Philosophical Aspects in Pattern Recognition Research

Teresa Scantamburlo

Politecnico di Milano, 5th May, 2014
Motivation

To discuss the philosophical (often tacit) notions or assumptions underlying much of contemporary pattern recognition research and to undertake a critical reflection of its current status.
Outline

1. Pattern recognition: from philosophy to cognitive science
   • Pattern recognition & induction
   • Pattern recognition as abstraction and generalization
   • Pattern recognition as categorization
   • Models of categorization

2. The disciplinary status of pattern recognition
   • A general intuition
   • Perspectives on pattern recognition research
   • Scientific and engineering aspects
   • Examples

3. Paradigms in pattern recognition research
   • Kuhn’s notion of a paradigm
   • Feature-based approach
   • Similarity-based approach
   • Paradigms or research programmes/traditions

4. Conclusions
Pattern Recognition: From Philosophy to Cognitive Science
The field of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data into different categories”.

(Bishop, Pattern Recognition and Machine Learning, p. 1)
Some Examples

Some common characterizations:

- supervised
- unsupervised
- semi-supervised
- reinforcement
Pattern Recognition as Induction

“An inductive method is a principle for finding a pattern in the data that can then be used to classify new cases or to estimate the value of a function for new arguments. So, the problem of finding a good inductive method is sometimes called a pattern recognition problem”

(Harman & Kulkarni, Reliable Reasoning: Induction and Statistical Learning Theory, p. 22)
Induction vs. Deduction

Usually induction is distinguished from deduction. Valid deductive rules are necessarily truth preserving, while inductive rules are not:

**Deduction**

a) All men are mortal

b) Socrates is a man

c) Therefore, Socrates is mortal

**Induction**

a) Many apples have been found to have seeds

b) Until now, no apples have been found not to have seeds

c) So, the next apple will have seeds
Induction as Abstraction & Generalization

Pattern recognition includes two meanings of induction:

• “Generalization”: It allows to infer laws (or rules) from some observed events. It characterized the method of natural science (Bacon)

PR = to estimate a function dividing objects into regions (K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 1990)

• “Abstraction” or “apprehension”: the passage from data perception to abstract knowledge (Aristotle)

PR = “to see something as something” (Watanabe, *Pattern Recognition: Human and Mechanical*, 1985)
Some implications

How can we apply such knowledge? Is the philosophical analysis just a form of erudition?

For instance, thanks to the previous disambiguation (i.e., between abstraction and generalization) we may approach some inner debates:

1. Is PR a subfield of AI?
2. Are PR and ML the same field?
Induction as Categorization

Mechanized induction → formal models of categorization

Artificial intelligence

Pattern recognition
(automatic discovery of regularities)

Cognitive psychology
(information processing activities)

categorization
## Unsupervised Categorization

<table>
<thead>
<tr>
<th>Cognitive psychology</th>
<th>Pattern recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A key goal is to model categorization task so as to determine <em>why</em> certain classifications are preferred, compared to others.</td>
<td>The objective of cluster analysis is simply to find <em>convenient</em> and valid organization of the data, not to establish rules for separating future data into categories.</td>
</tr>
<tr>
<td>The information gained through its models is used to evaluate the plausibility of specific theories (e.g., prototype, exemplar, developmental, etc.) or hypotheses about categorization.</td>
<td>The information gained through its methods should suggest new experiments, and provide fresh insight into the subjects matter.</td>
</tr>
<tr>
<td>The focus is on the agents.</td>
<td>The focus is on the objects.</td>
</tr>
<tr>
<td>It provides models of categorization (or models of phenomena).</td>
<td>It provides tools for categorization (or models of data).</td>
</tr>
</tbody>
</table>

E. Pothos & A. Wills (Eds.), *Formal approaches in categorization*, 2010

Example: The Rational Model

• Anderson’s (1991) rational model adopts a category utility approach. It assumes that categories are formed because they allow us to infer unknown information about novel instances.

