Similarity Metrics: A Formal Unification of Cardinal and Non-Cardinal Similarity Measures*

Hugh Osborne and Derek Bridge

University of York
U.K.

Abstract. In [9] we introduced a formal framework for constructing ordinal similarity measures, and suggested how this might also be applied to cardinal measures. In this paper we will place this approach in a more general framework, called similarity metrics. In this framework, ordinal similarity metrics (where comparison returns a boolean value) can be combined with cardinal metrics (returning a numeric value) and, indeed, with metrics returning values of other types, to produce new metrics.

1 Introduction

In this paper we present a formal framework for the construction of *similarity metrics*, which subsume a number of ways of measuring similarity. In particular similarity measures that return boolean values, numeric values and structured data can all be modelled by similarity metrics.

1.1 An Example Case Base

We shall use the small example case base in Fig. 1 throughout this paper. Each case represents a holiday and has three attributes — the destination, the price and the activities available. In practice, a case could be a more complex structure than the simple tuples of the example. Our framework anticipates this, and we explain, in Sect. 3.2, how more complex case representations are accommodated. We use tuples in the example only because it makes for an easier exposition.

Our presentation assumes that the similarity metric is applied to the whole case base. In those case-based systems where case base interrogation is a two-stage process [1, 7] (a retrieval step and then a case selection step that applies a similarity measure only to the subset of the case base that has been retrieved) our framework can be used in the second of the two stages.

1.2 Similarity Measures Returning Booleans

In [9], to complement numeric-valued similarity measures, we introduced a formal framework for constructing ordinal (boolean-valued) similarity measures, which

* To appear in the proceedings of the 2nd International Conference on Case-based Reasoning, ICCBR-97

Dest	Price	Act
And (*Andalucia*)	500	{*golf*,*swimming*}
Ben (*Benidorm*)	350	{*swimming*}
Cre (*Crete*)	750	{*swimming*,*golf*,*windsurfing*}
Cre (*Crete*)	500	{*swimming*,*windsurfing*}
Dor (*Dordogne*)	400	{*golf*}

Fig. 1. An example case base.

can be used in situations where cardinal information [13] is not available or is likely to be misleading.

A retrieval request was presented as a pair, comprising a similarity measure and a 'seed'. The values in the seed were the values against which elements of cases in the case base were compared by the similarity measure.

Most typically, constructing a retrieval request in our previous framework would begin by defining orders on the domains of each of the relevant attributes of the cases. Some domains might have an existing order (e.g. **Price**, being numeric, has the existing total order $\leq$); for other domains (e.g. ones with symbolic values) we showed ways in which orders could be defined (e.g. a user-defined order would be the most suitable way to rank values in **Dest** — see e.g. Fig. 2).

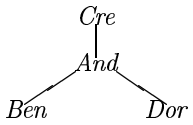

Fig. 2. A possible partial order — $\sqsubseteq_{\mathbf{Dest}}$ — for destinations.

We provided a set of operators for producing new orders from these 'underlying' orders. Orders on the domains of individual attributes could be extended to orders on whole tuples. A number of different orders on tuples, e.g. one per relevant attribute, would be combined using one of a number of connectives we defined (conjunction, disjunction and prioritisation), to give a new single order, and it would be this order that would be applied to the case base and the maxima taken.

Our new approach, using similarity metrics in place of orders, is similar in construction. Metrics on the domains of individual attributes can be imputed to whole tuples, and may be combined in various ways to give new metrics. These metrics can be applied to the case base and the maxima taken.

1.3 Similarity Measures Returning Other Types

The ordinal (boolean-valued) measures we introduced in [9] were never intended to supplant, but merely to complement cardinal measures. In recognition of this,

in [9] we defined numeric-valued similarity measures in a manner analogous to the way we had defined boolean-valued ones. The operators were redefined for the numeric case and different connectives were used (e.g. weighted addition and multiplication); there were no straightforward correspondences to the connectives of the ordinal framework (because cardinal and ordinal measures were not instances of a single framework). There were no connectives allowing cardinal and ordinal measures to be combined (not least since the result of such a combination would not have fallen within either the ordinal or the cardinal framework). Recent work [6, 11, 12] has proposed the use of similarity measures that are neither boolean- nor numeric-valued. It is desirable that these measures also be part of a common framework. This is also achieved by our framework. See [10] for details.

