Ordinal Clustering Algorithms

Melvin F. Janowitz
DIMACS, Rutgers University,
Piscataway, New Jersey 08854

This version dated July 9, 2011. This is work in progress.

Abstract

This paper will demonstrate the natural role played by weak orders in the development of clustering algorithms for data having only ordinal significance. Within this weak order setting, any discussion of continuity becomes superfluous. The results are based on a talk given at the Classification Society meeting June 16-18, 2011.

1 Background terminology

A working knowledge of cluster analysis is assumed, but the basic terminology and notation will be quickly presented here. In its basic form, the input to a cluster analysis problem might be a pair (P, A), where P is a finite nonempty set, and A is a collection of attributes (numerical, logical, ordinal) that the objects might or might not have, or might have in varying degrees. The goal is to discover any hidden structures that the attributes might impose on P. Typically this is expressed as a partition of P, or a nested sequence of partitions with the top one having only a single class. It is possible to proceed directly from attributes to the output partitions, but often there is an intermediate step: the construction of a dissimilarity coefficient (DC). We let $D(P)$ denotes the set of DCs on P. We shall not discuss the actual construction of a DC, but will content ourselves with defining such an object. Further background in cluster analysis may be obtained from sources like [2]. The symbol $\Re_0^+$ will denote the non-negative reals equipped with the usual ordering.

Definition 1.1 A *dissimilarity coefficient* is a mapping $d : P \times P \to \Re_0^+$ satisfying

- $d(x, x) = 0$ for all $x \in P$
- $d(x, y) = d(y, x) \geq 0$ for all $x, y \in P$.
- d is called a *definite* DC if $d(a, b) = 0$ implies $a = b$.

The DC d is called an *ultrametric* if also

- $d(x, y) \leq \max\{d(x, z), d(y, z)\}$ for all $x, y, z \in P$.

Notation Let $U(P)$ denote the set of ultrametrics on P, and partially order both $D(P)$ and $U(P)$ by the rule $d_1 \leq d_2$ if and only if $d_1(x, y) \leq d_2(x, y)$ for all $x, y \in P$. Note that $U(P)$ is closed under the formation of arbitrary existing joins.

Notation Let $\Sigma(P)$ denote the set of reflexive symmetric relations on P, partially ordered by $R_1 \subseteq R_2$ if and only if $aR_1b \Longrightarrow aR_2b$, and $E(P)$ the set of equivalence relations on P with the same ordering.

Definition 1.2 A *numerically stratified clustering* (NSC) is a mapping $C : P \times P \to \Sigma(P)$ that satisfies

- $h \leq k \Longrightarrow C(h) \subseteq C(k)$.
- $C(h) = P \times P$ for some $h \in \Re_0^+$.
- For each $h \in \Re_0^+$, there exists a $k > h$ such that $C(h) = C(k)$.

Corresponding to any DC d, there is an NSC Td defined by

$$Td(h) = \{(a, b) : d(a, b) \leq h\}.$$

We then have (See [6], Theorem 2.9, page 20):

The fundamental T-correspondence. The correspondence $d \mapsto Td$ is a one-one order inverting correspondence between DCs and NSCs whose inverse is given by $C \mapsto d_C$, where $d_C(a, b) = \min\{h \in \Re_0^+ : (a, b) \in C(h)\}$.

Definition 1.3 The NSC C is called a *dendrogram* if $C(h)$ is an equivalence relation for all h.

Restating Theorem 2.18 of [6], we have

Theorem 1.4 The DC d is an ultrametric if and only if Td is a dendrogram.

We may now formally define a *cluster method* to be a mapping $F : D(P) \to D(P)$ or when appropriate from $D(P)$ to $U(P)$. We shall concentrate on studying DCs d having only ordinal significance in that the actual values taken by d are not significant, only whether $d(a, b) < d(x, y)$. In such situations, it makes no sense to form things like a mean or a standard deviation or worry about things like ratios, but there remain many possible and useful cluster methods.

