N-Tuple Method

Robert M. Haralick

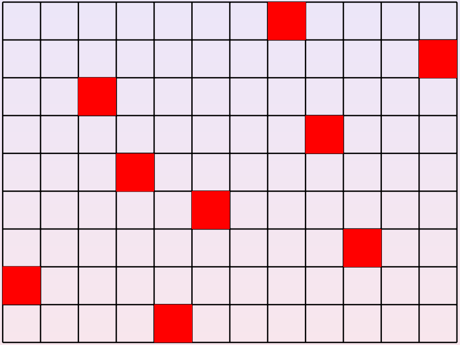
Computer Science, Graduate Center
City University of New York

Outline

N-Tuple Method

- Developed For Printed Character Recognition (1959)
- Each character is contained in an image of $M \times N$ pixels
- Each pixel is a binary 1 or a binary 0
- Designed for table lookup hardware

N-Tuple Method



L randomly chosen pixel positions

N-Tuple Method

- A small number of pixel positions are randomly selected
- Have multiple sets of such randomly selected pixel positions
- Each of the pixel positions has been thresholded and contains a binary 0 or a binary 1
- Concatenate all the binary values to form a binary number
- Use this number to access an address in a memory array
- For each character class
  - Have as many memory arrays as there are different randomly selected position sets

N-Tuple Method

- M pattern sets of L randomly selected pixel positions
- A printed character produces M L-digit binary numbers $b_1, \ldots, b_M$
- K character classes
- T_{mk} lookup table for pattern set m and class k
- $T_{mk}(b_m)$ holds the fraction of times a character in the training set of class k has the binary number b_m for the m^{th} pattern set
- Compute
  - $f_k = \min_{m=1}^{M} T_{mk}(b_m)$
  - $f_k = \sum_{m=1}^{M} T_{mk}(b_m)$
- Assign the character to unique class c_k, if there is one, for which $f_k > 0$ is highest
- Otherwise reserve decision

N-Tuple Method Recast

- M pattern sets of L randomly selected pixel positions
- A printed character produces M binary numbers $b_1, \ldots, b_M$
- K character classes
- T_m lookup table for pattern set m
- $T_m(b_m)$ holds the set of classes associated with the binary number b_m for the m^{th} pattern set
- Compute
  - $f = \cap_{m=1}^{M} T_m(b_m)$
  - Assign the character to unique class c_k, if there is one, where $c_k \in f$ and $|f| = 1$
  - Otherwise reserve decision

Relations

- Each of the possible pixel positions is a variable
- Let $X_1, \ldots, X_N$ be the N variables
- Let L_n be the possible values variable X_n can take
- Let R be the training set for one class

$$R \subseteq \bigtimes_{n=1}^{N} L_n$$

Index Sets

Definition

$I = \{i_1, \ldots, i_K\}$ is an index set if and only if for a given partial ordering $\leq$ on I, $i_1 < i_2 < \cdots < i_K$.

Definition

If $I = \{i_1, \ldots, i_K\}$ is an index set, we define the Cartesian product

$$\bigtimes_{i \in I} L_i = \bigtimes_{k=1}^{K} L_{i_k} = L_{i_1} \times L_{i_2} \times \cdots \times L_{i_K}$$

We adopt a convention that if the index set $I = \emptyset$, then $\bigtimes_{i \in I} L_i = \emptyset$.

N-ary Relation

Definition

If I is an index set with $|I| = N$ and $R \subseteq \bigtimes_{i \in I} L_i$, then we say (I, R) is an Indexed N-ary Relation on the range sets indexed by I. We also say that (I, R) has dimension N.

We can perform the usual set operations of union and intersection with the structures (I, R) and (J, S) and when $I = J$:

$$\begin{aligned}(I, R) \cup (I, S) &= (I, R \cup S) \\ (I, R) \cap (I, S) &= (I, R \cap S)\end{aligned}$$

Subset

The subset relation also has the usual meaning.

If $R \subseteq \bigtimes_{i \in I} L_i$ and $S \subseteq \bigtimes_{i \in I} L_i$, then

$$R \subseteq S \text{ if and only if } (I, R) \subseteq (I, S)$$

Index Function

Definition

Let J and M be index sets with

- $J = \{j_1, \ldots, j_{|J|}\}$
- $M = \{m_1, \ldots, m_{|M|}\}$
- $J \subset M$

The index function $f_{JM} : J \to [|M|]$ is defined by

$$f_{JM}(j) = k \text{ where } m_k = j$$

f_{JM} operates on an index j from the smaller index set and specifies where – place k – in the larger index set that the index j can be found; thus $m_k = j$.

