International Journal of Approximate Reasoning 71 (2016) 1–14

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Optimal approximations with Rough Sets and similarities in measure spaces

Ryszard Janicki *, Adam Lenarčič

Department of Computing and Software, McMaster University, Hamilton, L8S 4K1, Canada

ARTICLE INFO

Article history:
Received 21 April 2015
Received in revised form 13 November 2015
Accepted 30 December 2015
Available online 6 January 2016

Keywords:
Rough Sets
Optimal approximation
Similarity
Marczewski–Steinhaus index

ABSTRACT

When arbitrary sets are approximated by more structured sets, it may not be *possible* to obtain an exact approximation that is equivalent to a given set. Presented here, is a new proposal for a 'metric' approach to Rough Sets. We assume some *finite measure space* is defined on a given universe, and then use it to define various *similarity indexes*. A set of axioms and the concept of consistency for similarity indexes are also proposed. The core of the paper is a definition of the '*optimal*' or '*best*' *approximation* with respect to any particular similarity index, and an algorithm to find this optimal approximation by using the Marczewski–Steinhaus Index. This algorithm is also shown to hold for a class of similarity indexes that are consistent with the Marczewski–Steinhaus Index.

1. Introduction and motivation

When numerical data is generated empirically there is often some perturbation in the collection or recording of the data. Often, to accommodate the possibility of errors it is assumed that the true value for each piece of data is within some interval around the measured data. For example if the measured data is x, we might assume that the true value is in the range $(x-\varepsilon, x+\varepsilon)$. These values can be taken as the *lower* and *upper approximations*. When the measurement of data involves identifying or selecting a subset of items from a universe of possible items, and certain groups of the items are deemed equivalent under some criteria, Rough Sets are an appropriate tool for analysis [16,17].

The concepts of lower and upper approximation have long been defined in the context of Rough Sets, but the concept of an approximation is not restricted only to lower and upper approximations. Consider the well known *linear least squares approximation* of points in the two dimensional plane (credited to C. F. Gauss, 1795, cf. [2]). Here we know or assume that the points should be on a straight line and we are trying to find the line that fits the data best. However, this is not the case of an upper, or lower approximation in the sense of Rough Sets. The cases like the linear least squares approximation assume that there is a well defined concept of *similarity* (or *distance*) and some techniques for finding *maximal similarity* (*minimal distance*) between entities and their approximations.

What the Rough Sets approach seems to be missing is a feasible concept of the 'best' or 'optimal' approximation to a set. In this paper, which is a substantial generalization and extension of the ideas first proposed in [11], we provide a solution to this problem.

What we propose is a 'metric' or standard of measurement for comparison within the framework of Rough Sets [16] and a technique of using it.

* Corresponding author.
E-mail addresses: janicki@mcmaster.ca (R. Janicki), lenarcaj@mcmaster.ca (A. Lenarčič).

http://dx.doi.org/10.1016/j.ijar.2015.12.014
0888-613X/

We start with an assumption that our universe of elements is a *finite measure space*. This is a substantial extension of the model from [11] where cardinality was considered a measure of a set. Then we present axioms which we require for any similarity index to satisfy if it is to be used to find an optimal approximation. Next we introduce and analyze several similarity indexes for arbitrary sets. As most of these indexes were originally defined with cardinality as a measure of sets [4], we also provide their finite measure generalizations.

Later we define what it means for a Rough Sets approximation to be *optimal* (with respect to a given similarity index), and what it means for similarity indexes to be *consistent*. We also show that consistent similarity indexes yield to identical optimal approximations.

The main result of this paper is an efficient greedy algorithm which yields an *optimal rough sets approximation for any given set X, when the Marczewski–Steinhaus similarity index is used*, and some finite measure space is given. The algorithm is based on the properties of an index that quantifies the ratio of common to distinct elements of two given sets. We used the Marczewski–Steinhaus similarity index as an engine of our algorithm because it has a very natural and regular definition and convenient mathematical properties (it can be transformed into a metric).

The results are more general in that it appears it can be used with very minor changes for any similarity index that is consistent with the Marczewski–Steinhaus index, and this class is fairly large.

The paper is organized as follows. In Section 2 we recall the basic notions of Rough Sets and define the concept of border sets, while Section 3 adapts some ideas of measure theory for our purposes. In Section 4 we discuss several known similarity indexes in the context of measure theory, propose some axioms we think the similarity indexes should obey, define the concept of consistency for similarity indexes, and introduce the notion of optimal approximation in the framework of Rough Sets. Section 5 contains the main result of this paper, namely an efficient algorithm for finding optimal approximation when the very general and intuitive Marczewski–Steinhaus [13] index is used. We also show that this algorithm can be used with minor changes for any similarity index that is consistent with Marczewski–Steinhaus index. Section 7 analyses some examples that illustrate our concepts. A case of a similarity index that is not consistent with Marczewski–Steinhaus index is discussed in Section 8, and Section 9 contains final comments.

2. Rough Sets and borders

In this section we introduce, review, and also adapt for our purposes, some general ideas that are crucial to our approach.

The principles of Rough Sets [16,17] can be formulated as follows.

Let U be a finite and non-empty universe of elements, and let $E \subseteq U \times U$ be an *equivalence relation*. Recall that for each $E \subseteq U \times U$, $[x]_E$ will denote the equivalence class of E containing x, and U/E will denote the set of all equivalence classes of E.

The elements of $\mathfrak{Comp} = U/E$ are called *elementary sets* or *components* and they are interpreted as basic observable, measurable, or definable sets. We will denote the elements of $\mathfrak{Comp}$, i.e. equivalence classes of E, by bold symbols, and write for example $\mathbf{x} \in \mathbb{B} \subseteq \mathfrak{Comp}$.

The pair $\mathcal{AS} = (U, E)$ is referred to as a *Pawlak approximation space*.

A non-empty set $X \subseteq U$ is approximated by two subsets of U; $\underline{\mathbf{A}}(X)$ and $\overline{\mathbf{A}}(X)$, called the lower and upper approximations of X respectively, and are defined as follows:

Definition 1. (See [16,17].) For each $X \subseteq U$,

1. $\underline{\mathbf{A}}(X) = \bigcup\{\mathbf{x} \mid \mathbf{x} \in \mathfrak{Comp} \wedge \mathbf{x} \subseteq X\}$,
2. $\overline{\mathbf{A}}(X) = \bigcup\{\mathbf{x} \mid \mathbf{x} \in \mathfrak{Comp} \wedge \mathbf{x} \cap X \neq \emptyset\}$. □

Clearly $\underline{\mathbf{A}}(X) \subseteq X \subseteq \overline{\mathbf{A}}(X)$. There are many versions and many extensions of this basic model, see for example [10,21,22,26], as well as many various applications (cf. [9,20,22,23]). Even robotic locomotion can utilize this notion to ensure it remains within bounds, and could also use measures of similarity to move based on the best/optimal available (representable) approximation of its surroundings [6].

A set $A \subseteq U$ is *definable* (or *exact*) [17] if it is a union of some equivalence classes of the equivalence relation E. Let $\mathbb{D}$ denote the family of all definable sets defined by the space (U, E). Formally

$$\mathsf{A} \in \mathbb{D} \iff \exists \mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathfrak{Comp}.\ \mathsf{A} = \mathbf{x}_1 \cup \ldots \cup \mathbf{x}_n.$$

We would like to point out the duality of $\mathfrak{Comp}$ and $\mathbb{D}$. Each set of components $\mathsf{C} \subseteq \mathfrak{Comp}$ uniquely defines the *definable set* $\mathrm{dset}(\mathsf{C}) \in \mathbb{D}$, as $\mathrm{dset}(\mathsf{C}) = \bigcup_{\mathbf{x} \in \mathsf{C}} \mathbf{x}$, and each definable set $\mathsf{A} \in \mathbb{D}$ uniquely defines the *set of components* $\mathrm{comp}(\mathsf{A}) \subseteq \mathfrak{Comp}$, by $\mathrm{comp}(\mathsf{A}) = \{\mathbf{x} \mid \mathbf{x} \subseteq \mathsf{A}\}$.

Moreover, for each set of components $\mathsf{C} \subseteq \mathfrak{Comp}$ we have $\mathrm{comp}(\mathrm{dset}(\mathsf{C})) = \mathsf{C}$, and for each definable set $\mathsf{A} \in \mathbb{D}$ we have $\mathrm{dset}(\mathrm{comp}(\mathsf{A})) = \mathsf{A}$.

Clearly every lower and upper approximation is a definable set, i.e. $\underline{\mathbf{A}}(X) \in \mathbb{D}$ and $\overline{\mathbf{A}}(X) \in \mathbb{D}$ for every $X \subseteq U$. Furthermore, all definable sets are equal to their lower and upper approximations, as the below corollary shows.

Corollary 1. *For every* $X \subseteq U$, $X \in \mathbb{D} \iff \underline{\mathbf{A}}(X) = \overline{\mathbf{A}}(X) = X$. □

Since the definable sets in the area between the upper and lower approximations will play an important role in our model, we need to precisely define this area.

Definition 2. For every $X \subseteq U$, we define the set of components $\mathfrak{B}(X) \subseteq \mathfrak{Comp}$ called the **border** of X, and the set of **border sets** of X called $\mathbb{B}(X) \subseteq \mathbb{D}$, as follows:

1. $\mathbf{x} \in \mathfrak{B}(X) \iff \mathbf{x} \in \mathrm{comp}(\overline{\mathbf{A}}(X)) \setminus \mathrm{comp}(\underline{\mathbf{A}}(X))$,
2. $\mathsf{A} \in \mathbb{B}(X) \iff \mathsf{A} \subseteq \overline{\mathbf{A}}(X) \setminus \underline{\mathbf{A}}(X) \wedge \mathsf{A} \in \mathbb{D}$. □

The border and boarder sets are building blocks for our optimal approximation defined later. The corollary below describes basic properties of borders and border sets.

Corollary 2. *For every $X \subseteq U$,*

1. $\mathrm{dset}(\mathfrak{B}(X)) = \overline{\mathbf{A}}(X) \setminus \underline{\mathbf{A}}(X) \in \mathbb{B}(X)$ *and* $\mathfrak{B}(X) \subseteq \mathbb{B}(X)$,
2. $\mathsf{A} \in \mathbb{B}(X) \iff \exists \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \mathfrak{B}(X).\ \mathsf{A} = \mathbf{x}_1 \cup \ldots \cup \mathbf{x}_n$,
3. *if* $\mathsf{A} \in \mathbb{B}(X)$ *then* $\mathsf{A} \cap X \neq \emptyset$ *and* $\mathsf{A} \setminus X \neq \emptyset$.
4. *if* $X \in \mathbb{D}$ *then* $\mathbb{B}(X) = \emptyset$. □

Corollary 2(3) will often be used later in proofs of many important results of this paper. It states that if X is not definable, then it overlaps with each element of its border set.

3. Measures

In this section we recall some basic results from *measure theory*, adapted to our purposes (cf. [7,15]).

Let U be a set (not necessarily finite) and let $\mu : 2^U \to \mathbb{R}$, where $\mathbb{R}$ is a set of real numbers, be a function that satisfies the following properties:

1. for all $X \subseteq U$, $0 \leq \mu(X) < \infty$,
2. $\mu(\emptyset) = 0$,
3. if $X_i \subseteq U$ for $i = 1, \ldots, \infty$ and $X_i \cap X_j = \emptyset$ if $i \neq j$, then

$$\mu(\bigcup_{i=1}^{\infty} X_i) = \sum_{i=1}^{\infty} \mu(X_i).$$

Any such function is called a *finite measure* over 2^U, and a triple $(U, 2^U, \mu)$ is a *measure space* (cf. [7,15]). One can show that μ also satisfies:

- for all $X, Y \subseteq U$, if $X \subseteq Y$ then $\mu(X) \leq \mu(Y)$,
- for all $X_i \subseteq U$, where $i = 1, \ldots, \infty$ (and X_i are not necessarily disjoint), we have

$$\mu(\bigcup_{i=1}^{\infty} X_i) \leq \sum_{i=1}^{\infty} \mu(X_i).$$

A set X such that $\mu(X) = 0$ is called a *μ-null set* and all μ-null sets are called *negligible*.