2 Single-Linkage clustering

Using the Fundamental T-correspondence, we may define a cluster method F by

$$[T(Fd)](h) = \gamma([Td])(h) \text{ for all } h \tag{1}$$

where γ denotes the transitive closure operator. This is *single-linkage clustering* (denoted SL).

With this definition of cluster method, SL is the unique method that is idempotent and isotone, and whose image is the set of all ultrametrics on P. For a proof of this, see Theorem 2.27 of [6], and the discussion immediately preceding it.

- *idempotent*: $F = F \circ F$.
- *isotone*: $d_1 \leq d_2 \Longrightarrow F(d_1) \leq F(d_2)$.

This result shows that SL follows entirely from ordinal considerations. Yet SL became rather famous in [7] for the fact that it is *continuous*, and it is the only cluster method that produces an ultrametric and satisfies the conditions outlined in Chapter 9 of [7] (See page 91 of [7]). But this leaves us with a dilemma. Continuity involves a metric, and is concerned with the actual values of any input DC. But if the input DC only has ordinal significance, we are assuming that the values of the input DC are not significant, only whether one of them is smaller or larger than a second one. This led to an animated discussion in the cluster analysis literature shortly after the publication of [7]. We shall not rehash this now, but will find it useful to briefly discuss the nature of continuous cluster methods. We begin this by recalling the metric used by Jardine and Sibson. The distance between the DCs d and d' is given by

$$\Delta_0(d, d') = \max\{|d(a,b) - d'(a,b)| : a, b \in P\}. \tag{2}$$

Definition 2.1 Using a sequence definition of limits, we say that the cluster method F is continuous if $d_n \to d$ forces $F(d_n) \to F(d)$. We say that F is *right continuous* if $d_n \to d$ with $d \leq d_n$ forces $F(d_n) \to F(d)$. Left continuity has the expected definition.

Definition 2.2 A mapping θ on $\Re_0^+$ is said to be an *order automorphism* if it is one-one and onto, and $h \leq k \iff \theta(h) \leq \theta(k)$. The cluster method F is said to be *monotone equivariant* if $F(\theta d) = \theta F(d)$ for every order automorphism θ of $\Re_0^+$. For dealing with ordinal data, one of the primary assumptions of [7] is that one should use monotone equivariant (ME) cluster methods.

For ME cluster methods, the next theorem characterizes continuity. See the discussion preceeding Theorem 4.13 of [6] for a proof. Left continuity is characterized in Theorem 4.11 of [6].

Theorem 2.3 For ME cluster methods F, the following are equivalent:

1. F is flat in the sense that there is a mapping κ on relations such that $TF(d)) = \kappa \circ Td$ for all $d \in D(P)$.
2. F is continuous
3. F is right continuous
4. $F(\theta d) = \theta F(d)$ for all 0-preserving isotone mappings θ on $\Re_0^+$.

It follows that continuity is not that closely tied to single-linkage clustering, and follows from ordinal considerations. Though this is of interest, it still does not explain the connection between d and d' being close with respect to the Δ_0-metric, and various ordinal conditions. We turn now to such a consideration. Unfortunately, the basic ideas have not come to the author in any proper logical sequence. The key idea has come rather recently, but is basic to an earlier presentation. The author can only apologize to the reader for this.

Shortly after the book [7] was published, Robin Sibson published a paper [9] in which he described and advocated another more general method of dealing with data having only ordinal significance. We now introduce this idea.

Definition 2.4 The DCs d and d' are called *globally ordinal equivalent* if there is an order automorphism θ of $\Re_0^+$ such that $d = \theta(d')$. This defines an equivalence relation on $D(P)$, and is denoted $d \sim d'$. We say that a cluster method F preserves order equivalence if $d \sim d' \Longrightarrow F(d) \sim F(d')$. Two cluster methods F, G are called order similar and denoted $F \sim G$ if for each $d \in D(P)$, $F(d) \sim G(d)$. Note that no definite DC can be order equivalent to a DC that is not definite.

Theorem 2.5 Let F be a cluster method.

