or even arbitrary.

Any interpretation of probabilities or degrees of membership makes certain assumptions about rational behaviour concerning the specification of the weights. Here, the term weight is used, since, depending on the interpretation, these numerical values might represent probabilities, confidence or truth degrees and the term "'weight"' is neutral without referring to a specific interpretation. Even though the underlying justification for assuming a certain rational behaviour might be plausible, in most cases human experts are very often unable to specify the required weights in a consistent way, when the application becomes more complex. Human experts can often provide important prior or meta information on structures, dependencies and qualitative judgements. However, when exact quantifications are needed, the experts might not be able to specify unique values.

Therefore, it is very common to couple expert knowledge with data, so that the structural model information is provided by the expert, whereas the fine tuning of the model is carried out based on the available data. A very common way to handle this estimation of the model parameters is to formulate an optimization problem where the model parameters should be determined in such a way that the model fits[1] best to the data. Figure 1 illustrates this approach. The model parameters must be tuned in such a way that the given input data produce the desired output data with minimal error. Difficulties arise here, when there is no analytical or obvious way to optimize the model parameters parameters in case of a large number of variables. For more complex models this is almost always the case. Then parameter optimization can become an extremely complex or almost impossible task.

It is interesting to note that at least certain models have found a way out of this problem. Graphical models (for overviews see for instance (Borgelt and Kruse, 2002; Cowell et al., 2003; Cox and Wermuth, 1996)) and specifically Bayesian networks describe the dependence or independence structure of variables in the form of of an acyclic and directed graph. In this way a probability distribution over a high-dimensional variable space can be decomposed into a number of marginal and conditional distributions over low-

[1] Fitting the model should also include model validation by techniques like cross-validation or the minimum description length principle (see for instance Grünwald (2007)) in order to avoid overfitting.

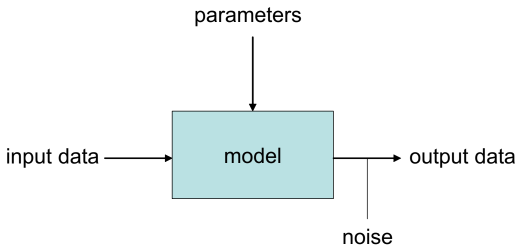

Figure 1. Open loop system for learning as an optimization problem.

dimensional spaces. Figure 2 illustrates the graphical structure of a Bayesian network. Given the dependency or model structure, the parameters of the Bayesian network are not learned according to the strategy illustrated in Figure 1 to optimize the input-output behaviour[2] of the Bayesian network with respect to the given data. Instead, the Bayesian network learns or estimates the low-dimensional probability distributions from the data. In this sense, the dependence or independence structure of the Bayesian network allows local computations of the parameters without taking their influence on the whole model into account.

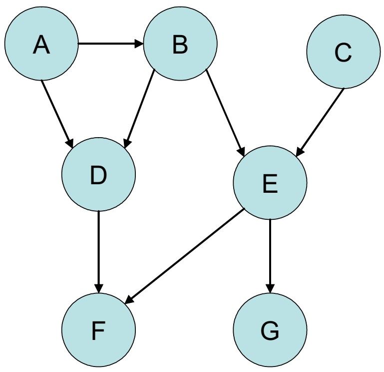

Figure 2. A Bayesian network for local learning.

The graphical structure of the Bayesian network specifies (conditional) independencies of variables. Therefore, using the information about inde-

[2] For a Bayesian network there are even no specific input or output variables, even though it can be used in this way.

pendence or conditional independence can lead to efficient and simplified parameter learning schemes. The remaining part of this paper explains the difficulties that arise, when similar ideas are applied in the context of similarity relations.

