Classification in Machine Learning
via Association Schemes

Katie Brodhead
Кэти Бродхед

Florida A&M University
Tallahassee, USA

Graphs, Groups, Representations, and Relations
Novosibirsk, Russia

August 6–19, 2018

Research Program
Goals, At a High Level

Our program is real-world computer science and we work to frame the work in the context of pure mathematics.

GOAL:

(1) **Do Machine Learning:** Learn a *pattern or labeling system* for given set of data, without having the rule for the data programmed ahead of time.

(2) **Be Rigorous:** Use *mathematical structures* such as association schemes, or orbits in permutation group theory, to shed light on this process.

Example. Cell phone users often label much of their data (e.g. photos with where a picture was taken, objects/persons in the photo). How a user will label a new piece of data?

Research Program
Goals, In More Formalized Form

Specifically, let C be a set of m colors ("labels"), and suppose X is a set of n data points in some space S. Let $L : X \to C$ be labeling map, known to us.

Question: If a new piece of data point $d = x_{n+1}$ is presented, does L encode enough information to *accurately* (via a computer learning process) determine its label d?

Typically, X is small so that L cannot. So, the objective is to expand X with a minimal amount of data points, along with knowledge about their labels, obtained by query.

But then, how do we determine whether to include the queried data point (i.e. that is useful)?

NOTE

Due to limitations of time, energy, expense, a queried data point will only be selected for computer re-training/learning if it is "useful".

We shortly formalize this in terms of non-conformity measures and p-values from statistics.

We need to, and will, do this first as this development allows us to partition our data into classes that indeed form a useful association scheme, dependent on the data presented to us.

Our Setting, Terminologically

A Supervised, Active, Online-based Learning Approach

Since the labeling map $L : X \to C$ with *domain* X is known to us, our work is called **supervised learning**. In a sense, we learn under the supervision of (knowledge provided by) L.

Because we are also allowed to expand X, via query, our learning is called **active**.

That is, on query, the learner receives a data point and knowledge about its label.

So, an active learner is able to ask questions, listen, and ask further questions based on answers, in contrast to a passive learner.

The points from whence X is expanded arise from unlabeled pool.

If the learner has knowledge about the entirety of the pool, minus the labels, prior to query, the work is called pool-based.

The alternative, called **online-based**, is where a learner must decide, on-the-fly, whether to request labels for each unlabeled point in sequence without knowledge about the pool.

Our Setting, Terminologically
is via Transductive Inference

Inductive Inference - occurs when empirical data is used to find the approximation of a functional dependency
(known as the "particular-to-the-general")

Deductive Inference - occurs when an approximation is used evaluate values of a function at point of interest
(known as the "general-to-the-particular")

Transductive Inference - the values of a function are estimated at points of interest in one step, using local information (e.g. k-Nearest Neighbor method)

There is good survey on past work done according to the different terminological mixes [Ho-Wechsler, 2008].

A Non-conformity Algorithm
and its tie to Association Schemes

We use a non-conformity measure to separate data into separate classes, which ultimately comprises an association scheme.

First, we describe the algorithm at a high level.

Second, we describe the measure, in preparation to develop the association scheme.

Third, we describe the association scheme.

Lastly, we consider future work, especially in association schemes or permutation group theory.

Non-conformity Algorithm

High-level Overview

Given:

- Training Set $T_r = \{(x_1, y_1), ..., (x_n, y_n)\}$
- Unlabeled Set $U = \{(u_1, v_1), ..., (u_m, v_m)\}$
- Test Set $T_e = \{(a_1, b_1), ..., (a_k, b_k)\}$

(1) Query an unlabeled point.

