Self-Confidence in Autonomous Systems
Papers from the AAAI 2015 Fall Symposium

An Approximation of Surprise Index as a Measure of Confidence

Adam Zagorecki,
Center for Simulation and Analytics, Cranfield University
Defence Academy of the United Kingdom, Shrivenham, SN6 8LA, UK, a.zagorecki@cranfield.ac.uk

Marcin Kozniewski and **Marek J. Druzdzel**
Decision Systems Laboratory, School of Information Sciences, University of Pittsburgh,
Pittsburgh, PA 15260, USA, {mak295, druzdzel}@pitt.edu
Faculty of Computer Science, Bialystok University of Technology, Wiejska 45A, 15-351 Bialystok, Poland

Abstract

Probabilistic graphical models, such as Bayesian networks, are intuitive and theoretically sound tools for modeling uncertainty. A major problem with applying Bayesian networks in practice is that it is hard to judge whether a model fits well a case that it is supposed to solve. One way of expressing a possible dissonance between a model and a case is the *surprise index*, proposed by Habbema, which expresses the degree of surprise by the evidence given the model. While this measure reflects the intuition that the probability of a case should be judged in the context of a model, it is computationally intractable. In this paper, we propose an efficient way of approximating the surprise index.

Introduction

Bayesian networks (BNs) (Pearl 1988) are a modeling tool widely used in systems performing reasoning under uncertainty. A BN is an acyclic graph, where nodes correspond to random variables in the model and graphical part models explicitly stochastic dependencies among variables. Uncertainty is modeled by means of associating a set of conditional probability distributions with each node. A BN as a whole encodes the joint probability distribution (JPD) over a set of discrete random variables using a relatively small number of parameters.

In this paper, we focus on determining if a given set of observations (observed variables) is likely to occur in the context of a given model. If a set of observations is deemed to be unlikely, it can be due to a rare event or the model being incorrect or unable to capture the mechanism that produced the observations. The *surprise index* (Habbema 1976) was proposed to measure the likelihood of an observation in the context of a model. A major problem with the surprise index is that calculating it in real life models is computationally infeasible. In this paper, we propose a method to approximate the surprise index by exploiting some properties of the JPD and sampling algorithms.

In terms of real-life applications in the context of autonomous systems, we can envisage several scenarios. The first example would be diagnostic context where there are several separate Bayesian networks, each of them responsible for modeling one sub-system (e.g., avionics, propulsion, sensor suite, etc.) – it is a common practice in diagnostic BNs in order to make modeling and inference tractable. In the diagnostic process, observations (error messages from on-board monitoring system) would be passed to all BN models, and then each model would determine how relevant evidence is for it's context and the model(s) with the lowest surprise index would be used, avoiding the use of models which are not suitable for the case at hand.

The second scenario would include a set of models that would model various scenarios of enemy intent (e.g., routine intercept, hostile action, etc.). Each model would capture typical behavior for a given intent and current state of information on enemy actions would be entered to each model to determine most likely model based on surprise index. In both cases, we assume that models can have different sets of variables and, therefore, direct comparison of probabilities is not possible. The goal is to determine if given observations are likely to be produced by the model only assuming the knowledge of the model and the set of observations, without any external validation schema.

Modeling Using Bayesian Networks

Let a *scenario* be a set of outcomes – assignments of values to all variables modeled by a BN. A scenario may be considered as a single observation of all variables in the model. Let a *case* (evidence) be a subset of a scenario (assignment of values to some variables in BN). Let *case variables* be a subset of all variables modeled by BN, that are instantiated by the case. Let a *case domain* be a set of all possible assignments of values to case variables.

