Pattern Recognition Pergamon Press 1975. Vol. 7, pp. 67–79. Printed in Great Britain

THE PATTERN DISCRIMINATION PROBLEM FROM THE PERSPECTIVE OF RELATION THEORY

R. M. HARALICK

Remote Sensing Laboratory, Department of Electrical Engineering, University of Kansas, Lawrence, Kansas, U.S.A.

(*Received* 21 *August* 1973 *and, in revised form* 26 *August* 1974)

Abstract—Pattern discrimination is viewed from the perspective of relation theory. Measurement data is a binary relation from the set of units to the set of measurements and the category identification data is a binary relation from the set of units to the set of categories. The decision rule is a binary relation from the set of measurements to the set of categories. First, no structure is assumed on the set of measurements or the set of units and the form of the optimal decision relation is determined for the case of unit independence. Then a binary relation dependence structure is assumed on the set of units and two approximating forms of decision relations are determined. The approximating decision relation for the unit dependence case has a distinctive form quite different from the decision rule which would result from the usual Markov dependence assumption. It is hoped that relation model for pattern discrimination can provide a useful complementary alternative to the statistical model currently in use.

Pattern discrimination/identification Binary relation Sets Decision rules Context Graphs

I. INTRODUCTION

We will pose some pattern recognition problems using the relation construct and then explore some alternative ways of examining these problems. Some of the deductions we make are illustrated in figures and the reader giving careful attention to them will understand the essence of the paper. Those desiring a more formal understanding will find that the lemmas which are mentioned in the body of the paper are formally stated and proved in the appendix.

The relation theoretic approach we take to pattern recognition is an interesting alternative to the more usual probablistic models. Readers who are interested in seeing another kind of set theoretic approach might examine the interval covering ideas of Michalski,[1] Michalski and McCormick,[2] and Read and Jayaramamurthy.[3]

The pattern discrimination problem we want to discuss usually occurs in a context in which there is an initial opportunity to be an omniscient observer discovering and formally recording some of the category identifications and pattern measurements made from some set of environmental objects.

The environmental objects can be plants, animals, insects, chemical compounds, printed characters, medical X-rays, photographs, images, words or sentences. The measurements can relate to shape, shape of the various parts, temperature, color, tone, texture or sequence dependencies.

We will call the objects *environmental units* (*units*, for short) and denote by the letter U the non-empty set of such units to be considered. We will call the measurements of the units *measurement patterns* (patterns, for short) and denote by the letter D the non-empty set of such patterns. Finally, we denote by the letter C the non-empty set of *category labels* with which the units may be identified.

The number of possible categories is usually orders of magnitude smaller than the number of possible measurement patterns or units but the number of units may be smaller or larger than the number of possible measurement patterns.

During the initial opportunity to be an omniscient observer, each unit in the unit training subset U_t of units is observed. We assume that in the measuring process (which may take only an instant or may take a long period of time) each unit measured gives rise to at least one measurement pattern and, for generality, possibly even more than one measurement pattern. We denote by Σ_t the relationships between the units measured and the measurement patterns. Σ_t is a binary relation from the subset of units U_t to the set of patterns D and hence a subset of $U_t \times D$; that is, $\Sigma_t \subseteq U_t \times D$. Σ_t is called the *measurement training relation* and consists of the set of all ordered pairs (u, d) where d is a measurement made of the unit u in the set U_t, (see Fig. 1).

At the same time that the units are measured, the investigator identifies the units by category. Each unit is assigned one category. We denote by T_t the relationships between the units observed and the category labels. T_t is a binary relation from the subset of units U_t to the set of categories C and hence a subset of $U_t \times C$; that is, $T_t \subseteq U_t \times C$. The relation T_t is called the *category identification training relation* and consists of the set of all ordered pairs (u, c) where c is the category identification made of the unit u

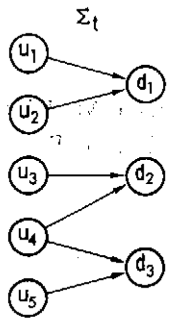


Fig. 1. Digraph of unit training relation Σ_t, where the unit training set is $U_t = \{u_1, u_2, u_3, u_4, u_5\}$, the measurement space is $D = \{d_1, d_2, d_3\}$, and the unit training relation is $\Sigma_t = \{(u_1, d_1), (u_2, d_1), (u_3, d_2), (u_4, d_2), (u_4, d_3), (u_5, d_3)\}$.

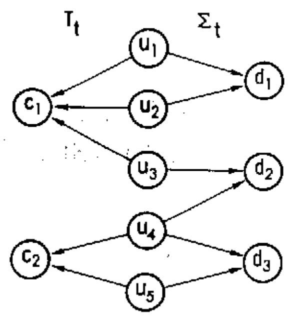


Fig. 3. Digraph of the training data consisting of the measurement training relation Σ_t and the true category identification training relation T_t of Figs. 1 and 2 respectively.

in the set U_t (see Fig. 2). The pair of relations Σ_t and T_t is often referred to as the *training data*. Figure 3 illustrates a digraph of an example pair of relations Σ_t and T_t.

After this initial opportunity of omniscient observation has passed, a new subset of units U_p becomes available for sensing or measuring. The measurement prediction relation $\Sigma_p \subseteq U_p \times D$, is then defined as the set of all ordered pairs (u, d) where d is a measurement made of the unit u in the set U_p. The category identification relation for the units in U_p we denote by T_p. The relations Σ_p and T_p are often referred to as the *prediction data*.

When no further information is available, the pattern discrimination problem becomes one of determining a rule to estimate or assign, in some optimal way, the proper category identifications to all units in U_p purely on the basis of the training data Σ_t and T_t and the measurement relation Σ_p.

When the problem is posed in this manner, the only basis for making category identifications for the units in U_p is by using the measurements made of them, that is Σ_p, since that is the only information we have about them. Hence, to solve the pattern discrimination problem, we need to find some relation Δ which pairs at least one category to each possible measurement subset of D. However, for simplicity, we shall assume that the relation Δ pairs at least

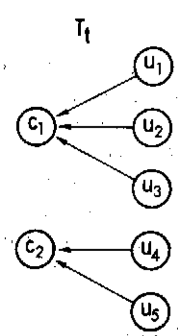


Fig. 2. Digraph of a true category identification training relation T_t, where the unit training set is $U_t = \{u_1, u_2, u_3, u_4, u_5\}$, the category set is $C = \{c_1, c_2\}$ and the true category identification training relation is $T_t = \{(u_1, c_1), (u_2, c_1), (u_3, c_1), (u_4, c_2), (u_5, c_2)\}$.

one category to each possible measurement in D. Note that implicit in this initial problem formulation is the feature of treating units independently i.e. assigning a category to one unit independent of a decision made about another unit.

The decision relation Δ must depend, of course, on the training data Σ_t and T_t. Once Δ is known, the *assigned category identifications* $\hat{T}_p$ of the units in U_p can be made in a straightforward way by the relation composition of Σ_p and Δ: $\hat{T}_p = \Sigma_p \circ \Delta$. $\hat{T}_p$ is a set having all the ordered pairs (u, c) where u is a unit of U_p, c is a category of C such that (1) there is some measurement d which is a measurement made of u, $(u, d) \in \Sigma_p$, and (2) d is paired by the decision relation with the category c, $(d, c) \in \Delta$.

2. THE ERROR-FREE CASE

Consider what happens when an error free solution to the pattern discrimination problem exists: in this case, the decision relation Δ allows all the units in the prediction set to be assigned to categories correctly and without error purely on the basis of the patterns measured from them. Hence, we must have that, for each unit u in the prediction set of units U_p, $(\Sigma_p \circ \Delta)(u) = T_p(u)$ since for any unit u, $(\Sigma_p \circ \Delta)(u)$ is the set of all categories relating to some measurment which has been taken of unit u in some instances and $T_p(u)$ is the set of all categories which have been the true category identification for unit u; the equality $(\Sigma_p \circ \Delta)(u) = T_p(u)$ then says that for each unit u, we may obtain its category identification $T_p(u)$, without error, from its pattern measurements $\Sigma_p(u)$ by linking its pattern measurements through the decision relation Δ, that is, $\Delta(\Sigma_p(u)) = (\Sigma_p \circ \Delta)(u) = T_p(u)$ for every unit u. Figure 4 illustrates an example set of relations Σ_p and T_p where an error free solution exists. Figure 5 illustrates an example set of relations Σ_p and T_p where an error free solution does not exist.

A measure indicating the identification accuracy of the decision relation Δ can be constructed by comparing $\hat{T}_p$, the decision relation assigned category

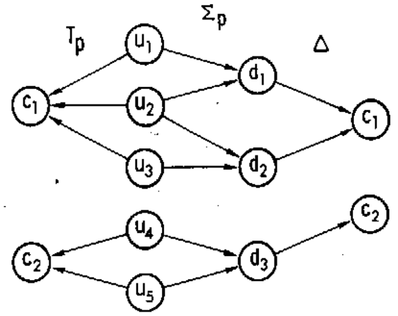


Fig. 4. An example set of relations Σ_p and T_p where an error free solution exists. The unit prediction set is $U_i = \{u_1, u_2, u_3, u_4, u_5\}$, the measurement space is $D = \{d_1, d_2, d_3\}$, the category set is $C = \{c_1, c_2\}$, the measurement prediction relation is $\Sigma_p = \{(u_1, d_1), (u_2, d_1), (u_2, d_2), (u_3, d_2), (u_4, d_3), (u_5, d_3)\}$, the true category identification prediction relation is $T_p = \{(u_1, c_1), (u_2, c_1), (u_3, c_1) (u_4, c_2), (u_5, c_2)\}$, the decision relation is $\Delta = \{(d_1, c_1), (d_2, c_1), (d_3, c_2)\}$, and the assigned category identification is the relation composition $\Sigma_p \bigcirc \Delta = \{(u_1, c_1), (u_2, c_1), (u_3, c_1), (u_4, c_2), (u_5, c_2)\}$.

identification with T_p, the true category identification of the prediction data.

There are three parts to the comparison of $\hat{T}_p = \Sigma_p \bigcirc \Delta$ with T_p:

$$\hat{T}_p \cap T_p \quad (1)$$

$$\hat{T}_p \cap T_p^c \quad (2)$$

$$\hat{T}_p^c \cap T_p. \quad (3)$$

Ordered pairs (u, c) which are in $\hat{T}_p \cap T_p$ are the unit-category pairs which are correctly linked through the decision relation. Ordered pairs (u, c) which are in $\hat{T}_p \cap T_p^c$ are those unit-category pairs which are linked through the decision relation but which are incorrect. Ordered pairs (u, c) which are in $\hat{T}_p^c \cap T_p$ are those unit-category pairs which should have been linked through the decision relation but which were not. Note that when $\Sigma_p \bigcirc \Delta = \hat{T}_p = T_p$, then $\hat{T}_p \cap T_p^c = \hat{T}_p^c \cap T_p = \phi$, the empty set, so that there are no errors. We shall take the best decision relation to be that one which minimizes the number of unit category pairs in the set $(\hat{T}_p \cap T_p^c) \cup (\hat{T}_p^c \cap T_p)$ or some quantity closely related to that number.