- A measure space $(U, 2^U, \mu)$ is *null set free* if the empty set, $\emptyset$, is the only μ-null set, i.e. if $\mu(X) = 0 \iff X = \emptyset$.

In the standard theory of measure, the property of null set freeness is not defined and not discussed (cf. [7,15]), however this is an important property for our approach. Note that for example cardinality is a null free measure.

If a set U is finite, the definition of a measure can be simplified.

- From (3) of the measure definition, we have that if $X = \{x_1, \ldots, x_m\}$, then $\mu(X) = \mu(\{x_1\}) + \ldots + \mu(\{x_m\})$.

This means that for finite sets we can define a measure element-wise, as $\mu : U \to \mathbb{R}$ and then just extend it for sets in a standard way as, for every $X \subseteq U$,

$$\mu(X) = \sum_{x \in X} \mu(x).$$

We will assume that if a set U is finite, a measure μ is element-wise defined. Discrete probability is an element-wise defined measure, with $\mu(U) = 1$.

- If U is finite then a measure space is null free if for every $x \in U$, $\mu(x) > 0$.

Cardinality is an example of null free element-wise defined measure given by $\mu(x) = 1$ for all $x \in U$.

4. Similarity indexes

The model extended in this paper requires us to have a way to conceive of, and consequently quantify, *similarity* between two sets. It is important to point out that in this context we evaluate similarity between *sets*, but *not* between *elements* (cf. [23]), and that these indexes do not assume any specific interpretation of sets as done in [20].

We only assume that we have a set U (not necessarily finite) and a *finite* measure space $(U, 2^U, \mu)$.

Suppose that we have a (total) function $sim : 2^U \times 2^U \to [0, 1]$ that measures *similarity* between sets. While such functions have been known since the beginning of the twentieth century [8], they do not have standard indisputable axiomatization [4]. Depending on the area of application, some desirable properties may vary [4,19,25].

In this paper we will assume that the function *sim* satisfies the following five, intuitive axioms. Namely, for all sets A, $B \subseteq U$, we have:

$S1$ (Maximum): $sim(A, B) = 1 \iff A = B$,
$S2$ (Symmetry): $sim(A, B) = sim(B, A)$,
$S3$ (Minimum): $sim(A, B) = 0 \iff A \cap B = \emptyset$,
$S4$ (Inclusion): if $a \in B \setminus A$ then $sim(A, B) < sim(A \cup \{a\}, B)$,
$S5$ (Exclusion): if $a \notin A \cup B$ and $A \cap B \neq \emptyset$ then $sim(A, B) > sim(A \cup \{a\}, B)$

Most similarities assume the axioms $S1$–$S3$ either explicitly or implicitly. The axioms $S4$ and $S5$, although satisfied by many known similarities, were only recently proposed in [11]. The axioms $S1$–$S5$ as formulated above also follow from [11].

In this paper we also propose a weakened version of $S5$, namely:

$S5'$ (Weak Exclusion): if $a \notin A \cup B$ then $sim(A, B) \geq sim(A \cup \{a\}, B)$

The first axiom ensures that if and only if a similarity measure returns one, the two sets are equal. The second axiom is the symmetry of similarity measures, meaning that one set is the same distance from a second set, as the second set is from the first, and the third axiom states that if two sets do not share any elements, their similarity is zero, and vice versa.

The axioms $S4$ and $S5$ deal with changing sizes of sets. We will call them *monotonicity axioms*. Axiom $S4$ dictates that if we add part of B to A, the result is closer to B than A alone, while axiom $S5$ reduces to the notion that if we add to A some new element not in B, then the result is more distant from B than A alone. The axiom $S5$ is only applicable when the sets being compared have at least one common element, i.e. $sim(A, B) > 0$. Otherwise $sim(A, B) = sim(A, B) = sim(A \cup \{a\}, B) = 0$, but since we allow for equality, this requirement is not present in the weakened version. We also provide a weakened fifth axiom, in which adding an element to A which was in neither set, may not necessarily decrease the similarity between them, allowing for the possibility of leaving the value unchanged.

We will also say that a measure of similarity *sim* is *metrical* (i.e. it is a suitable tool to evaluate distance between two sets), if the function $diff(A, B) = 1 - sim(A, B)$ is a proper *metric* or *distance*[1] which holds for all A, $B \subseteq U$ (cf. [3,4]).

The first similarity measure was proposed in 1901 by P. Jaccard [8]. It is still one of the most popular, however the following similarity measures are also prominent in the literature at this point in time:

- *Jaccard index* [8]: $sim_J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}$,
- *Dice–Sørensen index* [5,24]: $sim_{DS}(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|}$,
- *Marczewski–Steinhaus μ-index* [13,14]: $sim_{MS}(X, Y) = \frac{\mu(X \cap Y)}{\mu(X \cup Y)}$, where μ is a *finite measure* on some U such that X, $Y \subseteq U$,
- *Tversky index* [25]: $sim_T^{\alpha,\beta}(X, Y) = \frac{|X \cap Y|}{|X \cap Y| + \alpha|X \setminus Y| + \beta|Y \setminus X|}$, where α, $\beta \geq 0$ are parameters. Note that for $\alpha = \beta = 1$, $sim_T^{\alpha,\beta}(X, Y) = sim_J(X, Y)$ and for $\alpha = \beta = 0.5$, $sim_T^{\alpha,\beta}(X, Y) = sim_{DS}(X, Y)$.
- *Braun-Blanquet index*[2] [1,4]: $sim_{BB}(X, Y) = \frac{|X \cap Y|}{\max(|X|, |Y|)}$.

[1] This means if the function *diff* satisfies (cf. [3,4]):

1. $diff(A, B) \geq 0$,
2. $diff(A, B) = 0 \iff A = B$,
3. $diff(A, B) = diff(B, A)$,
4. $diff(A, C) \leq diff(A, B) + diff(B, C)$, i.e triangle inequality.

[2] This index was recently reinvented and analyzed in [19] in the context of Fuzzy Sets. The authors of [19] were probably unaware of its long existence.

In general, a measure space from Marczewski–Steinhaus μ-index may not be null set free, so it may not satisfy the axiom $S3$, however in practically all known applications, particular measure spaces are null set free (cf. [18]).

In our model the axiom $S3$, i.e. null set freeness, is important as for instance the results of Section 6 do not hold when $S3$ is not satisfied. *Hence, from now on, we assume that every measure space* $(U, 2^U, \mu)$ *discussed in this paper is* ***null set free***. A finite universe U is one of the principal assumptions of Rough Sets (see Section 2) and for finite U enforcing null set freeness is natural anyway, we just have to delete from U all elements a, such that $\mu(a) = 0$, and then deal with this new smaller universe.

In this paper we will also pay attention to a special version of Tversky index, when $\alpha = \beta$. We will call it *Symmetric Tversky index*, and define it formally as

- *Symmetric Tversky index*: $sim^{\alpha}_{sT}(X, Y) = \frac{|X\cap Y|}{|X\cap Y|+\alpha|X\setminus Y|+\alpha|Y\setminus X|} = \frac{|X\cap Y|}{|X\cap Y|+\alpha|(X\cup Y)\setminus(X\cap Y)|}$.

If $\mu(X) = |X|$, then $sim_J(X, Y) = sim_{MS}(X, Y)$, so the Jaccard index is a special case of more general Marczewski–Steinhaus μ-index.

When we replace the cardinality with *finite measure* in other indexes, we get:

- *Dice–Sørensen μ-index*: $sim_{\mu DS}(X, Y) = \frac{2\mu(X\cap Y)}{\mu(X)+\mu(Y)}$,
- *Tversky μ-index*: $sim^{\alpha,\beta}_{\mu T}(X, Y) = \frac{\mu(X\cap Y)}{\mu(X\cap Y)+\alpha\mu(X\setminus Y)+\beta\mu(Y\setminus X)}$,
- *Symmetric Tversky μ-index*:

$$sim^{\alpha}_{\mu T}(X, Y) = \frac{\mu(X \cap Y)}{\mu(X \cap Y) + \alpha\mu(X \setminus Y) + \alpha\mu(Y \setminus X)} = \frac{\mu(X \cap Y)}{\mu(X \cap Y) + \alpha\mu((X \cup Y) \setminus (X \cap Y))},$$

- *Braun-Blanquet μ-index*: $sim_{\mu BB}(X, Y) = \frac{\mu(X\cap Y)}{\max(\mu(X),\mu(Y))}$.

In the terminology introduced above, Jaccard μ-index and Marczewski–Steinhaus μ-index mean the same formula.

All the similarity indexes above clearly have values between 0 and 1 and all, except (general) Tversky index and Tversky μ-index, satisfy the similarity axioms $S1$–$S4$. The Tversky index is not, in general, symmetric, so it may not satisfy $S2$.[3] The Braun-Blanquet index does not satisfy $S5$, it only satisfies $S5'$.

Proposition 1 *(Similarity axioms and similarity indexes).*

1. *Marczewski–Steinhaus μ-index, Dice–Sørensen μ-index and Symmetric Tversky μ-index satisfy axioms $S1$–$S5$.*
2. *Tversky μ-index satisfies axioms $S1$ and $S3$–$S5$. It satisfies $S2$ if and only if $\alpha = \beta$. Tversky index also satisfies $S2$ if and only if $\alpha = \beta$.*
3. *Braun-Blanquet μ-index satisfies axioms $S1$–$S4$ and $S5'$.*

Proof. In the proof below we admit infinite or even uncountable sets, but a measure μ must be finite and null-free.

(*Marczewski–Steinhaus μ-index*) Only $S4$–$S5$ are not immediately obvious. If $a \in B \setminus A$, then $a \notin A \cap B$, so $\mu((A \cup \{a\}) \cap B) = \mu(A \cap B) + \mu(\{a\})$. On the other hand $A \cup B = (A \cup \{a\}) \cup B$, so $sim_{MS}(A, B) = \frac{\mu(A\cap B|}{\mu(A\cup B)} < \frac{\mu(A\cap B)+\mu(\{a\})}{\mu(A\cup B)} = sim_{MS}(A \cup \{a\}, B)$.

Hence $S4$ does hold.

If $a \notin A \cup B$ then $A \cap B = (A \cup \{a\}) \cap B$ and $(A \cup \{a\}) \cup B = (A \cup B) \cup \{a\}$, so

$$\mu((A \cup \{a\}) \cup B) = \mu(A \cup B) + \mu(\{a\}).$$

Thus $sim_{MS}(A, B) = \frac{\mu(A\cap B)}{\mu(A\cup B)} > \frac{\mu(A\cap B)}{\mu(A\cup B)+\mu(\{a\})} = sim_{MS}(A \cup \{a\}, B)$, so $S5$ holds too.

(*Dice–Sørensen μ-index*) Only $S4$–$S5$ are not immediately obvious. For $S4$ we have again $\mu(A \cup \{a\}) \cap B) = \mu(A \cap B) + \mu(\{a\})$. Define $n = \mu(A \cap B)$, $m = \mu(A) + \mu(B)$. Clearly $n < m$. Also note that $\mu(A \cup \{a\}) = \mu(A) + \mu(\{a\})$. Hence:

$$n < m \iff 2nm + 2n\mu(\{a\}) < 2nm + 2m\mu(\{a\}) \iff$$
$$sim_{\mu DS}(A, B) = \frac{2n}{m} < \frac{2n + 2\mu(\{a\})}{m + \mu(\{a\})} = sim_{\mu DS}(A \cup \{a\}, B),$$

which means that $S4$ holds.

If $a \notin A \cup B$, then $sim_{\mu DS}(A, B) = \frac{2n}{m} > \frac{2n}{m+\mu(\{a\})} = sim_{\mu DS}(A \cup \{a\}, B)$, so $S5$ holds too.

[3] Tversky index is an asymmetric by design similarity index on sets that compares a variant to a prototype. If we consider X to be the prototype and Y to be the variant, then α corresponds to the weight of the prototype and β corresponds to the weight of the variant. For the interpretation of X and Y as prototype and variant, α usually differs from β [25]. However for the interpretations used in this paper, the case $\alpha \neq \beta$ does not make much sense.

