An Introduction to Inductive Inference

Nan Fang
Institut für Informatik, Universität Heidelberg

2016.9.13

An introductory example

Learner reads more and more data about an r.e. set and outputs hypotheses what the set is.

Data	Hypotheses
2	Set of even numbers;
2,3	Set of all numbers;
2,3,5	Set of prime numbers;
2,3,5,13	Set of prime numbers;
2,3,5,13,1	Set of Fibonacci numbers;
2,3,5,13,1,8	Set of Fibonacci numbers;
2,3,5,13,1,8,21	Set of Fibonacci numbers.

Successful learning means that the conjectures converge to one correct concept.

Language acquisition

Think about language acquisition for children:

- Children can master any natural language in a few years' time on the basis of rather casual and unsystematic exposure to it.
- Many linguists (notably, Chomsky[1975],[1986]) believe that children are not genetically prepared to acquire any arbitrary language.
- Instead, a relatively small class $\mathcal{H}$ of languages may be singled out as "humanly possible".
- While linguists fail to propose a nontrivial description of $\mathcal{H}$.
- The branch of linguistics known as "comparative grammar" is the attempt to characterize the class of (biologically possible) natural languages through formal specification of their grammars.

A further example

- Let $\mathcal{L} = \{\omega \backslash \{n\} : n \in \omega\}$.
- Find an effective process to learn(indentify) every member of $\mathcal{L}$.

- How about the class $\mathcal{L} \cup \{\omega\}$?

A learning paradigm

A formal model of inductive inference should contain the following concepts:

1. a theoretically possible reality
2. intelligible hypotheses
3. the data available about any given reality, were it actual
4. a learner
5. successful behaviour by a learner working in a given, possible reality

Our models

1. A theoretically possible reality is a family of languages coded by a class of computably enumerable sets.
2. Hypotheses are symbolic representations of a real or fictitious world, for our reality we take a Gödel numbering.
3. The information about a given reality can be of different type, mainly two. One consist of only positive instances, the other contains negative ones as well.
4. A learner is a Turing machine converting the data he observed to hypotheses. Usually, we require it to be a total Turing machine. We could also have some other requirement on him.
5. One possible successful behaviour could be that, after many attempts, the learner's hypotheses finally converges to a code of the real object.

Notations

Definition

- we call a computably enumerable subset of ω a *language*.
- A *text* is a total function $t : \omega \to \omega \cup \{\#\}$.
- $content(t) = range(t) \setminus \{\#\}$.
- A text t is a *text for the set* $A \subseteq \omega$ if $A = content(t)$.
- $SEQ = (\omega \cup \{\#\})^*$
- A *learner* is a total computable function $M : SEQ \to \omega \cup \{?\}$.

Explanatory learning from text

Definition

- The learner M *TxtEx-learns the language L from the text t* for L if there is a number n_0 such that

$$\forall n \geq n_0 (M(t \restriction n) = M(t \restriction n_0) \in \omega) \& L = W_{M(t \restriction n_0)}$$

- The learner M *TxtEx-learns the language L* if M TxtEx-learns L from all texts t for L.
- The learner M *TxtEx-learns the family $\mathcal{L}$* of languages if M TxtEx-learns all languages in $\mathcal{L}$.
- A family $\mathcal{L}$ of languages is *TxtEx-learnable* if there is a learner M which TxtEx-learns the family $\mathcal{L}$. The class of all TxtEx-learnable families is denoted by TxtEx.

Some properties of TxtEx

Proposition (Monotonicity)

If $\mathcal{L}_0 \subset \mathcal{L}_1$ *and* $\mathcal{L}_1 \in \mathrm{TxtEx}$, *then* $\mathcal{L}_0 \in \mathrm{TxtEx}$.

Proposition

Any finite family $\mathcal{L}$ *of languages is in* TxtEx.

Proposition

The family $\mathcal{FIN}$ *of all finite sets is in* TxtEx.