Index Function Example

- $J = \{2, 5, 8\}$
- $M = \{2, 4, 5, 6, 7, 8, 9\} = \{m_1, m_2, m_3, m_4, m_5, m_6, m_7\}$
- Index function f_{JM}

f_{JM} operates on an index j from the smaller index set and specifies where – place k – in the larger index set that the index j can be found; thus $m_k = j$.

j	$f_{JM}(j)$
2	1
5	3
8	6

Projection

Definition

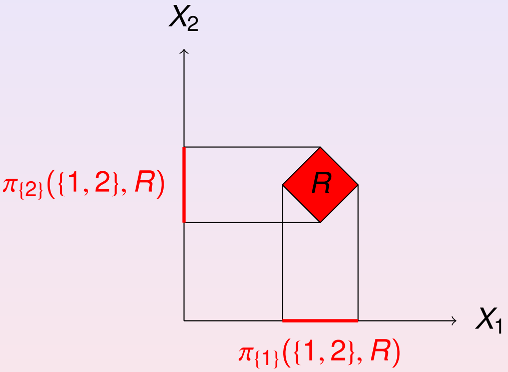
Let I and J be index sets with $I \subset J$. The projection operator projecting a relation on the range sets indexed by J onto the range sets indexed by I is defined by

$$\begin{aligned} \pi_I(J, R) &= (I, S) \text{ where} \\ S &= \{(x_1, \ldots, x_{|I|}) \in \bigtimes_{i \in I} L_i \mid \exists (a_1, \ldots, a_{|J|}) \in R,\ a_{f_{IJ}(i)} = x_i, i \in I\} \end{aligned}$$

We overload the definition of the projection operator so that it can operate on tuples as well as sets.

$$\begin{aligned} \pi_I(J, (a_1, \ldots, a_{|J|})) &= (I, (x_1, \ldots, x_{|I|}), \text{ where} \\ & (x_1, \ldots, x_{|I|}) = (a_{f_{IJ}(1)}, \ldots, a_{f_{IJ}(|I|)}) \\ &= (I, (a_{f_{IJ}(1)}, \ldots, a_{f_{IJ}(|I|)})) \end{aligned}$$

Projection



Projections

Proposition

Let I and J be an index sets with $I \subseteq J$. If $(J, A) \subseteq (J, B)$, then $\pi_I(J, A) \subseteq \pi_I(J, B)$.

Projections of Unions

Proposition

Let A and B be relations on the range sets indexed by index set J. Let $I \subset J$. Then

$$\pi_I(J, A \cup B) = \pi_I(J, A) \cup \pi_I(J, B)$$

Projections of Intersections

Proposition

Let A and B be relations on the range sets indexed by index set J. Let $I \subset J$. Then

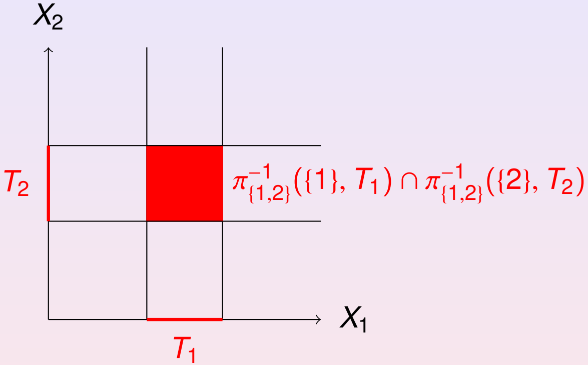
$$\pi_I(J, A \cap B) \subseteq \pi_I(J, A) \cap \pi_I(J, B)$$

The following example shows that this proposition cannot be made stronger.

- $J = \{1, 2\}$,$I = \{1\}$
- $A = \{(a, \alpha), (b, \beta)\}$, $B = \{(a, \gamma), (b, \delta)\}$

Then $A \cap B = \emptyset$ so that $\pi_I(J, A \cap B) = \emptyset$. But $\pi_I(J, A) = \{a, b\}$ and $\pi_I(J, B) = \{a, b\}$ so that $\pi_I(J, A) \cap \pi_I(J, B) = \{a, b\}$.