Figure 1 shows an example of a BN consisting of five binary variables. An example of scenario will be $s =$

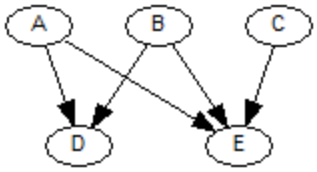

Figure 1: A Simple Bayesian network. There are five variables: A, B, C, D, and E. Each variable has possible outcomes: $\{a_1, a_2\}$, $\{b_1, b_2\}$, $\{c_1, c_2\}$, $\{d_1, d_2\}$, and $\{e_1, e_2\}$; additionally there are conditional probability tables $\Pr(A)$, $\Pr(B)$, $\Pr(C)$, $\Pr(D|A, B)$, and $\Pr(E|A, B, C)$.

$\{a_1, b_2, c_1, d_2, e_2\}$. An example case will be $c = \{a_1, c_1\}$. The case c variables are $\{A, C\}$ and case c domain contains $2^2 = 4$ elements.

A BN can be used to calculate the posterior probability distribution over variables of interest given the case evidence. As well, a BN enables to retrieve the marginal probability of observing a case (e.g., $\Pr(B = b)$) while keeping a simple representation of the JPD. This probability quantifies how often we may observe a given case. If this probability is low comparing to alternative cases, it may imply one of two situations: (1) a rare event, or (2) the case does not fit the model and, therefore, the model's predictions are unreliable. The former may be particularly relevant to the situations in which we have multiple models describing different domains and competing against each other. Examples of such, can be: a diagnostic system composed of many models for different subsystems, which would calculate probability of a fault in a subsystem given a set of observations; or various models designed to model different scenarios (such as enemy intent, describing various targets, etc.) which would attempt, based on available evidence, to determine which of these models is most relevant to the current situation.

Interpretation of Probability

Probability of observing a case is a quantity dependent on the context of a model, making it difficult to compare those quantities across different models. The bigger domain is covered by the model (number of variables, their outcomes), generally the probability of a case is lower. This is due to the fact that the probability over all cases must always add up to 1, and with the number of variables, the number of possible cases grows exponentially. That makes interpretation of a probability of a case between two different models challenging. Let us think about two models describing the probability of sequences of independent tosses of a fair coin. The first model describes 5 independent coin tosses, while the the second 20 independent tosses. In the first model, we have 2^5 possible equiprobable outcomes with probability $1/2^5 = 0.03125$, while the second model is defined by 2^{20} possible outcomes with probability $1/2^{20} \approx 10^{-6}$.

In both models the probability of getting all "heads" is different, but both models describe exactly the same mechanism and are equally correct. To address the problem of interpretation of a case probability in the context of different models we may use a *surprise index*.

Surprise Index

Surprise index of a case c (Habbema 1976) is defined as a sum of probabilities of less probable cases from the case domain. In particular:

$$\mathcal{SI}(c) \stackrel{\text{def}}{=} \sum_{c_i \,:\, \Pr(c_i) < \Pr(c)} \Pr(c_i) \; .$$

According to this definition, the more unexpected the case is, the smaller is the surprise index. In the example of coin tosses, in both cases we will have the same surprise index equal to 0.

The practical problem with using the surprise index is that its computation is intractable for regular models. It requires to iterate through all combinations of variables' outcomes (which is exponential in the number of variables) to calculate the exact value (Jensen et al. 1990).

In this paper, we show that it is still possible to reliably approximate the surprise index within reasonable computation time.

Approximating the Surprise Index

Based on Druzdzel's (1994) study of asymmetries in probabilities in the JPD, we propose a method to approximate surprise index. Druzdzel suggests that one can approximate the distribution of values in the JPD by the log-normal distribution (which is equivalent to approximation of logarithms of JPD values by the normal distribution) by applying central limit theorem.

A general method for calculating the properties of this distribution was presented by Bouckaert et al. (1996), however it is limited to cases where the conditional probability distributions in BN are strictly positive, i.e., do not contain parameters equal to zero.

JPD values distribution describes the probability of getting a given value of probability in a BN, when we pick a scenario from entire domain with uniform distribution – each scenario may be picked equiprobably. For our purpose we need a distribution of probability values of scenarios derived with a probability originating from the model. This distribution expresses the contribution of different values of probability to probability mass. Druzdzel (1994) has shown that this distribution can be reliably approximated by the tail of the log-normal distribution, assuming log-normality of probability values. Again, this is equivalent of normal distribution in logarithmic scale.