3 Similarity Relations and Fuzzy Systems

Here, we focus on a specific type of similarity relation. A very common definition of a similarity relation $s : X \times X \to [0,1]$ on a set X, where $s(x,y)$ expresses the similarity between elements x and y, requires the following properties:

(a) Reflexivity: $s(x,x) = 1$
(b) Symmetry: $s(x,y) = s(y,x)$
(c) Transitivity: $s(x,y) * s(y,z) \leq s(x,z)$

where $*$ is a suitable t-norm[3]. Besides the name similarity relation (Zadeh, 1971; Ruspini, 1991), depending on the choice of the operation $*$, s is also called an indistinguishability operator (Trillas and Valverde, 1984), fuzzy equality (relation) (Höhle and Stout, 1991; Klawonn and Kruse, 1993), fuzzy equivalence relation (Thiele and Schmechel, 1995) or proximity relation (Dubois and Prade, 1994). Reflexivity is an obvious property. Symmetry is very often, though not always, a canoncial property as well. Whether transitivity is required for similarity relations is sometimes questioned (De Cock and Kerre, 2003; Klawonn, 2003). However, in this paper we want to focus on similarity relations that satisfy a specific type of transitivity, namely transitivity with respect to the Łukasiewicz t-norm defined as $\alpha * \beta = \max\{\alpha + \beta - 1, 0\}$.

A similarity relation with respect to the Łukasiewicz t-norm can be viewed as a dual concept to a metric. In the following, we only consider similarity relations with respect to the Łukasiewicz t-norm and will not mention this fact explicitly each time. Such a similarity relation induces a (pseudo-)metric $\delta_s(x,y) = 1 - s(x,y)$ bounded by one and vice versa, any (pseudo-)metric δ_s bounded by one induces a similarity relation by $s_\delta(x,y) = 1 - \delta(x,y)$. The restriction that the metric is bounded by one is more or less neglectable, since any metric δ can be bounded by one, simply by defining $\bar{\delta}(x,y) = \min\{\delta(x,y), 1\}$ without affecting small distances that are usually of main interest.

[3] A t-norm is an associative and commutative operation on the unit interval that is nondecreasing in its arguments and has one as a unit element. For example, the product or the minimum are t-norms.

When dealing with real numbers real numbers metric distances is an elementary concept leading to the canonical metric $\delta(x,y) = |x-y|$.

Extensionality is a very simple concept to take similarity between elements into account during a reasoning process. Extensionality means that similar elements should lead to similar results. The extensionality property we need here, is the extensionality of sets. For an ordinary set M we have the trivial property

$$x \in M \wedge x = y \;\Rightarrow\; y \in M. \tag{1}$$

When we replace equality by similarity, this simple property translates to: If element x belongs to the set M and x and y are similar to a certain degree, then y should also belong to M to a certain degree. Since similarity is a matter of degree, the property that y belongs to M should also be a matter of degree. Therefore, it is necessary to let M be a fuzzy set, i.e. elements do not simply belong or do not belong to M, but have a membership degree to M. A fuzzy set $\mu : X \to [0,1]$ is said to be extensional with respect to the similarity relation s on X, if it satisfies

$$\mu(x) * s(x,y) \;\leq\; \mu(y)$$

for all $x, y \in X$. This extensionality property is an extension of the simple property (1) for equality to similarity relations. In the presence of a similarity relation intending to model indistinguishability between elements, a (fuzzy) set should be consistent, i.e. extensional with respect to the given similarity relation. When we refer to an element x in the presence of a similarity relation, we might actually refer to x or a similar element. The (fuzzy) set of elements similar to x is the extensional hull of the set $\{x\}$, i.e. the smallest extensional fuzzy set containing x, i.e. the fuzzy set that contains x and all elements similar to x:

$$\mu_x(y) \;=\; s(x,y).$$

More generally, the extensional hull $\hat{\mu}$ of a fuzzy set μ is the smallest extensional fuzzy set containing μ given by

$$\hat{\mu}(y) \;=\; \bigvee_{x \in X} (\mu(x) * s(x,y))$$

which can be read as

$$y \in \mu \;\Leftrightarrow\; (\exists x \in X)(x \in \mu \wedge x \approx y).$$

The extensional hull of an ordinary set is the extensional hull of its indicator function and can be understood as the (fuzzy) set of points that are similar to at least one element in the set.

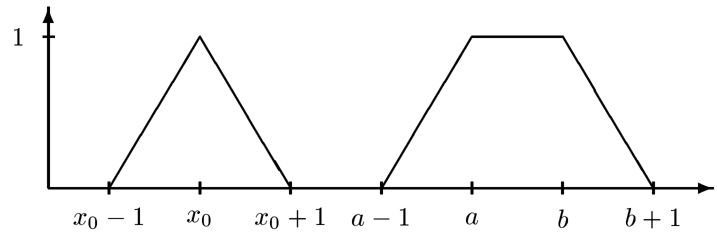

Figure 3. The extensional hulls of the point x_0 and of the interval $[a, b]$.