1. If F is monotone equivariant then the image of Fd is contained in the image of d for any DC d.

2. Let F preserve order equivalence. Then F is order similar to a monotone equivariant cluster method G if and only if F *compresses information* in the sense that for any $d \in D(P)$, $F(d)$ cannot have more levels in its image than does d.

Proof: Lemma 3.4 lf [6] and Theorem 4 of [3]. ∎

Remark 2.6 It is worth noting that every monotone equivariant cluster method preserves order equivalence, but the two concepts are not the same.

3 Metric and Ordinal Considerations

Let $d \in D(P)$, the DCs on P. A relation R is called a *threshold relation* for d if $R = Td(h)$ for some $h \in \Re_0^+$, and $R \subset P \times P$. Let

$$\text{image } (d) = h_0 = 0 < h_1 < \cdots h_t < h_{t+1}$$

with threshold relations $R_0 \subset R_1 \subset \cdots R_t \subset P \times P$. Then with

$$\sigma(d) = \{R_0, R_1, \ldots R_t\},$$

$\sigma(d)$ together with the image of d completely specifies d. Recall that $d \sim d'$ if $\sigma(d) = \sigma(d')$.

Note: The key to understanding ordinal properties of convergence is provided by the notion of $\precsim$. The idea is that $d \precsim d'$ ($d' \succsim d$) if $d(a,b) < d(x,y)$ implies $d'(a,b) < d'(x,y)$. Taking contrapositives, we have that

$$d \precsim d' \iff d'(a,b) \leq d'(x,y) \text{ implies } d(a,b) \leq d(x,y). \tag{3}$$

Definition 3.1 Let $d \in D(P)$ have image $h_0 = 0 < h_1 < \cdots < h_t < h_{t+1}$. The *mesh width* of d is defined by

$$\mu(d) = \frac{1}{2} \min\{h_i - h_{i-1} : 1 \leq t+1\} \tag{4}$$

Fact: Let $d \precsim d'$. Then $d'(a,b) = 0 \Longrightarrow d(a,b) = 0$.

Proof: Suppose $d(a,b) > 0$. Then $d(x,x) < d(a,b)$ would force $d'(x,x) < d'(a,b)$, a contradiction.

Fact: We note that $d \sim d'$ if both $d \precsim d'$ and $d' \precsim d$. Thus $d \sim d'$ forces $Td(0) = Td'(0)$ and $d(a,b) < d(x,y) \iff d'(a,b) < d'(x,y)$. Thus d is order equivalent to d', and there is no conflict of notation with that of global order equivalence as introduced by Sibson.

Lemma 3.2 *If $d \precsim d'$, then $Td(0) = Td'(k)$ for some k with $0 \leq k < \mu(d)$.*

Proof: Let $k = \max\{h : d'(a,b) = h \text{ and } d(a,b) = 0\}$. We may choose a, b so that $d'(a,b) = k$ and $d(a,b) = 0$. Let $d'(x,y) \leq k$. If $d(x,y) > 0$, then $d(a,b) < d(x,y)$ forces $k = d'(a,b) < d'(x,y)$, a contradiction. Thus $Td(0) = k$, as claimed. This also shows that $k < \mu(d)$. ∎

Fact: If $d \precsim d'$, then $Td(0) = R_\emptyset \Longrightarrow Td'(0) = R_\emptyset$; also $d'(a,b) = 0 \Longrightarrow d(a,b) = 0$, so $Td'(0) \subseteq Td(0)$. Here $R_\emptyset = \{(x,x) : x \in P\}$.

Theorem 3.3 *Let $d, d' \in D(P)$. Then $d \precsim d'$ is equivalent to every threshold relation of d being a threshold relation of d'.*

Proof: Assume first $d \precsim d'$. We have already noted that $Td(0) = Td(k)$ with $0 \leq k < \mu(d)$. Thus we must show that corresponding to each $h \geq 0$ such that $Td(h) \neq R_\emptyset$, there is a $k \geq 0$ such that $Td(h) = Td'(k)$. If $R = Td(h)$, let $k = \max\{d'(s,t) : sRt\}$, and choose $a, b \in P$ so that aRb and $d'(a,b) = k$. Evidently $wRz \Longrightarrow d'(w,z) \leq k$, so $Td(h) \subseteq Td'(k)$. If $d'(x,y) \leq k$, then $d'(x,y) \leq d'(a,b)$, so $d(x,y) \leq d(a,b)$. Since $d(a,b) \leq h$, this shows that $Td'(k) \subseteq Td(h)$.

