INDUCTIVE INFERENCE MACHINES FOR MATHEMATICAL STRUCTURES

Valentina S. Harizanov

Department of Mathematics
George Washington University
Washington, D.C.
harizanv@gwu.edu
http://home.gwu.edu/~harizanv/

1 Learning theory: inductive inference

E. M. Gold (1967): "Language identification in the *limit*"

Learning is a dialogue between a *teacher* and a *learner* who is trying to learn a class of *computably enumerable* sets (languages).

- A set is computably enumerable (c.e.) if there is an *algorithm*, a *computable* function, which enumerates it by listing its elements.

- $W_0, W_1, W_2, \ldots, W_e, \ldots$

  is an *algorithmic list* of all c.e. sets indexed by their Gödel codes.

- Every c.e. set has *infinitely* many Gödel codes.

2 Computable sets

- A set is *computable* (*decidable*) if both the set and its complement are c.e.

- Algorithmic enumerations of a set and its complement, give us a decision procedure for exactly determining the elements of the set.

- Not every c.e. set is computable.

  For example, the *halting set*, K, is c.e. but not computable.

- The set K consists of all natural numbers e on which the Turing machine whose Gödel code is e halts. Equivalently, e belongs to W_e; it appears in an algorithmic enumeration for W_e.

3 Algorithmic learner

- An algorithmic learner is a Turing machine that receives more and more data of a c.e. set to be learned, and outputs a sequence of possible codes that converges to the “description” of that set.

- A learner can also be nonalgorithmic.

- (1) The class of all finite sets of natural numbers is learnable.

  (2) The previous class enlarged by adding the set of all natural numbers is not learnable.

  (3) (Gold) The class of all *total computable functions* is not learnable by an algorithmic learner.

- (Blum and Blum)

  If a learner can learn a c.e. set W, then there is a finite part of W, called a *locking sequence* for W, on which the learner locks its conjecture.

4 Different convergence criteria

The learner's sequence of hypotheses for a set W is

$$e_0, e_1, e_2, \ldots$$

- EX learning (*syntactic* convergence)
  explanatory learning

  After finitely many steps, the hypotheses are the *same* and correct:

  $e_0, e_1, e_2, \ldots, e, e, e, \ldots$

  $W = W_e$

- BC learning (*semantic* convergence)
  behaviorally correct learning

  After finitely many steps, the hypotheses are correct, and allowed to be distinct:

  $W = W_{e_n} = W_{e_{n+1}} = \ldots$

5 EX-learnability is more restrictive than BC-learnability

Let K be the halting set.

- The class of sets

  $\{K \cup D : D \text{ is finite}\}$

  is BC-learnable, but not EX-learnable.

- (Case, Smith)

  The class EX^* is properly contained in the class BC, where EX^*-learnability of functions allows finitely many anomalous inputs.

6 Positive versus negative information

- Txt
  Learning from *text*: the learner requests only *positive* information (elements of the set), and the teacher eventually provides all elements of the set W to be learned.

- Inf
  Learning from an *informant*: the learner alternately requests *positive* information and *negative* information (elements of the complement), and the teacher eventually provides every element of the set W, and every element of its complement.

- Learning from text is more restrictive than learning from informant. (For example, the collection consisting of an infinite c.e. set together with all of its finite subsets can be learned from an informant, but not from text.)

7 Switching type of information

- Sw

  Learning from *switching* type of information: the learner can request positive or negative information about W, but when the learner almost always, that is, after finitely many switches, requests information of the same type, the teacher eventually gives all elements in W (if the type is positive), or all elements in its complement (if the type is negative).

- (Jain, Stephan)

  Switching type of information provides more learning power than giving positive information only; but it is weaker than providing information from an informant.

8 General criterion for non-$SwBC$-learnability

(Harizanov, Stephan)

Let L be a class of c.e. sets. Assume that there is some set W in the class L such that for every finite set D, there are U, U' in L with

(i) $U \subset W \subset U'$,

U approximates W from below; U' approximates W from above;

(ii) $D \cap U = D \cap U'$,

U and U' coincide on D.

Then the class L cannot be $SwBC$-learned.

9 Learning classes of algebraic structures

Stephan, Ventsov (2001): "Learning *algebraic structures* from text using semantical knowledge"

Learnability of classes of substructures of a certain algebraic structure

- the structure of integers Z with addition, subtraction and multiplication (*ring*),
- an abstract computable ring,
- the structure of rationals Q with addition and subtraction (*group*).

10 Vector space V_∞

- V_∞ is a "big" computable infinite-dimensional vector space whose scalars are the rationals.

- Can think of the elements of V_∞, the vectors, as infinite sequences of rational numbers with only finitely many nonzero components.