Inverse Projection



$$(\{1,2\}, T_1 \times T_2) = \pi^{-1}_{\{1,2\}}(\{1\}, T_1) \cap \pi^{-1}_{\{1,2\}}(\{2\}, T_2)$$

Inverse Projection

Definition

Let

- I and J be index sets with $I \subset J$
- $R \subseteq \bigtimes_{i \in I} L_i$

The Inverse Projection of (I, R) with respect to index set J is defined by

$$\pi_{IJ}^{-1}(I, R) = \{(J, (a_1, \ldots, a_{|J|})) \in (J, \bigtimes_{j \in J} L_j) \quad | \quad \pi_I(J, (a_1, \ldots, a_{|J|})) \in (I, R)\}$$

Inverse Projections

Proposition

Let I and J be index sets with $I \subseteq J$. Suppose $(I, R) \subseteq (I, S)$. Then,

$$\pi_{IJ}^{-1}(I, R) \subseteq \pi_{IJ}^{-1}(I, S)$$

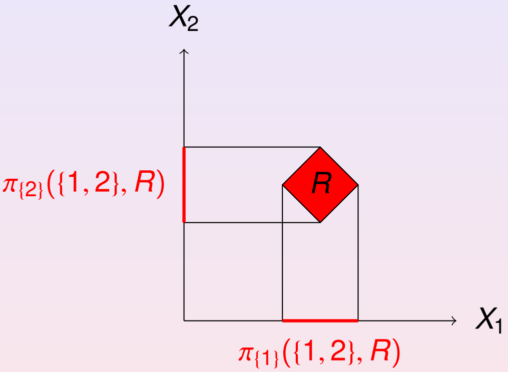
Projection and Inverse Projections

Proposition

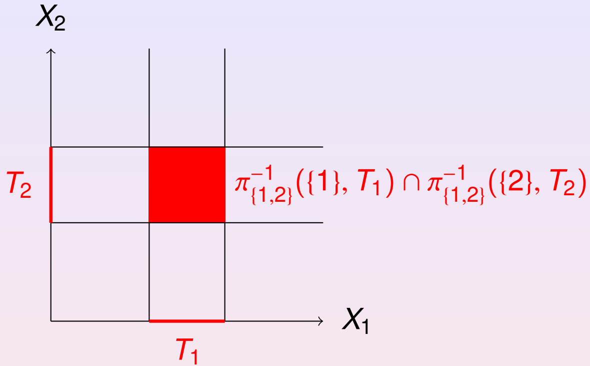
Let I, J be index sets with $I \subseteq J$. Then

$$\begin{aligned} \pi_I(\pi_{IJ}^{-1}(I, R)) &= (I, R) \\ \pi_{IJ}^{-1}(\pi_I(J, R)) &\supseteq (J, R) \end{aligned}$$

Projection



Inverse Projection



$$(\{1,2\}, T_1 \times T_2) = \pi^{-1}_{\{1,2\}}(\{1\}, T_1) \cap \pi^{-1}_{\{1,2\}}(\{2\}, T_2)$$

$$\pi_{\{1\}}\pi^{-1}_{\{12\}}(\{1\}, T_1) = (\{1\}, T_1)$$

Projections and Inverse Projections

Corollary

Let I and J be index sets with $I \subseteq J$. Suppose $\pi_I(J, S) = (I, R)$. Then $\pi_{JI}^{-1}(I, R) \supseteq (J, S)$.

Corollary

Let $J_1, \ldots, J_K, J$ be index sets with $J_1, \ldots, J_K \subseteq J$. Then

$$(J, R) \subseteq \pi_{J_1 J}^{-1} \pi_{J_1}(J, R) \cap \pi_{J_2 J}^{-1} \pi_{J_2}(J, R) \cap \cdots \cap \pi_{J_K J}^{-1} \pi_{J_K}(J, R)$$