Now assume that each threshold relation of d is a threshold relation of d'. Let $R_0 \subset R_1 \subset \cdots \subset R_t$ be the threshold relations of d, and $R'_0 \subset R'_1 \subset \cdots \subset R'_u$ the threshold relations of d'. Let R_i occur at level h_i, $d(a,b) = h_i$, and $d(a,b) < d(x,y)$. We are to show that $d'(a,b) < d'(x,y)$. By hypothesis, $R_i = R'_{\alpha(i)}$, with $R'_{\alpha(i)}$ occurring at level $k_{\alpha(i)}$. Since $(a,b) \in R_i = R'_{\alpha(i)}$, it is clear that $d'(a,b) \leq k_{\alpha(i)}$. On the other hand, $d(x,y) > h_i$ implies that $(x,y) \notin R_i$, so $(x,y) \notin R'_{\alpha(i)}$ and $d'(x,y) > k_{\alpha(i)}$. Thus $d'(a,b) \leq k_{\alpha(i)} < d'(x,y)$, thus completing the proof. ∎

Fact: For any 0-preserving isotone mapping θ on $\Re_0^+$, $\theta d \precsim d$.
Proof: This just says that $\theta d(a,b) < \theta d(x,y) \Longrightarrow d(a,b) < d(x,y)$.

Theorem 3.4 Let $d, d' \in D(P)$ and $0 < \varepsilon < \mu(d)$. If $d \precsim d'$, there exists $d'' \in D(P)$ such that $d' \sim d''$ and $\Delta_0(d, d'') < \varepsilon$. Thus if $[[d']]$ is the equivalence class of d' under $\sim$, the $d \precsim d'$ iff there is a sequence d_n of $[[d']]$ with limit d.

Example 3.5 Idea of Proof: Let $P = \{a, b, c, e, f\}$, and suppose d, d' are given as follows:

d	a	b	c	e	f
a	0	5	0	9	5
b	5	0	5	0	9
c	0	5	0	5	9
e	9	0	5	0	5
f	5	9	9	5	0

d'	a	b	c	e	f
a	0	1	0.45	5	1
b	1	0	3	0.8	5
c	0.45	3	0	1	5
e	5	0.8	1	0	1
f	1	5	5	1	0

Note that $\mu(d) = 2$ since the image of d is $\{0, 5, 9\}$. Evidently, $\Delta_0(d, d') = 4$. Let's construct d'' so that $d' \sim d''$ and $\Delta_0(d, d'') = 0.5$. We note that

$\sigma(d) = [R_0 \subset R_1 \subset R_2;\ h_0 = 0 < h_1 = 5 < h_2 = 9]$

$\sigma(d') = [R'_0 \subset R'_1 \subset R'_2 \subset R'_3 \subset R'_4 \subset R'_5;\ 0 < 0.45 < 0.8 < 1 < 3 < 5]$

We see that $d \precsim d'$ since $R_0 = R'_2$, $R_1 = R'_4$, $R_2 = R'_5$. We present an example of a possible d'' and indicate how it was constructed. This corrects a minor typo that occurs on p. 88 if [6].

d''	a	b	c	e	f
a	0	4.5	0.45	9	4.5
b	4.5	0	5	0.5	9
c	0.45	5	0	4.5	9
e	9	0.5	4.5	0	4.5
f	4.5	9	9	4.5	0

The idea now is to note that $d' \sim d''$ since there is an order automorphism θ of $\Re_0^+$ such that $d'' = \theta(d')$. Note that θ sends $(0, 0.45, 0.8, 1, 3, 5)$ to $(0, 0.45, 0.5, 4.5, 5, 9)$. The reader can verify that $\Delta_0(d, d'') = 0.5$. We could clearly have adjusted the levels in d'' to make it as close as we wanted to d. We could now present a proof for the Theorem, but will not do so. Instead, we leave it as an exercise for the reader. ∎

Fact: If d_n has limit d, then there exists a positive integer N such that $n \geq N$ forces $d \precsim d_n$.

