

AMiDST TOOLBOX

Session 2: Introduction to the Amidst Toolbox

Andrés R. Masegosa

University of Almeria
andres.masegosa@ual.es

Setting up AMIDST Toolbox

<https://github.com/andresmasegosa/GeiloWinterSchool2018>



Install AMIDST Toolbox

First, check whether you have installed Java 8:

```
$ java -version
```

If Java 8 is not installed download it from [here](#).

For compiling and running the toolbox you have two options:

1. **IntelliJ IDEA:** You can download it from [here](#).
2. **Maven:** Follow the [official web page](#) for instructions about how to install it.

Download GeiloWinterSchool2018 code repository

First, download the project code:

```
$ git clone https://github.com/andresmasegosa/GeiloWinterSchool2018.git
```

Enter in the downloaded folder:

```
$ cd GeiloWinterSchool2018/
```

If you have installed maven, you can compile and build the package from command line:

```
$ mvn clean package
```

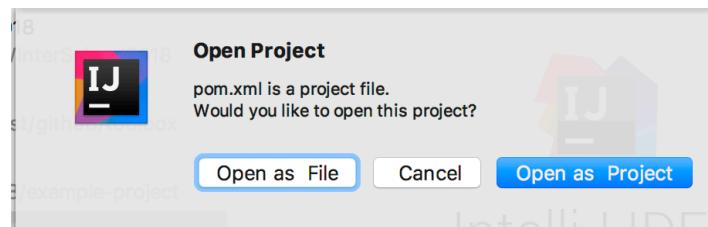
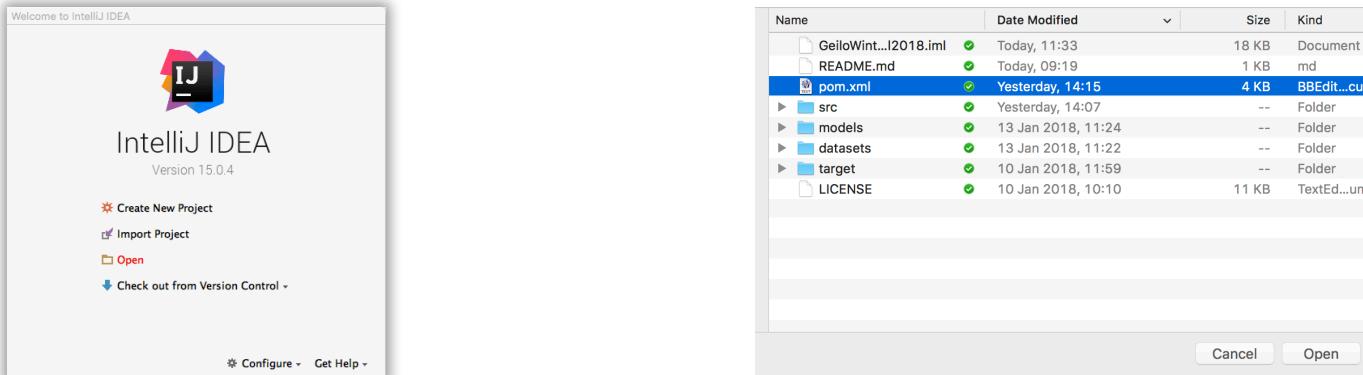
For running any Java file you should type:

```
$ java -cp target/example-project-full.jar NameOfFile
```