Projections of a Composite Relation

Proposition

Let I, J, K be index sets with $I, J \subseteq K$. Suppose

$$(K, R) = \pi_{IK}^{-1}(I, S) \cap \pi_{JK}^{-1}(J, T)$$

Then

- $\pi_I(K, R) \subseteq (I, S)$
- $\pi_J(K, R) \subseteq (J, T)$

Relation Join

Definition

- Index sets: J_1, J_2
- Index set: $K = J_1 \cup J_2$
- Relation: $R_1 \subseteq \bigtimes_{j \in J_1} L_j$
- Relation: $R_2 \subseteq \bigtimes_{j \in J_2} L_j$
- Relation: $S \subseteq \bigtimes_{k \in K} L_k$

(K, S) is the Relation Join of (J_1, R_1) and (J_2, R_2) if and only if

$$(K, S) = \pi^{-1}_{J_1 K}(J_1, R_1) \cap \pi^{-1}_{J_2 K}(J_2, R_2)$$

We write:

$$(K, S) = (J_1, R_1) \otimes (J_2, R_2)$$

Join Characterization Theorem

Theorem

Let I, J, K *be index sets with* $K = I \cup J$*. Let* $R \subset \bigtimes_{i \in I} L_i$ *and* $S \subset \bigtimes_{j \in J} L_j$ *Then* $(I, R) \otimes (J, S) = (K, T)$ *where*

$$T = \{(t_1, \ldots, t_{|K|}) \in \bigtimes_{k \in K} L_k \quad | \quad \pi_I(K, (t_1, \ldots, t_{|K|}))\} \in (I, R) \text{ and } \pi_J(K, (t_1, \ldots, t_{|K|})) \in (J, S)\}$$

Relation Join Example

	R		S
	1,3,5,6		2,3,4,5
1	(a,b,e,d)	1	(e,e,a,d)
2	(b,d,e,a)	2	(d,c,b,a)
3	(e,c,a,b)	3	(a,d,b,e)
4	(c,e,d,a)	4	(b,b,c,e)
		5	(e,d,c,e)

Then $R \otimes S$ is defined by

	$R \otimes S$
	1,2,3,4,5,6
(1,4)	(a,b,b,c,e,d)
(2,3)	(b,a,d,b,e,a)
(2,5)	(b,e,d,c,e,a)
(3,2)	(e,d,c,b,a,b)
(4,1)	(c,e,e,a,d,a)

Join Properties

Proposition

Let $I, J, K \subset [N]$ *be index sets. Let*

- $R \subseteq \bigtimes_{i \in I} L_i$
- $S \subseteq \bigtimes_{j \in J} L_j$
- $T \subseteq \bigtimes_{k \in K} L_k$

Then,

- $(I, R) \otimes (J, S) = (J, S) \otimes (I, R)$
- $((I, R) \otimes (J, S)) \otimes (K, T) = (I, R) \otimes ((J, S) \otimes (K, T))$
- $(I, R) \otimes (I, R) = (I, R)$
- $(I, R) \subseteq (I, T)$ *implies* $(I, R) \otimes (I, T) = (I, R)$

N-Tuple Method

- $[N] = \{1, \ldots, N\}$ indexes for all pixel positions
- $([N], T_c)$ training set for class c
- $J_1, \ldots, J_M$ M sets of random subsets of $[N]$, each having L elements
- $\pi_{J_m}([N], T_c) = (J_m, T_{mc}), m = 1, \ldots, M$ analogous to tables
- $r \in \bigtimes_{n=1}^{N} L_n$ is new measurement
- Assign r to unique class c, if there is one where $\pi_{J_m}([N], r) \in (J_m, T_{mc}), m = 1, \ldots, M$
- Otherwise reserve decision

N-Tuple Method

Theorem

The set of all tuples that could potentially be assigned to class c is

$$\{([N], x) \mid \pi_{J_m}([N], x) \in (J_m, T_{mc}), m = 1, \ldots, M\} = \otimes_{m=1}^{M}(J_m, T_{mc})$$

Acceptance Region

Definition

The Acceptance Region for a class is the set of all measurements that will be assigned to the class.

A_c: Acceptance Region for Class c

$$A_c = \otimes_{m=1}^{M}(J_m, T_{mc}) - \bigcup_{\{d \in C - \{c\}\}} \otimes_{m=1}^{M}(J_m, T_{md})$$

Reserve Decision Region

Definition

The Reserve Decision Region is the set consisting of measurements which do not belong to any acceptance region.