• In the rational model, the probability of classification of a novel instance into category $k$ depends on the product:

$$P(k)P(F|k)$$

• For example, a new object with many features of a cat would be assigned to the category of cats, as the feature structure of the object is most probable given this category membership.


The Rational Model
The Disciplinary Status of PR
A General Intuition

What emerges from the literature?
Engineering Perspective

“Pattern recognition is engineering because we try to design machines that read documents, count blood cells, inspect parts, etc.

We must understand the “physics” of the problem and select from amongst available tools the ones appropriate for the problem.

It is futile to look for general mathematical/computational techniques that can solve all problems”

(T. Pavlidis, ICPR, 2000)
Scientific Perspective

“Automatic pattern recognition is usually considered as an engineering area which focuses on the development and evaluation of systems that imitate or assist humans in their ability of recognizing patterns.

It may, however, also be considered as a science that studies the faculty of human beings (and possibly other biological systems) to discover, distinguish, characterize patterns in their environment and accordingly identify new observations.”

A Risk

• In the field of pattern recognition the engineering and the scientific aspects seems unrelated

• This recalls the traditional dichotomy between science and technology

• This opposition often results in a superior-subordinate relationship:

  “The spectre of technology as a subordinate exercise, the tedious and unexciting result of applying the results of science to practical ends is hard to exorcize.”

The main issue

How do these terms (i.e. science and engineering) interact/overlap?

What does it mean for PR being an engineering/scientific discipline?

How can PR be concerned with different types of investigation (e.g., descriptive, normative, applicative, etc.)?
Science and Engineering

- Traditionally science is associated with the knowledge of “causes” and engineering with the design of artefacts (note that in time this idea became controversial)

- A contemporary distinction:
  - “scientist is concerned with how things are”
  - “engineer with how things ought to be”

  (H. Simon *The sciences of the artificial* 1969)

- With the Scientific Revolution knowledge acquisition became inextricably tied to experimental practice and the development of technical tools.
Lessons from neural networks

- In 1943 McCulloch and Pitts proposed a model of artificial neurons

- In 1949 Hebb demonstrated a simple updating rule

- In 1960s Rosenblatt completed the perceptron

- In 1969 Minsky and Papert showed that perceptron can solve only linearly separable problems

Neural networks as a plausible model of brain activities

Neural networks fail as a model
Lessons from neural networks

- In 1982 Hopfield proposed asynchronous neural networks (associative memory)
- In 1985 Rumelhart et al applied back-propagation rule to neural networks
- In 1986 Hinton and Sejnowski invented Boltzmann machines
- In 1990 Poggio and Girosi introduced regularization networks
- In 2006 Hinton introduced deep learning

The return of neural networks (for applications, e.g., zip code recognition and speech synthesis)

Research on neural networks (applications and analogies with learning stages)
Paradigms in Pattern Recognition Research
How Mature is the Field of Pattern Recognition?

“The acquisition of a paradigm and of the more esoteric type of research it permits is a sign of maturity in the development of any given scientific field”

T. Kuhn, *The Structure of Scientific Revolutions*, 1970

See e.g., N. Cristianini, *On the current paradigm in artificial intelligence*, AICom, in press
Kuhn’s View of Science

“Normal” phase: the development of a science is driven by adherence to a paradigm which supports scientists in their “puzzle-solving” activity.

The crisis: worrying puzzles remain unsolved (“anomalies”) and the current approach loses progressively its original appeal.

Finally, the crisis is resolved by a scientific revolution leading to the replacement of the current paradigm by a new one (paradigm shift).
Narrow and Broad Paradigms

1. Narrow paradigms: examples or concrete achievements (e.g., Newton's mechanics and Franklin's theory of electricity)

2. Broad paradigms: they are described as a disciplinary matrix where we find:
   - Symbolic generalization: formal expressions, codified terms, fundamental definitions
   - World-views: beliefs, values, implicit knowledge (e.g., atomism in physics or structuralism in mathematics)
1. “Suppose the event we wish to classify can be represented by a set of $d$ numbers.