For conceptual simplicity, from this point on we will assume that the prediction measurement relation Σ_p consists of all the observable unit-measurement pairs; $\Sigma_p = \Sigma \subseteq U \times D$. (Note that the set of all observable unit-measurement pairs is most usually only a small subset of the set of all possible unit-measurement pairs $U \times D$). Similarly we assume that the true prediction category identification relation T_p consists of all the observed unit-category pairs, $T_p = T \subseteq U \times C$. (Note also that the set of all observed unit-category pairs is only a small subset of the set of all possible unit-category pairs $U \times C$.) It is natural to make the following assumptions about Σ and T: (1) at least one measurement is paired to each unit in U by Σ; that is, Σ is defined everywhere; (2) each measurement is paired with some unit; that is, Σ is onto the set of measurements; (3) at least one category is paired to each unit in U by T; that is, T is defined everywhere; (4) only one category is assigned to each unit in U by T; that is, T is single-valued; and (5) each category is paired with some unit in U by T; that is, T is onto the set of categories.

On the basis of the few assumptions made, it is possible to determine the decision relation when we know that some decision relation exists which will assign categories and make no errors. Furthermore, we shall be able to characterize and interpret the relationship between Σ and T which guarantees the existence of a no error decision relation.

The first observation we wish to make is that if there exists a decision relation with the property that the category assignment c given to any units having measurements d is correct, then it is only reasonable to expect that there should be at least some unit u, whose true category identification is c, $(u, c) \in T$, and which has d for one of its measurements, $(u, d) \in \Sigma$, for otherwise there would be no basis for including the pair (d, c) in the decision relation Δ. This observation can be stated in a more formal manner as in lemma 1: if, of those assignments made by the decision relation, there are to be no incorrect assignments in the sense of assigning a unit to the wrong category (the decision relation can also err by

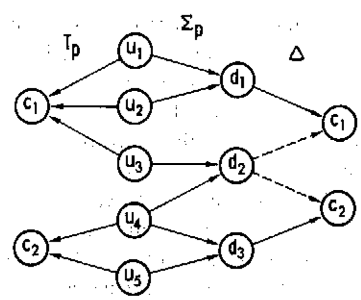


Fig. 5. An example set of relations Σ_p and T_p, where an error free solution does not exist. The unit prediction set is $U_p = \{u_1, u_2, u_3, u_4, u_5\}$, the measurement space is $D = \{d_1, d_2, d_3\}$, the category set is $C = \{c_1, c_2\}$, the measurement prediction relation is $\Sigma_p = \{u_1, d_1), (u_2, d_1), (u_3, d_2), (u_4, d_2), (u_4, d_3), (u_5, d_3)\}$, the category identification prediction relation is $T_p = \{(u_1, c_1), (u_2, c_1), (u_3, c_1), (u_4, c_2), (u_5, c_2)\}$. Consider three possible decision relations $\Delta_1 = \{(d_1, c_1), (d_2, c_1), (d_3, c_2)\}$, $\Delta_2 = \{(d_1, c_1), (d_2, c_2), (d_3, c_2)\}$, $\Delta_3 = \{(d_1, c_1), (d_2, c_1), (d_2, c_2), (d_3, c_2)\}$. The assigned categories for these relations are $\Sigma_p \bigcirc \Delta_1 = \{(u_1, c_1), (u_2, c_1), (u_3, c_1), (u_4, c_1), (u_4, c_2), (u_5, c_2)\}$, $\Sigma_p \bigcirc \Delta_2 = \{(u_1, c_1), (u_2, c_1), (u_3, c_2), (u_4, c_2), (u_5, c_2)\}$, and $\Sigma_p \bigcirc \Delta_3 = \{(u_1, c_1), (u_2, c_1), (u_3, c_1), (u_3, c_2), (u_4, c_1), (u_4, c_2), (u_5, c_2)\}$, respectively. The troublesome assignments are those for d_2, and they are shown in a dotted line. When the decision relation pairs d_2 with c_1 then u_4 will be assigned to c_1 incorrectly and u_3 will be assigned to c_1 correctly. When the decision relation pairs d_2 with c_2, then u_4 will be assigned to c_1 incorrectly and u_3 will be assigned to c_1 correctly. When the decision relation pairs d_2 with c_2, then u_4 will be assigned to c_2 correctly but u_3 will be assigned to c_2 incorrectly. When the decision relation pairs d_2 with both c_1 and c_2, then u_4 will be assigned to c_2 correctly and c_1 incorrectly and u_3 will be assigned to c_1 correctly and c_2 incorrectly.

leaving some units unassigned), then the decision relation Δ must be a subset of $\Sigma^{-1} \circ T$; that is, $\Delta \subseteq \Sigma^{-1} \circ T$. (The relation Σ^{-1} is the inverse of the relation Σ and is defined by the set of all measurement–unit pairs (d, u) such that $(u, d) \in \Sigma$).

The second observation we wish to make (lemma 2) is that if there exists a decision relation with the property that the category assignment given to any unit is correct ($\Sigma \circ \Delta \subseteq T$) and all units are given correct category assignments ($\Sigma \circ \Delta \supseteq T$), then it is necessary for the decision relation to contain all measurement-category pairs (d, c) such that for some unit u, u had measurement d, $(u, d) \in \Sigma$, and u had true category identification c, $(u, c) \in T$. Formally, $T = \Sigma \circ \Delta$ implies $\Delta \supseteq \Sigma^{-1} \circ T$.

Combining lemma 1 and 2 we obtain that if an error-free decision relation exists, that is, if there exists Δ such that $\Sigma \circ \Delta = T$, then Δ must be equal to the relation composition $\Sigma^{-1} \circ T$.

If we consider T to be defined everywhere and single-valued then $\{T^{-1}(c) | c \in C\}$ is a partition over the set of units. Lemma 3 characterizes the relationship between Σ and T for the error-free case. It basically states that a set of all units giving rise to the same measurement d must be contained in some cell $T^{-1}(c)$ of the partition of units. There are many equivalent ways of stating this idea. One is that $\Sigma^{-1} \circ T$ is single valued; another is that $\Sigma \circ \Sigma^{-1} \circ T \subseteq T$.

Having characterized interrelationships between Σ and T which are necessary and sufficient to allow the existence of an error-free decision relation, our next task is to determine how the unit training set is to be obtained so that the corresponding measurement training relation and category identification training relation can be used to calculate the error-free decision relation. Since in the error-free case, the error-free decision relation can be expressed by $\Sigma^{-1} \circ T$, we should like to find necessary and sufficient conditions on the training data Σ_t and T_t so that the decision rule calculated from the training data as $\Sigma_t^{-1} \circ T_t$ would be the same as that calculated by the full data as $\Sigma^{-1} \circ T$, i.e. $\Sigma^{-1} \circ T = \Sigma_t^{-1} \circ T_t$. Surprisingly, this characterization is quite easily obtained. Lemma 4 states that, if an error-free solution exists, the requirements that the unit training set U_t have such a diverse set of units such that (1) for each possible category there is some unit in U_t identified in that category and (2) for each possible measurement in D there is some unit in U_t giving rise to that measurement, are necessary and sufficient to guarantee that the decision relation calculated from the training data will in fact be the error-free decision relation.

The case when there exists an error-free decision relation is rather easy. What happens, however, when no such decision relation exists? Lemma 5 states that if the decision relation should contain all the measurement-category pairs (d, c) in $\Sigma^{-1} \circ T$, then at least one of the categories to which a unit gets assigned by the decision relation will be correct. Formally, $\Delta \supseteq \Sigma^{-1} \circ T$ implies $\Sigma \circ \Delta \supseteq T$. This means that it is possible to reduce the errors in the set $(\Sigma \circ \Delta)^c \cap T$ to zero. Unfortunately, setting $\Delta \supseteq \Sigma^{-1} \circ T$ can often lead to a relatively large number of errors in $(\Sigma \circ \Delta) \cap T^c$ as illustrated in the example of Fig. 6.

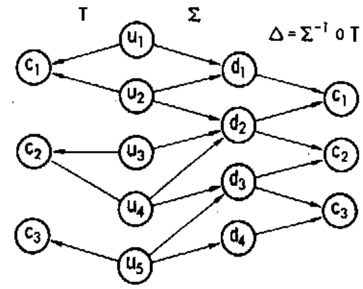


Fig. 6. The large number of incorrect category assignments in $(\Sigma \circ \Delta) \cap T^c$ which can occur when the decision relation $\Delta = \Sigma^{-1} \circ T$ and an error-free solution does not exist. The true category identification relation $T = \{(u_1, c_1), (u_2, c_1), (u_3, c_2), (u_4, c_2), (u_5, c_3)\}$, the assigned category identification relation $\Sigma \circ \Delta = \{u_1, c_1), (u_2, c_1), (u_2, c_2), (u_3, c_1), (u_3, c_2), (u_4, c_1), (u_4, c_2), (u_5, c_2), (u_5, c_3)\}$ and the incorrect assignments in $(\Sigma \circ \Delta) \cap T^c = \{(u_2, c_2), (u_3, c_1), (u_4, c_1), (u_4, c_3), (u_5, c_2)\}$. Note that there are no incorrect assignments in $(\Sigma \circ \Delta)^c \cap T$. Lemma 5 shows that this is always the case when $\Delta \supseteq \Sigma^{-1} \circ T$.

