

## An Architecture to Analyse Aviation Incidents

Jose Aguilar<sup>1</sup>, Donghui Shi<sup>2</sup>, José Gutiérrez de Mesa<sup>3</sup> and Danilo Chavez<sup>4</sup>

<sup>1</sup> CEMISID, Escuela de Ingeniería de Sistemas,  
Universidad de Los Andes, Mérida, Venezuela  
&  
Prometeo researcher, EPN-Quito, UTPL-Loja, Ecuador

<sup>2</sup> Department of Computer Engineering  
School of Electronics and Information Engineering  
Anhui Jianzhu University, Hefei, China

<sup>3</sup> Universidad de Alcalá, Dpto Ciencias de la Computación  
Edificio Politécnico-Campus Universitario, Alcalá de Henares (Madrid), Spain

<sup>4</sup>Escuela Politécnica Nacional -Quito, Ecuador

Copyright © 2017 Jose Aguilar et al. This article is distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

In this paper, we propose a hybrid intelligent model based on text mining and Fuzzy Cognitive Map (FCM), in order to determine the causes of aviation incidents. Our approach considers several dimensions of the problem, to define the possible causes of an aviation incident. It considers the human factors, one of the main causes of incidents, but also includes other aspects that normally are not considered in the analysis of Aviation incidents, such as the conditions of the flight or of the airplane. Particularly, we propose text mining tasks to extract the key information from the reports of the incidents, and a Multilevel FCM to integrate the different dimensions considered in our approach. We use a database about reports of aviation incidents to train and test our system. The preliminary results are very encouraging, because we can infer the reasons of the incidents.

**Keywords:** Aviation Incidents, Text mining, Fuzzy Cognitive Map

### 1 Introduction

The description of the airline operations is very complex due to the different dimensions to be considered: pilots, crew, maintenance, aircraft conditions, and

flight conditions, among others. The possible aviation incidents are vast, with a multitude of potential causal factors. In general, the pilot errors have been normally reported with one of the main causes of accidents [5, 12, 13], but this cause of the error has been trivialized, and does not contribute to the understanding of the real causal factors. There are other conditions leading to the flight anomalies or incidents, which normally are not considered.

Particularly, the Aviation Safety Reporting System (ASRS) has a collection of aviation incident reports, which provides information about unsafe situations that they have encountered during a flight or on the ground [5, 11]. These reports contain different types of data, and one of them is the textual description about each incident. There are several works that have used this data to identify the presence of human factors in aviation incidents. For example, in [5] analyzes anomalies voluntarily reported by pilots in civil aviation sector and identifies factors leading to such anomalies. They experiment with data obtained from the NASA ASRS database, to determine the associations between the contributing factors and the self-reported anomalies, using data mining. In [13] investigates the use of data streaming analytics to predict the presence of human factors in aviation incidents. They use four machine learning algorithms in their work. [11] analyzes the cause identification in the incident reports in the aviation domain. The cause identification system seeks to identify only those causes that can explain why the aviation incident described in a given report occurred. The difficulty of cause identification is because it is a multilabel categorization task, and the scarcity of annotated reports. In [4] is studied what factors are present leading to pilots submitting voluntary anomaly reports regarding their flight performance was conducted. They use the Diffusion Maps for performing dimensionality reduction on text records. High-dimensionality data were clustered and categorized. Dimensionality reduction techniques identified concepts or keywords within records, and allowed the creation of a framework for an unsupervised document classification system.

Normally, the previous works use modern methods of text mining, in order to extract useful information about the human factors in aviation incidents. But, we consider there is also a significant amount of aspects, which are not taken into account in order to reach conclusions about the causality of the aviation incidents. Each aspect defines a dimension of the possible cause of an aviation incident, specifically, the conditions of the airplane, and the condition of the flight. We propose the utilization of three dimensions, in order to analyze the airplane operations: human factors, airplane conditions, and flight conditions.

