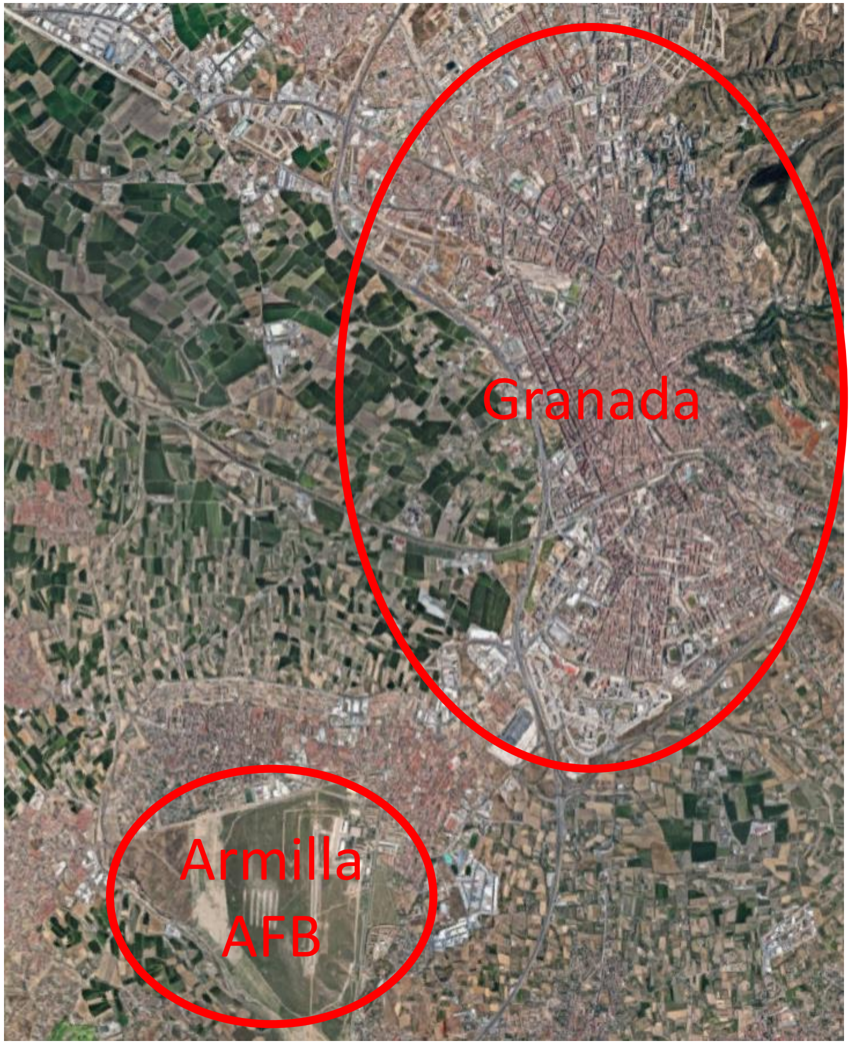
Regression Methods Applied to Flight Variables for Situational Awareness Estimation Using Dynamic Bayesian Networks

Carlos Morales, Serafín Moral

PGM 2016
International Conference on Probabilistic Graphical Models
Lugano, Switzerland
September 2016

University of Granada
Department of Computer Science and Artificial Intelligence
(Granada, SPAIN)

DECSAI



Armilla Air Force Base hosts the Spanish military helicopter school

Eurocopter EC120: Basic training

Sikorsky S76: Instrumental Flight Rules Search & Rescue

CONTENTS

- Motivation
- SA, IM & DBN in context
- The simulation environment
- SA estimation model
- Discretization & Regression
- Results
- Conclussions

MOTIVATION

- SA is widely recognized as a key indicator of crew performance in flight safety and aviation human factors.
- SA measurements are often applied to:
  - Aircraft systems design.
  - Flight procedures & training design.
  - No notice of crew performance rating during the flight.
- SA depends on multiple variables with complex relationships and dynamic behaviour.
  - Variables related to information management are specially relevant.
  - Some of these variables can be monitored using an Electronic Flight Bag (EFB).
- DBN are a good candidate to provide a measurement of certain SA-related aspects in real time.