3. THE DECISION RULE WHICH ERRS

Σ Single-valued

We now examine the case when an error-free decision relation may not exist. We consider first the special case when Σ is single-valued. Here, we may express the number of correct decisions minus the number of incorrect decisions (see lemma 6) by

$$\#((\Sigma \circ \Delta) \cap T) - \#((\Sigma \circ \Delta) \cap T^c)$$
$$= \sum_{d \in D} \left[2 \sum_{c \in \Delta(d)} P(d, c) - \#\Delta(d) \sum_{c \in C} P(d, c) \right],$$

where $P(d, c) = \#(\Sigma^{-1}(d) \cap T^{-1}(c))$.
The unique largest decision relation Δ which maximizes this is defined by

$$\Delta = \Bigg\{ (d, c) | \text{ for some } (d, c) [\ \#(\Sigma^{-1}(d) \cap T^{-1}(c)) \geq \#(\Sigma^{-1}(d) \cap T^{-1}(c')) \text{ for every } c' \in C] \text{ and } \left[\sum_{c' \in C} \#(\Sigma^{-1}(d) \cap T^{-1}(c')) \leq 2\,\#(\Sigma^{-1}(d) \cap T^{-1}(c)) \right] \Bigg\},$$

or equivalently,

$$\Delta = \Bigg\{ (d, c) | \text{ for some } (d, c), [P(d, c) \geq P(d, c') \text{ for every } c' \in C] \text{ and } \left[\sum_{c' \in C} P(d, c') \leq 2P(d, c) \right] \Bigg\}.$$

This is proven in three parts. Lemma 7 shows that if

$$P(d, c) < 1/2 \sum_{c' \in C} P(d, c')$$

for every $c \in C$ then

$$2 \sum_{c \in \Delta(d)} P(d, c) - \#\Delta(d) \sum_{c \in C} P(d, c)$$

is maximal if an only if $\Delta(d) = \phi$. Lemma 8 shows that if there exists a $c \in C$ such that

$$P(d, c) > 1/2 \sum_{c' \in C} P(d, c'),$$

then

$$2 \sum_{c \in \Delta(d)} P(d, c) - \#\Delta(d) \sum_{c \in C} P(d, c)$$

is maximal if and only if $\Delta(d) = \{c \mid P(d, c) \geq P(d, c')$ for every $c' \in C$. Lemma 9 takes care of the case of equality; that is, when there exists a $c \in C$ such that

$$P(d, c) = 1/2 \sum_{c' \in C} P(d, c').$$

In this case,

$$2 \sum_{c \in \Delta(d)} P(d, c) - \#\Delta(d) \sum_{c \in C} P(d, c)$$

is maximal if and only if $\Delta(d)$ is any subset of $\{c \mid P(d, c) \geq P(d, c')$ for every $c' \in C\}$. Hence, the unique largest decision relation Δ which maximizes

$$\sum_{d \in D} \left[2 \sum_{c \in \Delta(d)} P(d, c) - \#\Delta(d) \sum_{c \in C} P(d, c) \right]$$

is

$$\Delta = \left\{ (d, c) \| \text{ for some } (d, c)[P(d, c) \geq P(d, c') \text{ for every } c' \in C] \text{ and } \left[P(d, c) \geq 1/2 \sum_{c' \in C} P(d, c') \right] \right\}.$$

Σ *not single-valued*

When Σ is not single-valued, we know of no easy way of determining the optimal decision relation. The problem could be formulated so that it is equivalent to finding an optimal set of linear discriminant functions. Since we make no distributional assumptions, an iterative procedure to find such linear discriminant functions would have to be used in a rather high $\#$. D-dimensional Euclidean space. Therefore, we will take another approach and make some approximations.

The approximations we make are based on the following idea. If Σ is not single-valued, then some units may give rise to more than one measurement. In these cases, there may be more than one assignment given such units. Instead of considering each of these multiple assignments with equal weight, we should consider each unit category assignment pair weighted by the reciprocal number of times the unit can be assigned a category. This weight factor is $1/\#\Sigma(u)$. Thus those units which are "tight" and give rise to only one measurement d will have full weight in the counting for d. But those units which are "loose" and give rise to many measurements will have only small weight in the counting for one of its measurements d. The approximations we make are:

$$\#((\Sigma \circ \Delta) \cap T) \approx \sum_{c \in C} \sum_{d \in D} \sum_{u \in T^{-1}(c) \cap \Sigma^{-1}(d)} \frac{\#(\Delta(d) \cap \{c\})}{\#\Sigma(u)}$$

$$\#((\Sigma \circ \Delta) \cap T^c) \approx \sum_{d \in D} \sum_{c \in C} \sum_{\substack{c' \in C \\ c' \neq c}} \sum_{u \in T^{-1}(c) \cap \Sigma^{-1}(d)} \frac{\#(\Delta(d) \cap \{c\})}{\#\Sigma(u)}$$

These approximations are exact if all measurements from any given unit get paired by the decision relation with the same category. Lemma 10 shows that with these approximations, the number of correct decisions minus the number of incorrect decisions can be expressed by

$$\#((\Sigma \circ \Delta) \cap T) - \#((\Sigma \circ \Delta) \cap T^c) \approx \sum_{d \in D} \left[2 \sum_{c \in \Delta(d)} P(d, c) - \#\Delta(d) \sum_{c \in C} P(d, c) \right]$$

where

$$P(d, c) = \sum_{u \in T^{-1}(c) \cap \Sigma^{-1}(d)} \frac{1}{\#\Sigma(u)}.$$

Since the form

$$2 \sum_{c \in \Delta(d)} P(d, c) - \#\Delta(d) \sum_{c \in C} P(d, c)$$

is identical to the one obtained when Σ is single valued, Lemmas 7–9 show that, as before, the largest optimal decision relation Δ is defined by

$$\Delta = \left\{ (d, c) \| \text{ for some } (d, c)[P(d, c) \geq P(d, c') \text{ for every } c' \in C], \text{ and } \left[P(d, c) > 1/2 \sum_{c' \in C} P(d, c') \right] \right\}.$$

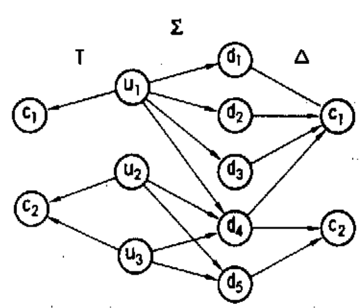
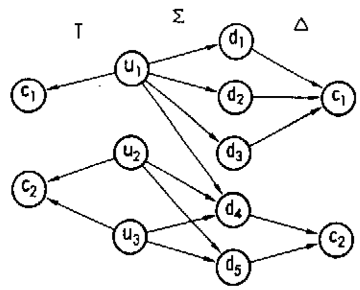


Fig. 7. A measurement relation Σ and true category identification relation T for which there exists no perfect or error-free decision relation Δ. The decision relation Δ which is shown is equal to $\Sigma^{-1} \circ T$. The set $(\Sigma \circ \Delta) \cap T^c = \{u_2, c_1), (u_3, c_1), (u_1, c_2)\}$ There are three incorrect identifications. The set $(\Sigma \circ \Delta) \cap T = \{(u_1, c_1), (u_2, c_2), (u_3, c_3)\}$. There are three correct identifications.



$\Sigma^{-1}(d_1) = \{u_1\}$
$\Sigma^{-1}(d_2) = \{u_1\}$
$\Sigma^{-1}(d_3) = \{u_1\}$
$\Sigma^{-1}(d_4) = \{u_1, u_2, u_3\}$
$\Sigma^{-1}(d_5) = \{u_2, u_3\}$

$T^{-1}(c_1) = \{u_1\}$
$T^{-1}(c_2) = \{u_2, u_3\}$

$\Sigma^{-1}(d_1) \cap T^{-1}(c_1) = \{u_1\}$
$\Sigma^{-1}(d_2) \cap T^{-1}(c_1) = \{u_1\}$
$\Sigma^{-1}(d_3) \cap T^{-1}(c_1) = \{u_1\}$
$\Sigma^{-1}(d_4) \cap T^{-1}(c_1) = \{u_1\}$
$\Sigma^{-1}(d_5) \cap T^{-1}(c_1) = \{\ \}$

$\Sigma^{-1}(d_1) \cap T^{-1}(c_2) = \{\ \}$
$\Sigma^{-1}(d_2) \cap T^{-1}(c_2) = \{\ \}$
$\Sigma^{-1}(d_3) \cap T^{-1}(c_2) = \{\ \}$
$\Sigma^{-1}(d_4) \cap T^{-1}(c_2) = \{u_2, u_3\}$
$\Sigma^{-1}(d_5) \cap T^{-1}(c_2) = \{u_2, u_3\}$

Fig. 8. The approximate largest optimal decision relation. $(\Sigma \bigcirc \Delta) \cap T^c = \{(u_1, c_2)\}$ and $(\Sigma \bigcirc \Delta) \cap T = \{(u_1, c_1), (u_2, c_2), (u_3, c_2)\}$.

$$\Delta = \left\{(d, c) \mid P(d, c) \geq P(d, c') \text{ for every } c' \in C \text{ and } P(d, c) > 1/2 \sum_{c' \in C} P(d, c')\right\}$$

where

$$P(d, c) = \sum_{u \in T^{-1}(c) \cap \Sigma^{-1}(d)} \frac{1}{\#\Sigma(u)}$$

$P(d_1, c_1) = \frac{1}{4}$
$P(d_2, c_1) = \frac{1}{4}$
$P(d_3, c_1) = \frac{1}{4}$
$P(d_4, c_1) = \frac{1}{4}$
$P(d_5, c_1) = 0$

$P(d_1, c_2) = 0$
$P(d_2, c_2) = 0$
$P(d_3, c_2) = 0$
$P(d_4, c_2) = 1$
$P(d_5, c_2) = 1$

4. DECISION RULES FOR RELATED UNITS

The discussion so far has provided no really new problems or solutions: only the set-relation orientation in which the material has been presented is new. In this section we shall use the relation orientation to pose a different kind of problem and suggest a solution to it. The problem we pose is this. Suppose that the units are related together because of spatial nearness (such as neighboring resolution cells on an image) or because of time (such as those signal time samples which occur one immediately after the other). In this case, the units are not independent and they should not be treated separately. Category assignments made for one unit should be dependent on the unit's measurements as well as measurements made from spatially related or time-sequentially related units. Specifically, we shall assume that units which are related together "tend" (we shall make this more precise) to have the same true category identification and "tend" to give rise to "similar" measurements. This assumption in a relation mode is somewhat analogous to a Markov dependence or tree dependence assumption.[4–6]

The idea of inter-unit relationships can be made precise when it is represented by a binary relation R on the set of units U, that is, $R \subseteq U \times U$. A pair of units (u_1, u_2) belongs to R if and only if unit u_1 is related to unit u_2.

Characteristic measurement decision algorithm

One possible way of using the relation R in the decision algorithm is to determine Δ as before and then, for each unit u, determine any measurement d which is linked to u by the most paths through R and through Σ and then pair u with any category linked to the characteristic measurement d by the decision relation Δ. This characteristic measurement decision algorithm is useful when D is small and R links many units (see Fig. 9). Let $A \subseteq U \times D$ be the binary relation pairing each unit with its characteristic measurement; that is, $A = \{(u, d) \mid \text{for some } (u, d),\ \#(R(u) \cap \Sigma^{-1}(d)) \geq \#(R(u) \cap \Sigma^{-1}(d')) \text{ for every } d' \in D\}$. Then the characteristic measurement decision algorithm assigns a unit u to category c by decision relation Δ if and only if $(u, d) \in A \bigcirc \Delta$.