As an example consider the similarity relation $s(x, y) = 1 - \min\{|x - y|, 1\}$ on the real numbers. The extensional hulls of a single point and an interval are shown in Figure 3.

It is noteworthy that these extensional hulls lead to triangular and trapezoidal fuzzy sets that are very common in fuzzy systems. Extensional hulls of points with respect to the similarity relation

$$s(x, y) = 1 - \min\{|x - y|, 1\}$$

always have a support of length two. In order to maintain the degrees of similarity (and the membership degrees), when changing the measurement unit (seconds instead of hours, miles instead of kilometres,...), we have to take a scaling into account:

$$s(x, y) = 1 - \min\{c \cdot |x - y|, 1\}$$

When we take a closer look at the concept of similarity relations, we can even introduce a more general concept of scaling. Similarity relations can be used to model indistinguishability. There are two kinds of indistinguishability, we have to deal with in typical similarity-based reasoning applications.

Enforced indistinguishability is caused by limited precision of measurement instruments, (imprecise) indirect measurements, noisy data, ...

Intended indistinguishability means that the human expert is not interested in more precise values, since a higher precision would not really lead to improved results.

Both kinds of indistinguishability might need a local scaling as the following example of designing an air conditioning system shows.

temperature (in °C)	scaling factor	interpretation
< 15	0.00	exact value meaningless (much too cold)
15-19	0.25	too cold, but not too far away from the desired temperature, regulation need not be too sensitive
19-23	1.50	very sensitive, near the optimal value
23-27	0.25	too warm, but not too far away from the desired temperature, regulation need not be too sensitive
> 27	0.00	exact value meaningless (much too hot)

When we apply these different scaling factors to our temperature domain, this has the following consequences, when we consider the similarity relation induced by the scaled distance. In order to determine how dissimilar two temperatures are, we do note compute their difference directly, but in the scaled domain, where the range up to 15 is shrunk to a single point, the range between 15 and 19 is shrunk by the factor 0.25, the range between 19 and 23 is stretched by the factor 1.5 and so on. The following table shows the scaled distances of some example values for the temperature.

pair of values (x, y)	scal. factor c	transformed distance $\delta(x, y) = \lvert c \cdot x - c \cdot y \rvert$	similarity degree $s(x, y) = 1 - \min\{\delta(x, y), 1\}$
(13,14)	0.00	0.000	1.000
(14,14.5)	0.00	0.000	1.000
(17,17.5)	0.25	0.125	0.875
(20,20.5)	1.50	0.750	0.250
(21,22)	1.50	1.500	0.000
(24,24.5)	0.25	0.125	0.875
(28,28.5)	0.00	0.000	1.000

Figures 4 and 5 show examples of extensional hulls of single points.

The idea of piecewise constant scaling functions can be extended to arbitrary scaling functions in the following way (Klawonn, 1994). Consider an integrable scaling function $c : R \to [0, \infty)$, where c is a function $c(x)$, not a constant like c before. If we assume that we have for small values $\varepsilon > 0$ that the transformed distance between x and $x + \varepsilon$ is given by

$$\delta(x, x + \varepsilon) \;\approx\; c(x) \cdot \varepsilon,$$

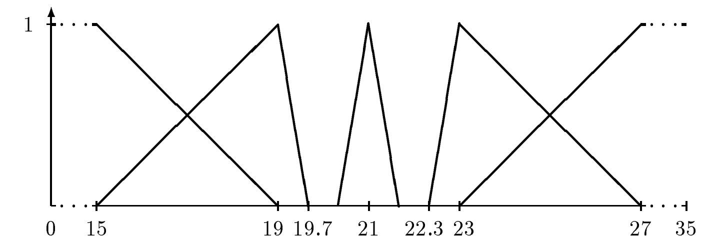

Figure 4. The extensional hulls of the points 15, 19, 21, 23 and 27.

