Bayesian method of uncertainty propagation in inference network

Martin Števlík, Vasyl Rebresh, Vladyslav Yablinskyi
Department of Cybernetics and Artificial Intelligence – Technical University of Košice
martin.stevlik@student.tuke.sk, vasyl.rebresh@student.tuke.sk, vladyslav.yablinskyi@student.tuke.sk

Report prepared for the final project of the course Knowledge-Based Systems
taught by Professor Kristína Machová from the Department of Cybernetics and Artificial Intelligence
Faculty of Electrical Engineering and Informatics

Abstract – A knowledge-based system is a part of artificial intelligence which aims to capture the knowledge of human experts to support decision making. Examples of knowledge-based systems include an expert system, which is so-called because they rely on human expertise. The expert systems are the computer applications developed to solve complex problems in a particular domain. In our work, we focused on Bayes Theorem which is one of the most important formulas in all of probability. This formula is a core tool in machine learning and artificial intelligence. And we evaluate its application in knowledge-based systems.

Key words: machine learning, expert systems, knowledge based systems, Bayesian method, inference network.

1 INTRODUCTION

Who was Bayes?

Thomas Bayes (1701 — 1761) was an English theologian and mathematician that belonged to the Royal Society (the oldest national scientific society in the world and the leading national organisation for the promotion of scientific research in Britain), where other eminent individuals have enrolled, like Newton, Darwin or Faraday.

Bayes Theorem is a method of calculating conditional probability. The traditional method of calculating conditional probability (the probability that one event occurs given the occurrence of a different event) is to use the conditional probability formula, calculating the joint probability of event one and event two occurring at the same time and then dividing it by the probability of event two occurring. However, conditional probability can also be calculated in a slightly different style by using Bayes Theorem. Bayes Theorem or the Theorem of Conditional Probability is used for calculating the probability of a hypothesis (A) being true given that a certain event (B) has happened. This calculation is described in the following steps: [1]

- Determine the probability of condition B being true, assuming that condition A is true.
- Determine the probability of event A being true.
- Multiply the two probabilities together.
- Divide by the probability of event B occurring.

This means that the formula for Bayes Theorem could be expressed like this:

$$P(A|B) = P(B|A) * P(A)/P(B) \tag{1}$$

Here, P(A) and P(B) are probabilities of observing A and B independently of each other. That's why we can say that they are marginal probabilities. P(B|A) and P(A|B) are conditional probabilities. P(A) is called Prior probability and P(B) is called Evidence. P(B|A) is called Likelihood and P(A|B) is called Posterior probability. [2]

2 THE PRINCIPLE OF BAYESIAN METHOD

What is indeterminate knowledge?

Indeterminate knowledge does not take the form of rigorous statements. Uncertain knowledge often represents various intuitions and enlightened practices. Uncertain knowledge is a natural part of the expert's solution and as such is an inseparable part of the knowledge system. In case of uncertainty in origin knowledge, it is necessary not only to derive new knowledge but also to derive its uncertainty. We are talking about spreading uncertainty through the inference network. This spread takes place from assumptions to conclusions according to certain rules that create a model of work with vague information. Numerical models of working with indeterminate information are divided into:

1. Intensional models these models solve the whole problem globally. The task is to find the best distribution of uncertainty that suits every partial - marginal knowledge.

2. Extensive models are based on the principle of a locality. That is spreading the uncertainty only in some local environment. They are also based on the principle of extensibility. Extensive models assume the existence of combinational functions that do not change from case to case.

For the symbolic and intensional model, there is no generally valid uncertainty propagation model. For extension models, there is an uncertainty propagation model in the form of a set of combination functions. The combined functions are prescribed for the manipulation of uncertainty during its spreading by the inference network. The inferential network is composed of basic network types and there is one combined function for each of them.

These include 3 types: negation, conjunction and disjunction. Uncertainty in the knowledge base represents the transfer of knowledge through indefinite production rules. We know the uncertainty of the assumption and the uncertainty of the production rule. We get either a direct uncertainty of the conclusion or a contribution to the uncertainty of the conclusion if there are more production rules with that conclusion. We will get the uncertainty of the conclusion or the contribution using the CTR functions and we will compose the contributions using the GLOB functions.

The subjective Bayesian method is one of the most frequently used methods of working with uncertainty. It takes into account both the uncertainty of the rules and an prior uncertainty of the statements. An prior uncertainty is given before the start of the uncertainty spreading process. In the process of spreading uncertainty in the inference network, it changes to a posterior.