(*Tversky and Symmetric Tversky μ-indexes*) If $X = Y$ then $sim_{\mu T}^{\alpha,\beta}(X, Y) = sim_{\mu T}^{\alpha,\beta}(Y, X)$. If $X \neq Y$ then $sim_{\mu T}^{\alpha,\beta}(X, Y) = sim_{\mu T}^{\alpha,\beta}(Y, X) \iff \alpha = \beta$. Let $n = \mu(A \cap B)$, $k = \mu(A \setminus B)$ and $l = \mu(B \setminus A)$.

If $a \in B \setminus A$ then $\mu(A \cup \{a\}) \cap B = \mu(A \cap B) + \mu(\{a\}) = n + \mu(\{a\})$, $(A \cup \{a\}) \setminus B = A \setminus B$, and $B \setminus A = (B \setminus (A \cup a)) \cup \{a\}$. Hence

$sim_{\mu T}^{\alpha,\beta}(A, B) = \frac{n}{n+\alpha k+\beta l} < \frac{n+\mu(\{a\})}{n+\alpha k+\beta(l-\mu(\{a\}))} = sim_{\mu T}^{\alpha,\beta}(A \cup \{a\}, B)$, so we are done with $S4$.

If $a \notin A \cup B$, then $sim_{\mu T}^{\alpha,\beta}(A, B) = \frac{n}{n+\alpha k+\beta l} > \frac{n}{n+\alpha(k+\mu(\{a\}))+\beta l} = sim_{\mu T}^{\alpha,\beta}(A \cup \{a\}, B)$, so $S5$ holds too.

(*Braun-Blanquet index*) It obviously satisfies $S1$–$S3$. Let $n = \mu(A \cap B)$, $r = \mu(A)$ and $s = \mu(B)$. If $a \in B \setminus A$ and $\mu(A) \geq \mu(B)$, then $n < r$ so

$sim_{\mu BB}(A, B) = \frac{n}{r} < \frac{n+\mu(\{a\})}{r+\mu(\{a\})} = sim_{\mu BB}(A \cup \{a\}, B)$.

If $\mu(A) + \mu(\{a\}) < \mu(B)$, then $sim_{\mu BB}(A, B) = \frac{n}{s} < \frac{n+\mu(\{a\})}{s} = sim_{\mu BB}(A \cup \{a\}, B)$.

If $\mu(A) < \mu(B) < \mu(A) + \mu(\{a\})$, then $sim_{\mu BB}(A, B) = \frac{n}{s} < \frac{n+\mu(\{a\})}{r+\mu(\{a\})} = sim_{\mu BB}(A \cup \{a\}, B)$.

Hence $S4$ does hold is this case.

If $a \notin A \cup B$ and $\mu(B) \leq \mu(A)$, then $sim_{\mu BB}(A, B) = \frac{n}{s} > \frac{n}{s+\mu(\{a\})} = sim_{\mu BB}(A \cup \{a\}, B)$. If $\mu(A) + \mu(\{a\}) > \mu(B) > \mu(A)$, then $sim_{\mu BB}(A, B) = \frac{n}{r} > \frac{n}{s+\mu(\{a\})} = sim_{\mu BB}(A \cup \{a\}, B)$. However if $\mu(B) > \mu(A) + \mu(\{a\})$, then $sim_{\mu BB}(A, B) = \frac{n}{r} = sim_{\mu BB}(A \cup a, B)$. This means that only $S5'$, not $S5$, is satisfied. □

Note that none of the above results holds if a measure μ is not finite, however Proposition 1 still holds for not null set free measures if the axiom $S3$ is not required.

From all the indexes analyzed above, only Marczewski–Steinhaus μ-index (i.e. also Jaccard index) is metrical as $diff_{MS}(X, Y) = 1 - sim_{MS}(X, Y)$ is a proper metric [13]. Also $diff_{MS}(X, Y) = \frac{\mu((X \setminus Y) \cup (Y \setminus X))}{\mu(X \cup Y)}$, appears to have a natural interpretation, while the other differences, $diff_{\mu DS}(X, Y)$, $diff_{\mu T}^{\alpha,\beta}(X, Y)$, and $diff_{\mu BB}(X, Y)$ look rather artificial.

The Symmetric Tversky index and μ-index are useful when one wants to express the difference of importance (w.r.t. similarity) between the intersection $X \cap Y$ and the rest of $X \cup Y$, i.e. $(X \cup Y) \setminus (X \cap Y)$. If $\alpha < 1$, the measure of $X \cap Y$ is more influential than that of the rest of $X \cup Y$, i.e. $(X \cup Y) \setminus (X \cap Y)$, if $\alpha > 1$ it is otherwise. Both Marczewski–Steinhaus and Dice–Sørensen μ-indexes are special cases of the Symmetric Tversky μ-index, the former with $\alpha = 1$ and the latter with $\alpha = 0.5$.

The Tversky μ-index with $\alpha \neq \beta$ implies $sim_{\mu T}^{\alpha,\beta}(X, Y) \neq sim_{\mu T}^{\alpha,\beta}(Y, X)$ for all $X \neq Y$, which is hard to justify and interpret in the setting of this paper. We will show that the concept of optimal approximation proposed later does not work for Tversky μ-index with $\alpha \neq \beta$.

5. Optimal approximations

Let $\mathcal{AS} = (U, E)$ be a *Pawlak approximation space*, i.e. U is a finite and non-empty set, called *universe*, and $E \subseteq U \times U$ is an *equivalence relation* on U.

We can now provide our *general* definition of optimal approximation.

Definition 3. For every set $X \subseteq U$, a definable set $\mathsf{O} \in \mathbb{D}$ is an **optimal approximation** of X (w.r.t. a given similarity measure *sim* that satisfies the axiom $S2$) if and only if:

$$sim(X, \mathsf{O}) = \max_{\mathsf{A} \in \mathbb{D}}(sim(X, \mathsf{A}))$$

The set of all optimal approximations of X will be denoted by $\mathbf{Opt}_{sim}(X)$. □

A specific optimal approximation depends on the precise definition of the similarity measure *sim*. If $sim_1 \neq sim_2$ then clearly $\mathbf{Opt}_{sim_1}(X)$ might differ from $\mathbf{Opt}_{sim_2}(X)$ for some $X \subseteq U$.

Note that Definition 3 does not make any sense for the Tversky μ-index with $\alpha \neq \beta$, as is such case, if $X \neq \mathsf{O}$, then $sim_{\mu T}^{\alpha,\beta}(X, \mathsf{A}) > sim_{\mu T}^{\alpha,\beta}(\mathsf{A}, X) \iff \alpha < \beta$. In the rough sets approach there is no reason why the set $X \setminus \mathsf{A}$ should be treated differently than $\mathsf{A} \setminus X$. While similarities without the axiom $S2$ have some applications (for example to make a distinction between prototypes and variants, cf. [25]), they are not part of this paper.

Axioms $S4$ and $S5$ imply that all optimal approximations reside between lower and upper approximations (inclusive), for all similarity indexes that satisfy them.

Proposition 2. *Assume that a similarity index sim(...) satisfies the axioms $S4$–$S5$. Then, for every set $X \subseteq U$, and every $\mathsf{O} \in \mathbf{Opt}_{sim}(X)$, we have*

$$\underline{\mathbf{A}}(X) \subseteq \mathsf{O} \subseteq \overline{\mathbf{A}}(X)$$

Proof. Suppose that $\underline{\mathbf{A}}(X) \not\subseteq \mathsf{O}$, i.e. $C = \underline{\mathbf{A}}(X) \setminus \mathsf{O} \neq \emptyset$. Since $C \subseteq X$, then by axiom $S4$, $sim(\mathsf{O}, X) < sim(\mathsf{O} \cup C, X)$, so O must not be optimal. Now suppose that $\mathsf{O} \not\subseteq \overline{\mathbf{A}}(X)$, i.e. $C = \mathsf{O} \setminus \overline{\mathbf{A}}(X) \neq \emptyset$. By axiom $S5$, $sim(\mathsf{O} \setminus C, X) > sim(\mathsf{O}, X)$, so O must not be optimal again. □

We also have to note here that the above result depends on the axiom $S5$ and its weakened version $S5'$ does not suffice, so Proposition 2 cannot be applied for Braun-Blanquet μ-index.

For example, consider the universe of elements $U = \{a_1, a_2, b_1, b_2, b_3, c_1, c_2, c_3, c_4, c_5, c_6, c_7, d_1\}$ with equivalence classes $\mathfrak{Comp} = \{A, B, C, D\}$, where $A = \{a_1, a_2\}$, $B = \{b_1, b_2, b_3\}$, $C = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7\}$, $D = \{d_1\}$ and select the set $X = \{a_1, b_1, b_2, c_1, c_2, c_3\}$. Also assume that $\mu(Y) = |Y|$ for all $Y \subseteq U$. When we compare X to all definable sets using the Braun-Blanquet index, the maximum similarity value we obtain is $\frac{1}{2}$. We can get this by evaluating $sim_{BB}(X, A \cup B)$, or $sim_{BB}(X, B \cup C)$, or $sim_{BB}(X, A \cup B \cup C)$. By our definition for the definable set $A \cup B \cup D$ we get $sim_{BB}(X, A \cup B \cup D) = \frac{1}{2}$ which makes it an optimal approximation, but $A \cup B \cup D \not\subseteq \overline{\mathbf{A}}(X) = A \cup B \cup C$.

One of the consequences of Proposition 2 is that any optimal approximation of X, is the union of the lower approximation of X and some element $\mathsf{A} \in \mathbb{B}(X) \cup \{\emptyset\}$.

Definition 4. Let $X \subseteq U$, and $\mathsf{O} \in \mathbb{D}$. We say that O is an ***intermediate approximation*** of X, if

$$\underline{\mathbf{A}}(X) \subseteq \mathsf{O} \subseteq \overline{\mathbf{A}}(X)$$

The set of all intermediate approximations of X will be denoted by $\mathbf{IA}(X)$. □

Note that it is independent of a similarity index that is used to find an optimal approximation $\mathbf{Opt}_{sim}(X)$.

From Proposition 2 we have:

Corollary 3. *For each set $X \subseteq U$,*

1. $\mathbf{Opt}_{sim}(X) \subseteq \mathbf{IA}_{sim}(X)$.
2. *If $\mathsf{O} \in \mathbf{Opt}_{sim}(X)$ then there exists $\mathsf{A} \in \mathbb{B}(X) \cup \{\emptyset\}$ such that $\mathsf{O} = \underline{\mathbf{A}}(X) \cup \mathsf{A}$.*
3. *If $\mathsf{O} \in \mathbf{Opt}_{sim}(X)$ then there exists $\mathsf{B} \in \mathbb{B}(X) \cup \{\emptyset\}$ such that $\mathsf{O} = \overline{\mathbf{A}}(X) \setminus \mathsf{B}$.* □

These are properties of optimal approximations. The set of them must be a portion of the intermediate approximations. Any optimal approximation must be the union of the lower approximation with some definable set which is in the upper but not the lower approximation (or is the empty set itself). It must also be possible to represent any optimal approximation by the upper approximation with some border set removed from it (or the empty set if the approximation is optimal).

The notion of optimal approximation also introduces some structure to the current available field of similarity measures, as certain different similarity indexes may generate the same optimal approximations.

Definition 5. We say that two similarity indexes sim_1 and sim_2 are ***consistent*** if for all sets A, B, C,

$$sim_1(A, B) < sim_1(A, C) \iff sim_2(A, B) < sim_2(A, C). \qquad \square$$

This clearly leads to the following result.

Corollary 4. *If sim_1 and sim_2 are consistent then for each $X \subseteq U$,*

1. $\mathbf{Opt}_{sim_1}(X) = \mathbf{Opt}_{sim_2}(X)$.
2. *sim_1 satisfies the axioms $S4$ and $S5$ if and only if sim_2 satisfies them.* □

These concepts will allow us to extend results and algorithms designed for specific similarity indexes, to larger classes of consistent indexes.

First we will show that Marczewski–Steinhaus μ-index and Symmetric Tversky μ-index are consistent.