1.4 Similarity Metrics: A Single Framework

The work reported in this paper has three inter-related advantages. It proposes a framework in which different types of similarity metrics (returning boolean values, numeric values or values of some other type) are all instances of a single framework. Secondly, the connectives we define for combining metrics to produce new metrics not only allow us to combine metrics that return values of the same type, e.g. two or more boolean-valued metrics, but also allow combination of metrics that return values of different types, e.g. a boolean-valued metric with a numeric-valued metric. In contrast to our earlier framework, our new framework can combine such measures without the need to inter-convert. And finally, the framework provides a richer set of connectives than we had in [9].

As explained in detail in the remainder of the paper, a metric is a function that returns values from some *lattice*. Appropriate lattices include those defined on booleans, on numbers, on structured data values, on pairs of booleans, on pairs of booleans and numbers, and so on. This generalises many definitions of similarity. We do not need special definitions of operators for, e.g., numeric-valued metrics: a single set of definitions applies to all metrics, irrespective of the result lattice. And a metric of any type may be combined, using connectives we define, with a metric of any other type, and the result will still be a metric.

Proofs of the work reported in this paper can be found in [10]. Since metrics are lattice-valued functions, we begin our detailed explanation by reminding the reader of some lattice definitions.

2 Lattices

Recall that a partial order is a reflexive, transitive, anti-symmetric relation.

Definition 1 *A* lattice *is a partially ordered set* $(S, \sqsubseteq)$ *with the property that for all elements* $x, y \in S$*,* x *and* y *have a* least upper bound*,* $x \sqcup y$*, and a*

greatest lower bound $x \sqcap y$*. The least upper bound and greatest lower bound are both unique*[2]*.*

A complete *lattice is a lattice where all subsets have a least upper bound and a greatest lower bound*[3]*.*

Figure 3 contains some examples of standard (complete) lattices. This figure also introduces some notation that we will use throughout, e.g. $\mathcal{L}(\mathbf{Bool})$ for the Boolean lattice. There are several ways of generating new lattices from lattices. These are summarised in Sect. 2.1.

	$\mathcal{L}(\mathbf{Bool})$	$\mathcal{L}(\mathcal{P}(\{a,b,c\}))$	$\mathcal{L}(\Re^{\infty}_{-\infty})$	$\mathcal{L}([n,m])$
		$\{a,b,c\}$	∞	m
		$\nearrow \uparrow \nwarrow$	$\vdots$	$\vdots$
	True	$\{a,b\}$ $\{a,c\}$ $\{b,c\}$	$\uparrow$	$\uparrow$
	$\uparrow$	$\nwarrow\!\!\nearrow$ $\nwarrow\!\!\nearrow$	0	0
	False	$\{a\}$ $\{b\}$ $\{c\}$	$\uparrow$	$\uparrow$
		$\nwarrow \uparrow \nearrow$	$\vdots$	$\vdots$
		$\emptyset$	$-\infty$	n
$\sqcup$	$\vee$	$\cup$	max	max
$\sqcap$	$\wedge$	$\cap$	min	min
$\sqsubseteq$	$\Rightarrow$	$\subseteq$	$\leq$	$\leq$
$\top$	*True*	$\{a,b,c\}$	∞	m
$\bot$	*False*	$\emptyset$	$-\infty$	n

Fig. 3. The lattices $\mathcal{L}(\mathbf{Bool})$, $\mathcal{L}(\mathcal{P}(\{a,b,c\}))$, $\mathcal{L}(\Re^{\infty}_{-\infty})$ and $\mathcal{L}([n,m])$.

2.1 Generating Lattices from Lattices

Inverses. If $\mathcal{L} = (\mathcal{C}, \sqcup, \sqcap)$ is a lattice, then the inverse $\mathcal{L}^{-1}$, defined as $(\mathcal{C}, \sqcap, \sqcup)$ is also a lattice. The ordering for $\mathcal{L}^{-1}$ can be derived, and satisfies $\sqsubseteq^{-1} = \sqsupseteq$ (and also $\sqsupseteq^{-1} = \sqsubseteq$, etc.). If $\mathcal{L}$ is a complete lattice then $\top^{-1} = \bot$ and $\bot^{-1} = \top$. To see the inverses of the lattices in Fig. 3, stand on your head!