Theorem 2.3 characterizes continuity for monotone equivariant cluster methods. It seems natural to see what can be said for cluster methods that preserve order equivalence. Theorem 3.6 begins this process.

Theorem 3.6 Let F preserve order equivalence. Then F continuous forces F to preserve $\precsim$, but the converse fails.

Proof: We assume F is a continuous cluster method that preserves order equivalence, and will prove that F necessarily preserves $\precsim$. Let $d, g \in D(P)$ with $d \precsim g$. By Theorem 3.4, there must be a sequence g_n of DCs having the property that $g_n \to d$ and each $g_n \sim g$. By continuity, $F(g_n)$ has limit $F(d)$, and by hypothesis, each $F(g_n) \sim F(g)$. It follows that $F(d) \precsim F(g)$. The proof would be complete if we could produce an example of a cluster method that preserves $\precsim$ but is not continuous. Such an example is presented on p. 90 of [6]. ∎

Definition 3.7 It now seems reasonable to call a cluster method F *order continuous* if it is true that $d \precsim d' \Longrightarrow F(d) \precsim F(d')$.

The reader might now assume that we are about to embark on a study of properties of order continuous cluster methods. Not so! We choose instead to change course in such a way as to demonstrate that we need not even be considering issues that involve continuity when we are faced with ordinal data.

4 A connection with weak orders

Consider the action of a cluster method F that preserves ordinal equivalence. If $d \sim d'$, then $F(d) \sim F(d')$, so we may think of F as acting on an equivalence class of $\sim$. It turns out that these equivalence classes can be endowed with an interesting and natural partial order. If we let $[[d]]$ denote the equivalence class of $\sim$ generated by the DC d, we may define $[[d]] \leq [[d']] \iff d \precsim d'$. one has to show that this is well-defined and a partial order. An indication of what is involved occurs in [6], Remark 4.23, p. 90. The point is that if $d_1 \precsim d_2 \sim d_3 \precsim d_4 \sim d_1$, then d_1, d_2, d_3, d_4 all lie in the same equivalence class. We leave the details to the reader. The point is that we now have a partially ordered set. To establish some notation, let W_d denote the weak order associated with $[[d]]$. In other words, $(a,b)W_d(x,y)$ whenever $d(a,b) \leq d(x,y)$. There is also what is called a *strict* weak order W_d^s associated with $[[d]]$. It is defined by $(a,b)W_d^s(x,y)$ whenever $d(a,b) < d(x,y)$. Thus $d \precsim d'$ is equivalent to $W_{d'} \leq W_d$, which in turn is equivalent to $W_d^s \leq W_{d'}^s$. The fact that the $\precsim$ partial ordering of equivalence classes of order similarity is equivalent to the usual ordering of associated strict weak orders is what leads us to believe that this might be the correct way of dealing with dissimilarities defined on data having ordinal significance.

Remark 4.1 The set $U(P)$ of ultrametrics has some important properties in $D(P)$ when they are each given their usual partial order. By [6], Lemma 2.21, p. 23, $U(P)$ is closed under the formation of arbitrary existing joins, and every $d \in D(P)$ is the meet of a family of ultrametrics. This suggests that a study of properties of W_u for u an ultrametric might be worthwhile.