MOTIVATION

PhD Thesis: Implement a simulation environment to:

Perform simulated flights

Emulate an EFB used by the pilot

Use DBN to measure SA

WORKING HYPOTHESES

1. **Modern Information Management standards have a key role in maintaining high levels of SA during the flight.**
2. **Computational power of EFB can monitor pilot activities.**
3. **We seek reflections of SA in aircraft control actions and information management performed by the pilot.**

SA, IM & DBN IN CONTEXT

Definition and measurement of SA

"*Situational Awareness is the field of study concerned with quantifying the perception of the environment critical to decision-makers in complex, dynamic areas*".

SA levels (Endsley, 1995)

Level 1 SA: Perception of the elements in the environment

- Monitor information & aircraft position queries performed by the pilot.
- Correlate them with pilot actions.

Level 2 SA: Comprehension of the current situation

- Level 1 + Analyze coherence of pilot actions with ideal flight parameters.
  - Expert knowledge required.

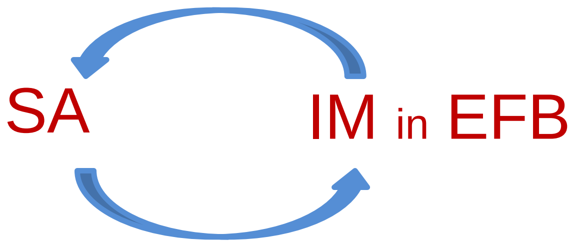
Level 3 SA: Projection of the near future

- Monitor if pilot actions (aircraft control and position & information queries) are timely.
  - Need to be careful with individual-dependent hypotheses.
- Detect when the pilot is "surprised" by flight events.
  - e.g. rough or wrong-sequenced corrections.

SA, IM & DBN IN CONTEXT

<u>Estimation of SA based related to cockpit information management:</u>

When a pilot operates a EFB to obtain information that should affect SA, there should be a relationship between SA and the way pilot makes use of information.



Consistency of the Aeronautical Information data model is enabled by SWIM

<u>SWIM (System Wide Information Management):</u>

Standards, infrastructure and governance enabling the management of Air Traffic Management (ATM) information and its exchange between qualified parties via interoperable services. Under development by Eurocontrol, FAA, etc.

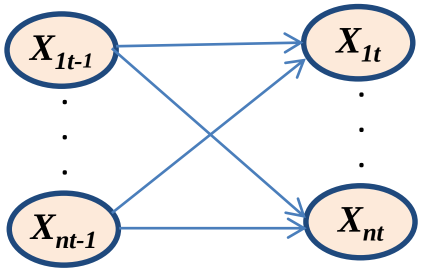
* We are not expecting to perform a global assessment of SA:
    e.g. No biometric sensors used at the current stage of research.
* Estimation is based on a "*<u>additional pilot workload = 0</u>*" principle:
    - Applicability to real flights.
    - Reduced bias compared to classical SA rating techniques (SAGAT, Endsley et al.,1998).

