6/5/2021

Luka Leko, Paul Knoll

Symbolic AI: Task 2

Group 01:

Contents

[1. Getting Started 2](#_Toc73898200)

[Basic Aleph File Structure 2](#_Toc73898201)

[2. Recursion 5](#_Toc73898202)

[2.1. Learn the member predicate: 5](#_Toc73898203)

[2.1.1. The mem/2 Background Knowledge File 5](#_Toc73898204)

[2.1.2. Executing Aleph on the mem files 6](#_Toc73898205)

[2.2. Learn the last/2 predicate 7](#_Toc73898206)

[3. Generalization 9](#_Toc73898207)

[3.1. Preparing the Aleph Files 9](#_Toc73898208)

[3.1.2. Results 11](#_Toc73898209)

[3.2. Optimizing prediction performance 13](#_Toc73898210)

# Getting Started

We first tried to understand what Aleph does by testing it out. For this we installed it as instructed in the task specification and then executed Aleph on the Train example from Ryszard Michalski. This is what we have learned:

## Basic Aleph File Structure

The whole Aleph Logic is contained within a single .pl file, which we will call aleph-swi.pl. For Aleph to be able to induce a set of rules (also called a theory), it needs the following three files contained within the same folder:

1. The .f file: It contains the set of positive examples. The examples are formatted as simple prolog facts. Aleph will consider those as correct during his learning procedure:

.f file for the Train example:

**eastbound(east1).**

**eastbound(east2).**

**eastbound(east3).**

**eastbound(east4).**

**eastbound(east5).**

1. The .n file: It contains the set of negative examples. It looks the same as .f file, except that Aleph will consider those facts as wrong during his learning procedure.
2. The .b file: This file is the most interesting out of the three. The Background knowledge is in the form of Prolog clauses that encode information relevant to the domain. This file also contains language and search restrictions for Aleph. Amongst other things, this file mostly contains:
   1. Modes. These declare the mode of call for predicates that can appear in any clause hypothesised by Aleph.
   2. Determinations. Determination statements declare the predicates that can be used to construct a hypothesis.
   3. Type definitions.
   4. You can also set certain variables used by Aleph with set(variableName, value).

When one wants to run Aleph to learn rules to distinguish between positive and negative examples, one has to:

1. Open swipl,
2. Consult the aleph-swi.pl file.
3. Read\_all(filename). This will make Aleph load the example.
4. Call Induce. This will initiate the learning step. Based on the configuration within the Background Knowledge file, this will produce a certain set of rules. During the learning, Aleph will also print out the theory that he has learned, including a matrix stating how many of the positive and negative examples one can correctly identify given that theory. Used on the trains example, Aleph is able to find a theory consisting of only a single rule which is able to correctly predict every positive example and exclude every negative example:

[theory]

[Rule 1] [Pos cover = 5 Neg cover = 0]

eastbound(A) :-

has\_car(A,B), short(B), closed(B).

[Training set performance]

Actual

+ -

+ 5 0 5

Pred

- 0 5 5

5 5 10

Accuracy = 1

The solution that we have found by hand to distinguish between eastbound trains and westbound trains looks as follows:

**eastbound(A) :-**

**has\_car(A, B), has\_car(A, C), has\_car(A, D), shape(C, triangle, 1).**

With this we can correctly predict all positive examples as correct, and only wrongly predict one negative example as correct.

We can also look at the bottom clauses, i.e. the clauses that contain all literals that we have allowed through the background knowledge. You can do so by calling **show(bottom)**. The bottom clause for the first example looks as follows:

eastbound(A) :-

has\_car(A,B), has\_car(A,C), has\_car(A,D), has\_car(A,E), short(E), short(C), closed(C), long(D), long(B), open\_car(E), open\_car(D), open\_car(B), shape(E,rectangle), shape(D,rectangle), shape(C,rectangle), shape(B,rectangle), wheels(E,2), wheels(D,3), wheels(C,2), wheels(B,2),

load(E,circle,1), load(D,hexagon,1), load(C,triangle,1), load(B,rectangle,3).

We can see that it does indeed contain all literals that were also included in the rule that Aleph has found during learning.

# Recursion

## 2.1. Learn the member predicate:

As always, the files needed for Aleph are the Background Knowledge file (mem.b), the positive examples (mem.f) and the negative examples (mem.n). The example files obviously contain only examples, so only the .b file is really interesting. It will now be described in more detail.