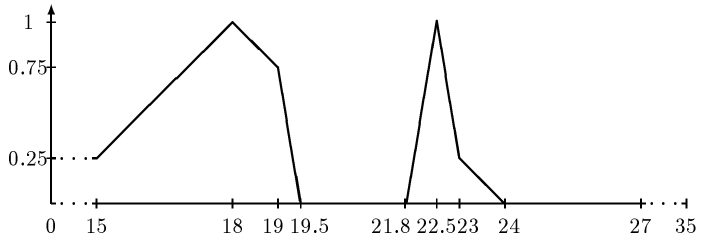

Figure 5. The extensional hulls of the points 18.5 and 22.5.

then the transformed distance induced by the scaling function c can be computed by

$$\delta(x,y) \;=\; \left|\int_x^y c(s)\,ds\right|.$$

This idea of scaling functions can exploited to derive a simplified learning scheme for fuzzy rule-based systems (Klawonn, 2006). The fuzzy sets in the body of a rule represent extensional hulls of single points and the fuzzy rule-based system constructs an interpolating function based on these nodes or sampling points, taking the underlying similarity relation into account. In this sense, the body of a fuzzy rule is nothing else than a single value including the similar values in terms of the similarity relation. Prototype-based fuzzy clustering (for an overview see for instance (Höppner et al., 1999)) is based on very similar ideas. A cluster is represented by a single point – the prototype – and the membership degree of a data point to the cluster decreases with increasing distance to the cluster. Specialized algorithms as proposed in (Klawonn and Kruse, 1997; Keller and Klawonn, 2000; Borgelt,

2005) use even a scaling concept for the attributes. However, these scalings are carried out more or less independently on the single variables. Some fuzzy clustering algorithms (Gustafson and Kessel, 1979; Gath and Geva, 1989) take also dependencies between the variables into account and use in addition to the scaling a rotation in order to compute similarities or distances.

Considering fuzzy rule-based systems in the context of similarity-based reasoning, a body of a rule represents a point in the – usually – multidimensional input space. When a concrete (multidimensional) measured input is given, the similarity of the measured value to the value representing the body of the rule must be determined. This is usually done by first determining the membership degrees of each variable value to the corresponding fuzzy set. These membership degrees correspond to the similarities of the values of the single variables. In order to compute the overall similarity degree – the degree to which the corresponding fuzzy rule is applicable – these membership degrees are normally aggregated by a suitable operation, usually a t-norm like the product or the minimum. In terms of the similarity relations this means that the similarity relations on the single variables are aggregated to an overall similarity relation on the product space of the variables. This kind of aggregation requires an independence assumption for the similarity relations, since it is assumed that the overall similarity can be derived solely from knowing the similarity degree of each variable without referring to single values.

In order to illustrate that this assumption is unrealistic in many cases, we consider a typical control task. The aim is to balance an inverted pendulum that is fixed on a cart that can be driven forward and backward to fulfil the task (see Figure 6). As input variables we use the deviation of the angle of the pendulum from the upright position e and the angle velocity Δe.

With this simple example, we can easily demonstrate the dependency between the similarity relation on the variables e and Δe. Let us first consider the situation on the left hand side of Figure 6. The inverted pendulum is almost in an upright position. In this case, it is very important to have a more or less precise value of Δe in order to know whether the inverted pendulum tends to fall down to the right, will overshoot to the left or remain more or less stable in its current position. This means that the similarity relation on the domain of the variable Δe must be fine granular.

The situation on the right hand side of Figure 6 is completely different. The inverted pendulum has fallen down almost completely and a strong control reaction has to be carried out in order to get it closer to the upright position. In this case, the actual value of Δe does not matter much at all. This means that a very coarse similarity relation on Δe would be sufficient.

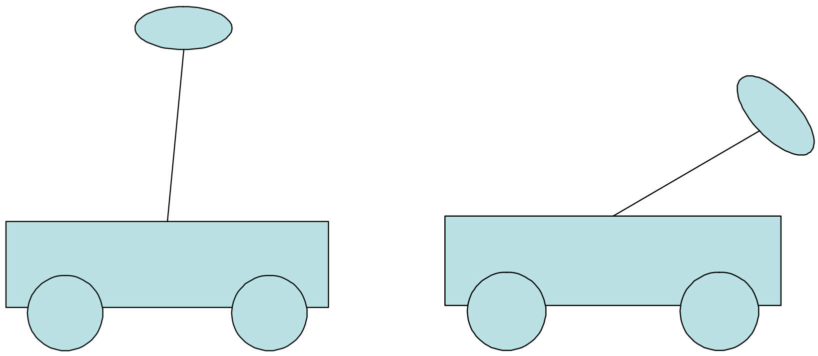

Figure 6. Balancing an inverted pendulum.

