
ABAGAIL Documentation

Release latest

April 19, 2016

1	Features	3
1.1	Hidden Markov Models	3
1.2	Feed-forward backpropagation neural networks of arbitrary topology	3
1.3	Fast training with the sequential minimal optimization algorithm	3
1.4	Information gain or GINI index split criteria	3
1.5	Fast kd-tree implementation for instance based algorithms of all kinds	4
1.6	Basic matrix and vector math, a variety of matrix decompositions based on the standard algorithms	4
1.7	Randomized hill climbing, simulated annealing, genetic algorithms, and discrete dependency tree MIMIC	4
1.8	Kruskals MST and DFS	4
1.9	EM with gaussian mixtures, K-means	4
1.10	PCA, ICA, LDA, Randomized Projections	4
2	Installation	5
3	Usage	7
4	Frequently Asked Questions	9
4.1	Table of contents	9
4.2	What is ABAGAIL?	10
4.3	How to get started with ABAGAIL?	10
4.4	How is ABAGAIL project organized?	11
4.5	What are the different ways to work with ABAGAIL?	11
4.6	How to use ABAGAIL with Eclipse?	12
4.7	Is there a tutorial or documentation for ABAGAIL?	12
4.8	Can you generate graphs out of ABAGAIL library?	12
4.9	How does ABAGAIL output results?	13
4.10	Does Abagail have any form of built in cross validation?	13
4.11	How can I find out the actual number of iterations performed in ABAGAIL?	14
4.12	Could someone help explain the training_ iterations in ABAGAIL? When I increase it, I get better accuracy.	14
4.13	My neural network classifier is performing very poorly, how to fix it?	14
4.14	How does the neural network get its weights updated?	14
4.15	How to change number of neurons in a hidden layer?	15
4.16	I am trying to understand what the count ones optimization function is doing, but I am not sure...	15
4.17	TravelingSalesmanRouteEvaluationFunction - why the 1/total distance?	15
4.18	How do I get multi-label classification working in AbaloneTest.java?	16
4.19	My dataset has categorical features, can I use ABAGAIL?	16

4.20	Getting ant build error: /home/unazi/ABAGAIL/build.xml:40: Unexpected attribute “additional-param”, how do I fix this?	16
4.21	What are some characteristics of problems simulated annealing works well on?	17
5	Supervised Learning Algorithms	19
6	Clustering Algorithms	21
7	Reinforcement Learning Algorithms	23
8	Random Search Algorithms	25
9	Optimization Algorithms	27
10	Examples	29

The *ABAGAIL* library contains a number of interconnected Java packages that implement machine learning and artificial intelligence algorithms. These are artificial intelligence algorithms implemented for the kind of people that like to implement algorithms themselves.

1.1 Hidden Markov Models

- Baum-Welch reestimation algorithm, scaled forward-backward algorithm, Viterbi algorithm
- Support for Input-Output Hidden Markov Models
- Write your own output or transition probability distribution or use the provided distributions, including neural network based conditional probability distributions
- Neural Networks

1.2 Feed-forward backpropagation neural networks of arbitrary topology

- Configurable error functions with sum of squares, weighted sum of squares
- Multiple activation functions with logistic sigmoid, linear, tanh, and soft max
- Choose your weight update rule with standard update rule, standard update rule with momentum, Quickprop, RPROP
- Online and batch training
- Support Vector Machines

1.3 Fast training with the sequential minimal optimization algorithm

- Support for linear, polynomial, tanh, radial basis function kernels
- Decision Trees

1.4 Information gain or GINI index split criteria

- Binary or all attribute value splitting
- Chi-square significance test pruning with configurable confidence levels
- Boosted decision stumps with AdaBoost

- K Nearest Neighbors

1.5 Fast kd-tree implementation for instance based algorithms of all kinds

- KNN Classifier with weighted or non-weighted classification, customizable distance function
- Linear Algebra Algorithms

1.6 Basic matrix and vector math, a variety of matrix decompositions based on the standard algorithms

- Solve square systems, upper triangular systems, lower triangular systems, least squares
- Singular Value Decomposition, QR Decomposition, LU Decomposition, Schur Decomposition, Symmetric Eigenvalue Decomposition, Cholesky Factorization
- Make your own matrix decomposition with the easy to use Householder Reflection and Givens Rotation classes
- Optimization Algorithms

