Olympic Aritificial Intelligence

Jaehyun Lee
Florida State University Panama City
Computer Engineering
Panama City, Florida
jl21bd@fsu.edu

Kameron Butler
Florida State University Panama City
Computer Engineering
Panama City, Florida
kjb21@fsu.edu

Tyler Coltrane
Florida State University Panama City
Electrical Engineering
Panama City, Florida, USA
tmc21@fsu.edu

***Abstract*— Bayesian Networks are probabilistic graphical models that use probability theory to model relationships between variables. They are particularly useful in situations where there are complex relationships between variables and it's difficult to determine causality. In this study, students used the GeNIe Modeler to create a Bayesian Network based on data sourced from an online community of data scientists and machine learning practitioners. The network was used to address three problems: the impact of hosting country on athletics' performance, the athletic events that produced the most USA women gold medalists, and the type of sport that will see the most Turkish medalists won. The GeNIe Modeler proved to be a versatile tool for probabilistic modeling and decision making. The results of the study showed that Bayesian Networks can provide efficient computation of probabilities and handle missing data.**

I. Introduction

Bayesian Networks are a type of probabilistic graphical model that uses probability theory to model relationships between variables. They are used in artificial intelligence to reason under uncertainty and make decisions based on uncertain information. The network consists of nodes representing variables and edges representing probabilistic dependencies between the variables. Bayesian Networks are particularly useful in situations where there are complex relationships between variables, and it's difficult to determine causality. One of the main advantages of Bayesian Networks is that they allow for efficient computation of probabilities, even when the number of variables is large. They can also handle missing data and incorporate new information as it becomes available. In summary, Bayesian Networks are a powerful tool in artificial intelligence for modeling complex relationships between variables and making decisions under uncertainty. The tool being used to model these Bayesian networks is the "GeNIe Modeler."[3] This tool can acquire a source of data and construct graphical models based on the data provided. The data source[1] of the data used in these networks is a from an online community of data scientists and machine learning practitioners that share their work for others to learn. Students will address the problems that they will complete, the contents of the data, the constraints put onto the data, and the imperfections of the Bayesian network in problem statement. When constructing the model of the Bayesian network, students simplified the data for the problems they wanted to solve. Students decide questions to respond to with the probabilities while using Bayesian networks to find the probability of each question defined.

II. Problem Statement

The problem statement consists of three outcomes the students are attempting to achieve. Problem one is wanting to seek if the hosting countries city impacts the athletics' performance. Problem two is seeing which athletic events will produce the most USA women gold medalists. Problem three is seeing which types of sport people will see the most Turkish medalists from. The data source [1] that students use for this contains an excel sheet of all the medal winners in the Summer Olympics from 1976 Montreal to 2008 Beijing. It includes each person, by name, which was awarded a medal within the period. Students decided that the data set could be grouped so that each person could be identified as a medalist. The Bayesian network that was built using the data compiled from the excel sheet is not perfect. The nodes of the network will have some effects on each other, such as the gender of the medalist can affect the country of the medalist. This may be due to the different point of views that each country has toward each sport. For example, a country that is known for softball will produce more female medalist from the country, causing a change in the country of the medalists. Within the excel sheet is also countries that do not exist anymore, which are not beneficial to the network. An example of this is west soviet, and east soviet. Another reason this network is not perfect, is that some of sports on the excel sheet are team sports, which mean that there will be more medals for the country if that sport is won. When students compiled the data in the Bayesian network, they focused on the highest probabilities to see the results clearer. Students realize that this data is not exact for the actual probabilities of these problems but is enough to find different conditional probabilities for future Olympics.

III. Tool Capabilities

The tool the students used to model the given data is called "GeNIe Modeler". GeNIe Modeler is a powerful software tool used for probabilistic modeling, decision making, and inference [3]. It is developed by the Decision Systems Laboratory at the University of Pittsburgh and is widely used in academia and industry for a range of applications, including risk analysis, medical diagnosis, and natural language processing. One of the key features of GeNIe Modeler is its

ability to construct graphical models. These models provide a visual representation of the relationships between variables and help users to understand the complex interactions between different factors. The software also includes a range of built-in tools for creating and editing graphical models, including support for Bayesian networks, decision trees, and influence diagrams. GeNIe Modeler also includes advanced inference algorithms that allow users to perform probabilistic reasoning on their models. This includes the ability to perform probabilistic inference, sensitivity analysis, and optimization. The software also includes support for machine learning techniques such as parameter estimation, feature selection, and model validation. This tool is a very versatile tool for probabilistic modeling and decision making. Its graphical interface and advanced inference algorithms make it an ideal choice for researchers and practitioners in a range of fields [3].

IV. Model

The model the students used for the Bayesian Network shown in *Fig. 1* below using the GeNIe Modeler[3]. This model shows how given the hosts country/city contributes to the country of the medalist and that contributes to sports. Sports and gender contribute to events which will then lead to the outcome of the medal color won.