In other words, the similarity on Δe depends on the value of the variable e. When e is small, a refined similarity relation, distinguishing more between values, is needed on the domain of the variable Δe than in the case, when e is large.

Although this simple example clarifies in an intuitive manner that interaction between similarity relations should not be ignored, it does, however, not provide a formal definition what dependence or independence of similarity relations means. The following section discusses the notion of independence from a more general point of view in order to apply it in Section 5 to the context of similarity relations.

4 Independence Concepts

In this section we take a more general view on the formal concept of independence, before we apply it to similarity relations in the following section. We consider the notion of independence in a more general framework of product spaces. Independence involves (at least) two variables or domains. The variables can be considered separately or in combination. Independence intuitively refers to the property that it is sufficient to measure the variables separately and then to combine the results of these marginal measurements. For instance, in probability theory independence of two random variables X and Y means the following. If we want to know the probability $P(X \in A, Y \in B)$ for two combined events, i.e. Borel-measurable sets A and B, in the case of independence we can first compute the marginal probabilities $p_A = P(X \in A)$ and $p_B = P(Y \in B)$ and then combine these two probabilities (by the product). However, when the variables are dependent, we cannot deduce the joint distribution $P_{X,Y}(X \in A, Y \in B)$ knowing only p_A and p_B for all Borel-measurable sets A and B.

The general view that we take here is the following. We have two sets, each of them endowed with a structure for modelling uncertainty or similarity. In the case of probability theory this structure would be a σ-algebra together with a probability measure. In the context of modelling similarity it would simply be a similarity relation. We also consider the cartesian product of these two sets. What would be the resulting structure on the cartesian product when we combine the (marginal) structures on the two sets in an "independent" manner?

Category theory provides one framework to formalize these ideas. A formal introduction of the notion of a category would be far beyond the scope of this paper. Therefore, we give a more informal description of the underlying concepts. A category can be considered as a class of objects and morphisms between the objects. Very often, the objects are sets with an additional structure (for instance, algebraic structures like groups, topological spaces, measure spaces or probability spaces). In this case the morphism are structure preserving mappings between the elements, homomorphisms in the case of algebraic structures, continuous mappings for topological spaces or measurable mappings for measure spaces. Such categories are called concrete categories (Adamek et al., 1990). The product $X = X_1 \times X_2$ together with two projections (morphisms) $\pi_i : X \to X_i$ $(i = 1, 2)$ of two objects X_1 and X_2 in a category is characterized by the following property. For any object Y and two morphisms $f_i : Y \to X_i$ $(i = 1, 2)$ there is a morphism f such that the diagram in Figure 7 commutes, i.e. $\pi_i \circ f = f_i$ $(i = 1, 2)$ holds. It can be shown that the product or product space X in a category is unique up to isomorphism in case of its existence.

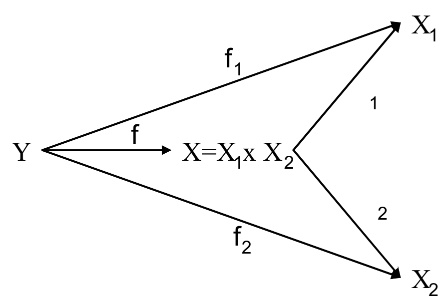

Figure 7. The product space as a limit in a category.

If product structures in the sense of category theory are available, one possible definition of independence is the following. A structure (probability measure or similarity relation) on a product space is considered to be composed of independent components, if the product of its projections yields the same structure on the product space.

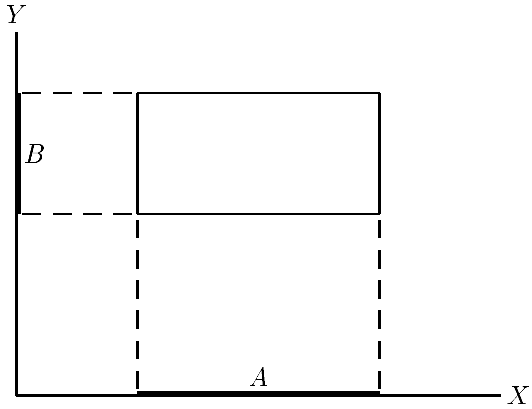

Figure 8. Reconstructing the original structure from its projections.