### The mem/2 Background Knowledge File

The Background Knowledge File for the mem/2 predicate looks as follows:

% Simple illustration of the learning of recursive predicates

% in Aleph

% To run do the following:

% a. Load Aleph

% b. read\_all(mem).

% c. induce.

:- modeh(\*,mem(+any,+list)).

:- modeb(\*,mem(+any,+list)).

:- modeb(1,((+list) = ([-any|-list]))).

:- set(i,3).

:- set(noise,0).

:- determination(mem/2,mem/2).

:- determination(mem/2,'='/2).

:- modeh(\*,mem(+any,+list))

says that in the head of a clause you only allow the mem() predicate with 2 parameters, the first parameter of type any and the second parameter of type list.

:- modeb(\*,mem(+any,+list))

allows that same structure to occur as a literal in the body of a clause.

:- modeb(1,((+list) = ([-any|-list])))

Allows another literal format to appear in the body of a clause, namely that a list equals another list where the head is of type any and the tail is another list.

With

:- set(variableName, value)

One can set variables for Aleph, which can steer how Aleph learns during Induction. For example,

:- set(noise, 0)

sets an upper bound on the number of negative examples allowed to be covered by an acceptable clause.

The last part in the member Background Knowledge are the determinations, with which one selects what predicates are allowed in what clauses in general. For example,

:- determination(mem/2,mem/2)

allows mem/2 predicates to appear in clauses where mem/2 is in the head.

### Executing Aleph on the mem files

With this very restricted search, Aleph is able to find the following implementation of the mem/2 predicate in Prolog:

[theory]

[Rule 1] [Pos cover = 12 Neg cover = 0]

mem(A,B) :-

B=[A|C].

[Rule 2] [Pos cover = 10 Neg cover = 0]

mem(A,B) :-

B=[C|D], mem(A,D).

[Training set performance]

Actual

+ -

+ 19 0 19

Pred

- 0 6 6

19 6 25

The mem/2 predicate it has found is indeed a correct implementation of the member predicate. Obviously, because of that, it is also able to predict all example correctly.

## Learn the last/2 predicate

In order to have Aleph learn the last/2 predicate, one has to provide Aleph with the three files as always. The two example files are trivial to construct, as one only has to think of a couple of examples for where the predicate is positive and where it is negative. One only has to watch out to also cover the base case in both files at least once.

To know how the Background Knowledge file should look like, we first defined the last/2 predicate ourselves, and then we looked at what kinds of predicates we needed there. The idea is that if we know one possible definition of the predicate, then we only have to make sure that Aleph is able to construct that particular clause, and then it will be able to find it. This was our recursive definition of the last/2 predicate:

mem(A, [A|\_]).

mem(A, [H|T]) :- mem(A, T).

Fortunately for us, this recursive definition is already very close to the recursive definition of the mem/2 predicate, for which we already have the Background Knowledge File. With that in mind, we created the following Background Knowledge File:

:- modeh(\*,last(+any,+list)).

:- modeb(\*,last(+any,+list)).

:- modeb(1,((+list) = ([-any]))).

:- modeb(1,((+list) = ([-any\_2|-list]))).

:- set(i,3).

:- set(noise,0).

:- determination(last/2,last/2).

:- determination(last/2,'='/2).

With this slight modification, Aleph was able to find the following clauses:

[theory]

[Rule 1] [Pos cover = 5 Neg cover = 0]

last(A,B) :-

B=[A].

[Rule 2] [Pos cover = 11 Neg cover = 0]

last(A,B) :-

B=[C|D], last(A,D).

[Training set performance]

Actual

+ -

+ 16 0 16

Pred

- 0 13 13

16 13 29

It has found our definition of the last/2 predicate. It does not look exactly the same, but when evaluated by the Prolog Interpreter it is equivalent to our definition.

The following bottom clause was the bottom clause of the first positive example (this example was a base case):

[bottom clause]

last(A,B) :-

B=[A], B=[C|D].

The following bottom clause is the bottom clause of the first positive example in the second iteration step (i.e. where all base cases were already taken care of):

5 ?- show(bottom).

[bottom clause]

last(A,B) :-

B=[C|D], last(A,D), D=[A], D=[C|E].