We use properties estimated from the JPD (mean and standard deviation) to create a normal approximation of values of probabilities given the model. So we have

$$\mu = E(X) = \sum_{i=1}^{M} \Pr(s_i) \log \Pr(s_i) \,, \qquad (1)$$

$$\begin{aligned} \sigma^2 &= E(X^2) - E^2(X) \\ &= \sum_{i=1}^{M} \Pr(s_i) \left(\log \Pr(s_i)\right)^2 - \mu^2 \,, \end{aligned} \qquad (2)$$

where M is the number of all scenarios and s_i is a particular scenario. Then the cumulative distribution function of the normal distribution (corresponding to the log-normal distribution) is the approximation of the surprise index (it integrates over all less probable cases by its definition). This may be expressed as

$$\mathcal{SI}(s) \approx F_{N(\mu,\sigma)}(\log(\Pr(s))) \ ,$$

where $F_{N(\mu,\sigma)}$ is the cumulative distribution function of the normal distribution with μ and σ as parameters.

Approximating Statistics by Sampling

In some cases, calculation of statistics necessary to compute an approximation of the surprise index may be impractical computationally. This problem arises when a BN model has large number of dependencies between variables or a case for which we wish to calculate the surprise index consists of a small number of variables. In those situations, we propose to use approximated values of the mean and the standard deviation obtained by means of stochastic sampling. As we are interested in estimating probability distribution over case probabilities, we must randomly sample cases from case domain using model. We estimate the mean and the standard deviation using the probabilities of cases calculated from the BN model.

To obtain a set of cases for the estimation of the mean and the standard deviation, we start with generating scenarios from the BN by means of probabilistic-logic sampling (Henrion 1988) and selecting only the values of the case variables.

The proposed algorithm takes as the input (1) a case c for which the surprise index is to be calculated and (2) a BN **B**. Below an outline of the algorithm is presented:

- based on input case c determine case domain
- generate n cases and calculate the marginal probability of each case using **B**
  - $s = \varnothing$
  - generate scenario s by traversing through **B**; for each variable X in **B**:
    - assign the value x_i of variable randomly using the conditional probability distribution for X
    - add x_i to the scenario s
  - collapse scenario s to case c_s that has the same case domain as c
  - calculate marginal probability for c_s and take it's logarithm ($\log \Pr(c_s)$)
- calculate mean and standard deviation of obtained logarithms of probabilities of c_ss (i.e. $\log \Pr(c_s)$s)
- calculate surprise index using cumulative distribution function of the normal distribution.

Acknowledgments

Partial support for this work has been provided by the National Institute of Health under grant number U01HL101066-01.

The algorithms described in this paper were implemented and tested using the SMILE$^{\textcircled{c}}$ reasoning engine for graphical probabilistic model, contributed to the community by the Decision Systems Laboratory, University of Pittsburgh, and available at https://genie.sis.pitt.edu/.

References

Bouckaert, R. R.; Castillo, E.; and Gutiérrez, J. 1996. A modified simulation scheme for inference in Bayesian networks. *International Journal of Approximate Reasoning* 14(1):55–80.

Druzdzel, M. J. 1994. Some properties of joint probability distributions. In *Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence*, 187–194. Morgan Kaufmann Publishers Inc.

Habbema, J. 1976. Models for diagnosis and detection of combinations of diseases. *Decision Making and Medical Care* 399–411.

Henrion, M. 1988. Propagating uncertainty in Bayesian networks by probabilistic logic sampling. In Lemmer, J., and Kanal, L., eds., *Uncertainty in Artificial Intelligence 2*. Elsevier Science Publishers B.V. (North Holland). 149–163.

Jensen, F. V.; Chamberlain, B.; Nordahl, T.; and Jensen, F. 1990. Analysis in HUGIN of data conflict. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, 519–528. Elsevier Science Inc.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. San Francisco, CA: Morgan Kaufmann Publishers Inc.