SETTING UP INTELLIJ IDEA

AMIDST
TOOLBOX

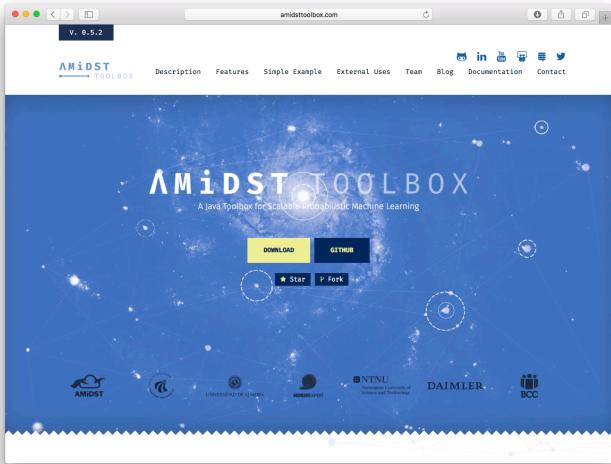


AMIDST Toolbox

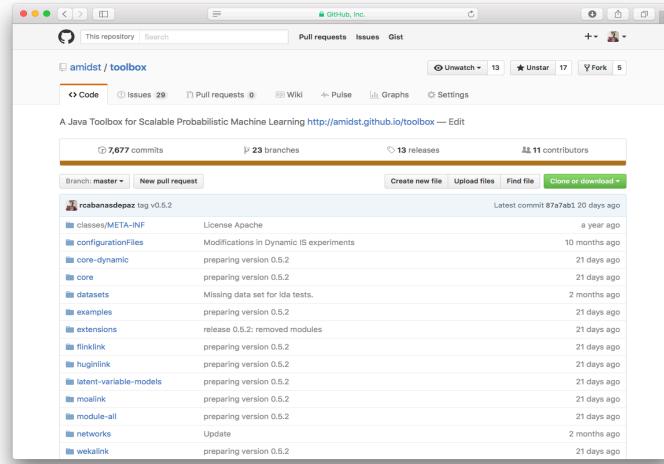


AMIDST TOOLBOX

AMIDST
TOOLBOX



www.amidsttoolbox.com

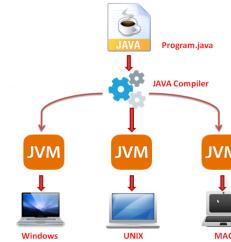


github.com/amidst/toolbox



Apache
License 2.0





- **Machine Learning on the Java Virtual Machine**
 - Integration with other big data technologies
 - Spark, Flink, Hadoop, Hbase, etc rely on JVM.
 - Many IT systems also rely on the JVM.

Code-project



Λ M i D S T TOOLBOX

TOOLBOX



core-dynamic



core



moalink



latent-variable-models



flinklink



wekalink



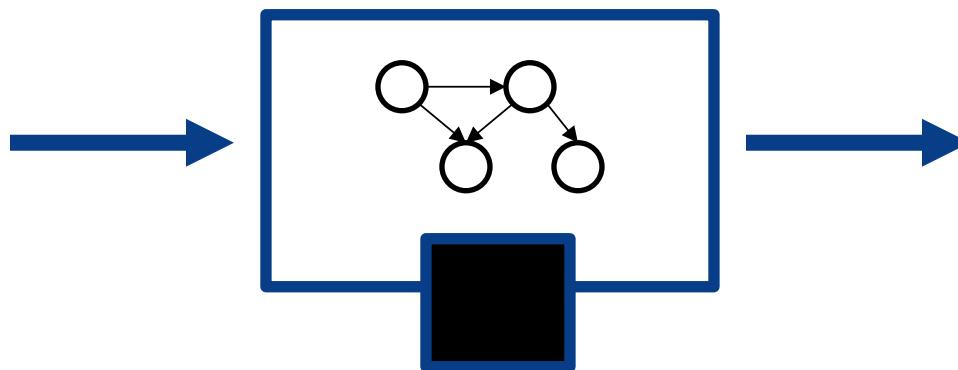
huginlink



Openbox Models

[Probabilistic Graphical Models]

[Big] Data
[+Prior Information]



[Scalable] Bayesian Inference Engine
(Powered by Variational Methods)