Proposition 3 *(Consistency of Marczewski–Steinhaus and sym. Tversky μ-indexes). For all A, B, C and $\alpha > 0$*

$$sim_{MS}(A, B) < sim_{MS}(A, C) \iff sim^{\alpha}_{\mu sT}(A, B) < sim^{\alpha}_{\mu sT}(A, C).$$

Proof. If $A = C$ then $sim_{MS}(A, C) = sim^{\alpha}_{\mu sT}(A, C) = 1$, so the equivalence holds. Assume $A \neq C$. Since $sim_{MS}(A, C) > 0$, then $A \cap C \neq \emptyset$. Moreover $A \setminus C \neq \emptyset$ or $C \setminus A \neq \emptyset$. Hence:

$$sim_{MS}(A,B) < sim_{MS}(A,C) \iff \frac{\mu(A\cap B)}{\mu(A\cup B)} < \frac{\mu(A\cap C)}{\mu(A\cup C)} \iff$$
$$\frac{\mu(A\cap B)}{\mu(A\cap B)+\mu(A\setminus B)+\mu(B\setminus A)} < \frac{\mu(A\cap C)}{\mu(A\cap C)+\mu(A\setminus C)+\mu(C\setminus A)} \iff$$
$$\mu(A\cap B)(\mu(A\setminus C)+\mu(C\setminus A)) < \mu(A\cap C)(\mu(A\setminus B)+\mu(B\setminus A)) \iff$$
$$\frac{\mu(A\cap B)}{\mu(A\cap C)} < \frac{\mu(A\setminus B)+\mu(B\setminus A)}{\mu(A\setminus C)+\mu(C\setminus A)} \iff \frac{\mu(A\cap B)}{\mu(A\cap C)} < \frac{\alpha\mu(A\setminus B)+\alpha\mu(B\setminus A)}{\alpha\mu(A\setminus C)+\alpha\mu(C\setminus A)} \iff$$
$$\frac{\mu(A\cap B)}{\mu(A\cap B)+\alpha\mu(A\setminus B)+\alpha\mu(B\setminus A)} < \frac{\mu(A\cap C)}{\mu(A\cap C)+\alpha\mu(A\setminus C)+\alpha\mu(C\setminus A)} \iff sim^{\alpha}_{\mu sT}(A,B)$$
$$< sim^{\alpha}_{\mu sT}(A,C). \quad \square$$

The above proposition immediately implies that the Dice–Sørensen and Marczewski–Steinhaus μ-indexes are consistent too.

Corollary 5 *(Consistency of Marczewski–Steinhaus and Dice–Sørensen μ-indexes). For all A, B, C,*

$$sim_{MS}(A,B) < sim_{MS}(A,C) \iff sim_{\mu DS}(A,B) < sim_{\mu DS}(A,C).$$

Proof. Since $sim_{\mu DS}(A,B) = sim^{0.5}_{\mu sT}(A,B)$. $\square$

In general the Braun-Blanquet μ-index is *not consistent* with the Marczewski–Steinhaus index. To show this, consider the case of $A=\{a_1,a_2,a_3,a_4\}$, $B=\{a_1,a_2,a_3,a_5,\ldots,a_{21}\}$, $C=\{a_1,a_4,a_{22},\ldots,a_{32}\}$, and μ is just cardinality, so $\mu(X)=|X|$. We have here $|A|=4$, $|B|=20$, $|C|=13$, $|A\cap B|=3$ and $|A\cap C|=2$. Hence $sim_{MS}(A,B)=\frac{3}{21}=\frac{1}{7}>sim_{MS}(A,C)=\frac{2}{15}$, while $sim_{BB}(A,B)=\frac{3}{20}<sim_{BB}(A,C)=\frac{2}{13}$.

The similarities that do not satisfy the axiom $S2$ are not covered by the theory presented in this paper, however, for curiosity reasons, we will show that Tversky index with $\alpha\neq\beta$ is not consistent with the Jaccard index. Consider $A=\{a_1,a_2,a_3,a_4\}$, $B=\{a_1,a_2,a_3,a_4,a_6,\ldots,a_{12}\}$, and $C=\{a_3,a_4,a_5\}$. Then $|A|=4$, $|B|=11$, $|C|=3$, $|A\cap B|=4$, $|A\cap C|=2$, $|A\cup B|=11$, and $|A\cup C|=5$. So in this case $sim_J(A,B)=\frac{4}{11}<sim_J(A,C)=\frac{2}{5}$, but for any α and β such that $\frac{\alpha}{\beta}>\frac{5}{4}$, we have $sim^{\alpha,\beta}_T(A,B)>sim^{\alpha,\beta}_T(A,C)$. For example for $\alpha=1.5$ and $\beta=1.0$ we have $sim^{\alpha,\beta}_T(A,B)=\frac{4}{11}>sim^{\alpha,\beta}_T(A,C)=\frac{1}{3}$. Hence, in general Marczewski–Steinhaus μ-index and Tversky μ-index are not consistent for $\alpha\neq\beta$.

So far we have not applied the concept of optimal approximation to any specific similarity measure. We only assumed that the function *sim* satisfies the axioms $S1$–$S5$. However to show more specific and detailed properties of optimal approximations, especially an efficient algorithm to find one, we need to choose a specific similarity measure. Due to Corollary 4(1), the results will hold for all other consistent similarity indexes.

6. Optimal approximations with Marczewski–Steinhaus similarity index

This section contains the main results of this paper.

Marczewski–Steinhaus μ-index is metrical, and has a natural and regular definition, which makes it perfect for discovering and proving mathematical results.

Let $\mathcal{AS}=(U,E)$ be a Pawlak approximation space (i.e. U is finite), and let $\mu: U\to\mathbb{R}$ be a given ***finite*** and ***null-free*** measure on U (defined element-wise).

For every $X, Y\subseteq U$, such that $X\setminus Y\neq\emptyset$, we define the index $\rho(X,Y)$, called the *ratio of common to distinct elements*, as follows

$$\rho(X,Y)=\frac{\mu(X\cap Y)}{\mu(X\setminus Y)}.$$

Note that $\rho(X,Y)$ is sound only if μ is finite and null-free.

By Proposition 1(1), the Marczewski–Steinhaus μ-index satisfies the axioms $S1$–$S5$, so the property specified by Proposition 2 and Definition 4 is satisfied.

Now, suppose that $\mathsf{O}\in\mathbf{IA}(X)$ is an intermediate approximation of X, and $\mathbf{x}\in\mathfrak{B}(X)$ is an element of the border of X which has no common element with O, i.e. $\mathsf{O}\cap\mathbf{x}=\emptyset$. To determine which definable set is a better approximation of X (more similar to X), O or $\mathsf{O}\cup\mathbf{x}$, we can use the lemma below.

Lemma 1. *Let* $X \subseteq U$, $\mathsf{O} \in \mathbf{IA}(X)$, $\mathsf{A}, \mathsf{B} \in \mathbb{B}(X)$, $\mathsf{A} \cap \mathsf{O} = \emptyset$, *and* $\mathsf{B} \subseteq \mathsf{O}$. *Then*

1. $sim_{MS}(X, \mathsf{O} \cup \mathsf{A}) \geq sim_{MS}(X, \mathsf{O}) \iff \rho(\mathsf{A}, X) \geq \frac{\mu(X \cap \mathsf{O})}{\mu(X \cup \mathsf{O})} = sim_{MS}(X, \mathsf{O})$
2. $sim_{MS}(X, \mathsf{O} \setminus \mathsf{B}) \leq sim_{MS}(X, \mathsf{O}) \iff \rho(\mathsf{B}, X) \geq \frac{\mu(X \cap \mathsf{O})}{\mu(X \cup \mathsf{O})} = sim_{MS}(X, \mathsf{O})$

Proof. (1) Let $\mu(X \cap \mathsf{O}) = n$, $\mu(X \cup \mathsf{O}) = m$, $\mu(\mathsf{A} \setminus X) = l$, and $\mu(\mathsf{A} \cap X) = k$, so $\rho(\mathsf{A}, X) = \frac{k}{l}$. By Corollary 2(3) and the fact that μ is null-free, the values of n, m, l, k are all bigger than zero.

We have $sim_{MS}(X, \mathsf{O}) = \frac{\mu(X \cap \mathsf{O})}{\mu(X \cup \mathsf{O})}$ and $sim_{MS}(X, \mathsf{O} \cup \mathsf{A}) = \frac{\mu(X \cap (\mathsf{O} \cup \mathsf{A}))}{\mu(X \cup (\mathsf{O} \cup \mathsf{A}))}$.

Because $\mathsf{A} \cap \mathsf{O} = \emptyset$, $\mu(X \cap (\mathsf{O} \cup \mathsf{A})) = \mu(X \cap \mathsf{O}) + \mu(X \cap \mathsf{A}) = n + k$ and

$$\mu(X \cup (\mathsf{O} \cup \mathsf{A})) = \mu(X \cup \mathsf{O}) + \mu(\mathsf{A} \setminus X) = m + l.$$

Hence,

$$sim_{MS}(X, \mathsf{O} \cup \mathsf{A}) \geq sim_{MS}(X, \mathsf{O}) \iff \frac{n+k}{m+l} \geq \frac{n}{m} \iff \frac{k}{l} \geq \frac{n}{m} \iff \rho(\mathsf{A}, X) \geq \frac{\mu(X \cap \mathsf{O})}{\mu(X \cup \mathsf{O})} = sim_{MS}(X, \mathsf{O}).$$

(2) Let $\mu(X \cap \mathsf{O}) = n$, $\mu(X \cup \mathsf{O}) = m$, $\mu(\mathsf{B} \setminus X) = l$, and $\mu(\mathsf{B} \cap X) = k$, so $\rho(\mathsf{B}, X) = \frac{k}{l}$. By Corollary 2(3) and the definition of a measure μ, the values of n, m, l, k are all bigger than zero.

We have here $sim_{MS}(X, \mathsf{O}) = \frac{\mu(X \cap \mathsf{O})}{\mu(X \cup \mathsf{O})}$ and $sim_{MS}(X, \mathsf{O} \setminus \mathsf{B}) = \frac{\mu(X \cap (\mathsf{O} \setminus \mathsf{B}))}{\mu(X \cup (\mathsf{O} \setminus \mathsf{B}))}$.

Because $\mathsf{B} \subseteq \mathsf{O}$, $\mu(X \cap (\mathsf{O} \setminus \mathsf{B})) = \mu(X \cap \mathsf{O}) - \mu(X \cap \mathsf{B}) = n - k$ and

$$\mu(X \cup (\mathsf{O} \setminus \mathsf{B})) = \mu(X \cup \mathsf{O}) - \mu(\mathsf{B} \setminus X) = m - l.$$

Thus,

$$sim_{MS}(X, \mathsf{O} \setminus \mathsf{B}) \leq sim_{MS}(X, \mathsf{O}) \iff \frac{n-k}{m-l} \leq \frac{n}{m} \iff \frac{k}{l} \geq \frac{n}{m} \iff \rho(\mathsf{B}, X) \geq \frac{\mu(X \cap \mathsf{O})}{\mu(X \cup \mathsf{O})} = sim_{MS}(X, \mathsf{O}). \qquad \square$$

Note that we cannot replace equations (1) and (2) of Lemma 1 by one equation, as the assumptions about A and B are entirely different. Moreover Lemma 1 does not hold if the measure μ is not null-free, as then the values on n, m, l, k from the proof of Lemma 1 are no longer bigger than zero.

Clearly the above lemma also holds for $\mathsf{A} = \mathbf{x} \in \mathfrak{B}(X)$. Intuitively, if more than half of the elements of $\mathbf{x}$ also belong to X, or equivalently, if more elements of $\mathbf{x}$ belong to X than do not, the rough set $\mathsf{O} \cup \mathbf{x}$ should approximate X better than O. The results below support this intuition.

Corollary 6 *('Majority rule'). Let* $X \subseteq U$, $\mathsf{O} \in \mathbf{IA}(X)$, $\mathbf{x} \in \mathfrak{B}(X)$, *and* $\mathbf{x} \cap \mathsf{O} = \emptyset$. *Then:* $\mu(\mathbf{x} \cap X) \geq \mu(\mathbf{x} \setminus X) \iff \frac{\mu(\mathbf{x} \cap X)}{\mu(\mathbf{x})} \geq \frac{1}{2} \Longrightarrow sim_{MS}(X, \mathsf{O} \cup \mathbf{x}) \geq sim_{MS}(X, \mathsf{O})$.