The uncertainty of the production rule can be in the Bayesian method of expression in two ways:

- absolutely
- relatively

In the absolute expression of uncertainty, conditional probabilities are used, namely P (H / E) as the probability of closing H if the assumption E is met.

The relative expression of uncertainty can take the form of a degree of adequacy and a degree of necessity. The degree of adequacy expresses the change in the chances of conclusion if the assumption is met. The degree of necessity expresses the change in the chances of conclusion in the event of non-fulfilment of the assumption. [3]

3 ALGORITHM DESCRIPTION

In our project, we are focusing on Bayesian method of uncertainty propagation in the inference network.

Suppose that in the inference process, the assumption E acquires some posterior probability P (E / É). É indicates a relevant comparison related to a specific case. The formula for P (E / É) has 2 variants where I is the interval and G is the estimate:

$$P(E/É) = P(E) + P(E)/I * G \tag{2}$$

$$P(E/É) = P(E) + (1 - P(E))/I * G \tag{3}$$

Based on a specific observation, the probability of conclusion can be composed of two components: support from fulfilment and support from non-fulfilment of the assumption, which after derivation can be imagined as a line, which is shown in the following figure.

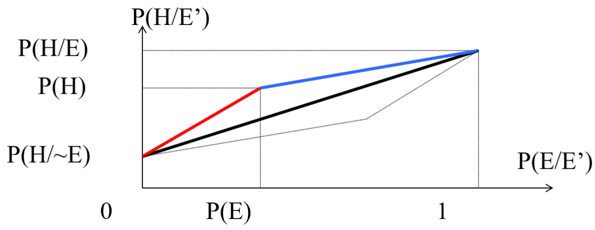

Figure 1: *Graph of the combined CTR function*

In such a way that this line also contains a prior probabilities of assumption and conclusion. By interpolating these three points, a linear polynomial function is created, which represents the definition of the combined CTR function in the following form:

$$0 <= P(E/É) <= P(E) => \tag{4}$$

$$P(H/É) = P(H/-E)+(P(H)-P(H/-E))/P(E)*P(E/É) \tag{5}$$

$$P(E) <= P(E/É) <= 1 => \tag{6}$$

$$P(H/É) = P(H)+(P(H/E)-P(H))/P(E)*[P(E/É)-P(E)] \tag{7}$$

Composing the contributions of individual rules with the same conclusion to the posterior probability of conclusion P (H / É1, É2, ...) is realized in a relative form. The weight of each rule is determined and defined as:

$$Lj = O(H/Ej)/O(H) \tag{8}$$

Where L represents the degree of adequacy of LS. If the assumption of statistical independence of individuals is assumed, the GLOB function can be defined as follows:

$$O(H/É) = P(H/É)/(1 - P(H/É)) \tag{9}$$

This definition of the GLOB function gives us the final a posterior chance of conclusion, which needs to be transformed into the final a posterior probability of conclusion.

$$P(H/É1, É2, ...) = O(H/É1, É2, ...)/1 + O(H/É1, É2, ...) \tag{10}$$

4 CONCLUSIONS

Bayesian belief networks and influence diagrams are attractive approaches for representing uncertain expert knowledge in coherent probabilistic form. Bayesian methods and classical methods both have advantages and disadvantages, and there are some similarities. When the sample size is large, Bayesian inference often provides results for parametric models that are very similar to the results produced by frequentist methods. Big advantage of Bayesian method is fact that method provides a natural and principled way of combining prior information with data, within a solid decision theoretical framework. You can incorporate past information about a parameter and form a prior distribution for future analysis.

When new observations become available, the previous posterior distribution can be used as a prior. All inferences logically follow from Bayes' theorem. Bayesian method and analysis has also many disadvantages where we can include silence of method, method does not tell you how to select a prior. There is no correct way to choose a prior. Bayesian inferences require skills to translate subjective prior beliefs into a mathematically formulated prior. If you do not proceed with caution, you can generate misleading results. In our specific task we found a great usage of Bayesian method for propagation in inference network.

REFERENCES

[1] J. Zornoza. (). "Probability learning i : Bayes' theorem," [Online]. Available: https://towardsdatascience.com/probability-learning-i-bayes-theorem-708a4c02909a. (accessed: 01.09.2019).

[2] J. Zornoza. (). "Probability learning ii: How bayes theorem is applied in machine learning," [Online]. Available: https://towardsdatascience.com/probability-learning-ii-how-bayes-theorem-is-applied-in-machine-learning-bd747a960962. (accessed: 01.11.2019).

[3] K. Machová, *Znalostné systémy v otázkach a odpovediac*, ser. International series of knowledge based systems. VEGA, 2005, ISBN: 9780198520115.