(2) Determine non-conformity measure of the data point when it is assigned different labels. [See picture]

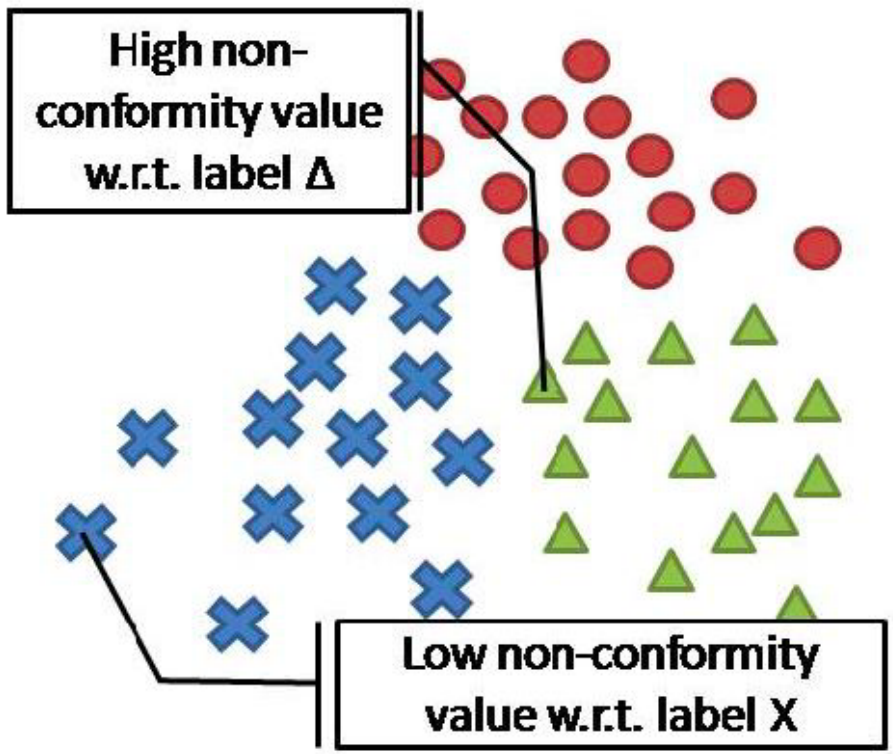
(3) If the point has a high non-conformity, include it. Reason: It contributes new information about its label.

(4) Determine the accuracy of the test set with each new included point

(5) The results should show increasing accuracy.

Nonconformity Algorithm

Pictorial view of a non-conformity measure



Note: Non-conformity measures exist for different classifiers, e.g. *k*-nearest neighbor, support vector machines.

Non-conformity Measure - Example 1

k-Nearest Neighbor (kNN)

[Vovk, et al., 2005]. Let x_i be a data point and define

D_i^y : list of sorted distances between a particular data point x_i and other data points with the same class label, say y.

D_i^{-y} : list of sorted distances between x_i and data points with label other than y.

D_{ij}^y : j^{th} shortest distance in the list of sorted distances, D_i^y.

The non-conformity measure $\nu_y(x_i)$ of data point x_i is :

$$\nu_y(x_i) = \frac{\sum_{j=1}^{k} D_{ij}^y}{\sum_{j=1}^{k} D_{ij}^{-y}}$$

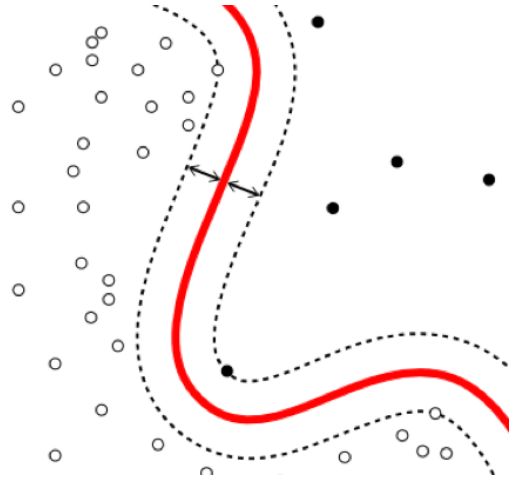
Non-conformity Measure - Example 2

Support Vector Machine (SVM)

The measure of x amounts to solving a quadratic programming problem of the form

$$\text{minimize} \quad (1/2)x^T Px + q^T x$$

$$\text{subject to} \quad Gx \leq h, Ax = b$$



A "Dual Problem" is then actually solved and the Lagrange multipliers which arise give a natural monotonic measure of non-conformity. A multiplier is zero outside the margin and lies in an interval $[0, C]$ otherwise. When $|C| > 2$, a "one-against-rest" approach is used.