Proof. Clearly $\mu(\mathbf{x} \cap X) \geq \mu(\mathbf{x} \setminus X) \iff \frac{\mu(\mathbf{x} \cap X)}{\mu(\mathbf{x} \setminus X)} \geq 1$. But $\frac{\mu(X \cap \mathsf{O})}{\mu(X \cup \mathsf{O})} \leq 1$, so by Lemma 1,

$$sim_{MS}(X, \mathsf{O} \cup \mathbf{x}) \geq sim_{MS}(X, \mathsf{O}). \qquad \square$$

However, the reciprocal of Corollary 6 does not hold. It may happen that $\frac{\mu(\mathbf{x} \cap X)}{\mu(\mathbf{x})} < \frac{1}{2}$, but the rough set $\mathsf{O} \cup \mathbf{x}$ still approximates X better than O. Consider the following example.

Example 1. Let $\mathsf{O} = \{a_1, a_2, a_3, a_4, a_5\}$, $\mathbf{x} = \{b_1, b_2, b_3, b_4, b_5\}$ and $X = \{a_1, a_2, a_3, a_4, a_5, b_1, b_2, c_1\}$. Here we assume that the measure μ is cardinality i.e. $\mu(A) = |A|$ for all finite A. Then we have $\frac{|\mathbf{x} \cap X|}{|\mathbf{x}|} = \frac{2}{5} = 0.4 < \frac{1}{2}$, but $sim_J(X, \mathsf{O} \cup \mathbf{x}) = \frac{7}{11} = 0.636 > sim_J(X, \mathsf{O}) = \frac{5}{8} = 0.6254$. $\square$

We know from Proposition 2 that if $\mathsf{O} \in \mathbf{Opt}(X)$, then $\mathsf{O} = \underline{\mathbf{A}}(X) \cup \mathbf{x}_1 \cup \ldots \cup \mathbf{x}_k$, for some k, where each $\mathbf{x}_i \in \mathfrak{B}(X)$, $i = 1, \ldots, k$. Lemma 1 allows us to explicitly define the components $\mathbf{x}_i \in \mathfrak{B}(X)$.

Theorem 1. *For every* $X \subseteq U$, *the following two statements are equivalent:*

1. $\mathsf{O} \in \mathbf{Opt}(X)$
2. $\mathsf{O} \in \mathbf{IA}(X) \wedge \forall \mathbf{x} \in \mathfrak{B}(X). \Big(\mathbf{x} \subseteq \mathsf{O} \iff \rho(\mathbf{x}, X) = \frac{\mu(\mathbf{x} \cap X)}{\mu(\mathbf{x} \setminus X)} \geq \frac{\mu(X \cap \mathsf{O})}{\mu(X \cup \mathsf{O})} = sim_{MS}(X, \mathsf{O})\Big)$.

Proof. (1) $\Rightarrow$ (2) By Proposition 2, $\mathsf{O} \in \mathbf{IA}(X)$. Let $\mathbf{x} \in \mathfrak{B}(X)$ and $\mathbf{x} \subseteq \mathsf{O}$. Suppose that $\frac{\mu(\mathbf{x}\cap X)}{\mu(\mathbf{x}\setminus X)} < \frac{\mu(X\cap \mathsf{O})}{\mu(X\cup \mathsf{O})}$. Then by Lemma 1, $sim_{MS}(X, \mathsf{O}\setminus \mathbf{x}) > sim_{MS}(X, \mathsf{O})$, so O is not optimal.

Let $\frac{\mu(\mathbf{x}\cap X)}{\mu(\mathbf{x}\setminus X)} \geq \frac{\mu(X\cap \mathsf{O})}{\mu(X\cup \mathsf{O})}$. Suppose that $\mathbf{x} \in \mathfrak{B}(X)$ and $\mathbf{x} \cap \mathsf{O} = \emptyset$. By Corollary 2(3), we have $\mathbf{x} \cap X \neq \emptyset$, so let $a \in \mathbf{x} \cap X$. Since $\mathbf{x} \cap \mathsf{O} = \emptyset$, then $a \in X \setminus \mathsf{O}$. Then by Proposition 1(1) and axiom $S4$, $sim_{MS}(X, \mathsf{O} \cup \{a\}) > sim_{MS}(X, \mathsf{O})$, so O is not optimal. Note that Lemma 1 gives only $sim_{MS}(X, \mathsf{O} \cup \mathbf{x}) \geq sim_{MS}(X, \mathsf{O})$ which is not strong enough.

(2) $\Rightarrow$ (1) Suppose O satisfies (2) but $\mathsf{O} \notin \mathbf{Opt}(X)$. Let $\mathsf{Q} \in \mathbf{Opt}(X)$. Hence, by the proof (1) $\Rightarrow$ (2), Q satisfies (2). We have to consider two cases $\mathsf{Q} \setminus \mathsf{O} \neq \emptyset$ and $\mathsf{O} \setminus \mathsf{Q} \neq \emptyset$.

(*Case 1*) Let $\mathsf{Q} \setminus \mathsf{O} \neq \emptyset$ and let $\mathbf{y} \in \mathfrak{B}(X)$ be such that $\mathbf{y} \subseteq \mathsf{Q} \setminus \mathsf{O}$. Since Q satisfies (2), we have $\frac{\mu(\mathbf{y}\cap X)}{\mu(\mathbf{y}\setminus X)} \geq \frac{\mu(X\cap \mathsf{Q})}{\mu(X\cup \mathsf{Q})} = sim_{MS}(X, \mathsf{Q})$, and because $\mathsf{Q} \in \mathbf{Opt}(X)$, $sim_{MS}(X, \mathsf{Q}) \geq sim_{MS}(X, \mathsf{O})$. But this means that $\frac{\mu(\mathbf{y}\cap X)}{\mu(\mathbf{y}\setminus X)} \geq \frac{\mu(X\cap \mathsf{O})}{\mu(X\cup \mathsf{O})}$. However O also satisfies (2) and $\mathbf{y} \in \mathfrak{B}(X)$, so by (2), $\mathbf{y} \subseteq \mathsf{O}$, a contradiction. Hence $\mathsf{Q} \setminus \mathsf{O} = \emptyset$.

(*Case 2*) Let $\mathsf{O} \setminus \mathsf{Q} = \{\mathbf{y}_1, \ldots, \mathbf{y}_p\} \subseteq \mathfrak{B}(X)$. Let $\mu(X \cap \mathsf{O}) = n$, $\mu(X \cup \mathsf{O}) = m$, and $\mu(\mathbf{y}_i \setminus X) = l_i$, $\mu(\mathbf{y}_i \cap X) = k_i$, for $i = 1, \ldots, p$. Since O satisfies (2), for each $i = 1, \ldots, p$, we have $\frac{\mu(\mathbf{y}_i\cap X)}{\mu(\mathbf{y}_i\setminus X)} \geq \frac{|X\cap \mathsf{O}|}{|X\cup \mathsf{O}|}$, or equivalently $\frac{k_i}{l_i} \geq \frac{n}{m}$. Hence

$$(k_1 + \ldots + k_p)m \geq (l_1 + \ldots + l_p)n.$$

On the other hand, $sim_J(X, \mathsf{Q}) = sim_J(X, \mathsf{O} \setminus (\mathbf{y}_1 \cup \ldots \cup \mathbf{y}_p)) > sim_J(X, \mathsf{O})$, so by Lemma 1, $\frac{\mu((\mathbf{y}_1\cup\ldots\cup\mathbf{y}_p)\cap X)}{\mu((\mathbf{y}_1\cup\ldots\cup\mathbf{y}_p)\setminus X)} < \frac{\mu(X\cap \mathsf{O})}{\mu(X\cup \mathsf{O})}$. Because $\mathbf{y}_i$ are components, we have $\mathbf{y}_i \cap \mathbf{y}_j = \emptyset$ when $i \neq j$. Thus $\mu((\mathbf{y}_1 \cup \ldots \cup \mathbf{y}_p) \cap X) = \mu(\mathbf{y}_1 \cap X) + \ldots + \mu(\mathbf{y}_p \cap X) = k_1 + \ldots + k_p$, and $\mu((\mathbf{y}_1 \cup \ldots \cup \mathbf{y}_p) \setminus X) = \mu(\mathbf{y}_1 \setminus X) + \ldots + \mu(\mathbf{y}_p \setminus X) = l_1 + \ldots + l_p$. This means that we have

$$\frac{\mu((\mathbf{y}_1 \cup \ldots \cup \mathbf{y}_p) \cap X)}{\mu((\mathbf{y}_1 \cup \ldots \cup \mathbf{y}_p) \setminus X)} < \frac{\mu(X \cap \mathsf{O})}{\mu(X \cup \mathsf{O})} \iff \frac{k_1 + \ldots + k_p}{l_1 + \ldots + l_p} < \frac{n}{m},$$

which yields

$$(k_1 + \ldots + k_p)m < (l_1 + \ldots + l_p)n,$$

a contradiction, i.e. $\mathsf{O} \setminus \mathsf{Q} = \emptyset$. Thus, $\mathsf{Q} \setminus \mathsf{O} = \emptyset$ and $\mathsf{O} \setminus \mathsf{Q} = \emptyset$, i.e., $\mathsf{Q} = \mathsf{O}$, so $\mathsf{O} \in \mathbf{Opt}(X)$. □

Theorem 1 gives the necessary and sufficient conditions for optimal approximations of X (with respect to the Marczewski–Steinhaus index and a given measure $\mu : U \to \mathbb{R}$) in terms of the elements of $\mathfrak{B}(X)$. We will use it to build an efficient algorithm for finding optimal approximations.

By Theorem 1, the value of $\rho(\mathbf{x}, X)$ will indicate if $\mathbf{x} \in \mathfrak{B}(X)$ is a part of an optimal approximation of X, or not. Since $\mathfrak{B}(X)$ is finite, its elements can be enumerated by natural numbers $1, \ldots, |\mathfrak{B}(X)|$.

- Assume that $r = |\mathfrak{B}(X)|$, $\mathfrak{B}(X) = \{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ and also

$$i \leq j \iff \rho(\mathbf{x}_i, X) \geq \rho(\mathbf{x}_j, X).$$

In other words, we sort $\mathfrak{B}(X)$ by decreasing values of $\rho(\mathbf{x}, X)$. We will use this sorting to build a special sequence of intermediate approximations.

Let $\mathsf{O}_0, \mathsf{O}_1, \ldots, \mathsf{O}_r \in \mathbf{IA}(X)$ be the sequence of intermediate approximations of X defined for $i = 0, \ldots, r-1$ as follows: $\mathsf{O}_0 = \underline{\mathbf{A}}(X)$ and

$$\mathsf{O}_{i+1} = \begin{cases} \mathsf{O}_i \cup \mathbf{x}_{i+1} & \text{if } sim_{MS}(X, \mathsf{O}_i \cup \mathbf{x}_{i+1}) \geq sim_{MS}(X, \mathsf{O}_i) \\ \mathsf{O}_i & \text{otherwise.} \end{cases}$$

Note that usually $\mathsf{O}_r \neq \overline{\mathbf{A}}(X) = \underline{\mathbf{A}}(X) \cup \mathbf{x}_1 \cup \ldots \cup \mathbf{x}_r$, since $\mathsf{O}_r = \mathbf{x}_1 \cup \ldots \cup \mathbf{x}_r$, only if $sim_{MS}(X, \mathsf{O}_i \cup \mathbf{x}_{i+1}) \geq sim_{MS}(X, \mathsf{O}_i)$ for all $i = 0, \ldots, r-1$, or equivalently, if $\mathsf{O}_i = \mathbf{x}_1 \cup \ldots \cup \mathbf{x}_i$ for $i = 1, \ldots, r$.

We claim that at least one of these O_i's is an optimal approximation. The following technical result is needed to prove this claim.