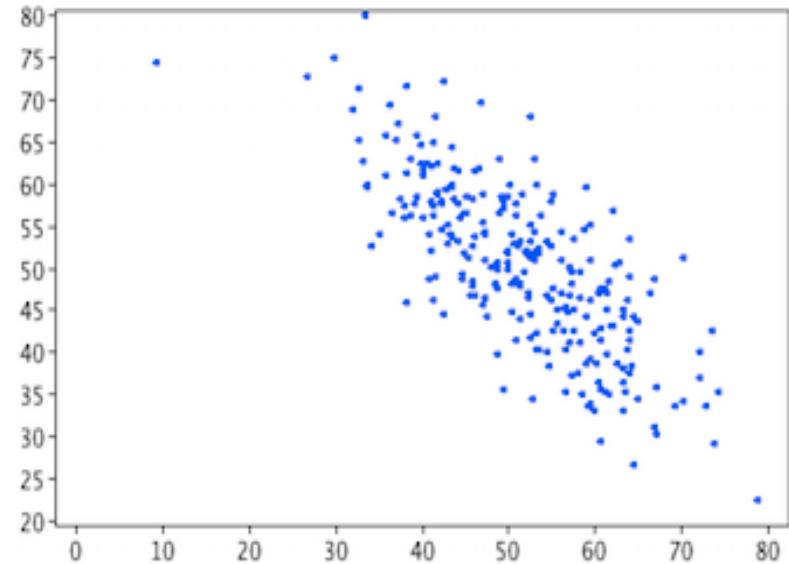
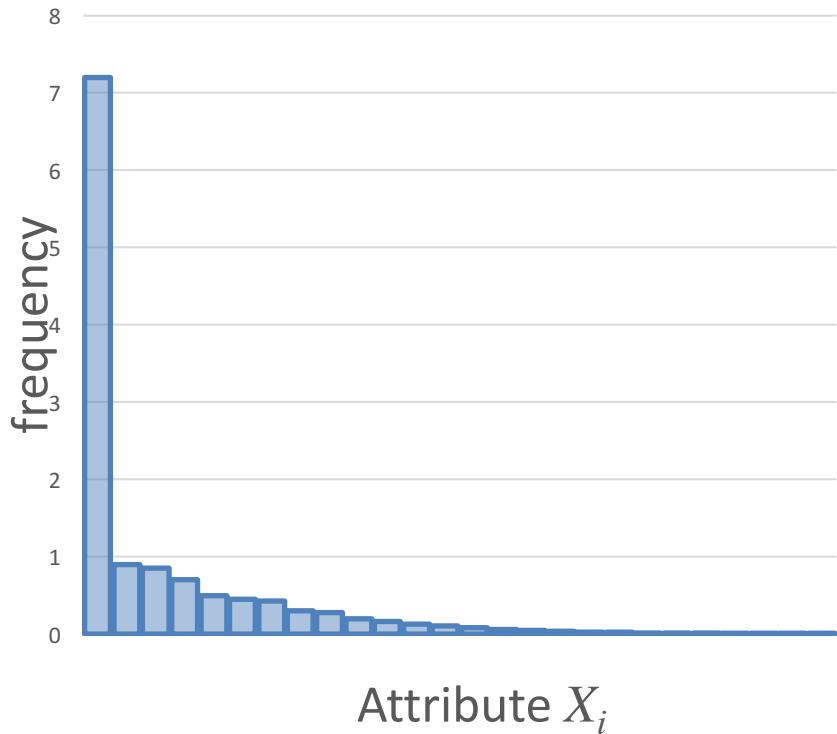
Use Case I





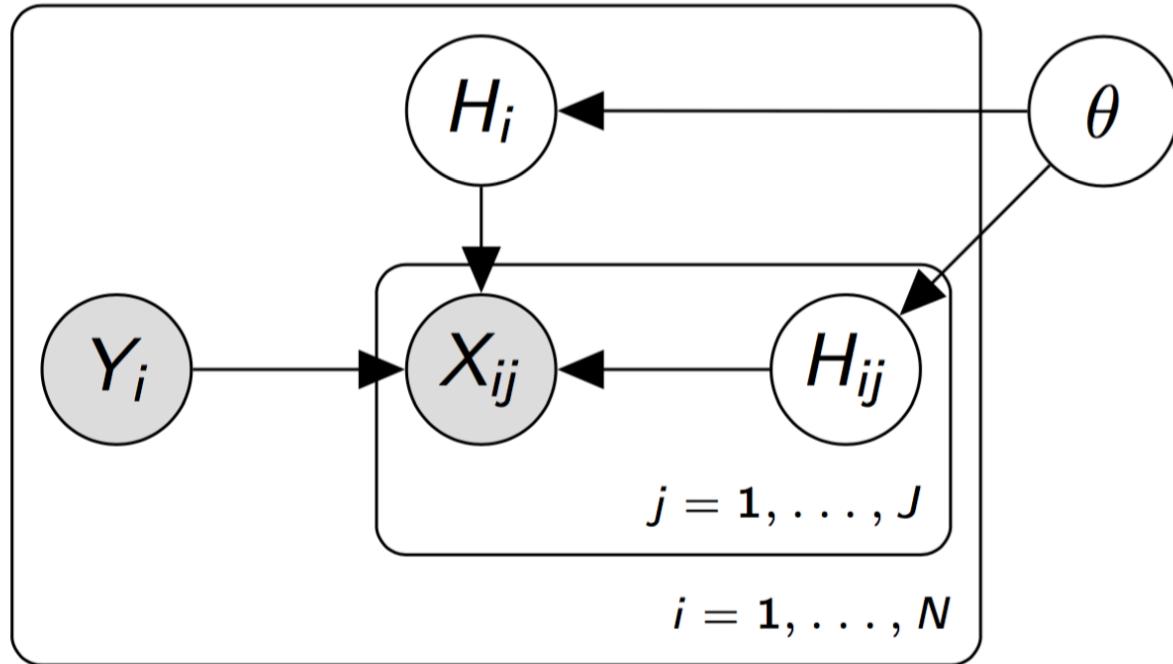
Predicting Defaulting Clients

Predicts probability a customer will default within 2 years



- Daily data for millions of clients
- Tons of missing data.
- Odd distributions.