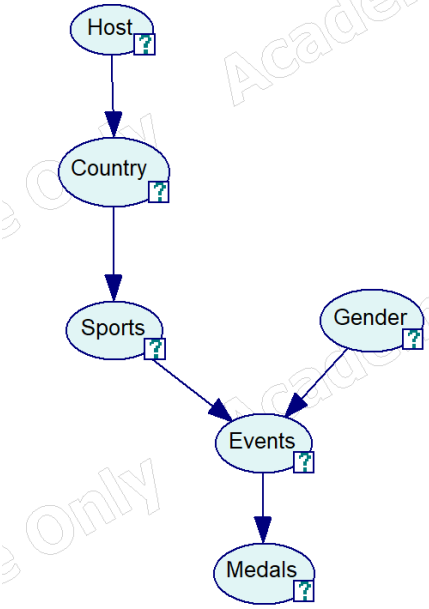

Fig. 1. The picture of the bayesian network students used.

V. Results

A. First Problem

The students first problem goal was to find out whether the hosting countries city actually impacted the athletics' performance. The main thing to consider was seeing if having a home advantage was really a thing. To test this out, the students looked into the relationship of the Host Nation and the Medalist from the hosting country. In the students results to support this statement, they investigated the Medalists from South Korea and how they did in the Seoul Olympic and compared them to their results in other Olympics. To calculate this, students searched for the probability of the medalists from Montreal (Canada) Olympics being South Korean, P(C = KOR, H = Montreal(CAN)). The calculated probability for that is shown below in *Fig. 2*.

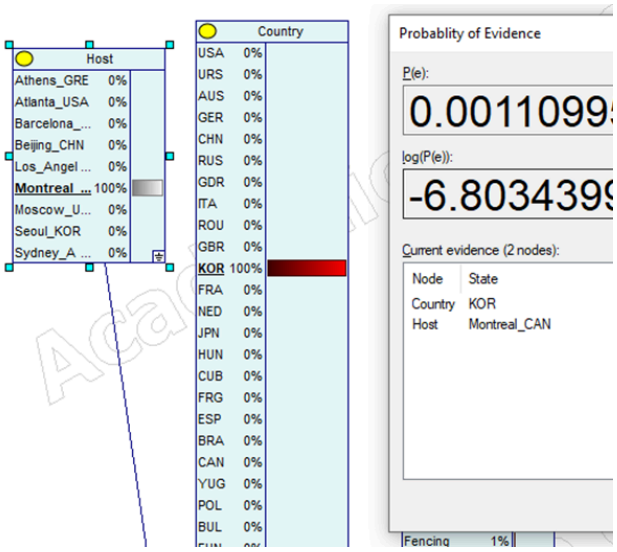

Fig. 2. Probability of the medalists from Montreal (Canada) Olympics being South Korean, P(C = KOR, H = Montreal(CAN)).

The next thing students calculated is the probability of the medalists from Seoul, Korea Olympics being South Korean, P(C = KOR, H = Seoul(KOR)). The calculated probability to those results is shown below in *Fig. 3*.

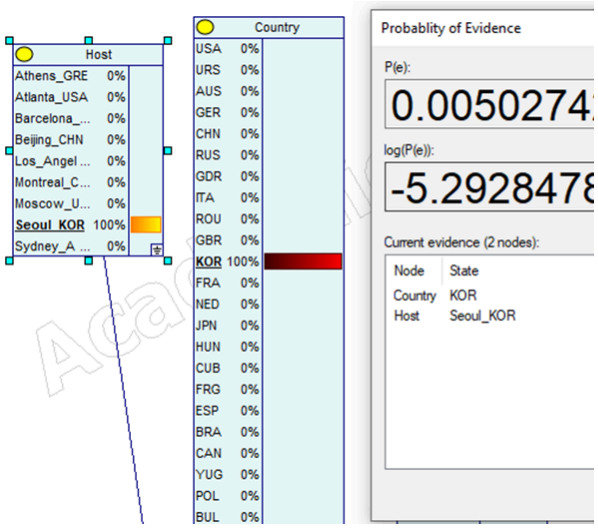

Fig. 3. Probability of the medalists from Seoul, Korea Olympics being South Korean, P(C = KOR, H = Seoul(KOR)).

To finalize supporting the home advantage probability, the last thing the students calculated is the probability of host nations given the country of the medalist is South Korea,

P(H|C = KOR). Shown below in *Fig. 4.* is the given probabilities for each host.

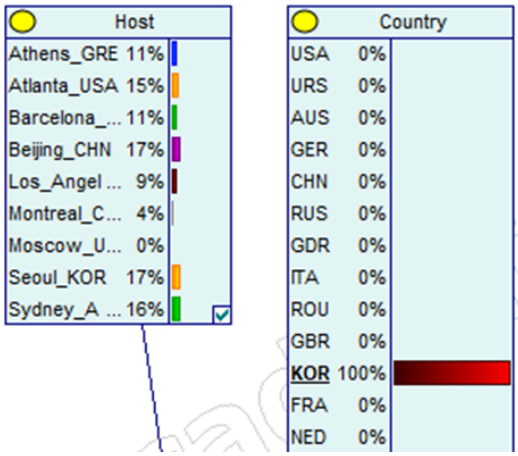

Fig. 4. Probability of host nations given the country of the medalist is South Korea, P(H|C = KOR).