Lemma 2. *Let $k_1, \ldots, k_n$ and $l_1, \ldots, l_n$ be positive numbers such that $\frac{k_1}{l_1} \geq \frac{k_i}{l_i}$ for $i = 1, \ldots, n$. Then $\frac{k_1}{l_1} \geq \frac{k_1+\ldots+k_n}{l_1+\ldots+l_n}$.*

Proof. $\frac{k_1}{l_1} \geq \frac{k_i}{l_i}$ implies $k_1 l_i \geq k_i l_1$ for $i = 1, \ldots, n$. Hence $k_1 l_1 + k_1 l_2 + \ldots + k_1 l_n \geq k_1 l_1 + k_2 l_1 + \ldots + k_n l_1 \iff \frac{k_1}{l_1} \geq \frac{k_1+\ldots+k_n}{l_1+\ldots+l_n}$, which ends the proof. □

The essential properties of the sequence $\mathsf{O}_0, \mathsf{O}_1, \ldots, \mathsf{O}_r$ are provided by the following theorem.

Theorem 2. *For every $X \subseteq U$, we set $r = |\mathfrak{B}(X)|$, and we have*

1. $sim_{MS}(X, \mathsf{O}_{i+1}) \geq sim_{MS}(X, \mathsf{O}_i)$, *for* $i = 0, \ldots, r-1$.
2. *If* $\rho(\mathbf{x}_1, X) \leq sim_{MS}(X, \underline{\mathbf{A}}(X))$ *then* $\underline{\mathbf{A}}(X) \in \mathbf{Opt}(X)$.
3. *If* $\rho(\mathbf{x}_r, X) \geq sim_{MS}(X, \overline{\mathbf{A}}(X))$ *then* $\overline{\mathbf{A}}(X) \in \mathbf{Opt}(X)$.
4. *If* $sim_{MS}(X, \mathsf{O}_p) \leq \rho(\mathbf{x}_p, X)$ *and* $sim_{MS}(X, \mathsf{O}_{p+1}) > \rho(\mathbf{x}_{p+1}, X)$, *then* $\mathsf{O}_p \in \mathbf{Opt}(X)$, *for* $p = 1, \ldots, r-1$.
5. *If* $sim_{MS}(X, \mathsf{O}_r) \leq \rho(\mathbf{x}_r, X)$ *then* $\mathsf{O}_r = \overline{\mathbf{A}}(X) \in \mathbf{Opt}(X)$.
6. *If* $\mathsf{O}_p \in \mathbf{Opt}(X)$, *then* $\mathsf{O}_i = \mathsf{O}_p$ *for all* $i = p+1, \ldots, r$. *In particular* $\mathsf{O}_r \in \mathbf{Opt}(X)$.
7. $\mathsf{O} \in \mathbf{Opt}(X) \Longrightarrow \mathsf{O} \subseteq \mathsf{O}_p$, *where* p *is the smallest one from (6).*

Proof. (1) Immediately from Lemma 1 and the definition of the sequence $\mathsf{O}_0, \ldots, \mathsf{O}_r$.

(2) From Proposition 2 we have that if $\mathsf{O} \in \mathbf{Opt}(X)$, then $\mathsf{O} = \underline{\mathbf{A}}(X) \cup \mathbf{x}_{i_1} \cup \ldots \cup \mathbf{x}_{i_s}$ for some $i_j \in \{1, \ldots, r\}$. Since $\rho(\mathbf{x}_1, X) \geq \rho(\mathbf{x}_{i_j}, X)$ for $j = 1, \ldots, s$, by Lemma 2, $\rho(\mathbf{x}_1, X) \geq \frac{\mu((\mathbf{x}_{i_1} \cup \ldots \cup \mathbf{x}_{i_s}) \cap X)}{\mu((\mathbf{x}_{i_1} \cup \ldots \cup \mathbf{x}_{i_s}) \setminus X)}$. Hence $sim_{MS}(X, \underline{\mathbf{A}}(X)) \geq \frac{\mu((\mathbf{x}_{i_1} \cup \ldots \cup \mathbf{x}_{i_s}) \cap X)}{\mu((\mathbf{x}_{i_1} \cup \ldots \cup \mathbf{x}_{i_s}) \setminus X)}$, so by Lemma 1, $sim_{MS}(X, \underline{\mathbf{A}}(X)) \geq sim_{MS}(X, \mathsf{O})$, which means $\underline{\mathbf{A}}(X) \in \mathbf{Opt}(X)$.

(3) Note that $\rho(\mathbf{x}_r, X) \geq sim_{MS}(X, \overline{\mathbf{A}}(X))$ implies $\rho(\mathbf{x}_i, X) \geq sim_{MS}(X, \overline{\mathbf{A}}(X))$ for all $i = 1, \ldots, r$. Hence by Theorem 1, $\overline{\mathbf{A}}(X) \in \mathbf{Opt}(X)$.

(4) We have $sim_{MS}(X, \mathsf{O}_0) \leq sim_{MS}(X, \mathsf{O}_1) \leq \ldots \leq sim_{MS}(X, \mathsf{O}_r)$ and $\rho(\mathbf{x}_1, X) \geq \rho(\mathbf{x}_2, X) \geq \ldots \geq \rho(\mathbf{x}_r, X)$. Hence the property $sim_{MS}(X, \mathsf{O}_p) \leq \rho(\mathbf{x}_p, X)$ implies $\mathsf{O}_i = \underline{\mathbf{A}}(X) \cup \mathbf{x}_1 \cup \ldots \cup \mathbf{x}_i$ for all $i = 1, \ldots, p$. Adding the property $sim_{MS}(X, \mathsf{O}_{p+1}) > \rho(\mathbf{x}_{p+1}, X)$ implies $\mathsf{O}_i = \mathsf{O}_p$ for all $i = p+1, \ldots, r$, which meant that $\mathsf{O}_p = \underline{\mathbf{A}}(X) \cup \mathbf{x}_1 \cup \ldots \cup \mathbf{x}_p$ satisfies (2) of Theorem 1. Hence $\mathsf{O}_p \in \mathbf{Opt}(X)$.

(5) Again the property $sim_{MS}(X, \mathsf{O}_r) \leq \rho(\mathbf{x}_r, X)$ implies $\mathsf{O}_r = \underline{\mathbf{A}}(X) \cup \mathbf{x}_1 \cup \ldots \cup \mathbf{x}_r$, so $\mathsf{O}_r = \overline{\mathbf{A}}(X)$. Additionally $\mathsf{O}_r = \underline{\mathbf{A}}(X) \cup \mathbf{x}_1 \cup \ldots \cup \mathbf{x}_r$ satisfies (2) of Theorem 1, so $\mathsf{O}_r \in \mathbf{Opt}(X)$.

(6) From the proofs of (4) and (5).

(7) We have to show that if $\mathsf{O} = \underline{\mathbf{A}}(X) \cup \mathsf{A} \in \mathbf{Opt}(X)$, where $\mathsf{A} \in \mathbb{B}(X)$, then $\mathsf{A} \subseteq \mathbf{x}_1 \cup \ldots \cup \mathbf{x}_p$. Suppose $\mathbf{x}_j \subseteq \mathsf{A}$ and $j > p$. Then $\rho(\mathbf{x}_j, X) < sim_{MS}(X, \mathsf{O}_p) = sim_{MS}(X, \mathsf{O})$, so O does not satisfy (2) of Theorem 1. Hence $\mathsf{A} \subseteq \mathbf{x}_1 \cup \ldots \cup \mathbf{x}_p$. □

Point (1) of Theorem 2 states that O_{i+1} is a better (or equal) approximation of X than O_i, (2) and (3) characterize the case when either $\underline{\mathbf{A}}(X)$ or $\overline{\mathbf{A}}(X)$ are optimal approximations, while (4) shows conditions when some O_p is an optimal approximation. Point (5) states that once O_p is found to be optimal, we may stop calculations as the remaining O_{p+i} are the same as O_p, and the last point, (6) indicates that O_p is the greatest optimal approximation.

Algorithm 1 *(Finding the greatest optimal approximation). Let $X \subseteq U$.*

1. *Construct* $\underline{\mathbf{A}}(X)$, $\overline{\mathbf{A}}(X)$, *and* $\mathfrak{B}(X)$. *Assume* $r = |\mathfrak{B}(X)|$.
2. *For each* $\mathbf{x} \in \mathfrak{B}(X)$, *calculate* $\rho(\mathbf{x}, X) = \frac{\mu(\mathbf{x} \cap X)}{\mu(\mathbf{x} \setminus X)}$.
3. *Order* $\rho(\mathbf{x}, X)$ *in decreasing order and number the elements of* $\mathfrak{B}(X)$ *by this order, so* $\mathfrak{B}(X) = \{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ *and* $i \leq j \iff \rho(\mathbf{x}_i, X) \geq \rho(\mathbf{x}_j, X)$.
4. *If* $\rho(\mathbf{x}_1, X) \leq sim_{MS}(X, \underline{\mathbf{A}}(X))$ *then* $\mathsf{O} = \underline{\mathbf{A}}(X)$.
5. *If* $\rho(\mathbf{x}_r, X) \geq sim_{MS}(X, \overline{\mathbf{A}}(X))$ *then* $\mathsf{O} = \overline{\mathbf{A}}(X)$.
6. *If neither (4) nor (5) is applied, calculate* O_p, *starting from* $p = 1$ *and increasing* p *by* 1, *until* $sim_{MS}(X, \mathsf{O}_{p+1}) > \rho(\mathbf{x}_{p+1}, X)$. *If* $sim_{MS}(X, \mathsf{O}_{p+1}) > \rho(\mathbf{x}_{p+1}, X)$ *holds, set* $\mathsf{O} = \mathsf{O}_p$. □

Note that the biggest p in (6) of the above algorithm is $r-1$. However, due to Theorem 2(5), step (5) of the above algorithm covers the case $\mathsf{O} = \mathsf{O}_r$. Theorem 2 also guarantees that one of (4), (5), or (6) with $1 \leq p < r$ always holds. The case (4) of the above algorithm means that the optimal approximation O satisfies $\mathsf{O} = \mathsf{O}_0$, the case (5) corresponds to $\mathsf{O} = \mathsf{O}_r = \overline{\mathbf{A}}(X) = \underline{\mathbf{A}}(X) \cup \mathbf{x}_1 \cup \ldots \cup \mathbf{x}_r$, and the case (6) corresponds to all other cases.

From Theorem 2 we also have that O is the greatest optimal approximation, i.e. $\mathsf{O} \in \mathbf{Opt}(X)$, and for all $\mathsf{O}' \in \mathbf{Opt}(X)$, $\mathsf{O}' \subseteq \mathsf{O}$. We also know that $sim_{MS}(X, \mathsf{O}') = sim_{MS}(X, \mathsf{O})$

This *greedy* algorithm (because of the choice of $\rho(\mathbf{x}, X)$, cf. [12]) has a complexity of $C_1 + C_2 + \mathrm{O}(r \log r)$, where C_1 is the complexity of constructing $\underline{\mathbf{A}}(X)$, $\overline{\mathbf{A}}(X)$, and $\mathfrak{B}(X)$; while C_2 is the complexity to assign $\mu(x)$ for each $x \in U$. Algorithms with $C = \mathrm{O}(|U|^2)$ can be found for example in [20], and clearly $C_2 = \mathrm{O}(|U|)$.

The most crucial line of the algorithm, line (6), runs in $\mathrm{O}(r)$, but line (3) involves sorting which has complexity $\mathrm{O}(r \log r)$. Since $r < |U|$, the total complexity is $\mathrm{O}(|U|^2)$.

Algorithm 1 gives us the greatest optimal approximation O, however the whole set $\mathbf{Opt}(X)$ can easily be derived from O just by subtracting appropriate elements of $\mathfrak{B}(X)$.

Note that because of Corollary 4(1), Algorithm 1 is also effective for *any* similarity measure *sim* that is *consistent* with the Marczewski–Steinhaus μ-index sim_{MS}, for any finite and null-free measure μ. In particular, by Proposition 3 and Corollary 5 we can use it for Symmetric Tversky μ-index $sim_{\mu T}$, Dice–Sørensen μ-index $sim_{\mu DS}$; and of course the classical Jaccard index sim_J.

Table 1
Pawlak's space of houses and their prices. The column 'Quality index' is used only in Example 3.