Homomorphisms. If $\mathcal{L}_1 = (\mathcal{C}_1, \sqcup_1, \sqcap_1)$ and $\mathcal{L}_2 = (\mathcal{C}_2, \sqcup_2, \sqcap_2)$ are two lattices, a lattice-homomorphism from $\mathcal{L}_1$ to $\mathcal{L}_2$ is a function $\textcircled{h}$ from $\mathcal{C}_1$ to $\mathcal{C}_2$ that preserves least upper bounds and greatest lower bounds — i.e. $x \sqcup_1 y = z \Rightarrow \textcircled{h}(x) \sqcup_2 \textcircled{h}(y) = \textcircled{h}(z)$ and $x \sqcap_1 y = z \Rightarrow \textcircled{h}(x) \sqcap_2 \textcircled{h}(y) = \textcircled{h}(z)$. A lattice-homomorphism is written $\textcircled{h}(\mathcal{L})$, so that, in this case, $\textcircled{h}(\mathcal{L}_1) = \mathcal{L}_2$. Some lattice

[2] In [10] we take the definitions of the binary operators $\sqcup$ and $\sqcap$ to be basic, and then derive the definition of the partial order: $x \sqsubseteq y \equiv x = x \sqcap y$. This explains why, in the rest of this paper, lattices are defined in terms of $\sqcup$ and $\sqcap$.

[3] See [3] for a more detailed introduction to lattice theory.

homomorphisms are illustrated in Fig. 4. The third and fourth homomorphisms in Fig. 4 are examples of homomorphisms from product lattices, which will be introduced below.

Homomorphism	Type	Definition
$\circledR$	$\mathbf{Bool} \to \Re^{\infty}_{-\infty}$	$\circledR(\mathit{True}) = 1$ $\circledR(\mathit{False}) = 0$
$\circledB$	$\Re^{\infty}_{-\infty} \to \mathbf{Bool}$	$\circledB(n) = (n \geq 0)$
$\circledwedge$	$(\mathbf{Bool}, \mathbf{Bool}) \to \mathbf{Bool}$	$\circledwedge(x, y) = x \wedge y$
$\circledvee$	$(\mathbf{Bool}, \mathbf{Bool}) \to \mathbf{Bool}$	$\circledvee(x, y) = x \vee y$

Fig. 4. Some lattice homomorphisms.

Products. If $\mathcal{L}_1 = (\mathcal{C}_1, \sqcup_1, \sqcap_1)$ and $\mathcal{L}_2 = (\mathcal{C}_2, \sqcup_2, \sqcap_2)$ are both lattices, then their product, $\mathcal{L}_1 \times \mathcal{L}_2$ is also a lattice $(\mathcal{C}_1 \times \mathcal{C}_2, \sqcup_\times, \sqcap_\times)$ with $(x_1, x_2) \sqsubseteq_\times (y_1, y_2)$ defined to be $x_1 \sqsubseteq_1 y_1 \wedge x_2 \sqsubseteq_2 y_2$. If $\mathcal{L}_1$ and $\mathcal{L}_2$ are both complete lattices, then $\mathcal{L}_1 \times \mathcal{L}_2$ is also a complete lattice, with $\top = (\top_1, \top_2)$ and $\bot = (\bot_1, \bot_2)$.

An Example. The product of the power set lattice and the boolean lattice — $\mathcal{L}(\mathcal{P}(\{a, b, c\}))$ and $\mathcal{L}(\mathbf{Bool})$ — is the lattice $(\mathcal{P}(\{a, b, c\}) \times \mathbf{Bool}, (\cup, \vee), (\cap, \wedge))$ (or, conventionally, $\mathcal{L}(\mathcal{P}(\{a, b, c\}) \times \mathbf{Bool})$).

This lattice is shown graphically in Fig. 5, and defines an order on the Cartesian product of $\mathcal{P}(\{a, b, c\})$ and **Bool** in which, for example, $(\{a, b, c\}, \mathit{False}) \sqsubseteq_\times (\{a, b, c\}, \mathit{True})$ because $\{a, b, c\} \sqsubseteq \{a, b, c\}$ and $\mathit{False} \sqsubseteq \mathit{True}$.

Prioritisations. Lattice inverses, homomorphisms and products are standard lattice operations. Prioritisation is an operation that we have defined for its usefulness to our framework.