High Non-Conformity

How to determine it

Assume as the so-called null-hypothesis from statistics that new point x_{n+1} it is assigned label L. Let ν_L be a non-conformity measure on X with assumption that each x_i is assigned its given labels, and x_{n+1} is assigned label L.

A p-value function P_L may be defined by

$$P_L(x_{n+1}) = \frac{\text{count}\{k : \nu_L(x_k) \geq \nu_L(x_{n+1})\}}{\text{card}(X)}$$

We obtain linear ordering $p_1 \leq ... \leq p_M$ of all p-values arising from the calculation P_y for each $y \in L$. One can use $p_M - p_{M-1} \approx 0$ [Ho-Wechsler], or even $\text{mean}_{ij}\{p_i - p_j\} \approx 0$ [Balasubramanian+] as determinations of high non-conformity, since either difference signifies low confidence.

An Association Scheme Approach
to Labeling

[Bannai-Ito, 1984]. Let X be a set of cardinality n and let R_i $(i = 0, 1, ..., d)$ be subsets of $X \times X$.

An **association scheme** of class d on X is a configuration $\mathcal{X} = (X, \{R_i\})_{0 \leq i \leq d}$ which satisfies (i)-(iv):

(i) $R_0 = \{(x, x) : x \in X\}$

(ii) $X \times X = R_0 \cup ... \cup R_d$, where $R_i \cap R_j = \emptyset$ if $i \neq j$

(iii) $(\forall i)(\exists j) R_j = \{(x, y) : (y, x) \in R_i\}$. Let R_i^t denote R_j.

(iv) $(\forall i, j, k)(\exists C)(\forall (x, y) \in R_k)$
$\{z \in X : (x, z) \in R_i \text{ and } (z, y) \in R_j\} = C$.
C is called the **intersection number** corresponding to i, j, k and denoted by p_{ij}^k.

$\mathcal{X}$ is **commutative** if $p_{ij}^k = p_{ij}^k$ for all i, j, k

$\mathcal{X}$ is **symmetric** if $R_i^t = R_i$ for every i.

An Association Scheme Approach
to Labeling

Let X be the set of data points, and $L = \{L_1, ..., L_M\}$ be set of labels for X. For ease of presentation, we write that $x \in L_i$ to mean that x has label L_i.

An association scheme arises as a disjoint union:

$$X \times X = \bigcup_{ij} R_0 \cup \hat{L}_i \cup \check{L}_i \cup C_{ij}^{\delta} \cup \overline{C_{ij}^{\delta}} \cup B_{ij}^{\delta} \cup \overline{B_{ij}^{\delta}}$$

The *basic idea* is that a pair $(x, y) \in X \times X$ of data points share a certain relationship, based on their labels and their non-conformity measures. The partition captures this.

This partitioning into classes s natural, along labels and relationships amongst labels, via non-conformity measures ν_i corresponding to labels L_i, respectively.

Further, this can be applied to any labeling system.

An Association Scheme Approach
to Labeling

A **numbering** (*a lá* Yurii L. Ershov, Sobolev Institute) of a collection C of objects is a surjective map $F : \mathbb{N} \to C$.

Fix ν_i-based numbering $n_i : L_i \to \text{card}(L_i)$ such that

$$n_i(x) < n_i(y) \implies \nu_i(x) \leq \nu_i(y).$$

We define:

$$R_0 = \{(x, x) : x \in X\}$$
$$\hat{L}_i = \{(x, y) : n_i(x) < n_i(y) \text{ and } x, y \in L_i\}$$
$$\check{L}_i = (L_i \times L_i) \setminus (\hat{L}_i \cup R_0)$$

$\hat{L}_i$ is to be understood to essentially be an "outgoing" conformity order imposed on L_i where $(x, y) \in \hat{L}_i$ means that "x" is more conform than "y".