Custom Gaussian Mixture Model

H_{ij} defines local mixture

H_i defines a global mixture.

```
//Set-up Flink session.  
final ExecutionEnvironment env = ExecutionEnvironment.getExecutionEnvironment();  
  
//Load the data stream  
String filename = "hdfs://dataFlink_month0.arff";  
DataFlink<DataInstance> data =  
    DataFlinkLoader.loadDataFromFolder(env, filename, false);  
  
//Build the model  
Model model = new CustomGaussianMixture(data.getAttributes());  
    .setClassIndex(2);
```



SCALABLE INFERENCE

RUNNING CODE EXAMPLE

```
//Set-up Flink session.  
final ExecutionEnvironment env = ExecutionEnvironment.getExecutionEnvironment();  
  
//Load the data stream  
String filename = "hdfs://dataFlink_month0.arff";  
DataFlink<DataInstance> data =  
    DataFlinkLoader.loadDataFromFolder(env, filename, false);  
  
//Build the model  
Model model = new CustomGaussianMixture(data.getAttributes())  
    .setClassIndex(2);  
  
//Learn the model  
model.updateModel(data);
```



DATA STREAMS

RUNNING CODE EXAMPLE

```
//Set-up Flink session.  
final ExecutionEnvironment env = ExecutionEnvironment.getExecutionEnvironment();  
  
//Load the data stream  
String filename = "hdfs://dataFlink_month0.arff";  
DataFlink<DataInstance> data =  
    DataFlinkLoader.loadDataFromFolder(env, filename, false);  
  
//Build the model  
Model model = new CustomGaussianMixture(data.getAttributes())  
    .setClassIndex(2);  
  
//Learn the model  
model.updateModel(data);  
  
//Update your model  
for(int i=1; i<12; i++) {  
    filename = "dataFlink_month"+i+".arff";  
    data = DataFlinkLoader.loadDataFromFolder(env, filename, false);  
    System.out.println(model.predict(data));  
    model.updateModel(data);  
}
```





Predicting Defaulting Clients

- Old BCC's models based on logistic regression got an AUC around 0.8
- AMIDST's models gets an AUC over 0.9
- Model will be in production this spring.