Since this decision algorithm is just like the decision algorithm $\Sigma \bigcirc \Delta$, using the approximations

$$\#(A \bigcirc \Delta \cap T) \approx \sum_{c \in C} \sum_{d \in D} \sum_{u \in T^{-1}(c) \cap A(d)} \frac{\#(\Delta(d) \cap \{c\})}{\#A(u)}$$

$$\#((A \bigcirc \Delta \cap T^c)) \approx \sum_{d \in D} \sum_{c \in C} \sum_{\substack{c' \in C \\ c' \neq C}} \sum_{u \in T^{-1}(c) \cap A^{-1}(d)} \frac{\#(\Delta(d) \cap \{c\})}{\#A(u)}$$

we obtain that $\#((A \circ \Delta) \cap T) - \#((A \circ \Delta) \cap T^c)$ is approximately maximized if and only

$$\Delta = \Big\{(d, c) | \text{ for some } (d, c)[P(d, c) \geq P(d, c') \text{ for every } c' \in C], \text{ and } \Big[P(d, c) \geq 1/2 \sum_{c' \in C} P(d, c')\Big]\Big\},$$

where

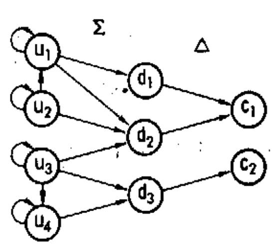
$$P(d, c) = \sum_{u \in T^{-1}(c) \cap A(d)} \frac{1}{\# A(d)}.$$

The approximation is exact when A is single valued or when all measurements linked to each unit are paired with the same category by Δ.

Characteristic category decision algorithm

When the set D is large and R does not link many units, the chance for a unique characteristic measurement to occur is small and the characteristic measurement decision rule is not effective. In this case we can formulate another decision algorithm such as: assign the unit u to that category c having a majority of the measurements in $(R \circ \Sigma)(u)$ linked to c by Δ. Figure 10 illustrates this measurement majority decision algorithm.

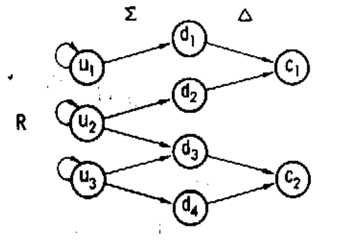
When for each unit u having true category identification c, a majority of the measurements linked to u through $R \circ \Sigma$ each have a majority of the units linked by $(R \circ \Sigma)^{-1}$ linked to category c by T, then



u_1, u_1, d_1
u_1, u_1, d_2
u_2, u_2, d_2
u_2, u_1, d_2
u_2, u_1, d_1
u_3, u_3, d_2
u_3, u_3, d_3
u_3, u_4, d_3
u_4, u_4, d_3

List of paths from u through R and then through Σ.

Fig. 9. Digraph of R, Σ and Δ, illustrating the characteristic measurement decision algorithm. Unit u_1 has two characteristic measurements d_1 and d_2 with one path to each. Since d_1 and d_2 are paired with category c_1, by Δ, the decision algorithm assigns u_1 to category c_1. Unit u_2 has two paths to characteristic measurement d_2 and d_2 is paired with category c_1, by Δ. Therefore, u_2 is assigned to category c_1. Unit u_3 has two paths to characteristic measurement d_3 and d_3 is paired with category c_2 by Δ. Therefore, u_3 is assigned to category c_2. Unit u_4 has its only path to characteristic measurement d_3 and hence is also assigned to category c_2.



$(R \circ \Sigma)(u_1) = \{d_1, d_2, d_3\}$
$(R \circ \Sigma)(u_2) = \{d_2, d_3, d_4\}$
$(R \circ \Sigma)(u_3) = \{d_3, d_4\}$

Fig. 10. The characteristic category decision algorithm. Unit u_1 is assigned to category c_1 since the majority of measurements in $(R \circ \Sigma)(u_1)$ are linked to c_1 by Δ. Units u_2 and u_3 are assigned to category c_2 since a majority of measurements in $(R \circ \Sigma)(u_2)$ and $(R \circ \Sigma)(u_3)$ are linked to c_2 by Δ.

lemma 11 shows that a no error decision relation Δ exists. Formally, if $(u, c) \in T$ implies

$$\#\{d \in (R \circ \Sigma)(u) | \#((R \circ \Sigma)^{-1}(d) \cap T^{-1}(c)) > 1/2\, \#(R \circ \Sigma)^{-1}(d)\} > 1/2 \#(R \circ \Sigma)(u),$$

and the unit u is assigned to category c if and only if $\#(\Delta^{-1}(c) \cap (R \circ \Sigma)(u)) > 1/2\, \#(R \circ \Sigma)(u)$, then an error-free decision relation Δ exists and is defined by

$$\Delta = \{(d, c) | \#(R \circ \Sigma)^{-1}(d) \cap T^{-1}(c) > 1/2\, \#(R \circ \Sigma)^{-1}(d)\}.$$

That this condition is sufficient but not necessary is shown by an example illustrated in Fig. 11.

As usual, the case for the error-free decision relation has to be worked out using an approximation. When

$$\#(T \cap \hat{T}) - \#(T^c \cap \hat{T}) \approx \sum_{u \in U} \frac{\sum_{d \in D} \sum_{u' \in R(u) \cap \Sigma^{-1}(d)} \# \Delta(d) \cap T(u)}{\sum_{u' \in R(u)} \# \Sigma(u')} - \sum_{u \in U} \frac{\sum_{d \in D} \sum_{u' \in R(u) \cap \Sigma^{-1}(d)} \# \Delta(d) \cap T^c(u)}{\sum_{u' \in R(u)} \# \Sigma(u')}$$

Lemma 12 shows that

$$\#(T \cap \hat{T}) - \#(T^c \cap \hat{T}) \approx \sum_{d \in D} \Big[2 \sum_{c \in \Delta(d)} P(d, c) - \# \Delta(d) \sum_{c \in C} P(d, c)\Big].$$

Then, using Lemmas 7–9, we obtain that the approximate optimal decision rule Δ is

$$\Delta = \Big\{(d, c) | \text{ for some } (d, c)[P(d, c) \geq P(d, c') \text{ for every } c' \in C] \Big[P(d, c) \geq 1/2 \sum_{c' \in C} P(d, c')\Big]\Big\},$$

where

$$P(d,c) = \sum_{u \in T^{-1}(c)} \left(\frac{1}{\sum_{u' \in R(u)} \# \Sigma(u')} \right) \# R(u) \cap \Sigma^{-1}(d)$$

5. SUMMARY

We have discussed the pattern discrimination problem from the perspective of set-relation theory. The central idea of this perspective is to think of the measurement data as a binary relation $\Sigma \subseteq U \times D$ from the set of units U to the set of measurements D and to think of the category identification data as a binary relation $T \subseteq U \times C$ from the set of units U to the set of categories C. The decision relation $\Delta \subseteq D \times C$ is a binary relation from the set of measurements D to the set of categories U. The unit category assignments pair is then naturally given by the relation composition $\Sigma \circ \Delta$. We have shown that when there exists an error-free decision relation Δ, the training data Σ_t and T_t must be chosen so that $\Sigma_t(U) = D$ and $T_t(U) = C$. In this case $\Sigma_t^{-1} \circ T_t = \Sigma^{-1} \circ T = \Delta$ so that the decision relation Δ which is computed from Σ_t and T_t will be optimal. In the case when no error-free decision relation exists we have shown that some of the possibly multiple category identification assignments made by the decision relation $\Sigma_t^{-1} \circ T_t$ are incorrect assignments and that one of the assignments must be correct. To reduce the errors of commission (assigning units to incorrect categories) and yet keep the errors of omission (not assigning units to any category) we examined a decision relation which tended to minimize the sum of these two kinds of errors. This decision relation Δ was defined by

$$\Delta = \left\{ (d,c) \mid \text{for some } (d,c)[P(d,c) \geq P(d,c') \text{ for every } c' \in C] \text{ and } P(d,c) > 1/2 \sum_{c' \in C} P(d,c') \right\},$$

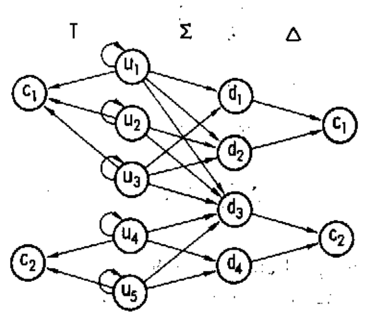


Fig. 11. An example where $\#((R \circ \Sigma^{-1})(d_3) \cap T^{-1}(c_1)) > \#((R \circ \Sigma^{-1})(d_3) \cap T^{-1}(c_2))$ yet $\Delta(d_3) = c_2$ is the best pairing for d_3 when the decision algorithm is "assign" the unit u to each category c such that $\#(\Delta^{-1}(c) \cap (R \circ \Sigma)(u)) \geq (\Delta^{-1}(c') \cap (R \circ \Sigma)(u))$ for every $c' \in C$. Note that if d_3 and c_1 were paired together by Δ, then there would still be perfect category assignments.

where

$$P(d,c) = \sum_{u \in T^{-1}(c) \cap \Sigma^{-1}(d)} \frac{1}{\# \Sigma(u)}.$$

It does indeed minimize the error when Σ is single valued but only tends to when Σ is not single-valued.

We then examined the case when units are related by a reflexive binary relation R on the set of units U. We suggested two possibilities for the definition of an appropriate decision relation.

The first possibility is based on a situation where D is a small set, many units are related together, and Σ may not be single-valued. In this case, we might expect that each unit is linked through R and then through Σ to many measurements in D. Among these measurements some of them are linked to a unit u by the most paths. Call any such measurement d a characteristic measurement for unit u. The characteristic measurement algorithm would assign unit u to any categories linked to d by the decision relation Δ. Formally, let $A \subseteq U \times D$ be the binary relation pairing each unit with its characteristic measurement,

$$A = \{(u,d) \mid \text{for some } (u,\phi)\ \#(R(u) \cap \Sigma^{-1}(d)) \geq \#(R(u) \cap \Sigma^{-1}(d')) \text{ for every } d' \in D\}$$

The characteristic measurement decision algorithm defines the unit category assignment pairs by the relation composition $A \circ \Delta$ where the decision relation Δ is defined by

$$\Delta = \left\{ (d,c) \mid \text{for some } (d,c)[P(d,c) \geq P(d,c') \text{ for every } c' \in C] \text{ and } P(d,c) \geq 1/2 \sum_{c' \in C} P(d,c') \right\},$$

where

$$P(d,c) = \sum_{u \in T^{-1}(c) \cap A(d)} \frac{1}{\# A(d)}.$$

The second possibility is based on a situation where D is a large set and few units are related together. In this case there is not likely to be unique representative characteristic measurements for each unit although each unit may be linked to many measurements in D through $R \circ \Sigma$. The decision relation Δ links each one of the measurements in $(R \circ \Sigma)(u)$ to a category in C. Among these categories consider that category linked to u by the majority paths. Call such category a characteristic category for u. The characteristic category algorithm would assign unit u to such a characteristic category for u.