Thus, in this paper we analyze these multiple dimensions, and propose an intelligent model based on the FCM technique and text mining, to infer the possible causes of the incidents. The FCMs have been used in different domains [1, 2, 3, 9, 10], to model systems based on the representation of concepts that describe the main aspects of the modeled system (states, variables or characteristics of the system), and the causal relationships between them. FCMs use fuzzy logic theory to describe their structure and to infer answers of the map from a given data input. We propose to use this technique, because it allows developing a descriptive model about the

causal relationships among the variables that define a system. The text mining technique is used to extract the information, which is used to instance the FCM.

## 2 Context of the Aviation Incidents

To describe an aviation incident can be considered several dimensions, each one defines a set of concepts about one specific aspect, to analysis the aviation incidents from different point of views [4, 5, 7]. In this paper, we consider the next dimensions:

- Human Factors of the Pilots of airplane: it includes the concepts that describe the different aspects of the pilot of the aircraft, considering that they are one of the first causes of incidents.
- Conditions of the airplane: in this group are included the concepts about the condition of the airplane, its maintenance, and other information that can affect the airplane conditions.
- Condition of the flight: in this group is defined the set of specific conditions of the flight, such as the weather, the crew behavior, among other things.

### 2.1 *Human Factors of the Pilots or of the Supervision of the airplane*

It includes the aspects linked to human factors in an aviation incident [8, 13]:

- Unsafe Acts: it is divided into two categories: errors and violation. Errors are unintentional behaviors, while violations are a willful disregard of the rules. a) Errors: Skill-Based Errors, Decision Errors, Perceptual Errors; b) Violations: Routine Violations, Exceptional Violations.
- Preconditions for Unsafe Acts: is divided into three categories – environmental factors, condition of the operators, and personnel factors. a) Environmental Factors: physical and technological factors; b) Condition of the Operators: mental state, physiological state, and physical/mental limitations that affect practices or actions; c) Personnel Factors: crew resource management and personal readiness factors.
- Unsafe Supervision: is divided into four categories. a) Inadequate Supervision: The role of any supervisor is to provide their staff with the opportunity to succeed, and they must provide guidance, training, leadership, etc. to ensure the task is performed safely; b) Plan Inappropriate Operation: operations unacceptable during normal operation (e.g., risk management, crew pairing); c) Fail to Correct Known Problem: instances when deficiencies are known to the supervisor, yet are allowed to continue unabated; d) Supervisory Violation: instances when existing rules and regulations are willfully disregarded by supervisors.
- Organizational Influences: is divided into three categories. a) Resource Management: organizational-level decision-making regarding the allocation and maintenance of organizational assets; b) Organizational Climate: working atmosphere within the organization; c) Operational Process: organizational decisions and rules that govern the everyday activities within an organization (e.g. operations, procedures).

## 2.2 *Conditions of the airplane*

In this group, we define the main aspects that determine the conditions of the airplane, mainly linked to the maintenance [5, 7, 8, 11, 12]:

- General Information: describes general information about the airplane.
- Operational events: contains a list of potential events, which can be caused by maintenance errors, i.e. Flight delays or cancellations, Aircraft damage.
- Maintenance Error: lists the errors that could occur and lead to an event. i.e. improper installation or servicing, or repair.
- Contributing Factors: contains situational variables that could contribute to the maintenance errors.
- Error Prevention Strategies (Corrective actions): Define aspects that require to be investigated, such as: Procedures in the maintenance organization, Processes in the maintenance organization.
- Human factors for aircraft maintenance proficiency: Defines the aspects about the human errors in maintenance tasks, such as: Lack of Communication, Lack of Teamwork, Fatigue, Pressure, and Distraction.

## 2.3 *Conditions of the flight*

These concepts define the specific conditions of the flight, which have not been considered by the previous levels [5, 8, 12]:

- The Airline Cost: concepts about the cost of flight for the company: Flight operating expenses, Ground operating expenses, System operating expenses
- Operational Data: concept about the operation of the flight: Fuel Consumption (Kg), Aircraft Kilometers Flown, Aircraft Landings.
- Operational Performance Data: concepts about the quality of the operation of the flight: Number of Revenue Passengers Carried, Mishandled Bags.

The interest with this set of dimensions is to define the possible causes of incidents.

# 3 **System for the Analysis of Aviation Incidents**

We consider the dimensions defined in the section 2.1, in order to have a large view about the problem.