We can see that the first literals from the first bottom clause are a superset of the literals used in Rule 1. The same counts also for the second bottom clause and Rule 2.

# Generalization

## Preparing the Aleph Files

This exercise was about inducing a set of clauses which can distinguish between good and bad training examples in the dataset that we were provided with. The dataset has the following shape:

First 10 rows of the training\_dataset.csv:

A13,18,A32,A42,A61,A73,2,A93,A101,A121,A143,A152,A172,2,A191,A201,Bad

A11,12,A32,A40,A61,A72,4,A93,A101,A121,A143,A152,A172,2,A191,A201,Bad

A14,18,A32,A42,A61,A72,4,A94,A102,A122,A143,A152,A172,2,A191,A201,Bad

A12,18,A34,A42,A61,A72,2,A93,A101,A121,A143,A152,A172,1,A191,A201,Bad

A11,18,A32,A46,A61,A71,4,A92,A101,A121,A143,A152,A171,1,A191,A201,Bad

As one can see, this is not yet in the correct format for Aleph to be able to process those examples. With the script generalization\_convert.ipynb we converted this data into the correct format, spread out over the three files generalization.b, generalization.f and generalization.n. Because the script is not relevant for Aleph, it is omitted at this point.

First 5 rows of generalization.f:

generalization(a13,21,a32,a40,a62,a73,1,a92,a101,a123,a141,a152,a174,1,a192,a201).

generalization(a11,12,a34,a40,a61,a73,3,a93,a101,a121,a143,a151,a172,2,a191,a202).

generalization(a14,9,a34,a43,a65,a73,1,a93,a101,a121,a143,a152,a173,2,a191,a201).

generalization(a14,24,a34,a42,a64,a73,4,a93,a101,a121,a143,a152,a173,1,a192,a201).

generalization(a14,24,a32,a42,a61,a74,2,a92,a101,a122,a143,a151,a173,1,a192,a201).

We can see that one example consists of 16 features, three of which are numerical features, and the other 13 are categorical features. For the categorical features, only the equality predicate really makes sense to compare two values that belong to the same feature. For the numerical features, the predicates “≤”, “≥”, “<”, “>” are also plausible. Comparing between two different features obviously does not make sense. Because of that, we defined the Background Knowledge the following way:

Generalization.b:

% ------------ variable settings ------------

:- set(clauselength, 6).

:- set(minpos, 5).

:- set(minacc, 0.9).

:- set(verbosity,0).

:- set(evalfn,laplace).

:- set(search,heuristic).

:- set(test\_pos,'generalization\_test.f').

:- set(test\_neg,'generalization\_test.n').

% ------------ modes ------------

:-modeh(1,generalization(+att1,+att2,+att3,+att4,+att6,+att7,+att8,+att9,+att10,+att12,+att14,+att15,+att17,+att18,+att19,+att20)).

:- modeb(1,(+att1 = #att1)). % the same for all other features

:- modeb(1,gteq(+att2,#att2)).

:- modeb(1,lteq(+att2,#att2)).

:- modeb(1,gt(+att2,#att2)).

:- modeb(1,lt(+att2,#att2)). % the same for all other numeric features

% ------------ determinations ------------

:- determination(generalization/16,lt/2).

:- determination(generalization/16,gt/2).

:- determination(generalization/16,lteq/2).

:- determination(generalization/16,gteq/2).

:- determination(generalization/16,'='/2).

:- determination(generalization/16,not/2).

% ------------ type definitions ------------

att1(a13). att1(a14). att1(a12). att1(a11). %the same for all other features.

The file contains four different sections:

1. The variables section, where we steer Aleph’s Learning procedure.
2. The modes section
3. The determinations section
4. The type definitions. Those were automatically generated by the generalization\_convert.ipynb file.

## 3.1.2. Results

With this Background Knowledge, Aleph constructed 87 Rules for the training dataset.

[Rule 87] [Pos cover = 13 Neg cover = 0]

generalization(A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P) :-

E=a65, J=a121, K=a143, L=a152,

O=a191.