Formally, the characteristic category algorithm assigns the unit u to category c when

$$\#(\Delta^{-1}(c) \cap (R \circ \Sigma)(u)) > 1/2\ \#(R \circ \Sigma)(u).$$

The decision relation Δ is defined by.

$$\Delta = \left\{(d, c) | \text{ for some } (d, c)[P(d, c) \geq P(d, c') \text{ for every } c' \in C] \text{ and } \left[P(d, c) \geq 1/2 \sum_{c' \in C} P(d, c')\right]\right\},$$

where

$$P(d, c) = \sum_{u \in T^{-1}(c)} \left(\frac{1}{\sum_{u' \in R(u)} \# \Sigma(u')}\right) \# R(u) \cap \Sigma^{-1}(d).$$

Both the characteristic measurement algorithm and the characteristic category algorithm provide procedures for taking into account unit to unit dependence without making the typical Markov type assumption for the form of the probability distribution used in the usual statistical decision procedure. In fact, the notation $P(d, c)$ which we have used has suggested that when pattern discrimination is viewed from relation theory probability-like functions can arise but in a form peculiar to the set-relation context.

Although this paper is theoretical in nature, offering no practical example pattern recognitions problems which are solved by the suggested procedures, it is the hope of the author that the paper will open some new alternatives to investigators.

REFERENCES

1. R. S. Michalski, A geometric model for the synthesis of internal covers, Report 461, Department of Computer Science, University of Illinois, Urbana, Illinois (1971).
2. R. S. Michalski and B. H. McCormick, Interval generalization of switching theory, Report 442, Department of Computer Science, University of Illinois, Urbana, Illinois (1971).
3. J. S. Read and S. N. Jayaramamurthy, Automatic generation of texture feature detectors, *IEEE Trans Comp.* Vol. 803–812 (1972).
4. C. K. Chow and C. N. Lin, Approximating discrete probability distributions with dependence trees. *IEEE Trans Inf. Theor.* **IT-14** (3) 462–467 (1968).
5. C. K. Chow, A recognition method using neighbor dependence, *IRE Trans Electronic Comp.* **EC–11**, 683–690 (1962).
6. K. Abend, T. J. Harley and L. N. Kanal, Classification of binary random patterns. *IEEE Trans. Inf. Theor.* **IT–11**, (4) (1965).

APPENDIX

Lemma 1. Let Σ be onto. If $\Sigma \circ \Delta \subseteq T$, then $\Delta \subseteq \Sigma^{-1} \circ T$.

Proof. Let $(d, c) \epsilon \Delta$. Since Σ is onto, $\exists\ u \epsilon U \ni (u, d) \epsilon \Sigma$. Now, $(u, d) \epsilon \Sigma$ and $(d, c) \epsilon \Delta$ imply $(u, c) \epsilon \Sigma \circ \Delta$. Since $\Sigma \circ \Delta \subseteq T$, $(u, c) \epsilon T$ also. But $(u, d) \epsilon \Sigma$ and $(u, c) \epsilon T$ imply $(d, c) \epsilon \Sigma^{-1} \circ T$. Hence, $\Delta \subseteq \Sigma^{-1} \circ T$.

Lemma 2. Let T be single-valued and Δ be defined everywhere. If $T = \Sigma \circ \Delta$, then $\Delta \subseteq \Sigma^{-1} \circ T$.

Proof. Let $(d, c) \epsilon \Sigma^{-1} \circ T$. Then $\exists\ u \epsilon U \ni (u, d) \epsilon \Sigma$ and $(u, c) \epsilon T$. Since $(u, c) \epsilon T$ and $T = \Sigma \circ \Delta$, then $(u, c) \epsilon \Sigma \circ \Delta$ also. Since Δ is defined everywhere, $\exists\ \hat{c} \ni (d, \hat{c}) \epsilon \Delta$. But $(u, d) \epsilon \Sigma$ and $(d, \hat{c}) \epsilon \Delta$ imply $(u, \hat{c}) \epsilon \Sigma \circ \Delta = T$, and T single valued implies $\hat{c} = c$. Hence $(d, c) \epsilon \Delta$ so that $\Delta \subseteq \Sigma^{-1} \circ T$.

Lemma 3. Let Σ be defined everywhere, and T be defined everywhere and single valued. Then the following statements are equivalent:

$$\Sigma^{-1} \circ T \text{ is single-valued}$$
$$\Sigma \circ \Sigma^{-1} \circ T \subseteq T$$
$$\Sigma \circ \Sigma^{-1} \circ T = T$$
$$(d, c) \epsilon D \times C \text{ implies } \Sigma^{-1}(d) \subseteq T^{-1}(c) \text{ or } \Sigma^{-1}(d) \cap T^{-1}(c) = \phi.$$

Proof. (1) Suppose $\Sigma^{-1} \circ T$ is single-valued. Certainly,

$$(\Sigma^{-1} \circ T)^{-1} \circ (\Sigma^{-1} \circ T) \subseteq I.$$

Hence

$$T^{-1} \circ (\Sigma \circ \Sigma^{-1} \circ T) \subseteq I.$$

Letting T^{-1} play the role of Σ, $(\Sigma \circ \Sigma^{-1} \circ T)$ play the role of Δ, and I play the role of T, in lemma 1 we obtain

$$\Sigma \circ \Sigma^{-1} \circ T \subseteq T.$$

(2) Suppose $\Sigma \circ \Sigma^{-1} \circ T \subseteq T$. Since Σ is defined everywhere, $\Sigma \circ \Sigma^{-1} \supseteq I$. Thus,

$$\Sigma \circ \Sigma^{-1} \circ T \supseteq I \circ T = T.$$

But $\Sigma \circ \Sigma^{-1} \circ T \supseteq T$ and the supposition $\Sigma \circ \Sigma^{-1} \circ T \subseteq T$ imply $\Sigma \circ \Sigma^{-1} \circ T = T$.

(3) Suppose $\Sigma \circ \Sigma^{-1} \circ T \subseteq T$. Let $(d, c) \epsilon D \times C$ and $\hat{u} \epsilon \Sigma^{-1}(d)$. Either $\Sigma^{-1}(d) \cap T^{-1}(c) = \phi$ or not. If so then $\Sigma^{-1}(d) \cap T^{-1}(c) = \phi$. If not, then there exists a $u \epsilon \Sigma^{-1}(d) \cap T^{-1}(c)$.

Hence

$$(\hat{u}, d) \epsilon \Sigma,\ (d, u) \epsilon \Sigma^{-1} \text{ and } (u, c) \epsilon T.$$

This implies

$$(\hat{u}, c) \epsilon \Sigma \circ \Sigma^{-1} \circ T.$$

By supposition, $\Sigma \circ \Sigma^{-1} \circ T \subseteq T$ so that

$$(\hat{u}, c) \epsilon T \text{ or } \hat{u} \epsilon T^{-1}(c).$$

(4) Suppose $(d, c) \epsilon D \times C$ implies $\Sigma^{-1}(d) \subseteq T^{-1}(c)$ or $\Sigma^{-1}(d) \cap T^{-1}(c) - \phi$. Let $(c, \hat{c}) \epsilon (\Sigma^{-1} \circ T)^{-1} \circ (\Sigma^{-1} \circ T)$. Then there exists u_1 and $u_2 \epsilon U$ and $d \epsilon D$ such that

$$(c, u_1) \epsilon T^{-1},\ (u_1, d) \epsilon \Sigma,\ (d, u_2) \epsilon \Sigma^{-1} \text{ and } (u_2, \hat{c}) \epsilon T.$$

Hence,

$$u_1 \epsilon \Sigma^{-1}(d) \cap T^{-1}(c)$$
$$\text{so that } \Sigma^{-1}(d) \cap T^{-1}(c) \neq \phi.$$

But by supposition $\Sigma^{-1}(d) \subseteq T^{-1}(c)$ or $\Sigma^{-1}(d) \cap T^{-1}(c) = \phi$. Since $\Sigma^{-1}(d) \cap T^{-1}(c) \neq \phi$, $\Sigma^{-1}(d) \subseteq T^{-1}(c)$. Now $u_2 \epsilon \Sigma^{-1}(d)$. Hence

$$u_2 \epsilon T^{-1}(c).$$

Now T being single-valued and $(u_2, \hat{c}) \epsilon T$ and $(u_2, c) \epsilon T$ imply $\hat{c} = c$. Therefore, $(c, \hat{c}) = (c, \hat{c}) \epsilon I$ and $\Sigma^{-1} \circ T$ is single-valued.

Lemma 4. Let $\Sigma_t \subseteq U_t \times D$, $\Sigma_t \subseteq \Sigma \subseteq U \times D$, $T_t \subseteq U_t \times C$, $T_t \subseteq T \subseteq U \times C$ and suppose $\Sigma \circ \Sigma^{-1} \circ T \subseteq T$. Then $T_t(U_t) = C$ and $\Sigma_t(U_t) = D$, if and only if $\Sigma_t^{-1} \circ T_t = \Sigma^{-1} \circ T$.

Proof. Suppose T_t and Σ_t are onto. Let $(d, c) \epsilon \Sigma^{-1} \circ T$. Then $\exists\ u \epsilon U \ni (d, u) \epsilon \Sigma^{-1}$ and $(u, c) \epsilon T$. Since $\Sigma_t(U_t) = D$, $\exists\ u_t \epsilon U_t . (u_t, d) \epsilon \Sigma_t \subseteq \Sigma$. But $(u_t, d) \epsilon \Sigma$ and $(u, d) \epsilon \Sigma$ imply $(u_t, u) \epsilon \Sigma \circ \Sigma^{-1}$. And $(u_t, u) \epsilon \Sigma \circ \Sigma^{-1}$ combined with $(u, c) \epsilon T$ imply $(u_t, c) \epsilon \Sigma \circ \Sigma^{-1} \circ T$. By assumption $\Sigma \circ \Sigma^{-1} \circ T \subseteq T$ so that $(u_t, c) \epsilon T$.