## 3.1 *Our Architecture*

Our system is composed by two modules (see Figure 1), one of the module is responsible by the analysis of the different sources of information around the flight, for example, the incident reports collected by the ASRS; and the other infers possible causes of the incidents, according to the information defined for the first one. The first module extracts information from the incident reports using text mining tasks. The second module uses a FCM to model the causal relationships among the different aspects considered in our approach, which is instanced with the information extracted from the first module.



We created the Categories defined in [5] (see Table 1). The descriptors defined in the previous sections, are added into a category. In this way, the narrative textual data are automatically assigned to the categories if they include a descriptor of a category's definition.

Code	Categories	Descriptor
1	Aircraft issues	hydraulic failure, valve, disconnected, vibration, light, leak.
2	Perceptual errors	disoriented, spatial disorientation, visual perception, misjudge, illusion, not heard
3	Rule-based errors	maneuver, emergency, decision, exceeded ability, low fuel, wrong response, wrong, inappropriate, misdiagnose
4	Skill-based errors	wrong direction, did not notice, stalled, inadvertently, spin, forgot, conditions, step, control, checklist, technique, visual scan
5	Unsafe conditions	misinterpreted, incapacitated, wake turbulence, fatigued, not ready, unsafe, briefing, complacency, situational awareness, attention, misinterpretation, illness, mental fatigue, violation, violated, sleep,
6	Unsafe supervision	hurrying, not qualified, failed to enforce, known problem, unqualified, fatigue, conflict, no training, oversight, unauthorized
7	Violations	command, directive, procedure, low altitude, aggressive maneuver
8	Weather	twilight, haze

**Table 1:** Categories

### 3.3 Determination of the Causes of the incidents using a FCM model

In this section, we use FCM to infer the causes of the incidents. The FCMs are based on the Cognitive Maps (CMs) theory, a tool to model systems based on the representation of concepts that describe the main aspects of the modeled system (states, variables or characteristics of the system), and the causal relationships between them. FCMs use fuzzy logic theory to describe their structure and to infer answers of the map from a given data input. For the design of the initial structure of the maps, we can use the knowledge of the experts or historical data on the phenomenon to model.

In general, each concept represents a state or a characteristic of the system, and is represented by nodes in the oriented graph. The graph's edges are the causal influences between the concepts. The causal relationships are expressed by both positive or negative signs, and different weights. The value of a node reflects the activation degree of a concept in a given time. This value is a function of the sum of all incoming edges and the value of the concept of the immediately preceding state. By each input state  $C(0)$  there is a trajectory that normally finishes in a global stability of the system. When CMs converge to an attractor, it is an exit state that answers to a question of the type, What happens if  $C(0)$  occurs?. For more details about FCMs, see [1, 2, 3, 10].

Our FCM is composed by a set of concepts that describe the different dimensions analyzed in our models about the aviation incidents. This multilevel approach follows

the classical structure of the multilevel FCM approaches presented in the literature [3, 6]. Specifically, in this work, our concepts are the categories defined in the previous section: a) *Human Factors of the Pilots of airplane*: Perceptual errors, Rule-based errors, Skill-based errors, Unsafe conditions, Unsafe conditions, Violations; b) *Conditions of the airplane*: Aircraft issues; c) *Condition of the flight*: Weather.

Additionally, we define a set of concepts to be inferred, called “Indicators about Aviation Incidents”. It is composed of the concepts that we are going to infer. They describe the possible cause of the aviation incidents. This group defines the set of concept to be inferred from our model, considering the different possible causes of incidents. Our model can infer a combination of possible causes of incidents, which is a more robust response to determine the causes of aviation incidents. These causes of incidents are: a) *Pilot faults*: in this case, when the concept is 1 is because the incidents are faults of the pilot; b) *Airplane faults*: in this case, when the concept is 1 is because the incidents are faults of the Airplane; c) *Environmental faults*: in this case, when the concept is 1 is because the incidents are faults of the company

The relationship among the concepts have been defined using the data from ASRS [12, 13], which has been processed in the previous phase. Some of these data have been used to train the FCM (to learn the weights of the relationship), and others to test the model. The learning mechanism used by our FCM to adjust the weights of the relationships is a classical supervised learning approach described in [1, 2, 3].