Remark 4.2 Weak orders on a finite set. The poset of weak orders on a finite set is investigated and characterized in [5]. References to earlier related work by other authors are given therein. A thorough treatment may be found in [8]. We mention that the weak orders on a finite set form what is called a semiBoolean algebra. These are defined and investigated in [1]. Essentially, we are dealing with a finite join semilattice in which every principal filter is a Boolean algebra. Let's see how this applies to the structure under consideration. We are given a finite nonempty set P. We will be looking at a structure of the form $2 \times L$, where $2 = \{0,1\}$ is a two element chain, and L is the poset of weak orders on the two element subsets of P. The idea is to map d into $(0, W_d)$ if d is not definite and into $(1, W_d)$ d is definite. Note that $d \precsim d+1$, where $[d+1](a,b) = d(a,b)+1$. For example, if d is not definite, and $\sigma(d) = \{R_0, R_1\}$, then $d+1$ is definite with $\sigma(d+1) = \{R_\emptyset, R_0, R_1\}$. The DCs d and $d+1$ are not order equivalent, $d \precsim d+1$, but not $d+1 \precsim d$. The model we have just described is just a theoretical model in which properties of clustering algorithms might be developed. What remains is the development of efficient scalable cluster algorithms within the model. It turns out that obvious modifications of some of the algorithms described some time ago in [4] can be used to start the process. The reader should note that issues involving continuity are no longer present — even though there is a fundamental connection with them.

Definition 4.3 Let's make sure we agree on terminology. A *weak order* on a set P is a binary relation W on P having the property that it is

- **reflexive** in that xWx for all $x \in P$.
- **transitive** in that xWy, yWz together imply that xWz for all $x, y, z \in P$.

- **complete** xWy or yWx for every pair $\{x, y\}$ of members of P.

A *strict* weak order also has the property that

- For any $x, y \in P$, xWy implies that yWx fails .

Weak orders are partially ordered by the rule $W_1 \leq W_2 \iff xW_1y \implies xW_2y$. Corresponding to any weak order W, there is a strict weak order W^s defined by $xW^sy \iff xWy$ but yWx fails. Note that $W_1 \leq W_2 \iff W_2^s \leq W_1^s$. The ordering induced by $\precsim$ corresponds to the usual ordering of the induced strict weak orders.

We need to at least indicate how cluster methods might be used with dissimilarities of the form d_W. If $d \in D(P)$ has only ordinal significance, we can easily obtain d_W by simply rank ordering the image of the restriction of d to the two element subsets of P. We also need to remember whether there is a pair $\{a, b\}$ with $a \neq b$ and $d(a, b) = 0$. It turns out that a number of possible algorithms were outlined in [4], and called type 1 cluster methods therein. These are all agglomerative cluster methods, though the connection with preservation of $\precsim$ was not made in this earlier paper. At the ith stage we have constructed a partition $\mathfrak{P}_i$ of P, and we need to merge classes of $\mathfrak{P}_i$ to form a partition $\mathfrak{P}_{i+1} \succ \mathfrak{P}_i$ by means of *links* from a reflexive symmetric relation L_i. An L_i-link between disjoint subsets A, B of P is a pair x, y with $x \in A$, $y \in B$ and xL_iy. The biggest change to be made involves the output list of partitions. Since the input is simply a weak order, there seems little point in keeping track of output levels. Here are some typical methods as outlined in [4].

1. k-clustering (k a positive integer): Merge clusters A, B if at least k or all possible L_i-links have been made between A and B. For k =1, this is single-linkage clustering, and for k sufficiently large it is a version of complete linkage.

2. u-clustering ($0 < u \leq 1$): Merge clusters A, B if at least a portion $u|A||B|$ of the links between A and B have been made. For u close to 0 this is single linkage, and for $u = 1$ a version of complete linkage. This is commonly called proportional-linkage clustering. It was apparently due originally to [10], but did not gain wide acceptance due to its tendency to produce what are commonly called reversals in the literature. In the present context, this does not appear to be a problem.

For the moment, we close by presenting an example that will hopefully illustrate the potential for u–clustering to play a role somewhat similar to that played traditionally by average linkage clustering. The data appears as part of the Clustan clustering. It consists of percentage contents of various mammal milk samples. Note the rather dramatic differences in scale for the attributes. We proceeded by normalizing the data so that each attribute had smallest value 0 and largest value 1. We then used squared Euclidean distance as the dissimilarity. It is hoped that this DC has ordinal significance, but it is difficult to assign an exact meaning to each numerical value. We present the 4 cluster cutoff for: A. average linkage, B. complete linkage, and u-clustering with C. $u = .5$ and D. $u = .6$.