Since T_t is defined everywhere, $\exists\ \hat{c} \epsilon C\ (u_t, \hat{c}) \epsilon T_t$. But $T_t \subseteq T$ so that $(u_t, \hat{c}) \epsilon T$. Now, $(u_t, c) \epsilon T$, $(u_t, \hat{c}) \epsilon T$ and T single-valued imply $c = \hat{c}$. Hence $(u_t, \hat{c}) \epsilon T_t$.

Finally, $(u_t, c) \in T_t$ and $(u_t, d) \in \Sigma_t$ imply $(d, c) \in \Sigma_t^{-1} \circ T_t$ so that $\Sigma^{-1} \circ T \subseteq \Sigma_t \circ T_t$. Also, $\Sigma_t^{-1} \circ T_t \subseteq \Sigma^{-1} \circ T$ since $\Sigma_t \subseteq \Sigma$, $T_t \subseteq T$. Hence, $\Sigma_t^{-1} \circ T_t = \Sigma^{-1} \circ T$.

Suppose $\Sigma_t^{-1} \circ T_t = \Sigma^{-1} \circ T$. Let $c \in C$ be given. Since T is onto C, $\exists\ u \in U \ni (u, c) \in T$. Since Σ is defined everywhere $\exists\ d \in D \ni (u, d) \in \Sigma$. Hence $(d, c) \in \Sigma^{-1} \circ T$ But $\Sigma^{-1} \circ T = \Sigma_t^{-1} \circ T_t$ so that $(d, c) \in \Sigma^{-1} \circ T_t$. So $\exists\ u_t \in U_t \ni (u_t, d) \in \Sigma_t$ and $(u_t, c) \in T_t$. Therefore T_t is onto C.

Now let $d \in D$ be given. Since Σ is onto D, $\exists\ u \in U \ni (u, d) \in \Sigma$. Since T is defined everywhere $\exists\ c \in C \ni (u, c) \in T$. Hence $(d, c) \in \Sigma^{-1} \circ T$. But $\Sigma^{-1} \circ T = \Sigma_t^{-1} \circ T_t$ so that $(d, c) \in \Sigma_t^{-1} \circ T_t$. So $\exists\ u_t \in U_t \ni (u_t, d) \in \Sigma_t$ and $(u_t, c) \in T$. Therefore Σ_t is onto D.

Lemma 5. Let Σ be defined everywhere. If $\Delta \supseteq \Sigma^{-1} \circ T$, then $\Sigma \circ \Delta \supseteq T$.

Proof. Let $(u, c) \in T$. Since Σ is defined everywhere, there exists a d such that $(u, d) \in \Sigma$. Hence $(d, c) \in \Sigma^{-1} \circ T$. But $\Sigma^{-1} \circ T \subseteq \Delta$ so that $(d, c) \in \Delta$. Now, $(u, d) \in \Sigma$ and $(d, c) \in \Delta$ imply $(u, c) \in \Sigma \circ \Delta$.

Lemma 6. If Σ and T are single-valued, and defined everywhere, then $\#(\Sigma \circ \Delta) \cap T - \#(\Sigma \circ \Delta) \cap T^c$

$$= \sum_{d \in D} \left[2 \sum_{c \in \Delta(d)} P(d, c) - \#\Delta(d) \sum_{c \in C} P(d, c) \right],$$

where $P(d, c) = \#T^{-1}(c) \cap \Sigma^{-1}(d)$.

Proof.

$$\begin{aligned} \#(\Sigma \circ \Delta) \cap T &= \#\{(u, c) \mid \text{for some } u, d, c, (u, d) \in \Sigma, (d, c) \in \Delta, \text{ and } (u, c) \in T\} \\ &= \sum_{u \in U} \#\{(u, c) \mid \text{for some } d, c, (u, d) \in \Sigma, (d, c) \in \Sigma, \text{ and } (u, c) \in T\} \\ &= \sum_{c' \in C} \sum_{u \in T^{-1}(c')} \#\{(u, c) \mid \text{for some d}, c, (u, d) \in \Sigma, (d, c) \in \Delta, \text{ and } (u, c) \in T\} \\ &= \sum_{c' \in C} \sum_{c \in C} \sum_{u \in T^{-1}(c') \cap T^{-1}(c)} \#\{(u, c) \mid \text{for some } d, (u, d) \in \Sigma, (d, c) \in \Delta\}. \end{aligned}$$

Since T is single valued, $u \in T^{-1}(c') \cap T^{-1}(c)$ implies $c' = c$ so that

$$\#(\Sigma \circ \Delta) \cap T = \sum_{c \in C} \sum_{u \in T^{-1}(c)} \#\{(u, c) \mid \text{for some } d, (u, d) \in \Sigma, (d, c) \in \Delta\}.$$

Since Σ is single-valued,

$$\begin{aligned} \#(\Sigma \circ \Delta) \cap T &= \sum_{c \in C} \sum_{d \in D} \sum_{u \in T^{-1}(c) \cap \Sigma^{-1}(d)} \#\{(u, c) \mid (d, c) \in \Delta\} \\ &= \sum_{c \in C} \sum_{d \in D} \sum_{u \in T^{-1}(c) \cap \Sigma^{-1}(d)} \#\Delta(d) \cap \{c\} \\ &= \sum_{d \in D} \sum_{c \in \Delta(d)} \#T^{-1}(c) \cap \Sigma^{-1}(d). \end{aligned}$$

Letting $P(d, c) = \#T^{-1}(c) \cap \Sigma^{-1}(d)$, we obtain

$$\#(\Sigma \circ \Delta) \cap T = \sum_{d \in D} \sum_{c \in \Delta(d)} P(d, c).$$

$$\begin{aligned} \#(\Sigma \circ \Delta) \cap T^c &= \#\{(u, c) \mid \text{for some } u, d, c, (u, d) \in \Sigma, (d, c) \in \Delta, \text{ and } (u, c) \notin T\} \\ &= \sum_{u \in U} \#\{(u, c) \mid \text{for some } d, c, (u, d) \in \Sigma, (d, c) \in \Delta, \text{ and } (u, c) \notin T\} \\ &= \sum_{c' \in C} \sum_{u \in T^{-1}(c')} \#\{(u, c) \mid \text{for some } d, c, (u, d) \in \Sigma, (d, c) \in \Delta, \text{ and } (u, c) \notin T\} \\ &= \sum_{c' \in C} \sum_{c \in C} \sum_{u \in T^{-1}(c') - T^{-1}(c)} \#\{(u, c) \mid \text{for some } d, (u, d) \in \Sigma, (d, c) \in \Delta\}. \end{aligned}$$

But $u \in T^{-1}(c') \cap T^{-1}(c)$ if and only if $c' \neq c$.

$$\begin{aligned} \#(\Sigma \circ \Delta) \cap T^c &= \sum_{c' \in C} \sum_{\substack{c \in C \\ c \neq c'}} \sum_{u \in T^{-1}(c')} \#\{(u, c) \mid \text{for some } d, (u, d) \in \Sigma, (d, c) \in \Delta\} \\ &= \sum_{c' \in C} \sum_{\substack{c \in C \\ c \neq c'}} \sum_{d \in D} \sum_{u \in T^{-1}(c')\Sigma^{-1}(d)} \#\{(u, c) \mid (d, c) \in \Delta\} \\ &= \sum_{c' \in C} \sum_{\substack{c \in C \\ c \neq c'}} \sum_{d \in D} \sum_{u \in T^{-1}(c') \cap \Sigma^{-1}(d)} \#\Delta(d) \cap \{c\} \\ &= \sum_{d \in D} \sum_{c' \in C} \sum_{\substack{c \in \Delta(d) \\ c \neq c'}} \#T^{-1}(c') \cap \Sigma^{-1}(d) \\ &= \sum_{d \in D} \sum_{c' \in C} \sum_{\substack{c \in \Delta(d) \\ c \neq c'}} P(d, c'). \end{aligned}$$

$$\begin{aligned} \#(\Sigma \circ \Delta) \cap T^c &= \sum_{d \in D} \sum_{c' \in C} P(d, c') \sum_{\substack{c \in \Delta(d) \\ c \neq c'}} 1 \\ &= \sum_{d \in D} \sum_{c' \in C} P(d, c') \#(\Delta(d) - \{c'\}) \\ &= \sum_{d \in D} \sum_{c' \in \Delta(d)} P(d, c') \#(\Delta(d) - \{c'\}) + \sum_{d \in D} \sum_{c' \in \Delta(d)^c} P(d, c') \#(\Delta(d) - \{c'\}) \\ &= \sum_{d \in D} [\#\Delta(d) - 1] \sum_{c \in \Delta(d)} P(d, c) + \sum_{d \in D} \#\Delta(d) \sum_{c \in \Delta(d)^c} P(d, c) \\ &= \sum_{d \in D} \#\Delta(d) \sum_{c \in C} P(d, c) - \sum_{d \in D} \sum_{c \in \Delta(d)} P(d, c) \end{aligned}$$

Finally

$$\begin{aligned} \#(\Sigma \circ \Delta) \cap T - \#(\Sigma \circ \Delta) \cap T^c &= \sum_{d \in D} \sum_{c \in \Delta(d)} P(d, c) - \left[\sum_{d \in D} \#\Delta(d) \sum_{c \in C} P(d, c) - \sum_{d \in D} \sum_{c \in \Delta(d)} P(d, c) \right] \\ &= 2 \sum_{d \in D} \sum_{c \in \Delta(d)} P(d, c) - \sum_{d \in D} \#\Delta(d) \sum_{c \in C} P(d, c) \end{aligned}$$

Lemma 7. Suppose

$$P(d, c) < \frac{1}{2} \sum_{c' \in C} P(d, c') \quad c \in C.$$

Then $\Delta(d) = \phi$ if and only if

$$2 \sum_{c \in \Delta(d)} P(d, c) - \#\Delta(d) \sum_{c \in C} P(d, c)$$

is maximal.