Other learning criteria

Definition (finite learning)

The learner M *TxtFin-learns the language* L *from the text* t for L if there is a number n_0 such that

$$\forall n \geq n_0 (M(t \upharpoonright n) = M(t \upharpoonright n_0) \in \omega)$$

$$\& \forall n < n_0 (M(t \upharpoonright n) = ?) \& L = W_{M(t \upharpoonright n_0)}.$$

Definition (behaviourally correct learning)

The learner M *TxtBC-learns the language* L *from the text* t for L if there is a number n_0 such that

$$\forall n \geq n_0 (M(t \upharpoonright n) \in \omega \& L = W_{M(t \upharpoonright n)}).$$

Just as in the case of TxtEx-learning, we define the classes TxtFin and TxtBC accordingly.

Properties of learning from text

Proposition (Monotonicity)

Let $\mathrm{X} \in \{\mathrm{TxtFin}, \mathrm{TxtEx}, \mathrm{TxtBC}\}$. *If* $\mathcal{L}_0 \subset \mathcal{L}_1$ *and* $\mathcal{L}_1 \in \mathrm{X}$*, then* $\mathcal{L}_0 \in \mathrm{X}$.

Proposition

$$\mathrm{TxtFin} \subseteq \mathrm{TxtEx} \subseteq \mathrm{TxtBC}$$

Proposition

Let L_0 *and* L_1 *be languages such that* $L_0 \subset L_1$*. Then* $\{L_0, L_1\} \notin \mathrm{TxtFin}$

Learning from informant

Definition

- An *informant for a language L* is its characteristic sequence $L(0)L(1)L(2)\ldots$.
- The learner *M InfEx-learns the language L* if there is a number n_0 such that

$$\forall n \geq n_0 (M(L \restriction n) = M(L \restriction n_0) \in \omega) \& L = W_{M(L \restriction n_0)}$$

- The learner *M InfEx-learns the family $\mathcal{L}$* of languages if *M* InfEx-learns all languages in $\mathcal{L}$.
- A family $\mathcal{L}$ of languages is *InfEx-learnable* if there is a learner *M* which InfEx-learns the family $\mathcal{L}$. The class of all InfEx-learnable families is denoted by InfEx.

Similarly, we can define the classes InfFin and InfBC.

Properties of learning from informant

Proposition

$$\text{InfFin} \subseteq \text{InfEx} \subseteq \text{InfBC}$$

Proposition

For $X = Ex, Fin, BC$,

$$\text{TxtX} \subseteq \text{InfX}.$$

Questions: What's the relationship among these six major classes? Are their inclusion relations strict?

Learnability Hierarchy

The answer is yes, and there is a full hierarchy of them:

$$\mathrm{InfBC}$$
$$\cup$$
$$\mathrm{TxtBC} \cup \mathrm{InfEx}$$
$$\subset \qquad\qquad\qquad \supset$$
$$\mathrm{TxtBC} \qquad\qquad\qquad\qquad \mathrm{InfEx}$$
$$\supset \qquad\qquad\qquad \subset$$
$$\mathrm{TxtBC} \cap \mathrm{InfEx}$$
$$\cup$$
$$\mathrm{TxtEx}$$
$$\cup$$
$$\mathrm{InfFin}$$
$$\cup$$
$$\mathrm{TxtFin}$$

Other variants of learning model

- A learner may be not totally computable; learner(machine) may have some time or space requirement.
- The hypotheses space is not a Gödel numbering.
- Oracle learning
- Probabilistic learning

Computational Learning VS. Statistical Learning

Computational Learning	VS.	Statistical Learning
What can machines learn, *exactly*, in *theory*?		*How* can machines learn, *correctly(less error and more efficiently)*, in *practice*?
Linguist		Language teacher

References

- S.Jain, D.Osherson, J.S.Royer, A.Sharma. *Systems That Learn, An Introduction to Learning Theory(2nd Edition)*. MIT Press, 1999.
- K. Ambos-Spies. *Lecture Notes*. 2016.
- N. Chomsky. *Reflections on language*. Pantheon, 1975.
- N. Chomsky. *Knowledge of language*. Praeger, 1986.