Universidad de Granada

SA, IM & DBN IN CONTEXT

<u>Applicability of DBN</u>

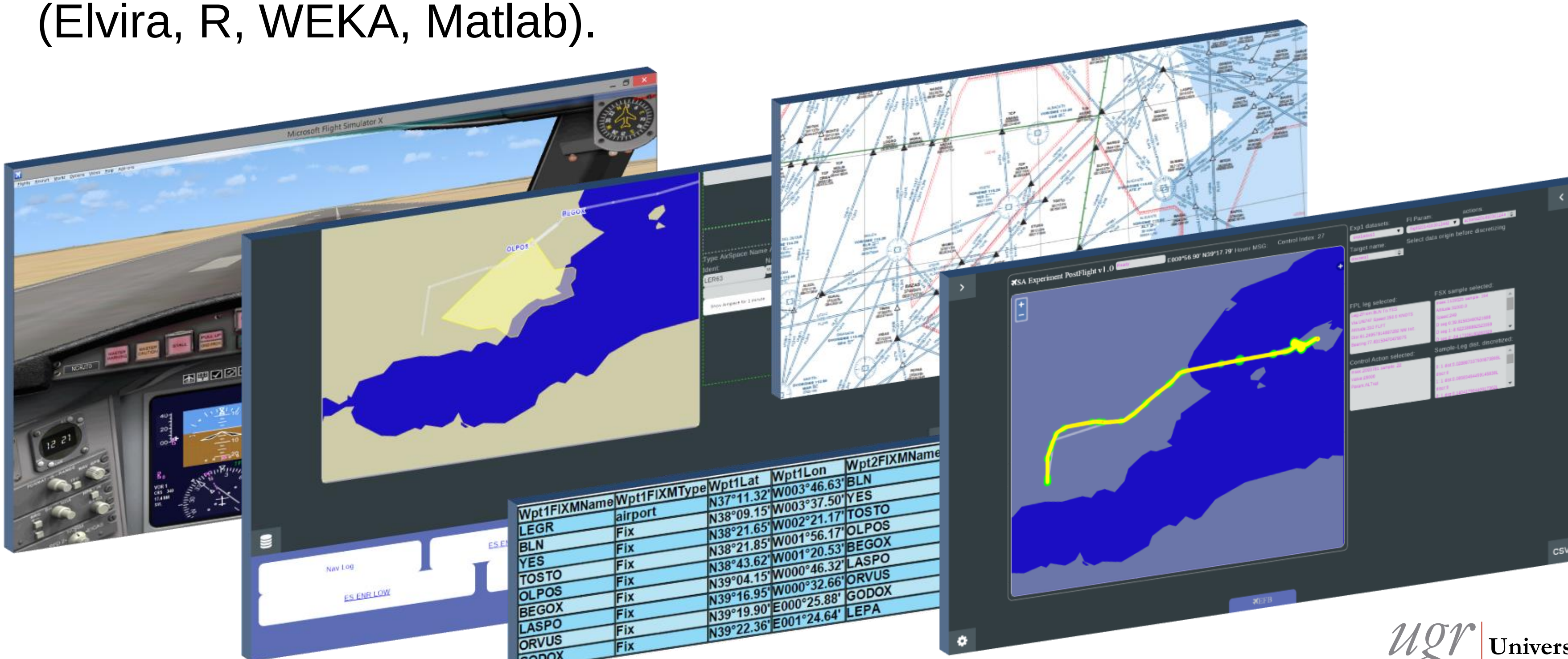
- Bayesian networks are specially appropriate when we have a large number of variables with non-deterministic dependencies among them.
- In real-time applications, DBN can explicitly represent temporal relations between the measured variables.
- DBNs can be learned from data and, at the same time, include expert human knowledge when available:
  - **<u>Combine effectively expert & data-inferred knowledge</u>**.
- Inference in DBNs can be done in a short time, even in the presence of a large number of variables and observations.
- But it is necessary to adapt to DBN restrictions:
  - Only consider 1 former time step.
  - Need to discretize continuous variables
    - Without losing information.
    - Avoid EK dependency.