Figure 8 illustrates this concept in the context of sets without structure. A subset of a Cartesian product space can be reconstructed by its projections, if and only if it is a rectangle.

In the case of a probability measure on the product space $X \times Y$, this independence property is equivalent to $P(A \times B) = P(A \times Y) \cdot P(X \times B)$.

Another formalization of independence is the following one: No matter at which point the projection is carried out, the projection will always be the same. Figure 9 illustrates this idea for sets without structure. It means that for a given subset $S \subset X \times Y$ of a product space, the sets $\{y \in Y \mid (x_0, y)\}$ are either empty or identical independent of the choice of x_0. The same applies to the projection sets $\{x \in X \mid (x, y_0)\}$.

In the probabilistic setting, this means that $P(Y|X = x_0)$ is independent of the choice of x_0 and vice versa, $P(X|Y = y_0)$ is independent of the choice of y_0. In the case of probability this leads again to the classical definition of independence.

Note that this definition contains two parts: $P(Y|X = x_0)$ does not depend on any x_0 means that Y is independent of X, whereas $P(X|Y = y_0)$ is independent of the choice of y_0 says that X is independent of Y. In the context of probability theory we can only have both properties or none of the two properties. In the next section we will see that this is not the case

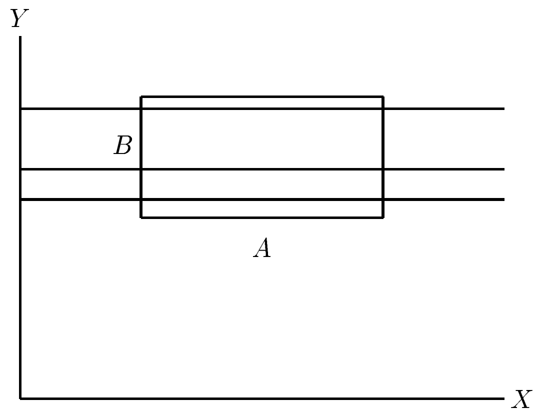

Figure 9. Projections at different positions.

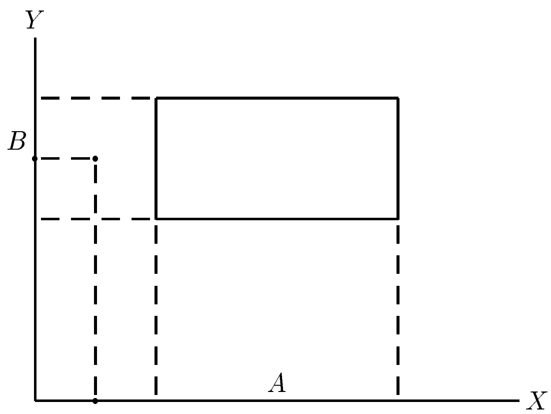

Figure 10. Deriving information from the projections.

for similarity relations.

Another possible notion of independence is the following one: The complete information for every point in the product space can be reconstructed from the projections. Figure 10 illustrates this idea for sets without structure.

In a probabilistic setting this can be viewed as a more general independence concept, at least when cumulative distribution functions are considered. Using copulas[4] (see for instance (Nelsen, 1999)), arbitrary two-

[4]A copula is a cumulative distribution function $C : [0,1]^n \to [0,1]$ with uniform distri-

dimensional cumulative distribution functions can be constructed. In this sense, independence would hold for any two cumulative distribution functions.

5 Similarity Relations and Independence

Since fuzzy rule-based systems usually use multiple inputs, it is necessary to combine the similarity relations on the single domains to an overall similarity relation in the product space. The canonical similarity relation on a product space – at least in the sense of category theory – is given by

$$s((x_1,\dots,x_p),(x'_1,\dots,x'_p)) \;=\; \min_{i\in\{1,\dots,p\}}\{s_i(x_i,x'_i)\}.$$

In terms of fuzzy rule-based systems this means that for a single rule, the membership degrees of an input would be combined using the minimum.