R: Reserve Decision Region

$$R = \bigtimes_{n=1}^{N} L_n - \bigcup_{c \in C} A_c$$

Marginal Class Conditional Probabilities

Let the training set for class $c \in C$ be $([N], R_c)$, where

$$R_c \subseteq \bigtimes_{n=1}^{N} L_n$$

Estimate marginal class conditional probabilities

$$P_{mc}((J_{mc}, x) \mid c) = \frac{|\{r \in R_c \mid \pi_{J_{mc}}(r) = x\}|}{|R_c|}$$

Acceptance region for class c

$$\begin{aligned} A_c &= \{x \mid P_{mc}((J_{mc}, x) \mid c) > 0, m = 1, \ldots, M \text{ and for every } c' \neq c, \\ &\quad \text{there exists } m' \text{ such that } P_{m'c'}((J_{m'c'}, x) \mid c) = 0\} \end{aligned}$$

Aggregating Evidence

- Let A and B be two events
- Let $P(A)$ be the probability of the event A
- Let $P(B)$ be the probability of the event B
- What can we say about the probability $P(A \cap B)$?

Proposition

Let A and B be events. Then

$$P(A \cap B) \leq \min\{P(A), P(B)\}$$

Proof.

Since $A \cap B \subseteq A$, $P(A \cap B) \leq P(A)$
Since $A \cap B \subseteq B$, $P(A \cap B) \leq P(B)$
Therefore $P(A \cap B) \leq \min\{P(A), P(B)\}$ □

Aggregating Evidence

Let A and B be events with probabilities a and b respectively.

	b	$1 - b$
a	$\min\{a, b\}$	$a - \min\{a, b\}$
$1 - a$	$b - \min\{a, b\}$	$1 + \min\{a, b\} - b - a$

Aggregating Evidence

Suppose $a \geq b$. Then $\min\{a, b\} = b$

	b	$1 - b$
a	$\min\{a, b\}$	$a - \min\{a, b\}$
$1 - a$	$b - \min\{a, b\}$	$1 + \min\{a, b\} - b - a$

	b	$1 - b$
a	b	$a - b$
$1 - a$	0	$1 - a$

Aggregating Evidence

Suppose $a \leq b$. Then $\min\{a, b\} = a$

	b	$1 - b$
a	$\min\{a, b\}$	$a - \min\{a, b\}$
$1 - a$	$b - \min\{a, b\}$	$1 + \min\{a, b\} - b - a$

	b	$1 - b$
a	a	0
$1 - a$	$b - a$	$1 - b$

Aggregating Evidence

Suppose $P(A \cap B) = 0$

	b	$1 - b$
a	0	a
$1 - a$	b	$1 - b - a$

But this requires that $a + b < 1$

Aggregating Evidence

Suppose $a + b > 1$

	b	$1 - b$
a	$a + b - 1$	$1 - b$
$1 - a$	$1 - a$	0

Proposition

Let A and B be events. Then

$$\max\{0, P(A) + P(B) - 1\} \leq P(A \cap B) \leq \min\{P(A), P(B)\}$$

Aggregating Evidence

Corollary

Let $A_1, \ldots A_N$ be events. Then for any m, n,

$$\max\{0, P(A_m) + P(A_n) - 1\} \leq P(A_m \cap A_n)\} \leq \min\{P(A_m), P(A_n)\}$$

Corollary

$$P(\cap_{n=1}^{N} A_n) \leq \min_{n=1,\ldots,N} P(A_n)$$

Aggregating Evidence

If the gain matrix is the identity matrix, then

Assign $x = (x_1, \ldots, x_K)$ to category c when

$$P(x_1, \ldots, x_K \mid c)P(c) \quad \geq \quad P(x_1, \ldots, x_K \mid c')P(c'), \text{ for every } c'$$

This motivates, Assign $x = (x_1, \ldots, x_K)$ to category c when

$$\min_k P_k(x_k \mid c)P(c) \quad \geq \quad \min_k P_k(x_k \mid c')P(c'), \text{ for every } c'$$

This motivates, Assign $([N], x)$ to category c when

$$\min_k P_k(\pi_{I_k}([N], x) \mid c)P(c) \geq \min_k P_k(\pi_{I_k}([N], x) \mid c')P(c')$$