	Object	Water	Protein	Fat	Lactose	Ash
1.	Bison	86.9	4.8	1.7	5.7	0.9
2.	Buffalo	82.1	5.9	7.9	4.7	0.78
3.	Camel	87.7	3.5	3.4	4.8	0.71
4.	Cat	81.6	10.1	6.3	4.4	0.75
5.	Deer	65.9	10.4	19.7	2.6	1.4
6.	Dog	76.3	9.3	9.5	3	1.2
7.	Dolphin	44.9	10.6	34.9	0.9	0.53
8.	Donkey	90.3	1.7	1.4	6.2	0.4
9.	Elephant	70.1	3.6	17.6	5.6	0.63
10.	Fox	81.6	6.6	5.9	4.9	0.93
11.	GuineaPig	81.9	7.4	7.2	2.7	0.85
12.	Hippo	90.4	0.6	4.5	4.4	0.1
13.	Horse	90.1	2.6	1	6.9	0.35
14.	Llama	86.5	3.9	3.2	5.6	0.8
15.	Monkey	88.4	2.2	2.7	6.4	0.18
16.	Mule	90	2	1.8	5.5	0.47
17.	Orangutan	88.5	1.4	3.5	6	0.24
18.	Pig	82.8	7.1	5.1	3.7	1.1
19.	Rabbit	71.3	12.3	13.1	1.9	2.3
20.	Rat	72.5	9.2	12.6	3.3	1.4
21.	Reindeer	64.8	10.7	20.3	2.5	1.4
22.	Seal	46.4	9.7	42	0	0.85
23.	Sheep	82	5.6	6.4	4.7	0.91
24.	Whale	64.8	11.1	21.2	1.6	1.7
25.	Zebra	86.2	3	4.8	5.3	0.7

Table 1: Mammal milk data table

Method	Cluster 1	Cluster 2	Cluster 3	Cluster 4
A	1 2 3 7 8 9 10 12 13 14 15 16 17 23 25	4 6 11 18 20	5 19 21 24	7 22
B	1 3 8 9 12 13 14 15 16 17 25	2 4 6 10 11 18 20 23	5 19 21 24	7 22
C	1 2 3 9 10 11 14 18 23 25	4 5 6 19 20 21 24	8 12 13 15 16 17	7 22
D	1 2 3 7 8 9 10 12 13 14 15 16 17 23 25	4 6 11 18 20	5 19 21 24	7 22

References

[1] Abbott, J. C. , *Sets, Lattices and Boolean Algebras*, Allyn and Bacon, Boston, 1976.

[2] Jain, A. K. and Dubes, R. C. *Algorithms for Clustering Data*, Prentice-Hall, Englewood Cliffs, 1988.

[3] Janowitz, M. F., *Preservation of global order equivalence*, Journal of Mathematical Psychology **20**, 1979, 78–88

[4] Janowitz, M. F., *Monotone equivariant cluster methods*, SIAM Journal on Applied Mathematics **37**, 1979, 148–165.

[5] Janowitz, M. F., *On the semilattice of weak orders of a set*, Mathematical Social Sciences **8**, 1984, 229–239.

[6] Janowitz, M. F., *Ordinal and relational clustering*, World Scientific, Singapore, 2010.

[7] Jardine, N. and Sibson, R., *Mathematical Taxonomy,* Wiley, New York, 1971.

[8] Roberts, F. S. *Measurement Theory, with Applications to Decisionmaking, Utility, and the Social Sciences, Encyclopedia of Mathematics and its Applications*, **7**, Addison-Wesley, 1979.

[9] Sibson, R., *Order invariant methods for data analysis*, Journal of the Royal Society Series B **34**, 1972. 311–349

[10] Sneath, P. H. A., *A comparison of different ckustering methods as applied to randomly placed points*, Classification Society Bulletin **1**, 1966, 2-18.