UNIVERSITY OF GLASGOW

Safety Critical Systems AX Report (H)

Andrei-Mihai Nicolae
2147392n@student.gla.ac.uk

March 3, 2017

CONTENTS

1 INTRODUCTION

Artificial Intelligence has been one of the hot topics of this decade as it is surrounding our lives more as days go by. We can see it in objects ranging from aircrafts to smart home controllers.

However, as AI has become so powerful, it may come with a cost for the majority of us. Because of the large number of risks when letting Artificial Intelligence and Machine Learning robots/devices perform certain actions for us (Mannino [2015]), we need to devote a huge amount of effort in designing better safety-oriented architectures, well-crafted and thorough testing as well as regular check-ups and revisions. Therefore, we need to see a considerable increase in tools that assess and mitigate risks when introducing AI into systems of any type. Such a tool will be discussed in further detail in this report.

1.1 AI-Driven Technologies

The importance of introducing AI into a field with safety as a top priority is crucial. Because there is no human involved, the goal of the whole AI community is to let the machines actually take our place in performing certain actions, so that our tasks would be simplified. As good examples where it has become more and more developed, here is a list of examples:

- Transportation (driverless cars, subways)
- Game playing machines (Deepmind's Go playing machine that beat the en-titre champion) Silver [2016]
- Medical robotics
- Manufacturing machines
- Education

As the list can go on, we can see how AI spans throughout most of the major aspects of our lives, thus the need of careful monitoring its development.

1.2 How AI Can Go Wrong

Going past the many fields driven by Artificial Intelligence nowadays, we need to also take a close look at how many technologies have proven to be very prone to failure.

One interesting case is something that happened only a few months ago with an Uber driverless car going through a red light in front of San Francisco's Museum of Modern Art (Wakabayashi [2015]).
As it was recorded on camera, the car just rushes through a busy street on red light. This is a clear sign of how developers are not placing enough testing and robust checks before launching such a safety critical systems into an open environment.

We can see that the cause of the previous example could have been harmful for us humans. However, there have been cases where AI was involved and it was even deadly. Such an event is the killing of a Volkswagen employee who was grabbed and killed by one of the manufacturing robots in the plant (Dockterman [2015]). Moreover, a robot in one Silicon Valley mall struck a child on the head by mistake (Rocha [2016]).

In conclusion, after discussing various developments in the Artificial Intelligence world and how these safety critical systems can fail drastically, a tool for assessing and mitigating such risks will be presented in the rest of this report.

2 Tool

After extensive research, I came to the conclusion that for such an application the best technique that should be used is Model Checking adapted specifically for AI systems.

2.1 Reasons behind Choice

Firstly, it's needed to be pointed out that various techniques (e.g. Fault Tree Analysis, Effects and Criticality Analysis) work as well for some sub-fields of AI-driven systems, but I believe that Model Checking is eventually the optimal choice. This is because of various reasons which will be exposed below.

2.1.1 AI is Young

Because of the relatively young age of AI and ML in mainstream technology, there is quite a significant amount of improvements needed to be implemented. As shown before, there are many cases where they failed and lead to possible catastrophic outcomes.

Model Checking is a solution to the problem: having a model system, check automatically (and exhaustively) if that particular system meets some given specification. I believe this is a very good approach to the tool needed to be implemented because the model checking applies to finite state systems (Wah [2009]). As AI is young, we need to make sure (even at the cost of time and not so optimal design decisions) that the system behaves correctly and gives the correct output regardless of the situation.
As such, a technique such as Model Checking would automatically check all possible combinations and determine whether the system is prepared to use AI/ML safely or not.

2.1.2 State of the Art Usage of the Model

Mention here about NASA and SPIN (slides are in Downloads folder)

2.1.3 Available Tools

Talk about the large number of tools available for implementing model checking

2.2 Case Study

Add images here and talk about a detailed use case (linking to other studies showing that model checking is used throughout the world)

2.3 Implementation of the Tool

2.4 Example of list (enumerate)

1. First item in a list
2. Second item in a list
3. Third item in a list

3 Evaluation

4 Results

5 Conclusion

References

Eliana Dockterman. Robot kills man at volkswagen plant. 2015. URL `http://time.com/3944181/robot-kills-man-volkswagen-plant/`.

Adriano Mannino. Artificial intelligence: Opportunities and risks. 2015. URL `https://foundational-research.org/wp-content/uploads/2016/06/AI-Policy-Paper.pdf`.

Veronica Rocha. Crime-fighting robot hits, rolls over child at silicon valley mall, 2016. URL `http://www.latimes.com/local/lanow/la-me-ln-crimefighting-robot-hurts-child-bay-area-201`

David Silver. Mastering the game of go with deep neural networks and tree search, 2016. URL `https://gogameguru.com/i/2016/03/deepmind-mastering-go.pdf`.

Benjamin W. Wah. *Wiley Encyclopedia of Computer Science and Engineering*. 2009.

Daisuke Wakabayashi. A lawsuit against uber highlights the rush to conquer driverless cars. 2015. URL `https://www.nytimes.com/2017/02/24/technology/anthony-levandowski-waymo-uber-goo`