Use Case II

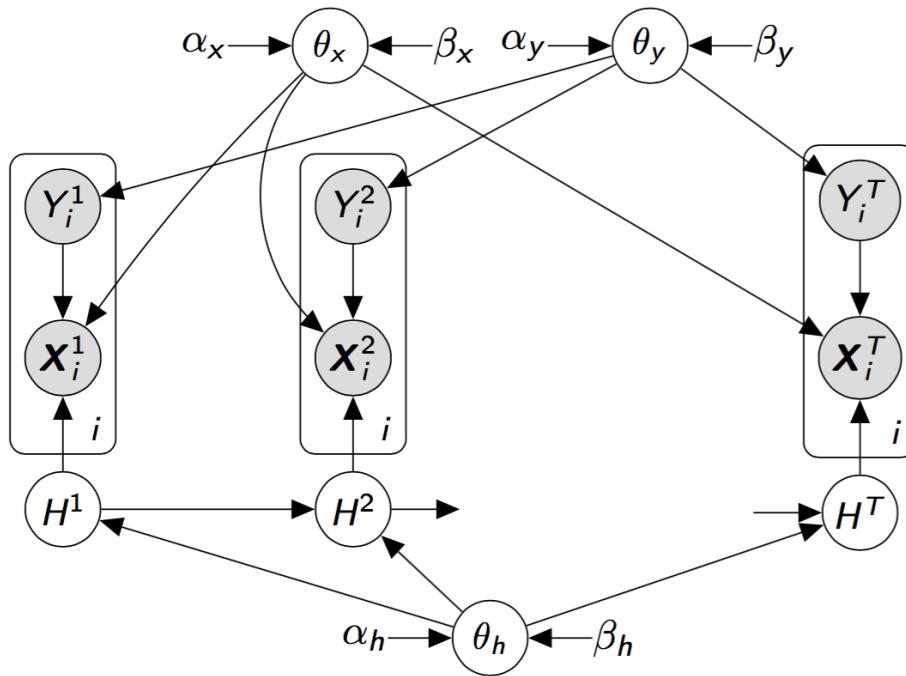




Tracking Concept Drift

Detects changes in customer profiles during Spanish financial crisis

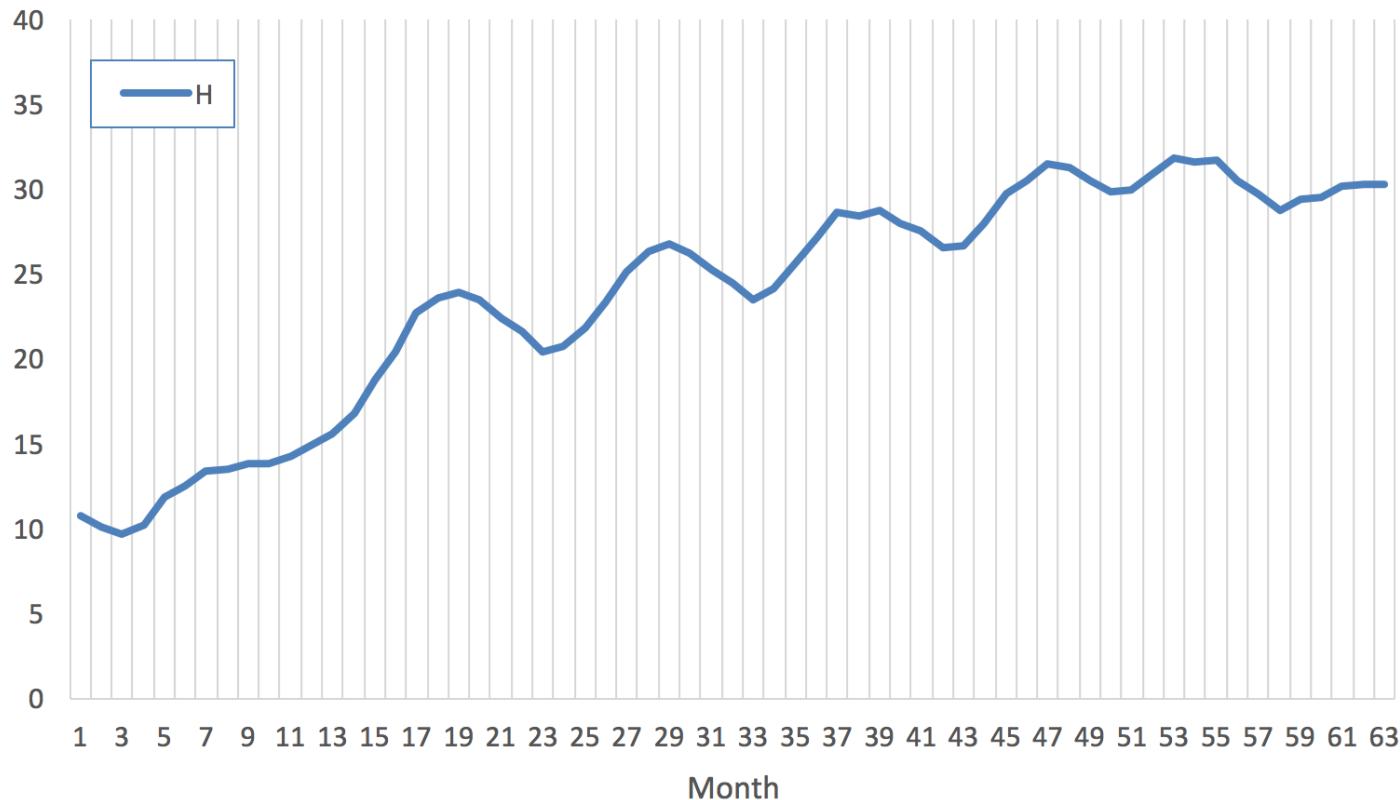




Hidden Variables are used to capture changes in customer profile

CONCEPT DRIFT DETECTION RESULTS

AMIDST
TOOLBOX