House	Price ($)	Equiv. class	Qual. index μ
h_1	289,000	e_1	502
h_2	389,000	e_5	869
h_3	319,000	e_2	611
h_4	333,000	e_3	723
h_5	388,000	e_5	937
h_6	284,000	e_1	399
h_7	339,000	e_3	585
h_8	336,000	e_3	650
h_9	345,000	e_4	834
h_{10}	311,000	e_2	366
h_{11}	319,000	e_2	512
h_{12}	312,000	e_2	622

Class	Elements	Range ($)
e_1	h_1, h_6	280–299,999
e_2	h_3, h_{10}, h_{11}, h_{12}	300–319,999
e_3	h_4, h_7, h_8	320–339,999
e_4	h_9	340–359,999
		360–379,999
e_5	h_2, h_5	380–400,000

Algorithm 1 requires the measure μ to be finite and null-free. The assumption of finiteness of μ is essential (cf. [13]), but null-freeness is merely technical. If μ is not null-free, we can use the algorithm presented below.

Algorithm 2 *(μ is not null-free). Let $\mathcal{AS} = (U, E)$ be a Pawlak approximation space, $\mu : U \to \mathbb{R}$ be a given measure that is finite but not null-free, and $X \subseteq U$.*

1. *Define $U' = U \setminus \{x \mid \mu(x) = 0\}$, $E' = E \cap (U' \times U')$, $X' = X \setminus \{x \mid \mu(x) = 0\}$ and $\mathcal{AS}' = (U', E')$.*
2. *Apply Algorithm 1 for X' and $\mathcal{AS}'$. Let O' be the outcome of this application.*
3. *Pick any $\mathsf{O} \in \mathbf{IA}(X)$ such that $\mathsf{O}' \subseteq \mathsf{O}$.*

Since $\mathsf{O}' \in \mathbf{Opt}(X')$ then $\mathsf{O} \in \mathbf{Opt}(X)$. Moreover $\mu(X \setminus X') = \mu(\mathsf{O} \setminus \mathsf{O}') = 0$. □

7. Examples

Our first example will use Jaccard index, i.e. a special case of Marczewski–Steinhaus index with $\mu(X) = |X|$.

Example 2. We define our universe of elements labeled $U = \{h_1, \ldots, h_{12}\}$ to be an assortment of houses, each with a price or value associated with it, as shown in Table 1. Based on its price, each house belongs to a representative equivalence class as demonstrated in the second table. Our classes will be defined by each range of $20,000, starting from $280,000 and ending with $400,000 (empty classes are excluded because $\emptyset \notin \mathfrak{Comp}$). We could say that all of the houses in each class are roughly equivalent in price.

Suppose we wish to select a subset which we are interested in. If houses $\mathsf{H} = \{h_1, h_3, h_8, h_9\}$ meet our requirements we could say that we have the financing available for each of the equivalence classes those houses belong to. Clearly $\underline{\mathbf{A}}(\mathsf{H}) = e_4$ and $\overline{\mathbf{A}}(\mathsf{H}) = e_1 \cup e_2 \cup e_3 \cup e_4$. Moreover, $\mathfrak{B}(\mathsf{H}) = \mathrm{comp}(\overline{\mathbf{A}}(\mathsf{H})) \setminus \mathrm{comp}(\underline{\mathbf{A}}(\mathsf{H})) = \{e_1, e_2, e_3\}$, and $\mathbf{IA}(\mathsf{H}) = \{\underline{\mathbf{A}}(\mathsf{H}), A_1, A_2, A_3, A_4, A_5, A_6, \overline{\mathbf{A}}(\mathsf{H})\}$ where $A_1 = e_1 \cup e_4$, $A_2 = e_2 \cup e_4$, $A_3 = e_3 \cup e_4$, $A_4 = e_1 \cup e_2 \cup e_4$, $A_5 = e_1 \cup e_3 \cup e_4$, and $A_6 = e_2 \cup e_3 \cup e_4$. We also have $sim_J(\mathsf{H}, \underline{\mathbf{A}}(\mathsf{H})) = \frac{|\mathsf{H} \cap \underline{\mathbf{A}}(\mathsf{H})|}{|\mathsf{H} \cup \underline{\mathbf{A}}(\mathsf{H})|} = \frac{1}{4}$, $sim_J(\mathsf{H}, \overline{\mathbf{A}}(\mathsf{H})) = \frac{|\mathsf{H} \cap \overline{\mathbf{A}}(\mathsf{H})|}{|\mathsf{H} \cup \overline{\mathbf{A}}(\mathsf{H})|} = \frac{2}{5}$, and $sim_J(\mathsf{H}, A_1) = \frac{2}{5}$, $sim_J(\mathsf{H}, A_2) = \frac{2}{7}$, $sim_J(\mathsf{H}, A_3) = \frac{1}{3}$, $sim_J(\mathsf{H}, A_4) = \frac{3}{8}$, $sim_J(\mathsf{H}, A_5) = \frac{3}{7}$, $sim_J(\mathsf{H}, A_6) = \frac{2}{7}$. From all these Jaccard index values, $\frac{3}{7}$ is the biggest number, so $\mathbf{Opt}(\mathsf{H}) = \{A_5\} = \{e_1 \cup e_3 \cup e_4\}$.

What about our algorithm? We have $\mathfrak{B}(\mathsf{H}) = \{e_1, e_2, e_3\}$, and $\rho(e_1, \mathsf{H}) = 1$, $\rho(e_2, \mathsf{H}) = \frac{1}{3}$, and $\rho(e_3, \mathsf{H}) = \frac{1}{2}$. Hence $\rho(e_1, \mathsf{H}) > \rho(e_3, \mathsf{H}) > \rho(e_2, \mathsf{H})$, so we rename the elements of $\mathfrak{B}(\mathsf{H})$ as $e_1 = \mathbf{x}_1$, $e_3 = \mathbf{x}_2$, $e_2 = \mathbf{x}_3$. Clearly $\rho(\mathbf{x}_1, \mathsf{H}) = 1 > sim_J(\mathsf{H}, \underline{\mathbf{A}}(\mathsf{H})) = \frac{1}{4}$ and $\rho(\mathbf{x}_3, \mathsf{H}) = \frac{1}{3} < sim_J(\mathsf{H}, \overline{\mathbf{A}}(\mathsf{H})) = \frac{2}{5}$, so neither step (4) nor (5) hold, so we go to the step (6), which is the most involved.

We begin by setting $\mathsf{O}_0 = \underline{\mathbf{A}}(\mathsf{H}) = e_4$. Since $sim_J(\mathsf{H}, \mathsf{O}_0) = \frac{1}{4} < sim_J(\mathsf{H}, \mathsf{O}_0 \cup \mathbf{x}_1) = \frac{2}{5}$, we have $\mathsf{O}_1 = \mathsf{O}_0 \cup \mathbf{x}_1 = e_1 \cup e_4$, and since $sim_J(\mathsf{H}, \mathsf{O}_1) = \frac{2}{5} < sim_J(\mathsf{H}, \mathsf{O}_1 \cup \mathbf{x}_2) = \frac{3}{7}$, we have $\mathsf{O}_2 = \mathsf{O}_1 \cup \mathbf{x}_2 = e_1 \cup e_3 \cup e_4$. However $sim_J(\mathsf{H}, \mathsf{O}_2) = \frac{3}{7} < \rho(\mathbf{x}_2, \mathsf{H}) = \frac{1}{2}$, so $\mathsf{O}_1 \notin \mathbf{Opt}(\mathsf{H})$. Since $sim_J(\mathsf{H}, \mathsf{O}_2) = \frac{3}{7} > sim_J(\mathsf{H}, \mathsf{O}_2 \cup \mathbf{x}_3) = \frac{2}{5}$, we set $\mathsf{O}_3 = \mathsf{O}_2$. Now we have $sim_J(\mathsf{H}, \mathsf{O}_3) = sim_J(\mathsf{H}, \mathsf{O}_2) = \frac{3}{7} > \rho(\mathbf{x}_3, \mathsf{H}) = \frac{1}{3}$, which means that $\mathsf{O}_2 = \{h_1, h_4, h_6, h_7, h_8, h_9\} \in \mathbf{Opt}(\mathsf{H})$.

Note also that $\mathsf{O}_1 = A_1$, and $\mathsf{O}_2 = A_5$, and $\mathbf{Opt}(\mathsf{H}) = \{O_2\}$. □

Our second example uses Marczewski–Steinhaus μ-index which is not cardinality.

Example 3. Consider the same universe $U = \{h_1, \ldots, h_{12}\}$, the same equivalence classes $\{e_1, e_2, e_3, e_4\}$, and the same set $\mathsf{H} = \{h_1, h_3, h_8, h_9\}$ as in the previous example. Realizing that price is only one of the factors (even though often the most important), a real estate agency 'Best Choice' introduced a service for customers where they will determine a quality index μ ranging from 0 to 1000, which takes into account price, age, type of house, look, and the special customer preferences.

Suppose that the index values for a particular customer are described in the right column of the left part of Table 1. The index μ is extended to sets of houses X so we can use it to calculate the intermediate similarity values. It is defined as $\mu(X) = \sum_{h\in X} \mu(h)$. Clearly the index μ is an element-wise null free measure as discussed in Section 3, so it can be used in formulas describing similarity indexes.

What is an optimal approximation of H with Marczewski–Steinhaus index $sim_{MS}(X,Y) = \frac{\mu(X\cap Y)}{\mu(X\cup Y)}$? To measure the similarity between H and its lower approximation, we have $sim_{MS}(\mathsf{H}, \underline{\mathbf{A}}\mathsf{H}) = \frac{\mu(\mathsf{H}\cap\underline{\mathbf{A}}(\mathsf{H}))}{\mu(\mathsf{H}\cup\underline{\mathbf{A}}(\mathsf{H}))} = \frac{\mu(h_9)}{\mu(\{h_1,h_3,h_8,h_9\})} = \frac{\mu(h_9)}{\mu(h_1)+\mu(h_3)+\mu(h_8)+\mu(h_9)} = \frac{834}{2897} = 0.28788$. The rest of the similarity values calculated in the same manner are as follows: $sim(\mathsf{H}, A_1) = 0.49636$, $sim(\mathsf{H}, A_2) = 0.32863$, $sim(\mathsf{H}, A_3) = 0.33689$, $sim(\mathsf{H}, A_4) = 0.468515$, $sim(\mathsf{H}, A_5) = 0.47585$, $sim(\mathsf{H}, A_6) = 0.35478$, and $sim(\mathsf{H}, \overline{\mathbf{A}}(\mathsf{H})) = 0.4595$. By inspection, we see the largest value is a result of comparing H to $A_1 = e_1 \cup e_4$, which is clearly different from our previous example where A_5 returned the largest value.

What about our algorithm? We have $\mathfrak{B}(\mathsf{H}) = \{e_1, e_2, e_3\}$, and now with different measure μ we calculate ρ for each element e in the border as

$$\rho(e, H) = \frac{\mu(e \cap \mathsf{H})}{\mu(e \setminus \mathsf{H})} = \frac{\Sigma_{h\in e\cap\mathsf{H}}\mu(h)}{\Sigma_{h\in e\setminus\mathsf{H}}\mu(h)}.$$

We get $\rho(e_1, \mathsf{H}) = \frac{\mu(h_1)}{\mu(h_6)} = 2.0100$, $\rho(e_2, \mathsf{H}) = \frac{\mu(h_3)}{\mu(h_{10})+\mu(h_{11})+\mu(h_{12})} = 0.4073$, and $\rho(e_3, \mathsf{H}) = \frac{\mu(h_4)}{\mu(h_7)+\mu(h_8)} = 0.5038$. Hence, $\rho(e_1, \mathsf{H}) > \rho(e_3, \mathsf{H}) > \rho(e_2, \mathsf{H})$, as in the previous example so we again rename the elements of $\mathfrak{B}(\mathsf{H})$ as $e_1 = \mathbf{x}_1$, $e_3 = \mathbf{x}_2$, $e_2 = \mathbf{x}_3$. Since $\rho(\mathbf{x}_1, \mathsf{H}) > sim_{MS}(\mathsf{H}, \underline{\mathbf{A}}(H))$ and $\rho(\mathbf{x}_3, \mathsf{H}) < sim_{MS}(\mathsf{H}, \overline{\mathbf{A}}(H))$, steps (4) and (5) are not satisfied, so we move to step (6).