B. Second Problem

The students second problem which dealt with finding which athletic events produced the most USA women gold medalists. What students achieved to find this is they knew they had to calculate P(E|C = USA, S = Athletic, G = Women, Medal = Gold). In *Fig. 5.* below, the probability results are shown.

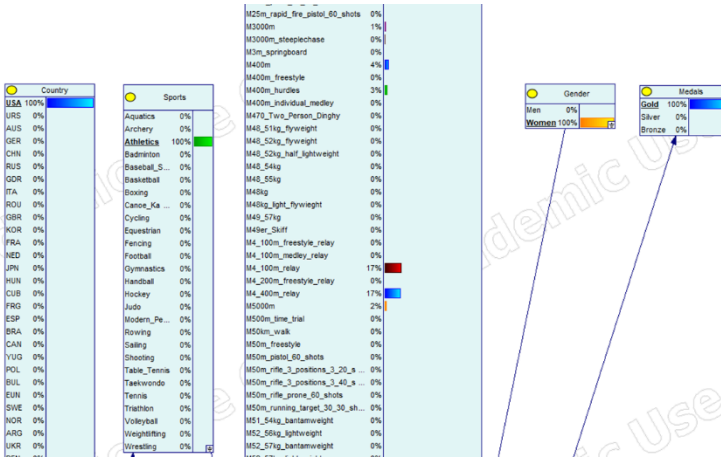

Fig. 5. Probability of athletic events that produced the most USA women gold medalists P(E|C = USA, S = Athletic, G = Women, Medal = Gold).

C. Third Problem

Problem three was to find which type of sport will see the most Turkish medalists won. To find this result, students seek to find the probability of sports given the Turkish medalists, P(S|C = TUR). The results are shown referenced in *Fig. 6.* below.

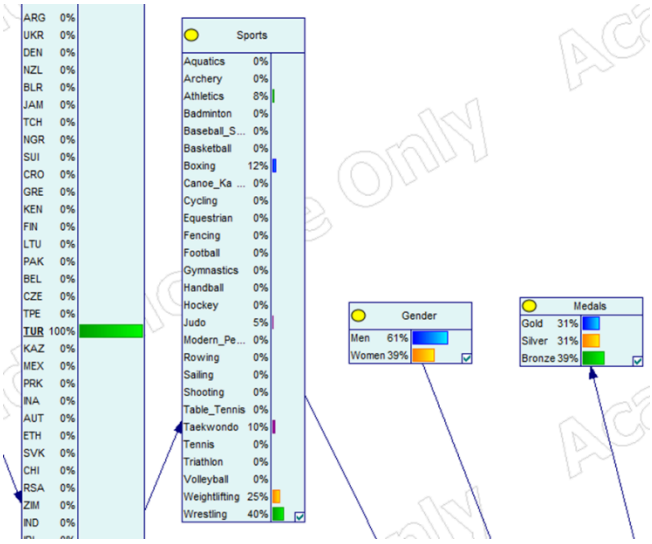

Fig. 6. Probability of sports given the Turkish medalists, P(S|C = TUR).

VI. Conclusion

Bayesian Networks are powerful tools in artificial intelligence for modeling complex relationships between variables and making decisions under uncertainty. The GeNIe Modeler is a robust software tool for probabilistic modeling and decision making, which can be used in a range of fields, including risk analysis, medical diagnosis, and natural language processing. The study showed that the Bayesian Network created by students using the GeNIe Modeler could successfully address complex problems such as the impact of hosting country on athletics' performance, the athletic events that produced the most USA women gold medalists, and the type of sport that will see the most Turkish medalists won. Despite imperfections in the data, the Bayesian Network provided efficient computation of probabilities, allowing for accurate results to be obtained. The study demonstrates the potential of Bayesian Networks and the GeNIe Modeler in aiding decision-making processes in a range of fields.

References

[1] Agrawal, Divyansh. "Summer Olympics Medals (1976-2008)." *Kaggle*, 3 Feb. 2020, https://www.kaggle.com/datasets/divyansh22/summer-olympics-medals?resource=download.

[2] "Introduction to Bayesian Networks: Bayes Server." *Introduction to Bayesian Networks | Bayes Server*, https://www.bayesserver.com/docs/introduction/bayesian-networks/.

[3] Druzdzel, Marek J., and Tomek Sowinski. "Genie Modeler." *BayesFusion*, https://www.bayesfusion.com/genie/#:~:text=GeNIe%20Modeler%20is%20a%20graphical,macOS%20and%20Linux%20under%20Wine.