Proof. The largest

$$\sum_{c \in \Delta(d)} P(d, c)$$

can be is

$$\min\left\{ \#\Delta(d) P_{\max}(d), \sum_{c \in C} P(d, c) \right\}$$

where $P_{\max}(d) = \max_c P(d, c)$. Since

$$P(d, c) < \frac{1}{2} \sum_{c' \in C} P(d, c') \forall c \in C, \quad P_{\max}(d) < \frac{1}{2} \sum_{c' \in C} P(d, c').$$

Hence,

$$\left. \begin{aligned} &\min\left\{ \#\Delta(d) P_{\max}(d), \sum_{c \in C} P(d, c) \right\} \\ &\leq \min\left\{ \frac{\#\Delta(d)}{2} \sum_{c \in C} P(d, c), \ \sum_{c \in C} P(d, c) \right\} \\ &\leq \frac{1}{2} \sum_{c \in C} P(d, c) \quad \text{if} \quad \#\Delta(d) = 1 \\ &\leq \sum_{c \in C} P(d, c) \quad \text{if} \quad \#\Delta(d) \geq 2 \\ &= 0 \quad \text{if} \quad \#\Delta(d) = 0 \end{aligned} \right\}$$

Now,

$$\left.\begin{aligned}&2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c) \le 2\left(\frac{1}{2}\sum_{c\in C} P(d,c)\right)\\&-\sum_{c\in C} P(d,c) \quad \text{if} \quad \#\Delta(d) = 1\\&\le 2\sum_{c\in C} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c) \quad \text{if} \quad \#\Delta(d) \ge 2\\&= 0 \quad \text{if} \quad \#\Delta(d) = 0\end{aligned}\right\}$$

so that

$$2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c) \le 0.$$

$$\Rightarrow \text{If } \Delta(d) = \phi, \quad \text{then} \quad \#\Delta(d) = 0$$

and

$$2\sum_{c\in\phi} P(d,c) - \#\phi\sum_{c\in C} P(d,c)$$
$$= 0 \ge 2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c)$$

for all $\Delta(d)$.

$$\Leftarrow \text{Suppose } 2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c)$$

is maximal, then

$$0 \ge 2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c)$$
$$\ge 2\sum_{c\in\Delta'(d)} P(d,c) - \#\Delta'(d)\sum_{c\in C} P(d,c)$$

for any $\Delta'(d)$.

0 Is the least upper bound and is achievable when $\Delta(d) = \phi$. The question is whether $\Delta(d) = \phi$ uniquely provides this upper bound. The answer is yes if

$$0 = 2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c)$$

for then

$$\#\Delta(d)\sum_{c\in C} P(d,c) = 2\sum_{c\in\Delta(d)} P(d,c)$$

or

$$\frac{\#\Delta(d)}{2} = \frac{\sum_{c\in\Delta(d)} P(d,c)}{\sum_{c\in C} P(d,c)} \le 1$$

so that $\#\Delta(d) \le 2$. If $\#\Delta(d) = 1$ the largest

$$\sum_{c\in\Delta(d)} P(d,c)$$

can be is $P_{max}(d)$ so that

$$2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c)$$
$$= 2P_{max}(d) - \sum_{c\in C} P(d,c)$$

which is less than zero by assumption. If $\#\Delta(d) = 2$ the largest

$$\sum_{c\in\Delta(d)} P(d,c)$$

can be is $2P_{max}(d)$. Hence,

$$2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c) \le 2(2P_{max}(d))$$
$$-2\sum_{c\in C} P(d,c) \le 2\left[2P_{max}(d) - \sum_{c\in C} P(d,c)\right],$$

which is less than zero by assumption.

Lemma 8. Suppose

$$\exists c \in P(d,c) > \frac{1}{2}\sum_{c'\in C} P(d,c')$$

for some $c \in C$. Then over all possible subsets $\Delta(d)$ of C,

$$T_\Delta = 2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c)$$

is maximal if and only if

$$\Delta(d) = \{c \mid P(d,c) \ge P(d,c')\forall c' \in C\}.$$

Proof. Define $\Delta(d) = \{c \mid P(d,c) \ge P(d,c')\forall c' \in C\}$. Then $P(d,c)$ must be maximal for any category in $\Delta(d)$; that is, for any $c \in \Delta(d)$,

$$P(d,c) = P_{max}(d) = \max_{c'\in C} P(d,c').$$

Since

$$\exists c \ni P(d,c) > \frac{1}{2}\sum_{c'\in C} P(d,c'), \quad P_{max}(d) = P(d,c)$$

and c is unique. Hence $\#\Delta(d) = 1$.

Now,

$$T_\Delta = 2\#\Delta(d)P_{max}(d) - \#\Delta(d)\sum_{c\in C} P(d,c)$$
$$= \#\Delta(d)\left[2P_{max}(d) - \sum_{c\in C} P(d,c)\right].$$
$$= 2P_{max}(d) - \sum_{c\in C} P(d,c).$$

Define for any subset S of C

$$T_S = 2\sum_{c\in S} P(d,c) - \#S\sum_{c\in C} P(d,c).$$

$T_\Delta - T_S > 0$ for any subset S of C other than $\Delta(d)$ if and only if $\Delta(d)$ uniquely maximizes T_Δ. Let S be any subset of C. Consider $T_\Delta - T_S$.

$$T_\Delta - T_S = 2P_{max}(d) - \sum_{c\in C} P(d,c) - 2\sum_{c\in S} P(d,c)$$
$$+ \#S\sum_{c\in C} P(d,c).$$

Either S is empty or not. If S is empty $T_S = 0$ and

$$T_\Delta - T_S = 2P_{max}(d) - \sum_{c\in C} P(d,c)$$

which is greater than zero since

$$P_{max}(d) > \frac{1}{2}\sum_{c\in C} P(d,c').$$

Hence empty S will not maximize T_S.

Now suppose S is not empty. Certainly,

$$\sum_{c\in C} P(d,c) \ge \sum_{c\in S} P(d,c).$$

$$\begin{aligned}T_\Delta - T_S &= 2P_{max}(d) - \sum_{c\in C} P(d,c) - 2\sum_{c\in S} P(d,c)\\&+ \#S\sum_{c\in C} P(d,c) \ge 2P_{max}(d) - \sum_{c\in C} P(d,c)\\&- 2\sum_{c\in C} P(d,c) + \#S\sum_{c\in C} P(d,c) \ge 2P_{max}(d)\\&- 3\sum_{c\in C} P(d,c) + \#S\sum_{c\in C} P(d,c) \ge \sum_{c\in C} P(d,c)\\&- 3\sum_{c\in C} P(d,c) + \#S\sum_{c\in C} P(d,c)\\&\ge \left[\sum_{c\in C} P(d,c)\right][\#S - 2] > 0 \quad \text{if} \quad \#S > 2.\end{aligned}$$

Hence any $S \ni \#S > 2$ will not maximize T_S. Consider what happens if $\#S = 1$. Either $c \in S$ implies $P(d,c) = P_{max}(d)$ or $c \in S$ implies $P(d,c) < P_{max}(d)$. If $c \in S$ implies $P(d,c) = P_{max}(d)$, then $S = \Delta(d)$. If $c \in S$ implies $P(d,c) < P_{max}(d)$ then

$$\begin{aligned}T_\Delta - T_S &= 2P_{max}(d) - \sum_{c\in C} P(d,c) - 2\sum_{c\in S} P(d,c)\\&+ \#S\sum_{c\in C} P(d,c) > 2P_{max}(d) - \sum_{c\in C} P(d,c)\\&- 2P_{max}(d) + \sum_{c\in C} P(d,c) = 0.\end{aligned}$$

Hence such an S will not maximize T_S. Consider what happens if $\#S = 2$.

$$\begin{aligned} T_S &= 2\sum_{c\in S} P(d,c) - \#S\sum_{c\in C} P(d,c) \\ &= 2\sum_{c\in S} P(d,c) - 2\sum_{c\in C} P(d,c) \\ &= -2\sum_{c\in S^c} P(d,c). \end{aligned}$$

Then,

$$T_\Delta - T_S = 2P_{\max}(d) - \sum_{c\in C} P(d,c) + 2\sum_{c\in S^c} P(d,c).$$

By assumption,

$$P_{\max}(d) > \frac{1}{2}\sum_{c\in C} P(d,c)$$

so that,

$$2P_{\max}(d) - \sum_{c\in C} P(d,c) > 0.$$

Since for any $c\in C$,

$$P(d,c) \geq 0, \quad 2\sum_{c\in S^c} P(d,c) \geq 0.$$

Hence $T_\Delta - T_S > 0$ and $S\ \#S = 2$ will not maximize T_S.

Lemma 9. Suppose

$$\exists c\in C \ni P(d,c) = \frac{1}{2}\sum_{c'\in C} P(d,c').$$

Then over all possible subsets $\Delta(d)$ of C,

$$T_\Delta = 2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c)$$

is maximal if and only if $\Delta(d)$ is any subset of $\{c \mid P(d,c) \geq P(d,c')\forall c\in C\}$.

Proof. Let $\Delta(d)$ be any given subset of $\{c \mid P(d,c) \geq P(d,c')\forall c'\in C\}$. Let S be any subset of $\{c \mid P(d,c) \geq P(d,c')\forall c'\in C\}$. Consider

$$\begin{aligned} T_\Delta - T_S &= 2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c) \\ &\quad - \left\{2\sum_{c\in S} P(d,c) - \#S\sum_{c\in C} P(d,c)\right\} \\ &= [\#S - \#\Delta(d)]\sum_{c\in C} P(d,c) + 2\left[\sum_{c\in\Delta(d)} P(d,c) - \sum_{c\in S} P(d,c)\right] \\ &= [\#S - \#\Delta(d)]\sum_{c\in C} P(d,c) \\ &\quad + 2[\#\Delta(d)P_{\max}(d) - \#SP_{\max}(d)] \\ &= [\#S - \#\Delta(d)]\sum_{c\in C} P(d,c) \\ &\quad + 2P_{\max}(d)[\#\Delta(d) - \#S] \\ &= [\#S - \#\Delta(d)]\left[\sum_{c\in C} P(d,c) - 2P_{\max}(d)\right] \end{aligned}$$

But

$$P_{\max}(d) = \frac{1}{2}\sum_{c\in C} P(d,c)$$

so that

$$\sum_{c\in C} P(d,c) - 2P_{\max}(d) = 0.$$

Hence $T_\Delta - T_S = 0$ and $\Delta(d)$ and S are equivalent with respect to criterion T.

Let S be any subset of C not a subset of $\{c \mid P(d,c) \geq P(d,c')\forall c'\in C\}$. Then $\exists \overset{*}{c}\in S \ni P(d,\overset{*}{c}) < P_{\max}(d)$ so that

$$\sum_{c\in S} P(d,c) < \#SP_{\max}(d).$$

Consider

$$\begin{aligned} T_\Delta - T_S &= 2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c) \\ &\quad - \left\{2\sum_{c\in S} P(d,c) - \#S\sum_{c\in C} P(d,c)\right\} \\ T_\Delta - T_S &= 2\#\Delta(d)P_{\max}(d) - \#\Delta(d)\sum_{c\in C} P(d,c) \\ &\quad - \left\{2\sum_{c\in S} P(d,c) - \#S\sum_{c\in C} P(d,c)\right\} \\ &= \#\Delta(d)\left[2P_{\max}(d) - \sum_{c\in C} P(d,c)\right] \\ &\quad - \left[2\sum_{c\in S} P(d,c) - \#S\sum_{c\in C} P(d,c)\right] \\ &> \#\Delta(d)\left[2P_{\max}(d) - \sum_{c\in C} P(d,c)\right] \\ &\quad - \left[2\#SP_{\max}(d) - \#S\sum_{c\in C} P(d,c)\right] \\ &> \#\Delta(d)\left[2P_{\max}(d) - \sum_{c\in C} P(d,c)\right] \\ &\quad - \#S\left[2P_{\max}(d) - \sum_{c\in C} P(d,c)\right] \\ &\quad [\#\Delta(d) - \#S]\left[2P_{\max}(d) - \sum_{c\in C} P(d,c)\right] \end{aligned}$$

But

$$P_{\max}(d) = \frac{1}{2}\sum_{c\in C} P(d,c)$$

so that

$$2P_{\max}(d) - \sum_{c\in C} P(d,c) = 0.$$

Hence $T_\Delta - T_S > 0$.