We begin with $\mathsf{O}_0 = \underline{\mathbf{A}}(\mathsf{H}) = e_4$, and $\mathsf{O}_1 = \mathsf{O}_0 \cup \mathbf{x}_1 = e_1 \cup e_4$. Note $\rho(\mathbf{x}_1, \mathsf{H}) = 2.0100 > sim_{MS}(\mathsf{H}, \mathsf{O}_1) = 0.49636$. So we stop here.

If we continued, we would examine $\mathsf{O}_2 = \mathsf{O}_1 \cup \mathbf{x}_2 = e_1 \cup e_3 \cup e_4$ and find $sim_{MS}(\mathsf{H}, \mathsf{O}_2) = 0.47585$ and $\rho(\mathbf{x}_2, \mathsf{H}) = 0.5038$, so the outcome would be the same.

Hence for this measure μ, $\mathbf{Opt}(\mathsf{H}) = \{O_1\}$. □

8. The case of Braun-Blanquet index

At the end of Section 5 we have shown that the Braun-Blanquet index [19] is inconsistent with the Jaccard index, and we argued that it is also inconsistent with Marczewski–Steinhaus index for almost any μ if the universe U is sufficiently large.

We will show that in general, the equivalence of Lemma 1 does not hold and Algorithm 1 does not work for Braun-Blanquet index.

Consider the following two examples.

Example 4. Consider a universe $U = \{a_1, a_2, b_1, b_2, \ldots, b_9, c_1, c_2, \ldots, c_{11}\}$, three equivalence classes covering U, $\mathsf{A}_1 = \{a_1, a_2\}$, $\mathsf{A}_2 = \{b_1, b_2, \ldots, b_9\}$ and $\mathsf{A}_3 = \{c_1, c_2, \ldots, c_{11}\}$, and the set $X = \{a_1, a_2, b_1, b_2, c_1, c_2\}$. We have $\underline{\mathbf{A}}(X) = \mathsf{A}_1 = \{a_1, a_2\}$, $\overline{\mathbf{A}}(X) = \mathsf{A}_1 \cup \mathsf{A}_2 \cup \mathsf{A}_3 = U$. One may check by inspection that $\mathbf{Opt}_{sim_J}(X) = \underline{\mathbf{A}}(X) = \mathsf{A}_1$, while $\mathbf{Opt}_{sim_{BB}}(X) = \mathsf{A}_1 \cup \mathsf{A}_2$. When applying Algorithm 1 with sim_J replaced by sim_{BB} we will get A_1 as an optimal approximation. The reason is that $sim_{BB}(X, \mathsf{A}_1 \cup \mathsf{A}_2) = \frac{|X\cap(\mathsf{A}_1\cup\mathsf{A}_2)|}{\max(|X|,|\mathsf{A}_1\cup\mathsf{A}_2|)} = \frac{4}{11} = 0.364 > sim_{BB}(X, \mathsf{A}_1) = \frac{|X\cap\mathsf{A}_1|}{\max(|X|,|\mathsf{A}_1|)} = \frac{2}{6} = 0.333$, but $\frac{|\mathsf{A}_2\cap X|}{|\mathsf{A}_2\setminus X|} = \frac{2}{7} = 0.286 < sim_{BB}(X, \mathsf{A}_1) = 0.333$, so the equivalent of Lemma 1 is not satisfied. Hence the first step of a modified Algorithm 1 would be faulty. Note also that $sim_{BB}(X, \mathsf{A}_1 \cup \mathsf{A}_2) > sim_{BB}(X, \mathsf{A}_1)$ while $sim_J(X, \mathsf{A}_1 \cup \mathsf{A}_2) = \frac{4}{11} = 0.364 < sim_J(X, \mathsf{A}_1) = \frac{2}{5} = 0.4$, so this is also a case of inconsistency between Jaccard and Braun-Blanquet indexes. □

Example 5. Consider a universe $U = \{a_1, a_2, b_1, b_2, \ldots, b_6, c_1, c_2, \ldots, c_{30}\}$, three equivalence classes covering U, $\mathsf{A}_1 = \{a_1, a_2\}$, $\mathsf{A}_2 = \{b_1, b_2, \ldots, b_6\}$ and $\mathsf{A}_3 = \{c_1, c_2, \ldots, c_{30}\}$, and the set $X = \{a_1, a_2, b_1, c_1, c_2, \ldots, c_5\}$. We have $\underline{\mathbf{A}}(X) = \mathsf{A}_1 = \{a_1, a_2\}$, $\overline{\mathbf{A}}(X) = \mathsf{A}_1 \cup \mathsf{A}_2 \cup \mathsf{A}_3 = U$. One may check by inspection that $\mathbf{Opt}_{sim_J}(X) = \underline{\mathbf{A}}(X) = \mathsf{A}_1$, while $\mathbf{Opt}_{sim_{BB}}(X) = \mathsf{A}_1 \cup \mathsf{A}_2$. When applying Algorithm 1 with sim_J replaced by sim_{BB} we will get A_1 as an optimal approximation. The reason is that $sim_{BB}(X, \mathsf{A}_1 \cup \mathsf{A}_2) = \frac{|X\cap(\mathsf{A}_1\cup\mathsf{A}_2)|}{\max(|X|,|\mathsf{A}_1\cup\mathsf{A}_2|)} = \frac{3}{9} = 0.333 > sim_{BB}(X, \mathsf{A}_1) = \frac{|X\cap\mathsf{A}_1|}{\max(|X|,|\mathsf{A}_1|)} = \frac{2}{9} = 0.222$, but $\frac{|\mathsf{A}_2\cap X|}{|\mathsf{A}_2\setminus X|} = \frac{1}{5} = 0.2 < sim_{BB}(X, \mathsf{A}_1) = 0.222$, so the equivalent of Lemma 1 is not satisfied either. Hence the first step of a modified Algorithm 1 would be faulty in this case as well. Note also that $sim_{BB}(X, \mathsf{A}_1 \cup \mathsf{A}_2) > sim_{BB}(X, \mathsf{A}_1)$ while $sim_J(X, \mathsf{A}_1 \cup \mathsf{A}_2) = \frac{3}{14} = 0.214 < sim_J(X, \mathsf{A}_1) = \frac{2}{9} = 0.222$, so this is another example of inconsistency between Jaccard and Braun-Blanquet indexes. □

Both examples can easily be adapted to other finite and null-free measures if U is sufficiently large. Hence, for Braun-Blanquet index we need a different algorithm, or a ratio different than $\rho(X, Y)$.

9. Final comments

In the above we have proposed a novel approach to rough set approximation. In addition to lower and upper approximations, we introduced and analyzed the concept of *optimal* approximation, which required the concept of a similarity index,

a measure space on the universe of elements, and a notion of border and border sets. We provided five simple similarity measure axioms (one axiom in two versions), and then analyzed several similarity indexes with respect to these axioms.

Only the Marczewski–Steinhaus index [13] (and popular Jaccard index [8], which is its special case) however, can naturally be interpreted as a measure of distance as well, so with this in mind, we used the index to design an algorithm which accepts a non-empty universe of elements (with an equivalence relation) and a subset $X \subseteq U$, and returns the optimal approximation. The algorithm is based on the properties of the index $\rho(X, Y)$, which is called the ratio of common to distinct elements. The algorithm runs in $O(r \log r)$ time where r is the number of elements in the 'border set', and thus has total time complexity $O(|U|^2)$. While the basic version of the algorithm requires that the measure is null free, i.e. $\mu(X) = 0 \iff X = \emptyset$, it was lated adapted for measures that are not null free.

We also introduced the concept of consistent similarity measures. Since consistent similarity indexes have identical optimal approximations, many results obtained for one index can be applied to all consistent indexes. In particular our algorithms can be applied, with very minor changes, to the Dice–Sørensen index and the Symmetric Tversky μ-index for any parameter ρ. The latter result is especially important as the Symmetric Tversky μ-index covers many more specialized similarity indexes [4].

However if a similarity index is not consistent with the Marczewski–Steinhaus index, our algorithm may not work, as we illustrate it for a fairly popular Braun-Blanquet index. At this point we do not know how to fix this problem.

One property of the optimal approximation definition that has not been sufficiently explored in this paper is its asymmetry, the second argument of *sim* in Definition 3 is always a definable set, i.e. a subset of $\mathbb{D}$. This asymmetry results from the Rough Sets setting, it is not the property of a similarity index per se, but its exploitation may lead to new results and better solutions.

Another intriguing problem, suggested by one of the reviewers and related to the above one, is the relationship between the measure μ and the equivalence relation E. The relation E represents the knowledge of an approximation space (U, E) and defines the set $\mathbb{D}$. Definition 3 indicates that an optimal simulation is a function of both the measure μ and the equivalence relation E, but how μ and E relate is an open problem.

Acknowledgements

We would like to thank the anonymous referees for their comments, error corrections, and useful suggestions. This research was partially supported by NSERC Grant of Canada.

References

[1] J. Braun-Blanquet, Pflanzensoziologie, Springer, Berlin, 1928.
[2] O. Bretcher, Linear Algebra with Applications, Prentice Hall, Englewood Cliffs, 1995.
[3] V. Bryant, Metric Spaces: Iteration and Application, Cambridge University Press, Cambridge, 1985.
[4] M.M. Deza, E. Deza, Encyclopedia of Distances, Springer, Berlin, 2012.
[5] L.R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (3) (1945) 297–302.
[6] I. Düntsch, G. Gediga, A. Lenarčič, Affordance Relations, Lecture Notes in Artificial Intelligence, vol. 5958, Springer, 2009, pp. 1–11.
[7] P. Halmos, Measure Theory, Van Nostrand, New York, 1950.
[8] P. Jaccard, Étude comparative de la distribution florale dans une portion des Alpes et des Jura, Bull. Soc. Vaud. Sci. Nat. 37 (1901) 547–549.
[9] R. Janicki, Approximations of arbitrary binary relations by partial orders. Classical and rough set models, Trans. Rough Sets 13 (2011) 17–38.
[10] R. Janicki, Property-driven rough sets approximations of relations, in [22], pp. 333–357.
[11] R. Janicki, A. Lenarčič, Optimal approximations with Rough Sets, in: Lecture Notes in Artificial Intelligence, vol. 8171, Springer, 2013, pp. 87–98.
[12] J. Kleinberg, E. Tardos, Algorithm Design, Addison–Wesley, 2005.
[13] E. Marczewski, H. Steinhaus, On a certain distance of sets and corresponding distance of functions, Colloq. Math. 4 (1958) 319–327.
[14] E. Marczewski, H. Steinhaus, O odległości systematycznej biotopów, Zastos. Mat. 4 (1959) 195–203.
[15] M.E. Munroe, Introduction to Measure and Integration, Addison–Wesley, 1953.
[16] Z. Pawlak, Rough Sets, Int. J. Comput. Inf. Sci. 34 (1982) 557–590.
[17] Z. Pawlak, Rough Sets, Kluwer, Dordrecht, 1991.
[18] A.C. Pielou, The Interpretation of Ecological Data, J. Wiley, New York, 1984.
[19] H. Rezai, M. Emoto, M. Mukaidono, New similarity measure between two fuzzy sets, J. Adv. Comput. Intell. Intell. Inform. 10 (6) (2006) 946–953.
[20] J. Saquer, J.S. Deogun, Concept approximations based on Rough sets and similarity measures, Int. J. Appl. Math. Comput. Sci. 11 (3) (2001) 655–674.
[21] A. Skowron, J. Stepaniuk, Tolerance approximation spaces, Fundam. Inform. 27 (1996) 245–253.
[22] A. Skowron, Z. Suraj (Eds.), Rough Sets and Intelligent Systems, Intelligent Systems Reference Library, vol. 42, Springer, 2013.
[23] R. Słowiński, D. Vanderpooten, A generalized definition of rough approximations based on similarity, IEEE Trans. Knowl. Data Eng. 12 (2) (2000) 331–336.
[24] T. Sørensen, A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analysis of the vegetation on Danish commons, Biol. Skr. 5 (4) (1957) 1–34.
[25] A. Tversky, Features of similarity, Psychol. Rev. 84 (4) (1977) 327–352.
[26] Y.Y. Yao, T. Wang, On rough relations: an alternative formulations, in: Lecture Notes in Artificial Intelligence, vol. 1711, Springer, 1999, pp. 82–91.