Assume there is a similarity relation v (or any other structure) on a product space $X\times Y$ with

$$\begin{aligned}\pi_X(v) &= s, \qquad \pi_X(v)(x_1,x_2)=\bigvee_{y\in Y} v((x_1,y),(x_2,y))\\ \pi_Y(v) &= t\end{aligned}$$

where s and t are the projections of v onto X and Y, respectively. What is the meaning of s and t being independent?

The first definition of independence in the previous section based on category theory means that s and t are independent, if and only if $v = \min\{s,t\}$ holds.

The second definition was asymmetric and leads indeed to an asymmetric property for similarity relations. To see this, consider the product space $[0.5,1]\times[0.5,1]$ and the metric defined by the transformation

$$t(x,y):[0.5,1]\times[0.5,1]\to[0.5,1]\times[0.25,1], \qquad (x,y)\mapsto(x,xy)$$

The distance between two points (x_1,y_1) and (x_2,y_2) is the (product space metric) distance between the transformed points

$$\max\{t(x_1,y_1),t(x_2,y_2)\}.$$

- $\delta((x_1,y_1),(x_2,y_2)) \;=\; \max\{|x_1-x_2|,|x_1y_1-x_2y_2|\}$
- $\delta_X(x_1,x_2) \;=\; |x_1-x_2| \;=\; \delta^{(y)}(x_1,x_2)$

butions on the unit interval as its one-dimensional marginal distributions.

- $\delta_Y(y_1, y_2) \;=\; |y_1 - y_2|$
- $\delta^{(x)}(y_1, y_2) \;=\; x|y_1 - y_2|$

This example shows that the similarity relation (or, dually, the metric) on X is independent of the similarity relation on Y, but not vice versa.

Finally, let us consider the last independence definition of the previous chapter. For similarity relations this independence notion means that there is a function

$$h : [0,1] \times [0,1] \to [0,1]$$

satisfying

$$h(1,1) \;=\; 1 \quad \text{and} \quad h(\alpha, \alpha') * h(\beta, \beta') \leq h(\alpha * \alpha', \beta * \beta')$$

such that

$$v((x,y),(x',y')) \;=\; h(s(x,x'), t(y,y')).$$

holds.

Although this independence notion is weaker than the first one, it does not lead to independence in all cases as in the probabilistic setting with copulas. Consider the product space $[0.5, 1] \times [0.5, 1]$ and the metric defined by the transformation

$$t(x,y) : [0.5,1] \times [0.5,1] \to [0.25,1], \qquad (x,y) \mapsto xy$$

The metric is

$$\delta((x_1,y_1),(x_2,y_2)) \;=\; |x_1 y_1 - x_2 y_2|.$$

We obtain

$$\begin{aligned}
\delta((0.8,0.5),(0.9,0.6)) &= 0.14 \\
\delta((0.8,0.8),(0.9,0.9)) &= 0.17 \\
\delta_X(0.8,0.9) &= 0.1 \\
\delta_Y(0.5,0.6) &= 0.1 \\
\delta_Y(0.8,0.9) &= 0.1
\end{aligned}$$

This example shows that we cannot reconstruct the first two different similarity degrees (or, dually, the distances) in the product space based only on the similarities in the projection spaces, since the latter ones are all identical.

6 Conclusions

In this paper, we have discussed the need for independence concepts for similarity relations and have investigated different approaches that lead to surprising and different results, especially when compared to the probabilistic setting. Further research is needed to make use of the independence concepts within applications of similarity-based reasoning. Especially, tests for independence need to be developed outside of the scope of statistics.

Bibliography

J. Adamek, H. Herrlich, and G.E. Strecker. *Abstract and Concrete Categories: The Joy of Cats.* Wiley, Chichester, 1990.

C. Borgelt. *Prototype-based Classification and Clustering.* PhD thesis, University of Magdeburg, 2005.

C. Borgelt and R. Kruse. *Graphical Models – Methods for Data Analysis and Mining.* Wiley, Chichester, 2002.

B. Bouchon-Meunier, G. Coletti, and C. Marsala. Independence and possibilistic conditioning. *Annals of Mathematics and Artificial Intelligence*, 35:107–123, 2004.

