Investigation into the inner workings of Concept Learning and Decision Trees

Adriaan Louw (53031377)

April 3, 2018

Contents

1 Abstract **1**

2 Method **1**

3 Results **1**
- 3.1 Concept Learning 1
  - 3.1.1 What is Concept Learning? 1
  - 3.1.2 General-to-specific ordering of hypotheses 2
  - 3.1.3 The FIND-S algorithm 2
  - 3.1.4 Version Spaces 3
- 3.2 Decision Trees 3

4 Conclusion **3**

1 Abstract

2 Method

3 Results

3.1 Concept Learning

3.1.1 What is Concept Learning?

We can describe Concept Learning in terms of trying to teach a machine to classify animals as either dogs or not.

We assuming we classify leaves by the following parameters

- Is the animal large or small? (Size)
- Is it brown or black? (Colour)
- Is the tail long or short? (Tail length)
- Does it have 2 or 4 legs?
- Does it bark?

We can then describe animals according to these parameters {size,colour,tail length,no of legs,does it bark}. This is called our hypothesis space.

Then a particular animal can be described as $\{small, brown, short, 4, no\}$ for example. For our dataset we have a selection of animals and whether that animal is a dog or not.

The concept that we wish the machine to learn is which values of each parameter defines the Concept of a do. In our case that would be $\{?, ?, ?, 4, yes\}$ where ? means any value.

Formally we can say:

- We have X which is all the instances in the hypothesis space
- We also have $c(x)$ which is the function that returns whether an animal is a dog or not
- Training data $D = \{\langle x, c(x)\rangle : x \in X, c(x) \in \{0, 1\}\}$

(Chandola, 2018)

3.1.2 General-to-specific ordering of hypotheses

From the above example, the most general hypothesis would be $\{?, ?, ?, ?, ?\}$. This means any animal would satisfy the hypothesis. The most specific hypothesis is $\{\emptyset, \emptyset, \emptyset, \emptyset, \emptyset\}$.

We can sort hypotheses based on whether they are more general or specific than other hypotheses.

Consider the following hypotheses: $h_1 = \{?, ?, ?, ?, yes\}$ and $h_2 = \{?, brown, ?, 4, yes\}$. h_1 is more general than h_2. In other words it less specific on what the parameters have to be to satisfy the hypothesis.

(Riedmiller, 2009)

3.1.3 The FIND-S algorithm

The goal of the FIND-S Algorithm is to find the "maximally specific hypothesis". This means it is the most specific hypothesis that can be found that represents the concept.

Assuming we have a the following dataset :

$$\begin{aligned} D &= \sum_{i=1}^{n} \{x_i, c(x_i)\} \\ &= \{\{\{small, brown, short, 4, no\}, 0\}, \\ &\quad \{\{small, brown, long, 4, yes\}, 1\}, \\ &\quad \{\{large, black, short, 2, no\}, 0\}, \\ &\quad \{\{large, black, long, 4, yes\}, 1\}\} \end{aligned} \tag{1}$$

The algorithm works as follows. We start with a empty hypothesis $h = \emptyset$. Then we iterate through each hypothesis in our data set (see Equation 1). If $c(x_i) = 0$ then go to the next hypothesis. If not, then we do a pairwise and between h and the data x and update h. As in $h \leftarrow h \wedge x$. Then we go onto the next data point.

Given data set D from equation 1. We start with $h = \emptyset$. The first data point has $c(x) = 0$. So we ignore it. The second data point has $c(x) = 1$ so we do a pairwise and with h.

$$\begin{aligned} h \leftarrow & h \wedge x \\ \leftarrow & \emptyset \wedge \{small, brown, long, 4, yes\} \\ \leftarrow & \{small, brown, long, 4, yes\} \end{aligned} \tag{2}$$

Thus, $h = \{small, brown, long, 4, yes\}$. The next data point is also ignores because it does not define a positive match. Then we go onto the last data point.

$$\begin{aligned} h \leftarrow & h \wedge x \\ \leftarrow & \{small, brown, long, 4, yes\} \wedge \{large, black, long, 4, yes\} \\ \leftarrow & \{?, ?, long, 4, yes\} \end{aligned} \tag{3}$$

Thus we have the hypothesis we learned form the data set is $h = \{?, ?, long, 4, yes\}$. (Chandola, 2018)

3.1.4 Version Spaces

Definition: A hypothesis h is consistent with the set of training examples D of target concept c if and only if $h(x) = c(x)$ for each training example $\langle x, c(x) \rangle$ in D.

$$Consistent(h, D) \equiv (\forall \langle x, c(x) \rangle \in D) h(x) = c(x) \tag{4}$$

Definition: the *version space*, $VS_{H,D}$, with respect to hypothesis space H and training examples D, is the subset of hypotheses from H consistent with all training examples in D.

$$VS_{H,D} \equiv \{h \in H | Consistent(h, D)\} \tag{5}$$

(Chandola, 2018; Riedmiller, 2009)

3.2 Decision Trees

4 Conclusion

References

Chandola, V. (2018). *Introduction to machine learning.* Retrieved from https://www.cse.buffalo.edu// chandola/teaching/machinelearningdocs/parta2-handout.pdf

Riedmiller, M. (2009). *Concept Learning.* Retrieved from http://ml.informatik.uni-freiburg.de/former/_media/documents/teaching/ss09/ml/version.pdf