Universidad de Granada

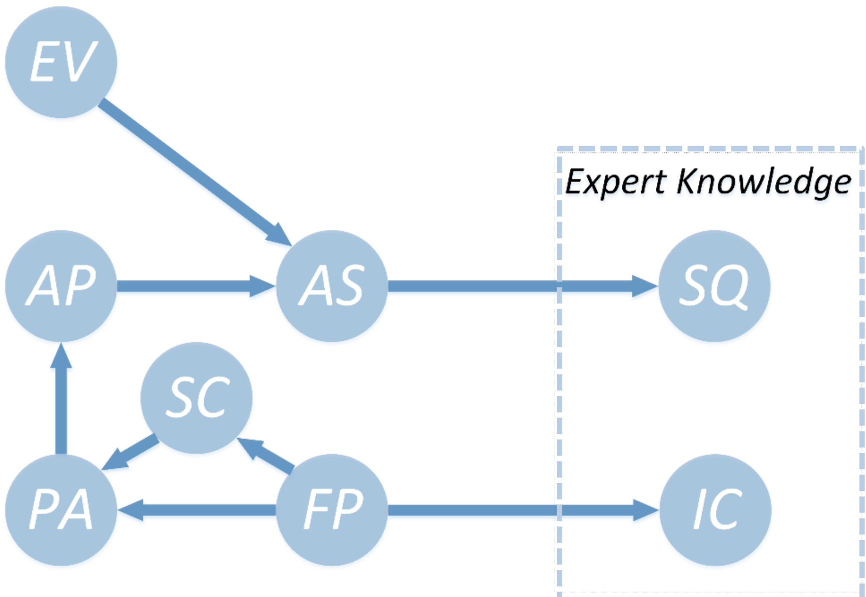
THE SIMULATION ENVIRONMENT

- Enables the performance of simulated flights.
- Provides the pilot with access to aeronautical information during the flight.
- Stores a dataset with the data collected during every experiment repetition.
- Performs certain post-flight transformations to the stored dataset and obtains data samples for exporting to external data processing tools (Elvira, R, WEKA, Matlab).



THE SIMULATION ENVIRONMENT

- The result is a dataset with different categories of variables, with an this expected simplified dependency model:



AS	Aircraft Situation	19
AP	Aircraft Parameters	4
EV	Environment Variable	4
PA	Pilot Action	21
IC	Information Check	37
SC	Situation Check	14
FP	Flight Plan	9
SQ	Situation Accuracy or Quality	11
119 variables, sampled every 7 seconds, 10 flights with durations from 54 to 64 minutes.		

- Additional expert knowledge can be added to the dataset.

SA ESTIMATION MODEL

- **<u>10-fold cross validation</u>**:
  - In this experiment, in order to learn a model, 10 flights were performed.
  - All are “correct” flights, without big deviations of the expected flight path and no “strange” pilot actions.
    - Different pilots have participated, different “flying styles”.
    - Same autopilot profile for all the flights. Limited automation.
  - The DBN probabilities for each flight are learned individually using the 9 remaining flights, using Elvira.
    - Further details: Morales and Moral, 2015

- Each flight contains the variables X_i (currently 119).
- Each variable X_i has a set of parents $\prod_{it}$
- **<u>Assumptions:</u>**
  - Markovian process: $\prod_{it}$ is included X_{t-1}
  - Stationary: $P(X_{it}|\prod_{it})$ is time independent

SA ESTIMATION MODEL

- Markovian assumption requires the creation of **summary variables**, some of them based on expert knowledge, that avoid loss of relevant information when only the previous time step is used to predict variables.
  - We are trying that these summary variables are route-independent.

- In future stages we plan to train the model to identify a **"low awareness" probability** that depends on how timely & accurate corrective actions are performed, in comparison with the learned "correct" pilot reaction.

- But before addressing this, it is necessary to optimize the discretization of continuous variables…

ugr Universidad de Granada

DISCRETIZATION & REGRESSION

1. Direct discretization approach:

- Continuous variables discretized with thresholds provided with expert knowledge.
- Most of these continuous variables are such that their value on a time *t* is a small variation of the value on previous time *t-1*.
- Model for the DBN learned using the discretization and an optimized set of parents for each variable, according to a comparison of different metrics: BIC, Akaike, K2.

DISCRETIZATION & REGRESSION

2. Learning a discretization from data:

- Based on a two step optimization:
  - First we compute an optimal discretization for each variable with the empty graph (no parents for any variable).
  - Then we learn the optimal DBN structure as in the previous method.
- To find an optimal discretization we use a greedy algorithm:
  - The possible interval limits are the middle points between the actual points of the variable in the learned data.
- Solves the problem of expert knowledge dependency.
- Performance not promising enough.