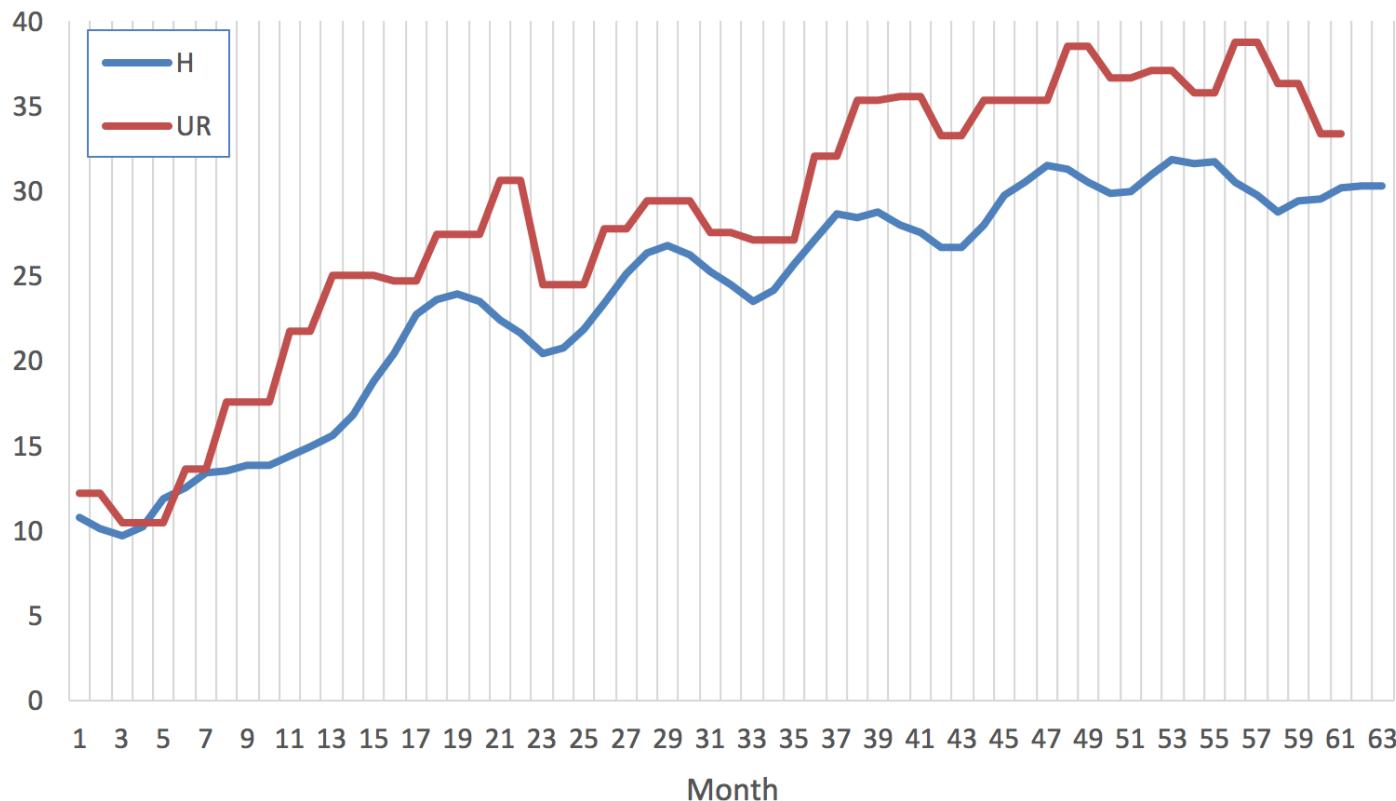
Hidden Variable Captures Concept Drift

Drift Pattern: Seasonal + Global trend



CONCEPT DRIFT DETECTION RESULTS

AMIDST
TOOLBOX



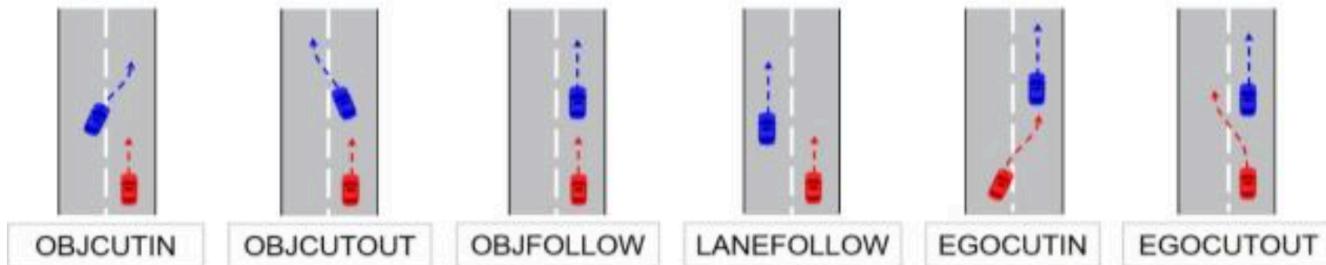
Unemployment Rate main driver of Concept Drift

Hidden Variable correlates with unemployment rate ($\rho = 0.961$)