## 4 Experiments

To test the quality of our system, we use some of the data from ASRS and consist of 18734 incident reports. We process all the data with the first module, and then, we divide the data into two groups, one to train the FCM, and others to test the quality.

### 4.1 Processing of the sources of Information

Each incident report contains structured data fields such as time, place, environment, aircraft, component, personnel, events, etc. The unstructured textual data contain narrative fields, which described the detailed information about the incidents. Only the narrative field was used as the input of FCM. The Primary Factors contain: Aircraft, Airport, ATC Equipment/Nav Facility/Buildings, Company Policy, Environment –Non Weather Related, Equipment /Tooling, Human Factors, Incorrect /Not Installed /Unavailable Part, Logbook Entry, Manuals, MEL, Procedure, Staffing, Weather.

Using the text mining tool, each incident report instances several categories, with a value that defines the intensity (occurrence) of each category in the incident report (see table 2). The value is normalized between 0 and 1.

Additionally, for each report we determine the possible cause of the incident, according to the classification made in ASRS. It is used during the learning procedure of the FCM.

Code	Category name	Intensity
1	Aircraft issues	0.91
2	Perceptual errors	0.12
3	Rule-based errors	0.55
4	Skill-based errors	0.83
5	Unsafe conditions	0.85
6	Unsafe supervision	0.6
7	Violations	0.21
8	Weather	0.03

**Table 2.** Example of categorization of an incident report by the first module

#### 4.2 FCM model for Aviation Incidents Prediction

Particularly, we propose a FCM based on two groups (developed with the FCM Design tool [3, 6]). In our case, the second group describes the concepts that are inferred (indicators about the causes of the incidents) using the knowledge discovered in the first step (module), and represents the output of our system. The information to instance the concepts in the first group is generated from the first module. The concepts of the first group are defined in the table 2, and the concepts of the second group are the inferred concepts of the second module (see Table 3).

Inferred Concepts
Pilot faults
Airplane faults
Environmental faults

**Table 3.** Inferred Concepts

#### 4.3 Tests

Remember that the weights of the relationships of our FCM model are learnt. We train our model using the reports, to instance the categories defined in the table 2, with their respective causes of incidents. We use different size of the data to train, and we compare the conclusion of the incident reports with our results learned. The table 4 shows these results (errors).

% reports	Error
40%	0.2
60%	0.13
80%	0.11
90%	0.12

**Table 4.** Results

Now, we test with the rest of the data do not used to train. An example of the inferred concepts is given in table 5. According to the table 5, for this case, the possible cause of the incident was due to the airplane faults (0.7), and in a less way,



due to the pilot faults (0.4). The result shown in table 5 correspond to the input concepts defined in Table 3.

Pilot faults	0.4
Airplane faults	0.7
Environmental faults	0.2

**Table 5.** Inferred Concepts

## 5 Conclusions

In this paper, we have proposed a hybrid intelligent model based on text mining and FCM, which determines the causes of aviation incidents. Our approach considers several dimensions of the problems, and determines several causes of an incident. Additionally, it proposes the possible combinations of different causes, and not only one.

The text mining tasks can be used to extract more information. Additionally, the multilevel FCM can use more concepts, in order to carry out a more detailed description of the phenomena. In this way, the database about reports of aviation incidents can be best exploited for our system. The preliminary results are very encouraging, but we need to carry out more simulations. It doesn't define a unique cause, and using its causal relationship among the concepts, we can study the variables that most affect a given incident.

We obtain similar results that previous works. The main difference in our approach with these works [5, 6, 12, 7, 14] is that we exploit all the information in the reports, and using this information we infer the possible causes of the incidents. Our model can be easily extended if we know more concepts due to that we extract more details from the report. There are multiple dimensions in the analysis of this type of incidents (see section 2), which can be added in our model. Additionally, our FCM has a causal inference that can be used to determine the effects of the different factors to be considered in the determination of the causes of the incidents. Previous works have not a holistic vision of the problem.

In the future work, we will exploit the different capabilities of the FCM, to define a methodology for an exhaustive analysis of the reports of aviation incidents (using the idea of the previous paragraph), in order to determine the specific factors that affects the incident.