1.7 Randomized hill climbing, simulated annealing, genetic algorithms, and discrete dependency tree MIMIC

- Make your own crossover functions, mutation functions, neighbor functions, probability distributions, or use the provided ones.
- Optimize the weights of neural networks and solve travelling salesman problems
- Graph Algorithms

1.8 Kruskals MST and DFS

- Clustering Algorithms

1.9 EM with gaussian mixtures, K-means

- Data Preprocessing

1.10 PCA, ICA, LDA, Randomized Projections

- Convert from continuous to discrete, discrete to binary
- Reinforcement Learning

Installation

- Install Java JDK

Download: <http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

Installation: http://docs.oracle.com/javase/8/docs/technotes/guides/install/install_overview.html

- Install Apache ant

Download: <http://ant.apache.org/bindownload.cgi>

Installation Instructions: <http://ant.apache.org/manual/index.html>

- Clone ABAGAIL Repository

```
cd <my_local_repo_directory>

git clone https://github.com/pushkar/ABAGAIL.git
```

Usage

To compile and run

```
cd ABAGAIL
ant
java -cp ABAGAIL.jar opt.test.XORTest
java -cp ABAGAIL.jar opt.test.TravelingSalesmanTest
```

To generate javadoc documentation:

```
ant javadoc
```

Note that for ant versions 1.9.3 or earlier, you need to remove

```
additionalparam="subpackages"
```

from the javadoc target tag.

To run the Abalone example

```
java -cp ABAGAIL.jar opt.test.AbaloneTest
```

For some examples of how to use the library for your own projects look for the test packages, such as `opt.test`, and `shared.test` inside of the project source directory.

Frequently Asked Questions

4.1 Table of contents

4.1.1 Getting Started

- *What is ABAGAIL?*
- *How to get started with Abigail?*
- *How is ABAGAIL project organized?*
- *What are the different ways can I work with ABAGAIL?*
- *How to use ABAGAIL with Eclipse?*
- *Is there a tutorial or documentation for ABAGAIL?*

4.1.2 Usage Questions

- *Can you generate graphs out of ABAGAIL library?*
- *How does ABAGAIL output results?*
- *Does Abigail have any form of built in cross validation?*
- *How can I find out the actual number of iterations performed in Abigail?*
- *Could someone help explain the training_iterations in ABAGAIL? When I increase it, I get better accuracy.*
- *My neural network classifier is performing very poorly, how to fix it?*
- *How does the neural network get its weights updated?*
- *How to change number of neurons in a hidden layer?*
- *I am trying to understand what the count ones optimization function is doing, but I am not sure...*
- *TravelingSalesmanRouteEvaluationFunction - why the 1/total distance?*
- *How do I get multi-label classification working in AbaloneTest.java?*
- *My dataset has categorical features, can i use ABAGAIL?*

4.1.3 Other questions

- *What are some characteristics of problems simulated annealing works well on?*
 - *Getting build error: /home/unazi/ABAGAIL/build.xml:40: Unexpected attribute “additionalparam”, how do I fix this?*
-

4.2 What is ABAGAIL?

ABAGAIL is an acronym of Absolute Best Andrew Guillory Artificial Intelligence Library.

from `src > abagail.html`:

This library is the result of close to a year of research and class work in artificial intelligence (AI). It contains a number of interconnected Java packages that implement machine learning and artificial intelligence algorithms. These are artificial intelligence algorithms implemented for the kind of people that like to implement algorithms themselves.

Back to top

4.3 How to get started with ABAGAIL?

Start by download the source code from github You can use git to clone the repository:

```
git clone https://github.com/pushkar/ABAGAIL.git
```

or download the .zip file from [github](#)

After downloading the source code, decide how you want to work with the library, the most common options are listed *here*.

VERY BASIC INSTRUCTIONS FOR COMMAND LINE

1. Install Java 8 SDK from here <http://www.oracle.com/technetwork/java/javase/downloads/index.html>
2. Install Ant <http://ant.apache.org/>
3. Clone or download source files from Git
4. Go with command line to where the build.xml file is and run: ant (note: the ant executable should be in your path somehow if you installed ant correctly.. so will java and javac)
5. Now run your scripts

Back to top

4.4 How is ABAGAIL project organized?

At a high-level, the project structure is organized as follows:

`dist`

- Contains classes to generate distributions

`func`

- Contains supervised learning algorithms
- Contains clustering algorithms

`opt`

- Contains random search algorithms
- Contains optimization test functions

`rl`

- Contains reinforcement learning algorithms

`shared`

- Contains filters
- Contains IO and test running functions

`util`

- Contains utility methods and data structures

Back to top

4.5 What are the different ways to work with ABAGAIL?

There are several options:

- use terminal or command line interface to compile and execute Java code
- use Java-compatible IDE with ABAGAIL as a library
examples: *Eclipse*
- use Scala console with ABAGAIL as a library
- use Jython

Back to top

4.6 How to use ABAGAIL with Eclipse?

1. fork `https://github.com/pushkar/ABAGAIL.git`
2. From the terminal clone your forked repo to a directory of your choice.
3. Create an Eclipse project, let's say it's "AbagailProject"
4. With the following steps you will convert the directory where you copied Abagail into an Eclipse Project
 - a. Go to File->New->Project.
 - b. On the menu double click on Java and select "Java Project from Existing Ant Build file". Click Next.
 - c. Browse to the Abagail folder and select the xml file. Click Finish.
 1. Now you need to include this new Eclipse project to the build folder of your project.
 - a. Right click on "AbagailProject" and select Properties.
 - b. On the left pane of the window select "Java Build Path."
 - c. Select the "Projects" tab.
 - d. Click Add...
 - e. Select the abagail project you just created (note that when you create the Eclipse project, the name of the project is "project" by default)
 1. To use any of the Abagail classes just use import. For instance, if you want to use the class `ArffDataSetReader`, you need to add the following import statement: `import shared.reader.ArffDataSetReader;`

Back to top

4.7 Is there a tutorial or documentation for ABAGAIL?

At this time the documentation is very limited. ABAGAIL comes with some documentation for you to look through when you build the project as well as a few examples. Users are actively encouraged to contribute to project documentation.

Back to top

4.8 Can you generate graphs out of ABAGAIL library?

Currently no. There are several external libraries and software you can use to visualize results:

- Microsoft Excel
- R
- matplotlib
- JFreeChart
- matlab
- mathematica

- ...etc.

[Back to top](#)

4.9 How does ABAGAIL output results?

The test functions output results to command line. Saving data to disk requires extending the source code.

Here is a simple extension methods for writing output to file, created by Yeshwant Dattatreya.

```
public void write_output_to_file(String output_dir, String file_path, String result, Boolean final_result) {
    // This function will have to be modified depending on the format of your file name.
    // Else the splits won't work.
    try {
        if (final_result) {
            String full_path = output_dir + "/" + new SimpleDateFormat("yyyy-MM-dd").format(new Date())
                + "/" + "final_result.csv";
            String[] params = file_path.split("_");
            String line = "";
            switch (params.length) {
                case 9:
                    line = params[0] + ",none," + params[6] + "," + params[8].split("\\.")[0];
                    break;
                case 10:
                    line = params[0] + "," + params[3] + "," + params[7] + "," + params[9].split("\\.")[0];
                    break;
                case 11:
                    line = params[0] + "," + params[3] + "_" + params[4] + "," + params[8] + "," +
                        + params[10].split("\\.")[0];
                    break;
            }
            PrintWriter pwtr = new PrintWriter(new BufferedWriter(new FileWriter(full_path, true)));
            synchronized (pwtr) {
                pwtr.println(line + result);
                pwtr.close();
            }
        } else {
            String full_path = output_dir + "/" + new SimpleDateFormat("yyyy-MM-dd").format(new Date())
                + "/" + file_path;
            Path p = Paths.get(full_path);
            Files.createDirectories(p.getParent());
            Files.write(p, result.getBytes());
        }
    } catch (Exception e) {
        e.printStackTrace();
    }
}
```

[Back to top](#)

4.10 Does Abagail have any form of built in cross validation?

Currently no. You will need to extend the source code to add cross-validation.

Back to top

4.11 How can I find out the actual number of iterations performed in ABAGAIL?

When you work with random search optimization problems, the number of iterations in the example problems is fixed. For example:

```
FixedIterationTrainer fit = new FixedIterationTrainer(algorithm, 200000);`
```

will train the algorithm 200,000 times.

Alternatively, you can use `ConvergenceTrainer` which will train until convergence, or up to set number of maximum iterations, whichever occurs earlier:

```
ConvergenceTrainer trainer = new ConvergenceTrainer(algorithm);`
```

Back to top

4.12 Could someone help explain the training_iterations in ABAGAIL? When I increase it, I get better accuracy.

In short, the more you train, the more accurate your weights become. This leads to better accuracy

Back to top

4.13 My neural network classifier is performing very poorly, how to fix it?

Make sure your dataset values are normalized. See [this article](#) for more details.