R.G. Cowell, A.P. Dawid, S.L. Lauritzen, and D.J. Spiegelhalter. *Probabilistic Networks and Expert Systems.* Springer, Berlin, 2003.

D.R. Cox and N. Wermuth. *Multivariate Dependencies: Models, Analysis and Interpretation.* Chapman & Hall/CRC, Boca Raton, 1996.

M. De Cock and E.E. Kerre. Why fuzzy t-equivalence relations do not resolve the poincaré paradox, and related issues. *Fuzzy Sets and Systems*, 133:175–180, 2003.

G. De Cooman. Possibility theory III: possibilistic independence. *International Journal of General Systems*, 25:353–371, 1997.

D. Dubois and H. Prade. Possibility theory, probability theory and multiple-valued logics: A clarification. *Annals of Mathematics and Artificial Intelligence*, 32:35–66, 2001.

D. Dubois and H. Prade. Similarity-based approximate reasoning. In J.M. Zurada, R.J. Marks II, and C.J. Robinson, editors, *Computational Intelligence Imitating Life*, pages 231–256. IEEE Press, New York, 1994.

I. Gath and A.B. Geva. Unsupervised optimal fuzzy clustering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11:773–781, 1989.

P.D. Grünwald. *The Minimum Description Length Principle.* MIT Press, Cambridge, MAssachusetts, 2007.

D.E. Gustafson and W.C. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *IEEE CDC*, pages 761–766, San Diego, 1979.

U. Höhle and L.N. Stout. Foundations of fuzzy sets. *Fuzzy Sets and Systems*, 40:257–296, 1991.

F. Höppner, F. Klawonn, R. Kruse, and T. Runkler. *Fuzzy Cluster Analysis.* Wiley, Chichester, 1999.

A. Keller and F. Klawonn. Fuzzy clustering with weighting of data variables. *Uncertainty, Fuzziness and Knowledge-Based Systems*, 8:735–746, 2000.

F. Klawonn. Should fuzzy equality and similarity satisfy transitivity? *Fuzzy Sets and Systems*, 133:175–180, 2003.

F. Klawonn. Reducing the number of parameters of a fuzzy system using scaling functions. *Soft Computing*, 10:749–756, 2006.

F. Klawonn. Fuzzy sets and vague environments. *Fuzzy Sets and Systems*, 66:207–221, 1994.

F. Klawonn and R. Kruse. Equality relations as a basis for fuzzy control. *Fuzzy Sets and Systems*, 54:147–156, 1993.

F. Klawonn and R. Kruse. Constructing a fuzzy controller from data. *Fuzzy Sets and Systems*, 85:177–193, 1997.

R.J. Nelsen. *An Introduction to Copulas.* Springer, Berlin, 1999.

A. O'Hagan and J. Forster. *Kendall's Advanced Theory of Statistics, Volume 2B: Bayesian Inference.* Arnold, London, 2004.

E.H. Ruspini. On the semantics of fuzzy logic. *Intern. Journ. of Approximate Reasoning*, 5:45–88, 1991.

G. Shafer. The game-theoretic framework for probability. In B. Bouchon-Meunier, J. Gutierrez-Rios, and L. Magdalena, editors, *Proceedings of the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 3–10, Paris, 2006. Éditions E.D.K.

G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, 1976.

P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.

H. Thiele and N. Schmechel. The mutual defineability of fuzzy equivalence relations and fuzzy partitions. In *Proc. Intern. Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and the Second International Fuzzy Engineering Symposium*, pages 1383–1390, Yokohama, 1995.

E. Trillas and L. Valverde. An inquiry into indistinguishability operators. In H.J. Skala, S. Termini, and E. Trillas, editors, *Aspects of Vagueness*, pages 231–256. Reidel, Dordrecht, 1984.

B.B. Yaghlane, P. Smets, and K. Mellouli. Independence concepts for belief functions. In B. Bouchon-Meunier, J. Gutierrez-Rios, and L. Magdalena, editors, *Technologies for constructing intelligent systems 2: tools*, pages 45–58. Physica, Heidelberg, 2002.

L.A. Zadeh. Similarity relations and fuzzy orderings. *Information Sciences*, 3:177–200, 1971.