$\check{L}_i$, then, is an "incoming" conformity imposed on L_i.

An Association Scheme Approach
to Labeling

Fix δ be a threshold of non-conformity for label L_i. That is, x_i is **very conform** with respect to label L_i if $\nu_{L_i}(x_i) \leq \delta$.

The remaining pairs (x, y) are such that x and y must necessarily have in different labels.

These are partitioned according to whether one ("B_{ij}^{δ}", "$\overline{B_{ij}^{\delta}}$"), both ("$C_{ij}^{\delta}$"), or neither ("$\overline{C_{ij}^{\delta}}$") are very conform with respect to their own labels. When a point is not very conform, it lies along a boundary B of a label. Define:

$$C_{ij} = \{(x, y) : x \in L_i \, y \in L_j, \nu_{L_i}(x) \leq \delta \text{ and } \nu_{L_j}(y) \leq \delta\}$$
$$\overline{C_{ij}^{\delta}} = \{(x, y) : x \in L_i \, y \in L_j, \nu_{L_i}(x) > \delta \text{ and } \nu_{L_j}(y) > \delta\}$$
$$B_{ij}^{\delta} = \{(x, y) : x \in L_i \, y \in L_j, \nu_{L_i}(x) \leq \delta \text{ and } \nu_{L_j}(y) > \delta\}$$
$$\overline{B_{ij}^{\delta}} = (B_{ij}^{\delta})^t$$

It is easy to verify conditions (i)-(iii) of the definition of an association scheme.

It is more challenging to see the that property of "intersection numbers" holds.

One key fact useful towards this end is the natural ordering imposed on elements, for elements of the same label. This guarantees that there is a precise fixed number of "z" whenever R_i and R_j correspond with $\hat{L}_i$ or $\check{L}_i$.

Also, whenever R_i (or similarly R_j) is not $\hat{L}_i$ or $\check{L}_i$, the number of connections that (x, z) encompasses for varying z, is also fixed. This is because each non-$\hat{L}_i$, non-$\check{L}_i$ class takes all connections to the other class, witnessing not only label information, but non-conformity information as well.

Future Work

We developed a natural framework for partitioning any labeled set of data into an association scheme.

Results in association schemes show the existence or non-existence of certain structures.

How would such structures arises in the presentation of data, and what combinatorial facts characterize these many types of structures?

It is clear that (1) a firm understanding of these and (2) when these arise in data can lead to a substantially (3) new set of tools for which to handle and predict data.

Our objective to work on each of these three key objectives as next steps.

We express our appreciation and recognize all conference organizers for their efforts, time, and attention.

We thank Akihiro Munemasa for his enthusiasm from whence I drew inspiration to consider association schemes.

References 1


V. Balasubramanian, S. Chakraborty, S. Panchanathan. *Generalized Query by Transduction for Online Active Learning*. *IEEE 12th International Conference on Computer Vision Workshops* Kyoto (2009) 1378–1385.


E. Bannai, T. Ito. *Algebraic Combinatorics I: Association Schemes*. *Benjamin/Cummings Publishing Company, Inc.* Menlo Park, CA, USA (1984).

We express our appreciation and recognize all conference organizers for their efforts, time, and attention.

References 2


S.-S. Ho and H. Wechsler. *Query by Transduction*. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(9) (2008) 1557–1571.


UCI machine learning repository. *http://archive.ics.uci.edu/ml/*. *Accessed: July 25, 2018.*


V. Vovk, A. Gammerman, G. Shafer. *Algorithmic Learning in a Random World*. *Springer-Verlag New York, Inc.* Secaucus, NJ, USA (2005)