Back to top

4.14 How does the neural network get its weights updated?

Q:I was looking through the `AbaloneTest.java` code, and notice that each of the randomized algorithm has a `train` function that, in the case of SA for e.g., picks a neighbor and its fitness value (or sticks with the current point). However, I wasn't able to figure out how the NN gets updated with the new weights? I don't see anything where the `train` the SA code passes anything back to the concrete class of `HillClimbingProblem`.

Under `ABAGAIL/src/func/nn/backprop/`: `BackPropagationNetworkFactory.java`

```
private static BackPropagationNetworkFactory factory = new BackPropagationNetworkFactory();

/*****/

// returns instance of BackPropagationNetwork which has updateWeights method.
networks[i] = factory.createClassificationNetwork(new int[] {inputLayer, hiddenLayer, outputLayer});

/*****/

networks[i].setWeights(optimalInstance.getData());
```

Networks is an instance of `BackPropagationNetwork.java`, which has a method `updateWeights()`, which in turn extends class `FeedForwardNetwork.java` > `LayeredNetwork.java` > `NeuralNetwork.java`. The last one has `setWeights()` method.

Back to top

4.15 How to change number of neurons in a hidden layer?

Q: I'm trying to use ABAGAIL to train a neural network with randomized hill climbing. It looks like the number of nodes in a hidden layer is assigned to be the same as the number of nodes in the input layer. How do I change the number of neurons in a hidden layer?

Look at `AbaloneTest.java` for an example. It shows you how to set up the layers using an int array. This file is under `opt > test`.

Back to top

4.16 I am trying to understand what the count ones optimization function is doing, but I am not sure...

Q: I am trying to understand what the count ones optimization function is doing. From the test class in ABAGAIL, I see that an array is created and filled up with all int=2. I think I understand the concept of counting all of the 1s in the vector, but I do not see how the array of all 2s turns into an array of 1s and 0s?

If you dig through the classes that uses the ranges variable, it will become clear. The 2 specifies how many different values are possible at any point in the vector (0 and 1).

Back to top

4.17 TravelingSalesmanRouteEvaluationFunction - why the 1/total distance?

The object of the Traveling Salesman problem is to minimize the distance in a route. However, the optimization algorithms maximize their fitness functions. So it returns the inverse of the distance, which gets larger as the distance gets smaller.

Returning the normal distance would result in the algorithms finding the longest routes, which isn't what we're interested in.

Back to top

4.18 How do I get multi-label classification working in AbaloneTest.java?

This is from the method `initializeInstances`.

You'll need to change whatever your classes are let's say ("cat" -> 0, "dog" -> 1, "other" ->2). This encoding will then correspond to the output layer. I.e. "cat" -> 0 will correspond to the 0th output layer in the array.

```
Instance[] instances = new Instance[attributes.length];

for(int i = 0; i < instances.length; i++) {
    instances[i] = new Instance(attributes[i][0]);

    // Read the digit 0-9 from the attribute array that was read from the csv
    int c = (int) attributes[i][1][0];

    // Create a double array of length 10, all values are initialized to 0
    double[] classes = new double[nClasses];

    // Set the i'th index to 1.0
    classes[c] = 1.0;
    instances[i].setLabel(new Instance(classes));
}
```

Back to top

4.19 My dataset has categorical features, can I use ABAGAIL?

Convert the categorical features to binary, e.g., if you have Country as a variable, then each value of country will have its own column with 1 for the country that's present, and 0 for values that isn't.

Back to top

4.20 Getting ant build error: /home/unazi/ABAGAIL/build.xml:40: Unexpected attribute "additionalparam", how do I fix this?

Make sure you are using a version where this issue has been fixed. See: <https://github.com/pushkar/ABAGAIL/pull/31>

Back to top

4.21 What are some characteristics of problems simulated annealing works well on?

Generally, simulated annealing works well if there are multiple minimum and you need to find the global minimum because SA will (hopefully) break out of the local minimum.

in the ABAGAIL library look for sample test problems under `opt > example` directory

Back to top

Supervised Learning Algorithms

Clustering Algorithms

Reinforcement Learning Algorithms

Random Search Algorithms

Optimization Algorithms

Examples

- AbaloneTest.java