[Training set performance]

Actual

+ -

+ 533 0 533

Pred

- 97 270 367

630 270 900

Accuracy = 0.8922222222222222

[Training set summary] [[533,0,97,270]]

[Test set performance]

Actual

+ -

+ 15 51 66

Pred

- 15 19 34

30 70 100

Accuracy = 0.34

We have tried different hyperparameters, and with this setting we achieve a solid prediction performance on the training dataset. However, the testing dataset has an incredibly poor prediction performance: 51 out of 70 negative examples were incorrectly classified as correct examples, and only 15 out of 30 positive examples were correctly classified (so altogether, this classifier was worse than Random on the Testing Data). Unfortunately, we were not able to get a better prediction performance on the test dataset. Just by simply playing around with the Aleph parameters listed at the top of the Background Knowledge File, we could not find a much better solution for this classification problem. We can ensure that Aleph has indeed done what we expected it to do by inspecting the bottom clause:

7 ?- show(bottom).

[bottom clause]

generalization(A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P) :-

lt(B,6), lt(G,4), lt(N,1), gt(B,6),

gt(G,4), gt(N,1), lteq(B,6), lteq(G,4),

lteq(N,1), gteq(B,6), gteq(G,4), gteq(N,1),

A=a11, B=6, C=a32, D=a40,

E=a65, F=a71, G=4, H=a92,

I=a101, J=a122, K=a143, L=a152,

M=a174, N=1, O=a192, P=a201.

We can see that it contains all the literals that we have wanted to allow. Because of the solid prediction performance on the training data, and because the bottom clause looks the way we expected it to, we justify that what we did was at least sensible, even though we have made no real progress on predicting unseen examples.

## Optimizing prediction performance

We have manually tried out different hyperparameter settings. Concretely, we have tested different values for the following variables:

1. Clauselength: upper bound on number of literals in an acceptable clause.
2. Minpos: a lower bound on the number of positive examples to be covered by an acceptable clause. The higher this number, the fewer clauses Aleph will find, and the worse its prediction performance will be on the training data, especially for the positive examples.
3. Minacc: a lower bound on the minimum accuracy of an acceptable clause. The higher this number, the fewer clauses.
4. Evalfn: Sets the evaluation function for a search.
5. Search: Sets the search strategy.

We have found that the most important parameters were minpos and minacc:

:- set(minpos, 15).

:- set(minacc, 0.85).

By manually trying out, we have found that minpos between 5 and 15 and minacc between 0.85 and 0.9 makes sense. The other hyperparameters are those that we were asked to try out and compare, which is what the following table does:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Evalfn | Seach | Train Accuracy | Test Accuracy | Rules |
| Coverage | bf | 0.648 | 0.5 | 38 |
| coverage | df | 0.66 | 0.44 | 77 |
| auto\_m | heuristic | 0.648 | 0.52 | 24 |
| gini | bf | 0.66 | 0.46 | 43 |
| laplace | heuristic | 0.61 | 0.52 | 28 |

# Try something interesting

## 4.1.

For the last task we chose the “Adult Data Set” from UCI (<https://archive.ics.uci.edu/ml/datasets/adult>) which is a mixed categorical and integer classification Dataset with 14 attributes. The goal is to predict whether a person earns more or less than 50K$ a year based on the 14 attributes. We chose to only use a subset of 8 of the attributes which are the following: age, workclass, years of education, occupation, race, sex ,working hours per week and native country. The first 2000 examples get used for the training and the next 1000 samples for testing. Samples with missing values get tossed out.

We again set up a Python script to convert the dataset into the positive and negative examples and create the background knowledge. The Background Knowledge is very similar to the Background Knowledge from Exercise 3. For the numerical features we again allowed the ≤, ≥, <, >, and = as comparison operators.

We set the following variables:

Set(minpos, 2).

Set(minacc, 0.7).

This led Aleph to construct 211 Rules that give an accuracy of 94% on the training set and 77% on the test set:

[Training set performance]

            Actual

         +            -

     +  381           0           381

Pred

     -  94         1367        1461

        475         1367        1842

Accuracy = 0.9489685124864278

[Training set summary] [[381,0,94,1367]]

[Test set performance]

           Actual

        +          -

     +  99          88         187

Pred

     - 126        600        726

       225        688        913

Accuracy = 0.7656078860898138

This led to a similar prediction performance as in Exercise 3. We were very good on the training dataset, but on the testing dataset, we only had a True Positive Rate of 99/225 = 0.44. Again, the bottom clause contains what we wanted, but in the end Aleph could not find a good set of clauses such that it is also accurate on the testing dataset.