Use Case III



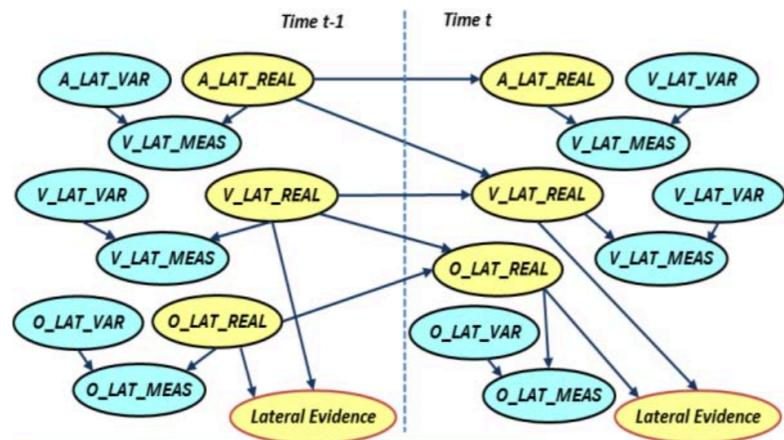
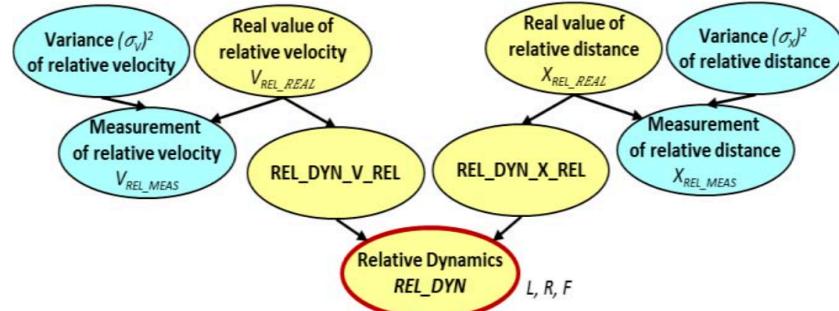
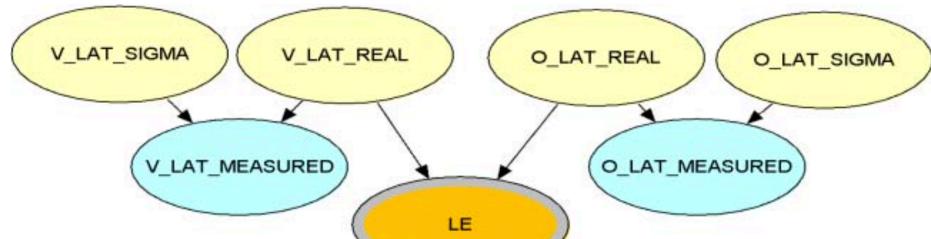
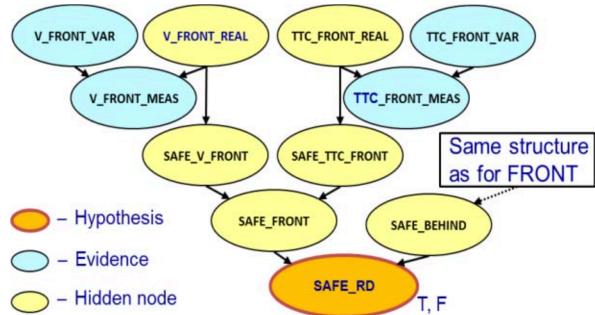


Weidl, Galia, et al. "Early Recognition of Maneuvers in Highway Traffic." *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*. Springer International Publishing, 2015.

Maneuver Recognition

Early detection of traffic maneuvers changes for intelligent cruise control (and autonomous driving).