If $\mathcal{L}_1 = (\mathcal{C}_1, \sqcup_1, \sqcap_1, \top_1, \bot_1)$ and $\mathcal{L}_2(\mathcal{C}_2, \sqcup_2, \sqcap_2, \top_2, \bot_2)$ are both *complete* lattices, then the *prioritisation* of $\mathcal{L}_1$ over $\mathcal{L}_2$, notation $\mathcal{L}_1 \gg \mathcal{L}_2$, is the complete lattice $(\mathcal{C}_1 \times \mathcal{C}_2, \sqcup_\gg, \sqcap_\gg, (\top, \top), (\bot, \bot))$ where $(x_1, x_2) \sqsubseteq_\gg (y_1, y_2)$ is defined to be $x_1 \sqsubset_1 y_1 \vee (x_1 = y_1 \wedge x_2 \sqsubseteq_2 y_2)$. This is an ordinary lexicographic ordering extended to lattices.

An Example. The graphical representation of $\mathcal{L}(\mathcal{P}(\{a, b, c\})) \gg \mathcal{L}(\mathbf{Bool})$ (or $\mathcal{L}(\mathcal{P}(\{a, b, c\}) \gg \mathbf{Bool})$) is shown in Fig. 6. Again the new lattice defines an order on the Cartesian product of $\mathcal{P}(\{a, b, c\})$ and **Bool**. However, whereas in the product lattice $(\{a\}, \mathit{True}) \not\sqsubseteq_\times (\{a, c\}, \mathit{False})$ because $\mathit{True} \not\sqsubseteq \mathit{False}$, here $(\{a\}, \mathit{True}) \sqsubseteq_\gg (\{a, c\}, \mathit{True})$ because $\{a\} \sqsubset \{a, c\}$. Also $(\{a, b, c\}, \mathit{False}) \sqsubseteq_\gg (\{a, b, c\}, \mathit{True})$ because $\{a, b, c\} = \{a, b, c\}$ and $\mathit{False} \sqsubseteq \mathit{True}$.

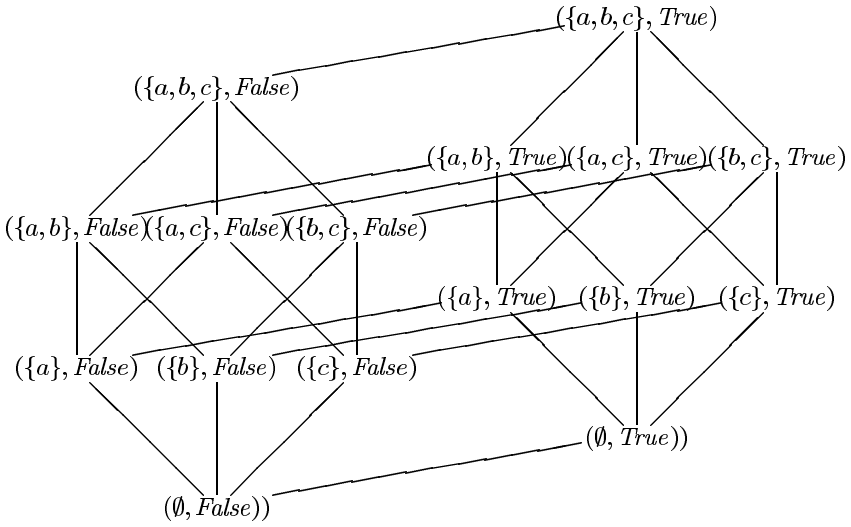

Fig. 5. The lattice $\mathcal{L}(\mathcal{P}(\{a,b,c\}) \times \mathbf{Bool})$.

3 Similarity Metrics

A *similarity metric* is a generalisation of a transitive relation. For some set α an $(\alpha, \mathcal{L})$ metric is a pair $(\trianglelefteq, \mathcal{L})$, where $\mathcal{L}$ is a complete lattice, with $\trianglelefteq :: \alpha \to \alpha \to \mathcal{L}$, such that

$$((x \trianglelefteq y) \sqcap_{\mathcal{L}} (y \trianglelefteq z)) \sqsubseteq_{\mathcal{L}} (x \trianglelefteq z) \ . \tag{1}$$

The expression $x \trianglelefteq y = v$ is pronounced "y exceeds x by v". The value of v is also called the "excess of y over x". Similarity metrics are inspired by the definition of *distance* in metric spaces [4], (1) being similar to a triangle inequality. In a metric space, however, the distance function is symmetric, and distances have a total order defined on them. In similarity metrics the order may be partial, and the distance function need not be symmetric. This partial ordering means that not all excesses need be comparable. The definition of maxima for similarity metrics must take account of this possible incomparability.