Lemma 10. Assumming

$$\#(\Sigma\bigcirc\Delta)\cap T \approx \sum_{c\in C}\sum_{d\in D}\sum_{u\in T^{-1}(c)\cap\Sigma^{-1}(d)} \frac{\#\Delta(d)\cap\{c\}}{\#\Sigma(u)}$$

$$\#(\Sigma\bigcirc\Delta)T^c \approx \sum_{d\in D}\sum_{c\in C}\sum_{\substack{c'\in C\\ c'\neq c}}\sum_{u\in T^{-1}(c')\cap\Sigma^{-1}(d)} \frac{\#\Delta(d)\cap\{c\}}{\#\Sigma(u)},$$

then

$$\begin{aligned} &\#(\Sigma\bigcirc\Delta)\cap T - \#(\Sigma\bigcirc\Delta)\cap T^c \\ &\qquad \approx \sum_{d\in D}\left[2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c)\right], \end{aligned}$$

where

$$P(d,c) = \sum_{u\in T^{-1}(c)\cap\Sigma^{-1}(d)} \frac{\#\Delta(d)\cap\{c\}}{\#\Sigma(u)}.$$

Proof.

$$\begin{aligned} \#(\Sigma\bigcirc\Delta)\cap T &\approx \sum_{d\in D}\sum_{c\in\Delta(d)}\left[\sum_{u\in T^{-1}(c)\cap\Sigma^{-1}(d)} \frac{1}{\#\Sigma(u)}\right] \\ &\approx \sum_{d\in D}\sum_{c\in\Delta(d)} P(d,c) \\ \#(\Sigma\bigcirc\Delta)\cap T^c &\approx \sum_{d\in D}\sum_{c\in C}\#\Delta(d)\cap\{c\}\sum_{\substack{c'\in C\\ c'\neq C}} \\ &\qquad \sum_{u\in T^{-1}(c')\cap\Sigma^{-1}(d)}\left(\frac{1}{\#\Sigma(u)}\right) \\ &\approx \sum_{d\in D}\sum_{c\in\Delta(d)}\sum_{\substack{c'\in C\\ c'\neq c}} P(d,c') \\ &\approx \sum_{d\in D}\sum_{c\in\Delta(d)}\left(\sum_{c'\in C} P(d,c') - P(d,c)\right) \end{aligned}$$

$$\approx \sum_{d\in D}\sum_{c\in\Delta(d)}\sum_{c'\in C} P(d,c') - \sum_{d\in D}\sum_{c\in\Delta(d)} P(d,c)$$

$$\approx \sum_{d\in D}\left[\#\Delta(d)\sum_{c\in C} P(d,c) - \sum_{c\in\Delta(d)} P(d,c)\right].$$

Hence $\#(\Sigma \circ \Delta) \cap T - \#(\Sigma \circ \Delta) \cap T^c$

$$\approx \sum_{d\in D}\left\{\sum_{c\in\Delta(d)} P(d,c) - \left[\#\Delta(d)\sum_{c\in C} P(d,c) - \sum_{c\in\Delta(d)} P(d,c)\right]\right\}$$

$$\approx \sum_{d\in D}\left[2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c)\right].$$

Lemma 11. Let $R \subseteq U \times U$, $T \subseteq U \times C$ be single valued and defined everywhere, $\Sigma \subseteq U \times D$. Suppose $(u,c)\in T$ implies $\#\{d \in (R \circ \Sigma)(u) \mid \#(R \circ \Sigma)^{-1}(d) \cap T^{-1}(c) > \frac{1}{2}\#(R \circ \Sigma)^{-1}(d)\} > \frac{1}{2}\#(R \circ \Sigma)(u)$. Define $\hat{T} = \{(u,c) \mid \#\Delta^{-1}(c) \cap (R \circ \Sigma)(u) > \frac{1}{2}\#(R \circ \Sigma)(u)\}$. Then $\Delta = \{(d,c) \mid \#(R \circ \Sigma^{-1})(d) \cap T^{-1}(c) > \frac{1}{2}\#(R \circ \Sigma)(d)\}$ implies $\hat{T} = T$.

Proof. Define $A(u) = \{(d,c) \mid (d,c)\in\Delta$ and $(u,d)\in R \circ \Sigma\}$. Suppose $(u,c)\in\hat{T}$. Then $\#\Delta^{-1}(c) \cap (R \circ \Sigma)(u) > \frac{1}{2}\#(R \circ \Sigma)(u)$. But $\#\Delta^{-1}(c) \cap (R \circ \Sigma)(u) = \#A(u) \cap (D \times \{c\})$. Also notice that $\#A(u) \le \#(R \circ \Sigma)(u)$. Hence $\#A(u) \cap (D \times \{c\}) > \frac{1}{2}\#(R \circ \Sigma)(u) \ge \frac{1}{2}\#A(u)$. This has to imply $(u,c)\in T$ for if not then $(u,c')\in T$ and $c' \ne c$. Since $(u,c')\in T$, $\#A(u) \cap (D \times \{c'\}) > \frac{1}{2}\#(R \circ \Sigma)(u) \ge \frac{1}{2}\#A(u)$. Thus $\#A(u) \ge \#A(u) \cap (D \times \{c,c'\}) > \#A(u)$, a contradiction. Hence, $(u,c)\in T$. Suppose $(u,c)\in T$. Then $\#A(u) \cap (D \times \{c\}) > \frac{1}{2}\#(R \circ \Sigma)(u)$ or $\#\Delta^{-1}(c) \cap (R \circ \Sigma)(u) > \frac{1}{2}\#(R \circ \Sigma)(u)$ so that $(u,c)\in\hat{T}$.

Lemma 12. If

$$\#T \cap \hat{T} - \#T^c \cap \hat{T} \approx \sum_{u\in U} \frac{\sum_{d\in D}\sum_{u'\in R(u)\cap\Sigma^{-1}(d)} \#\Delta(d) \cap T(u)}{\sum_{u'\in R(u)} \#\Sigma(u')} - \sum_{u\in U} \frac{\sum_{d\in D}\sum_{u'\in R(u)\cap\Sigma^{-1}(d)} (\#\Delta(d) \cap T^c(u))}{\sum_{u'\in R(u)} \#\Sigma(u')},$$

then

$$\#T \cap \hat{T} - \#T^c \cap \hat{T} \approx \sum_{d\in D}\left[2\sum_{c\in\Delta(d)} P(d,c) - \sum_{c\in C} P(d,c)\right],$$

where

$$P(d,c) = \sum_{u\in T^{-1}(c)} \left(\frac{1}{\sum_{u'\in R(u)} \#\Sigma(u')}\right) \#R(u) \cap \Sigma^{-1}(d)$$

Proof.

$$\#T \cap \hat{T} - \#T^c \cap \hat{T}$$

$$\approx \sum_{u\in U} \frac{1}{\sum_{u'\in R(u)} \#\Sigma(u')} \sum_{d\in D}\sum_{u'\in R(u)\cap\Sigma^{-1}(d)} (\#\Delta(d) \cap T(u) - \#\Delta(d) \cap T^c(u))$$

$$\approx \sum_{c\in C}\sum_{u\in T^{-1}(c)} \frac{1}{\sum_{u'\in R(u)} \#\Sigma(u')} \sum_{d\in D}\sum_{u'\in R(u)\cap\Sigma^{-1}(d)} [\#\Delta(d) \cap T(u) - \#\Delta(d) \cap T^c(u)]$$

$$\approx \sum_{c\in C}\sum_{u\in T^{-1}(c)} \frac{1}{\sum_{u'\in R(u)} \#\Sigma(u')} \sum_{d\in D}\sum_{u'\in R(u)\cap\Sigma^{-1}(d)} [\#\Delta(d) \cap \{c\} - \#\Delta(d) - \{c\}]$$

$$\approx \sum_{d\in D}\sum_{c\in\Delta(d)}\sum_{u\in T^{-1}(c)} \frac{1}{\sum_{u'\in R(u)} \#\Sigma(u')} \#R(u) \cap \Sigma^{-1}(d)$$

$$- \sum_{c\in C}\sum_{u\in T^{-1}(c)} \frac{1}{\sum_{u'\in R(u)} \#\Sigma(u')} [\#\Delta(d)]\#R(u) \cap \Sigma^{-1}(d)$$

$$+ \sum_{c\in\Delta(d)}\sum_{u\in T^{-1}(c)} \frac{1}{\sum_{u'\in R(u)} \#\Sigma(u')} \#R(u) \cap \Sigma^{-1}(d)$$

Letting

$$P(d,c) = \sum_{u\in T^{-1}(c)} \frac{1}{\sum_{u'\in R(u)} \#\Sigma(u')} \#R(u) \cap \Sigma^{-1}(d),$$

$$\#T \cap \hat{T} - \#T^c \cap \hat{T} \approx \sum_{d\in D}\left[2\sum_{c\in\Delta(d)} P(d,c) - \#\Delta(d)\sum_{c\in C} P(d,c)\right].$$

About the Author—ROBERT MARTIN HARALICK was born in Brooklyn, New York on 30 September 1943. He received the B.S. degree from the University of Kansas in 1966. He has worked with Autonetics and IBM. In 1965 he joined the Center for Research, University of Kansas as a research engineer and in 1969, when he completed his Ph.D. at The University of Kansas, he joined the faculty of the Electrical Engineering Department there. He is now an Associate Professor of Electrical Engineering at K.U. Dr. Haralick has done research in pattern recognition, multi-image processing, remote sensing, texture analysis, data compression, clustering, artificial intelligence, and general systems theory. He has developed KANDIDATS (Kansas Digital Image Data System) a multi-image processing package to run on a PDP-15 computer.

He is a member of the Institute of Electrical and Electronic Engineers, Association for Computer Machinery, Sigma Xi, Pattern Recognition Society, Society for General Systems Research and New York Academy of Sciences.