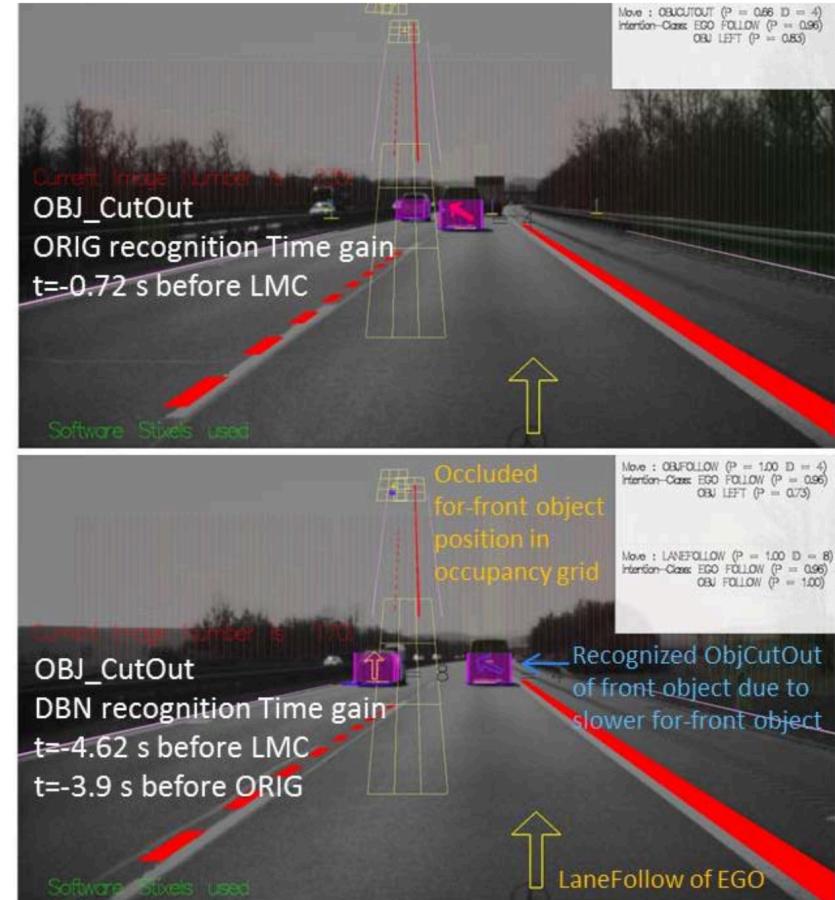
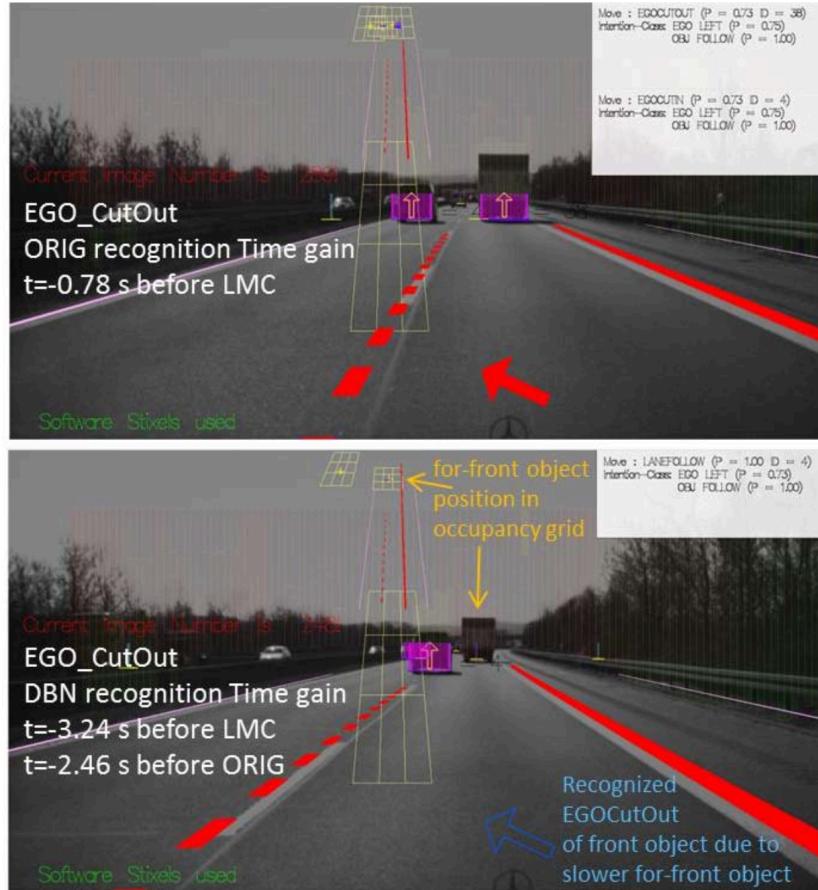
PROBABILISTIC MODEL



Weidl, Galia, et al. "Early Recognition of Maneuvers in Highway Traffic." *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*. Springer International Publishing, 2015.



PROTOTYPE



Weidl, Galia, et al. "Early Recognition of Maneuvers in Highway Traffic." *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*. Springer International Publishing, 2015.

Thanks for your attention

www

www.amidsttoolbox.com

@

contact@amidsttoolbox.com



[@AmidstToolbox](https://twitter.com/AmidstToolbox)

AMIDST
→ TOOLBOX