An element x of a set S is a maximum if, whenever the excess of x over y is comparable to the excess of y over x, then the excess of x over y is greater than or equal to the excess of y over x. An equivalent formulation is: if the excess of x over y is less than or equal to the excess of y over x, then the two excesses are equal. I.e. the maxima of a metric are given by:

$$\mathsf{max}\ S = \{x \in S | \forall y \in S : ((y \trianglelefteq x) \sqsubseteq_{\mathcal{L}} (x \trianglelefteq y)) \Rightarrow ((y \trianglelefteq x) = (x \trianglelefteq y))\} \ . \tag{2}$$

3.1 Some Examples

Similarity measures that return boolean values, numeric values, or values of other types can all be modelled as metrics, which shows that our framework subsumes

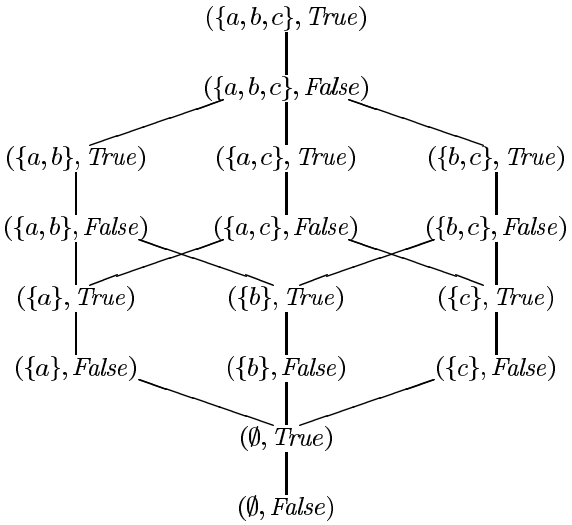

Fig. 6. The lattice $\mathcal{L}(\mathcal{P}(\{a,b,c\}) \gg \mathbf{Bool})$.

many other approaches to similarity measurement. We demonstrate this with examples relating to the holiday case base given in Fig. 1.

Destinations. In Fig. 2 we showed a partial order, $\sqsubseteq_{\mathbf{Dest}}$, on holiday destinations. This order (boolean-valued function) can be modelled by a metric $(\sqsubseteq_{\mathbf{Dest}}, \mathcal{L}(\mathbf{Bool}))$. Its result type is the lattice $\mathcal{L}(\mathbf{Bool})$. By this metric $Ben \sqsubseteq Cre = True$ — i.e. *Crete* exceeds *Benidorm* by *True* — and $And \sqsubseteq Dor = False$ — *Dordogne* exceeds *Andalucia* by *False*[4]. In fact, for any transitive relation $\sqsubseteq$ on α (i.e. a boolean valued function $\alpha \to \alpha \to \mathbf{Bool}$), the pair $(\sqsubseteq, \mathcal{L}(\mathbf{Bool}))$ is an $(\alpha, \mathcal{L}(\mathbf{Bool}))$ metric. It is easy to show that the maxima given by substituting in (2) are the usual maxima of a transitive relation.

Price. The most obvious metric for real numbers is the metric $\mathcal{M}_{\Re^{\infty}_{-\infty}}$ defined by $\mathcal{M}_{\Re^{\infty}_{-\infty}} = (\dot{-}^{-1}, \mathcal{L}(\Re^{\infty}_{-\infty}))$, where $\dot{-}$ is the operator $x \dot{-} y = \max(x - y, 0)$. This is a $(\Re^{\infty}_{-\infty}, \mathcal{L}(\Re^{\infty}_{-\infty}))$ metric. By this metric $500 \dot{-}^{-1} 750 = 250$ and $750 \dot{-}^{-1} 500 = 0$ — i.e. 750 exceeds 500 by 250, and 500 exceeds 750 by 0. The metric $(\dot{-}^{-1}, \mathcal{L}(\Re^{\infty}_{-\infty}))$ may be suitable for many numeric domains, and we can again show, by substituting in (2), that (2) would compute the maxima we would expect. Since this metric reflects the usual ordering $(\leq)$ on real numbers, it will rank larger numbers higher than smaller ones. Unless one is an elitist member of the super rich jet set, this is probably not the criterium one would apply to prices. A more sensible metric for prices would be $\mathcal{M}_{\mathbf{Price}} = (\dot{-}, \Re^{\infty}_{-\infty})$.