2. Let the values of these numbers be denoted by the symbols $x_1, x_2, \ldots x_d$. We shall call such a set of values a pattern.

3. It is convenient to think of a pattern as a point $X$ in $d$-dimensional space.

4. Once an event has been represented by the pattern $X$, we can speak of the problem of classifying the point $X$ and hence the event”.

Essentialism

Essentialism is based on the Aristotelian distinction between “essential” and “accidental” properties.

Such a distinction shaped the approach to the knowledge of natural phenomena (e.g., taxonomies) until the XIX century.

Essentialist tenet = essential properties provide the sufficient and necessary conditions for category membership.
Critics to Essentialism

“Look and see whether there is anything common to all. For if you look at them you will not see something that is common to all, but similarities, relationship, and a whole series of them at that. To repeat: don't think but look!”

L. Wittgenstein, *Philosophical Investigation*, n.66

Criticisms to essentialism in other fields:

- **Physics**: Bridgman’s operationalism
- **Biology**: Darwin’s evolution theory
- **Psychology**: Rosch’s prototype theory
Critical Aspects of Essentialism in PR

There are several contexts in which the feature-based approach is unsatisfactory:

- experts cannot define features in a straightforward way
- data are high dimensional
- features consist of both numerical and categorical variables
- missing or inhomogeneous data
- objects are described in terms of structural properties
Similarity-Based Approach

A relational or similarity-based approach tries to avoid a direct use of features exploiting contextual information. Within the relational paradigm objects are described in terms of relations instead of features.

Some examples of similarity-based approach/efforts are:

- Relaxation-labelling process
- Kernel trick
- Pairwise clustering (spectral or graph-based clustering)
- Collective classification approaches
- Markov Random Fields
The Relaxation Labelling Technique

A set of objects $B = \{b_1, ..., b_n\}$ and a set of labels $\Lambda = \{1, ... m\}$

Two sources of information:

1. local measurements which capture the salient features of each object viewed in isolation ($p_i(\lambda)$)

2. possible interactions among nearby labels which incorporate all the contextual knowledge ($q_i(\lambda; p)$)

\[
p_i(\lambda) \leftarrow \frac{p_i(\lambda)q_i(\lambda)}{\sum_{\mu=1}^{m} p_i(\mu)q_i(\mu)}
\]

\[
q_i(\lambda; p) = \sum_{j=1}^{n} \sum_{\mu=1}^{m} r_{ij}(\lambda, \mu)p_j(\mu)
\]
So, is PR Mature?

- Yes...There are many narrow paradigms (e.g. neural networks, kernel methods, spectral clustering) and a well-accepted broad essentialist paradigm

- But there are also signs of transitions from a essentialist to an anti-essentialist/relational paradigm

- Normal science, crisis or more paradigms? (Lakatos’ research programmes or Laudan’s research traditions)
Conclusion

- Conceptual tools to understand the structure of pattern recognition (specific characteristics, aims, relations to similar fields)

- Integration of Scientific and technological issues (distinctions and continuum)

- Essentialist and relational paradigms in pattern recognition research

- Future directions: to investigate the notion of progress from the standpoints of Lakatos’ “research programmes” and Laudan’s “research traditions”
Bibliography

• C. Bishop, Machine Learning and Pattern Recognition, Springer, New York, 2006


• T. Kuhn. The Structure of Scientic Revolutions. The University of Chicago Press, Chicago, 1970


• E. Pothos, A. Wills (Eds.), *Formal Approaches in Categorization*, Cambridge University Press, 2011


Conferences and Publications


• M. Pelillo, T. Scantamburlo, V. Schiaffonati, "Pattern Recognition between Science and Technology: A Red Herring?", ICPR 2014 (submitted)

Thanks!