- Pointwise vector addition and scalar multiplication, for example:

  $(1, \frac{1}{2}, -\frac{2}{3}, 0, 0, ...) + (-\frac{1}{2}, 2, 0, 0, 0, ...) =$
  $(\frac{1}{2}, \frac{5}{2}, -\frac{2}{3}, 0, 0, ...)$

- $6(1, \frac{1}{2}, -\frac{2}{3}, 0, 0, ...) = (6, 3, -4, 0, 0, ...)$

- $V_0, V_1, V_2, \ldots$
  is an *algorithmic list* of all c.e. subspaces of V_∞ indexed by their Gödel codes.

11 Dependence algorithm

- $6(1, \frac{1}{2}, -\frac{2}{3}, 0, 0, ...) + 2(-\frac{1}{2}, 2, 0, 0, 0, ...) = (5, 7, -4, 0, 0, ...)$

  The three vectors are linearly dependent.

- $(1, 0, 0, 0, ...)$,

  $(0, 1, 0, 0, ...)$,

  $(0, 0, 1, 0, ...)$,

  ...

  standard (computable) basis for V_∞

- A *basis* consists of the maximal set of linearly independent vectors.

- All bases have the same size, called the *dimension* of the space.

12 Computably enumerable vector spaces

- (Metakides-Nerode)

  Any c.e. (computable) vector space is isomorphic to the *quotient space* $\frac{V_\infty}{V}$, where V is a c.e. (computable) subspace of V_∞.

- In $\frac{V_\infty}{V}$, the equality of vectors is *modulo* V.

  That is, two vectors u and w are equal if their difference $u-w$ is 0 *modulo* V, that is, it belongs to V.

- Thus, the class of all c.e. subspaces (substructures) of $\frac{V_\infty}{V}$ can be viewed as the class $L(V)$ of the c.e. subspaces of V_∞ that contain V, the superspaces of V.

13 Characterizing $TxtEX$-learnable classes $L(V)$

(Harizanov, Stephan)

The following statements are equivalent for a c.e. subspace V of V_∞.

(i) The dimension of $\frac{V_\infty}{V}$ is finite.

(ii) The class $L(V)$ is $TxtEX$-learnable.

(iii) The class $L(V)$ is $TxtBC$-learnable.

(iv) The class $L(V)$ is $SwEX$-learnable.

What if the dimension of $\frac{V_\infty}{V}$ is infinite?

14 Thin vector spaces in computable algebra

Let V be a c.e. subspace of V_∞ such that the dimension of $\frac{V_\infty}{V}$ is infinite.

Let k be a natural number, possibly 0.

The space V is called *k-thin*

if for every c.e. subspace W of V_∞ such that W contains V, either

(i) the dimension of $\frac{V_\infty}{W}$ is at most k, or

(ii) the dimension of $\frac{W}{V}$ is finite,

and k is the least such number.

(Kalantari-Retzlaff)

For every $k \geq 0$, there exists a k-thin subspace of V_∞.

15 Characterizing $SwBC$-learnable classes $L(V)$

(Harizanov, Stephan)

The following statements are equivalent for a c.e. subspace V of V_∞.

(i) The class $L(V)$ is $SwBC$-learnable.

(ii) The dimension of $\frac{V_\infty}{V}$ is finite, or V is 0-thin, or V is 1-thin.

Proof of (i) $\Rightarrow$ (ii) uses our general result about $SwBC$-learnability.

16 Learning from an informant

(Harizanov, Stephan)

(i) There is a 0-thin vector space V such that the class $L(V)$ is $InfEX$-learnable.

(ii) There is a 0-thin vector space V such that the class $L(V)$ is not $InfEX$-learnable.

17 When there are only finitely many scalars

(Harizanov, Stephan)

Assume that V_∞ is over a *finite* field of scalars. Let V be a c.e. subspace of V_∞.

(i) $L(V)$ is $SwEX$-learnable exactly when $\frac{V_\infty}{V}$ is finite dimensional.

(ii) If V is k-thin, for any k, then $L(V)$ is $SwBC$-learnable.

18 Learning c.e. matroids

We also study learnability of the classes of c.e. substructures for computable *matroids*, which are abstract mathematical structures with *dependence* relations. They generalize linear dependence on vector spaces.

(Harizanov, Stephan)

(i) There is a computable matroid such that every c.e. submatroid is either finite or cofinite. The class of these submatroids can be $SwEX$-learned, but not with any bound on the number of switches.

(ii) There is a computable matroid whose class of all c.e. submatroids can be $SwBC$-learned with oracle K, the halting set, but not with any oracle A to which K is not Turing reducible.