[4] For metrics defined on $\mathcal{L}(\mathbf{Bool})$, simpler paraphrases are possible, and clearer: we can say that *Crete* exceeds *Benidorm*, and *Dordogne* does not exceed *Andalucia*.

Activities. Using w, s, and g as abbreviations for *windsurfing*, *swimming* and *golf*, a possible metric for holiday activities — $\mathcal{M}_{\mathbf{Act}}$ — would be the similarity metric $(\setminus^{-1}, \mathcal{L}(\mathcal{P}(\{g,s,w\})))$, where $\setminus^{-1}$ is the inverse of set difference — so that $\{g\} \setminus^{-1} \{g,s\} = \{s\}$, i.e. *golf* and *swimming* exceeds *golf* by *swimming*. Again (2) gives suitable maxima.

We see it as a strength of the framework that not only can boolean-valued and numeric-valued metrics be defined, but metrics of other types (e.g. set-valued as above, and feature-structure-valued as described in [10]) are also instances of the framework. That these different metrics can all be combined is another strength.

3.2 Generating Metrics from Metrics

Inverses. If $\mathcal{M} = (\trianglelefteq, \mathcal{L})$ is a metric, then its inverse, $\mathcal{M}^{-1} = (\trianglelefteq^{-1}, \mathcal{L})$ is also a metric. For example, $\mathcal{M}_{\mathbf{Price}} = \mathcal{M}^{-1}_{\Re^{\infty}_{-\infty}}$. Note that while it is certainly not the case that $(\trianglelefteq^{-1}, \mathcal{L}) = (\trianglelefteq, \mathcal{L}^{-1})$, it is the case that $\mathsf{max}_{(\trianglelefteq^{-1},\mathcal{L})} = \mathsf{max}_{(\trianglelefteq,\mathcal{L}^{-1})}$.

Compositions.

Left Composition. If $\mathcal{M} = (\trianglelefteq, \mathcal{L})$ is an $(\alpha, \mathcal{L})$ metric and f is a function of type $\beta \to \alpha$, then the left composition of the metric and f, notation $\mathcal{M} \circ f$, defined as $(\trianglelefteq \circ f, \mathcal{L})$, where $x \; (\trianglelefteq \circ f) \; y = f(x) \trianglelefteq f(y)$, is a $(\beta, \mathcal{L})$ metric.

Obvious candidates for left composition are *projection functions*. For example, π_1 will select the first element of a tuple. Then the left composition $\mathcal{M}_{\mathbf{Dest}} \circ \pi_1$ is a metric that applies to whole tuples. Specifically, since $\mathcal{M}_{\mathbf{Dest}} = (\sqsubseteq_{\mathbf{Dest}}, \mathcal{L}(\mathbf{Bool}))$ is a $(\mathbf{Dest}, \mathcal{L}(\mathbf{Bool}))$ metric, $\mathcal{M}_{\mathbf{Dest}} \circ \pi_1$ is a $(\mathbf{Holiday}, \mathcal{L}(\mathbf{Bool}))$ metric, in which, for cases c_1 and c_2, $c_1 \trianglelefteq c_2 = \pi_1(c_1) \sqsubseteq_{\mathbf{Dest}} \pi_1(c_2)$ — i.e. c_2 exceeds c_1 in the new metric if c_2's destination exceeds c_1's.

Although we are here using simple projection from a tuple, more complex case representations may require more complex projection functions. Projection functions might even implement inferencing [5, 8], perhaps to obtain "deep" features [2] from "surface" features.

Right Composition. If $\mathcal{M} = (\trianglelefteq, \mathcal{L})$ is an $(\alpha, \mathcal{L})$ metric, and $\textcircled{h}$ is a lattice homomorphism, then the right composition of these, $\textcircled{h} \circ \mathcal{M}$, defined as $(\textcircled{h} \circ \trianglelefteq, \textcircled{h}(\mathcal{L}))$, where $x \; (\textcircled{h} \circ \trianglelefteq) \; y = \textcircled{h}(x \trianglelefteq y)$, is an $(\alpha, \textcircled{h}(\mathcal{L}))$ metric. An example of a right composition could be the metric that simply compares prices — $\textcircled{B} \circ \mathcal{M}_{\mathbf{Price}}$, so that, for example, $750 \trianglelefteq 500 = \mathit{True}$, rather than 250.