Optimization Problem

Find the index sets $I_1, \ldots, I_K \subset I$, $|I_k| = \theta$ to optimize the classification accuracy

- Corresponding to each I_k is $P_k(\pi_{I_k}([N], x) \mid c)$
- Each class conditional P_k has entropy $E_k(c)$
- Find I_k's for which $P_k's$ have small entropy $E_k(c)$
- Small entropy means
    - There are many small probability values
    - A few large probability values
- With many small probability values, the upper bounds must move lower

Optimization Solution

- Go through all subsets of $[N]$ having size θ
- Evaluate the entropy of each of the corresponding class conditional probabilities
- Take the K subsets $I_1, \ldots, I_K$ whose class conditional probabilities have smallest entropy

Possible Final Project: Probability Generation

- Number of dimensions $N = 100$
- Range set for each dimension is the same
  - $L = \{0, 1, \ldots, 9\}$, $|L| = 10$
- Measurement Space is L^N, $|L^N| = 10^{100}$
  - Estimated number of elementary particles in the observable universe is 10^{85}
- Number of classes $|C| = 2$
- Choose equal prior class probabilities
- Number of dimensions of each projection index set is $\theta = 4$
- Each table corresponding to a projection has $10^4 = 10,000$ locations
- Number of projection index sets $K = 200$
- Memory associated with each class is then $200 \times 10^4 = 2 \times 10^6$
- Select at random K, N choose θ combinations
  - Try to make $|\{k \mid n \in I_k\}| >= 3, n = 1, \ldots, N$

Possible Final Project: Probability Generation

- For each I_k there are 10,000 probabilities
  - $P_{I_k}(\pi_{I_k}([N], x) \mid c_1)$
  - $P_{I_k}(\pi_{I_k}([N], x) \mid c_2)$
  - $P_{I_k}(\pi_{I_k}([N], x)$

$$P_{I_k}(\pi_{I_k}([N], x) = P_{I_k}(\pi_{I_k}([N], x) \mid c_1)P(c_1) + P_{I_k}(\pi_{I_k}([N], x) \mid c_2)P(c_2)$$

$$\sum_{\pi_{I_k}([N], x)} P_{I_k}(\pi_{I_k}([N], x) = 1$$

- Choose probability values so that
  - $P_{I_k}(\pi_{I_k}([N], x) \mid c_1)$ has small entropy
  - $P_{I_k}(\pi_{I_k}([N], x) \mid c_2)$ has small entropy
  - $P_{I_k}(\pi_{I_k}([N], x)$ has large entropy

Possible Final Project: Labelled Data Set Generation

- Choose at random 10^8 100-tuples
- Label each tuple with its ground truth category
- Label tuple $([N], x)$ with category c when

$$\min_k P_k(\pi_{I_k}([N], x) \mid c)P(c) \geq \min_k P_k(\pi_{I_k}([N], x) \mid c')P(c')$$

Possible Final Project: Decision Rule 1

- Choose at random K, N choose θ combinations
- Use the first half of the labeled data set to estimate $P_{I_k}(\pi_{I_k}([N], x) \mid c)$ (about 70,000 probabilities)
- Use the estimated probabilities $P_{I_k}(\pi_{I_k}([N], x) \mid c)$ as the basis of a decision rule
- The Training Set: Apply the decision rule based on the estimated probabilities on the first half of the data set
  - Use the ground truth labels to calculate the confusion matrix and probability of correct classification
- The Test Set: Apply the decision rule based on the estimated probabilities on the second half of the data set
  - Use the ground truth labels to calculate the confusion matrix and probability of correct classification
- Repeat 100 times and calculate the mean and standard deviation of the probability of correct classification for the training data and the test data

Alternate Final Project: Decision Rule 2

- Go through all subsets of $[N]$ having size θ
- Training Set: Evaluate the entropy of each of the corresponding class conditional probabilities using the first half of the data set
- Take the K subsets $I_1, \ldots, I_K$ whose class conditional probabilities have smallest entropy
- Using the selected class conditional probabilities, compute the confusion matrix and the probability of correct classification for the Training Set
- Test Set: Using the training set class conditional probabilities, compute the confusion matrix and the probability of correct classification for the Test Set
- Compare your results with those of Decision Rule 1
- Based on your results, Is Decision Rule 1 or Decision Rule 2 better