

Exp 1 - Find-S Algorithm

February 20, 2022

Find-S is a **basic concept-based machine learning algorithm** that tries to find the **maximally specific hypothesis** that fits a set of **training examples**. The algorithm starts with the most specific hypothesis and gradually becomes more general.

1 Experiment Description

1.1 Objective

To explore how **FIND-S algorithm** can be used to find the **maximally specific hypothesis** that fits the given set of training examples.

1.2 Algorithm

1. Initialize h to the **most specific hypothesis** in H , where H is the set of all hypotheses that the learner can consider for the **target concept**.
2. In each **positive training instance** x in X :
 - For each **attribute constraint** a_i in h :
 - If the **constraint** a_i is satisfied by x .
 - Then do nothing.
 - Else replace a_i by the next **more general constraint** (?) that is satisfied by x .
3. Output **hypothesis** h .

1.3 Procedure

1. Representation of **hypothesis**.
2. Application of the **Find-S algorithm** on the given data to get the **maximally specific hypothesis**.
3. Interpretation of the output obtained from **Find-S**.

1.4 System Requirements

Windows/Linux OS with R. No extra R library is required to be installed.

1.5 Dataset Summary

The dataset is a weather dataset containing attributes that determine whether an individual will play a sport on that day or not. The target concept is the **Sport** attribute which is a binary value.

- Number of instances: 8
- Number of attributes: 9
- Attribute information:
 1. Time: Morning or Evening (Time for playing)
 2. Sky: Sunny or Rainy
 3. AirTemp: Warm or Cold
 4. Humidity: Normal or High
 5. Wind: Strong or Mild
 6. Weather: Warm or Cool
 7. Company: Yes or No
 8. Forecast: Same or Change
 9. Sport: Yes or No

2 Code and Output

```
In [1]: rm(list = ls())  
        version$version.string
```

```
"R version 4.0.2 (2020-06-22)"
```

```
In [2]: # Importing CSV file containing weather data  
        data <- read.csv('weather.csv');  
        head(data);
```

A data.frame: 6 × 9

	Time <chr>	Sky <chr>	AirTemp <chr>	Humidity <chr>	Wind <chr>	Weather <chr>	Company <chr>	Forecast <chr>	Sport <chr>
1	Morning	sunny	warm	normal	strong	warm	yes	same	yes
2	Morning	sunny	warm	high	strong	cool	no	same	yes
3	Evening	rainy	cold	high	mild	warm	yes	change	no
4	Morning	sunny	warm	high	mild	cool	yes	same	yes
5	Morning	rainy	cold	normal	mild	warm	yes	same	yes
6	Evening	rainy	cold	high	strong	cool	no	change	no

```

# Initializing hypothesis array based on the length of the imported data
num_attributes = dim(data)[2]-1;
hypothesis <- c(length<-(num_attributes))
for(i in 1:num_attributes){
  hypothesis[i] = '0';
}

# Initializing hypothesis for the first data row
hypothesis <- array(data[1,1:num_attributes]);

```

```

In [3]: # Applying Find-S Algorithm
for(i in 1:dim(data)[1]){
  isequal <- data[i,num_attributes+1] == "yes";
  if(isequal){
    for(j in 1:num_attributes){
      isequal = data[i,j] == hypothesis[[j]];
      if(!isequal){
        hypothesis[[j]] = '?';
      }
      else{
        hypothesis[[j]] = data[i,j];
      }
    }
  }
}

```

```

In [4]: # Most specific hypothesis based on the given set of training data
hypothesis

```

A data.frame: 1 × 8 1 | Morning ? ? ? ? ? ? same

3 Conclusion

FIND-S algorithm got successfully implemented over the given dataset.