Products. If $\mathcal{M}_1 = (\trianglelefteq_1, \mathcal{L}_1)$ and $\mathcal{M}_2 = (\trianglelefteq_2, \mathcal{L}_2)$ are two metrics, then their product, $\mathcal{M}_1 \times \mathcal{M}_2$, is the metric $(\trianglelefteq_1 \times \trianglelefteq_2, \mathcal{L}_1 \times \mathcal{L}_2)$, with $\trianglelefteq_1 \times \trianglelefteq_2$ defined by $(x_1, x_2) \; (\trianglelefteq_1 \times \trianglelefteq_2) \; (y_1, y_2) = (x_1 \trianglelefteq_1 y_1, x_2 \trianglelefteq_2 y_2)$. A metric that will compare both destination and price is the metric $\mathcal{M}_{\mathbf{Dest}} \times \mathcal{M}_{\mathbf{Price}}$. In this metric, for example, $(\mathit{Ben}, 350) \trianglelefteq (\mathit{Cre}, 750) = (\mathit{True}, 0)$ because $\mathit{Ben} \sqsubseteq_{\mathbf{Dest}} \mathit{Cre}$, and $350 \dot{-} 750 = 0$. This is a new metric whose result type is the product of $\mathcal{L}(\mathbf{Bool})$

and $\mathcal{L}(\Re^{\infty}_{-\infty})$. Crucially, the combination of an ordinal and a cardinal metric results in a new metric, which could be applied to a case base, and the maxima taken. If this is done it will select not only the cheaper Cretan holiday (being the cheaper of the two in the preferred location), but also the cheapest holiday — i.e. the one in Benidorm. Our goal of combining different types of metric has been achieved.

It is also possible to take the $(\mathbf{Price}, \mathbf{Bool})$ metric defined above, i.e. $Ⓑ \circ \mathcal{M}_{\mathbf{Price}}$, and take its product with $\mathcal{M}_{\mathbf{Dest}}$ — a $(\mathbf{Dest}, \mathbf{Bool})$ metric. The resulting metric will be a $((\mathbf{Dest}, \mathbf{Price}), (\mathbf{Bool}, \mathbf{Bool}))$ metric, $\mathcal{M}_{\mathbf{Dest}} \times (Ⓑ \circ \mathcal{M}_{\mathbf{Price}})$ This metric is one whose result is a lattice defined on pairs of truth values, e.g. $(Ben, 350) \trianglelefteq (Cre, 750) = (True, False)$. The disjunctive homomorphism can now be applied. This will result in a $((\mathbf{Dest}, \mathbf{Price}), \mathbf{Bool})$ metric — $Ⓥ \circ (\mathcal{M}_{\mathbf{Dest}} \times (Ⓑ \circ \mathcal{M}_{\mathbf{Price}}))$, in which, for example, $(Ben, 350) \trianglelefteq (Cre, 750) = True$. Note also that — in contrast to the $((\mathbf{Dest}, \mathbf{Price}), (\mathbf{Bool}, \mathbf{Bool}))$ metric above, where $(Cre, 750) \trianglelefteq (Cre, 500)$ and $(Cre, 500) \trianglelefteq (Cre, 750)$ are not equal, since $(Cre, 750) \trianglelefteq (Cre, 500) = (True, True)$, while $(Cre, 500) \trianglelefteq (Cre, 750) = (True, False)$ — in this new metric $(Cre, 750) \trianglelefteq (Cre, 500)$ and $(Cre, 500) \trianglelefteq (Cre, 750)$ *are* equal (both being *True*). As a consequence, the maxima according to this metric will be both the Cretan holidays and the holiday in Benidorm (and the mutual excesses between all three cases will be comparable).

Prioritisations. If $\mathcal{M}_1 = (\trianglelefteq_1, \mathcal{L}_1)$ and $\mathcal{M}_2 = (\trianglelefteq_2, \mathcal{L}_2)$ are two metrics, then the prioritisation of $\mathcal{M}_1$ over $\mathcal{M}_2$, notation $\mathcal{M}_1 \gg \mathcal{M}_2$, is the metric $(\trianglelefteq_1 \times \trianglelefteq_2, \mathcal{L}_1 \gg \mathcal{L}_2)$. If the destination is more important than the price, this can be reflected in the metric — e.g. $\mathcal{M}_{\mathbf{Dest}} \gg \mathcal{M}_{\mathbf{Price}}$. In contrast to the product metric, this metric will prefer the cheaper Cretan holiday to the Benidorm holiday — the excesses are the same, but the lattice is different. I.e. in both $\mathcal{M}_{\mathbf{Dest}} \times \mathcal{M}_{\mathbf{Price}}$ and $\mathcal{M}_{\mathbf{Dest}} \gg \mathcal{M}_{\mathbf{Price}}$, the two excesses between these two holidays are:

$$(Ben, 350) \trianglelefteq (Cre, 500) = (True, 0)$$
$$(Cre, 500) \trianglelefteq (Ben, 350) = (False, 150)$$

but in $\mathcal{M}_{\mathbf{Dest}} \times \mathcal{M}_{\mathbf{Price}}$ it is not the case that $(False, 150) \sqsubseteq (True, 0)$ (since $False \sqsubseteq True$, but $150 \not\sqsubseteq 0$), while this relation does hold in $\mathcal{M}_{\mathbf{Dest}} \gg \mathcal{M}_{\mathbf{Price}}$.

To consider a more complex example, if the activities available are less important than the destination and price, the following metric can be used:

$$(Ⓥ \circ (\mathcal{M}_{\mathbf{Dest}} \times (Ⓑ \circ \mathcal{M}_{\mathbf{Price}}))) \gg \mathcal{M}_{\mathbf{Act}} \ .$$

The first part of this metric will, as shown above, select the two Cretan holidays and the Benidorm holiday. The second part will then maximise the activities available, and therefore recommend the more expensive Cretan holiday.

4 Conclusions

We have presented a formal framework for the specification of similarity measures — ordinal or cardinal, and also similarity measures returning other types, such as

sets or feature structures. We can not only define individual similarity measures, but also combine similarity measures, possibly returning values of different types, to produce new similarity measures. This, in combination with the large set of connectives available within the formalism, gives a powerful formalism for the construction of a wide range of similarity measures applicable to many problem domains.

References

1. A. Aamodt and E. Plaza. Case based reasoning: Foundational issues, methodological variations and system approaches. *AI Communications*, 7(1):39–59, 1994.
2. K.D. Ashley and E.L. Rissland. A case-based approach to modeling legal expertise. *IEEE Expert*, 3(3):70–77, 1988.
3. G. Birkhoff. *Lattice Theory.* American Mathematical Society, 1967.
4. M. Fréchet. Sur quelques points du calcul fonctionel. *Rendiconti del Circolo Matematico di Palermo*, 22:1–74, 1906.
5. T.R. Hinrichs. *Problem Solving in Open Worlds: A Case Study in Design.* Lawrence Erlbaum, 1992.
6. K.P. Jantke. Nonstandard concepts of similarity in case-based reasoning. In H.H. Bock, W. Lenski, and M.M. Richter, editors, *Proceedings der Jahrestagung der Gesellschaft für Klassifikation,* Information Systems and Data Analysis: Prospects, Foundations, Applications, pages 29–44. Springer Verlag, 1994.
7. J.L. Kolodner. *Case Based Reasoning.* Morgan Kaufmann, 1993.
8. P. Koton. Reasoning about evidence in causal explanations. In *Proceedings of AAAI-88*, pages 256–261, 1988.
9. Hugh Osborne and Derek Bridge. A case base similarity framework. In Ian Smith and Boi Faltings, editors, *Advances in Case-Based Reasoning,* Proceedings of EWCBR'96, pages 309–323, 1996.
10. Hugh Osborne and Derek Bridge. Similarity metrics: A formal approach to case base retrieval using general similarity metrics. Technical report, Department of Computer Science, University of York, 1997. In preparation.
11. E. Plaza. Cases as terms: A feature term approach to the structured representation of cases. In *Proceedings of the First International Conference on Case-Based Reasoning.* Springer Verlag, 1995.
12. E. Plaza. On the importance of similitude: An entropy-based assessment. In Ian Smith and Boi Faltings, editors, *Advances in Case-Based Reasoning,* Proceedings of EWCBR'96, pages 324–338, 1996.
13. M.M. Richter. Classification and learning of similarity measures. In *Proceedings der Jahrestagung der Gesellschaft für Klassifikation,* Studies in Classification, Data Analysis and Knowledge Organisation. Springer Verlag, 1992.

This article was processed using the LaTeX macro package with LLNCS style