**Acknowledgments.** Jose Aguilar has been partially supported by the Prometeo Project of the Ministry of Higher Education, Science, Technology and Innovation of the Republic of Ecuador.

## References

[1] J. Aguilar, Dynamic fuzzy cognitive maps for the supervision of multiagent sys-

tems, Chapter in *Fuzzy Cognitive Maps*, Vol. 247, Springer Berlin Heidelberg, 2010, 307-324. [https://doi.org/10.1007/978-3-642-03220-2\\_13](https://doi.org/10.1007/978-3-642-03220-2_13)

[2] J. Aguilar, Different Dynamic Causal Relationship Approaches for Cognitive Maps, *Applied Soft Computing*, **13** (2013), 271–282. <https://doi.org/10.1016/j.asoc.2012.08.037>

[3] J. Aguilar, J. Hidalgo, F. Osuna, Fernando, N. Pérez, Multilayer cognitive maps to model problems, *2016 IEEE International Conference on Fuzzy Systems*, (2016), 1547-1553. <https://doi.org/10.1109/fuzz-ieee.2016.7737874>

[4] C. Andrzejczak, W. Karwowski, P. Mikusinski, Application of diffusion maps to identify human factors of self-reported anomalies in aviation Work, *Journal of Prevention, Assessment and Rehabilitation*, **41** (2012), 188-197.

[5] C. Andrzejczak, W. Karwowski, W. Thompson, The identification of factors contributing to self-reported anomalies in civil aviation, *International Journal of Occupational Safety and Ergonomics*, **20** (2014), 3-18. <https://doi.org/10.1080/10803548.2014.11077029>

[6] J. Contreras, J. Aguilar, The FCM Designer Tool, Chapter in *Fuzzy Cognitive Maps*, Vol. 247, Springer Berlin Heidelberg, 2010, 71-88. [https://doi.org/10.1007/978-3-642-03220-2\\_4](https://doi.org/10.1007/978-3-642-03220-2_4)

[7] F. Gürbüz, L. Özbakir, H. Yapici, Classification rule discovery for the aviation incidents resulted in fatality, *Knowledge-Based Systems*, **22** (2009), 622-632. <https://doi.org/10.1016/j.knosys.2009.06.013>

[8] D. O'Hare, Cognitive Functions and Performance Shaping Factors in Aviation Accidents and Incidents, *The International Journal of Aviation Psychology*, **16** (2006), 145-156. [https://doi.org/10.1207/s15327108ijap1602\\_2](https://doi.org/10.1207/s15327108ijap1602_2)

[9] E. Papageorgiou, C. Stylios, P. Groumpos, An Integrated Two-Level Hierarchical System for Decision Making in Radiation Therapy Based on Fuzzy Cognitive Maps, *IEEE Transactions on Biomedical Engineering*, **50** (2003), 1326-1339. <https://doi.org/10.1109/tbme.2003.819845>

[10] Z. Peng, L. Wu, Two-Level Fuzzy Cognitive Map Mining for Text Categorization, *International Journal of Digital Content Technology and its Applications*, **6** (2012), 296-302. <https://doi.org/10.4156/jdcta.vol6.issue2.35>

[11] I. Persing, V. Ng, Semi-supervised cause identification from aviation safety reports, *4th International Joint Conference on Natural Language Processing*, Vol. 2, 2009. <https://doi.org/10.3115/1690219.1690265>

[12] A. Srivastava, R. Akella, V. Diev, S.P. Kumaresan, D.M. McIntosh, E.D. Pontikakis, Zuobing Xu, Yi Zhang, Enabling the discovery of recurring anomalies in aerospace problem reports using high-dimensional clustering techniques, *2006 IEEE Aerospace Conference*, (2006).  
<https://doi.org/10.1109/aero.2006.1656136>

[13] D. Shi, J. Zurada, J. Guan, Identification of Human Factors in Aviation Incidents Using a Data Stream Approach, *50th Hawaii International Conference on System Sciences*, (2017), 1073- 1078.  
<https://doi.org/10.24251/hicss.2017.127>

**Received: April 23, 2017; Published: May 25, 2017**