3. Learning the discretization + Linear Regression:

- Initial numerical estimation of each continuous variable X_{it} using linear regression and with variables in previous time as predictors Y^{i}_{t-1}.
  - Using ***R*** package *bestglm*
  - Number of predictors is limited in order to avoid overfitting:
    - number chosen: 4 predictors.
    - the best model is chosen using ***BIC*** criterion.
    - deterministic variable based on known values of variables at time step *t-1:*

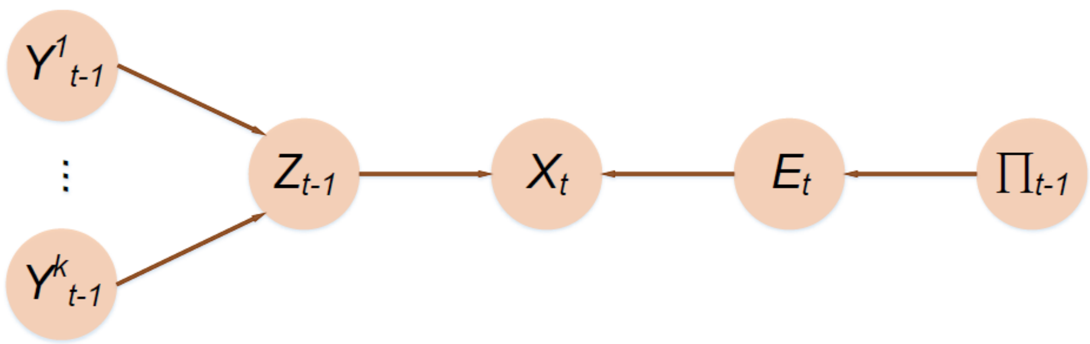
$$Z_{i(t-1)} = a + b_1 Y^1_{t-1} + \cdots b_k Y^k_{t-1}$$

    - error in the prediction:

$$E_{it} = X_{it} - Z_{i(t-1)}$$

3. Learning the discretization + Linear Regression:

- Once the regression is computed, the dataset is expanded by including the artificial variables Z_{it} and $E_{it:}$
- The full model is represented by:



Where Π_{t-1} are the parents of variable X_t.

RESULTS

Different discretization strategies are compared using the Logarithm of Probability (LP):

$$LP = \log(P((X_t)_{t\in T} \mid M)) = \log(P(X_{t_0})) + \sum_{t\in T\setminus\{t_0\}} \log(P(X_t \mid M, X_{t-1}))$$

Where M is the learned model and $t_o \in T$ is the initial time.

Higher values of LP are better. The best result corresponds to K2 score using summary variables.

Using regression always provides a higher LP:

	min Thr	10 Thr	20 Thr	50 Thr	100 Thr	regression
BIC Score	$-9.300e5$	$-6.690e5$	$-5.699e5$	$-6.360e5$	$-5.676e5$	$-2.475e5$
Akaike Score	$-9.310e5$	$-6.682e5$	$-5.655e5$	$-4.271e5$	$-5.075e5$	$-2.450e5$
K2 Score	$-9.305e5$	$-6.694e5$	$-5.653e5$	$-4.188e5$	$-3.602e5$	$-1.989e5$

Table 1: LP values for the estimation of all variables and using summary variables.

CONCLUSSIONS

- SA estimation is a relevant topic in flight safety & aviation human factors.
- We have implemented a simulation environment to collect relevant variables related with SA.
  - In accordance with state of the art aviation information management standards.
- Due to the limitations of DBN, it was necessary to perform a careful approach to the discretization of continuous variables and the creation of summary variables.
  - Regression methods allow us to obtain an accurate discretization that does not require specific expert knowledge.
- In return, we obtain a powerful tool to perform dynamic assessments based in large number of variables with complex relationships.
  - Future research to find a relevant measurement of SA.

Thanks for your attention

Questions?