Title: Agent's Belief: A Stochastic Approach

Author: Nuno Cavalheiro Marques
Author: Veska Noncheva:

Postal Address: Departamento de Informática, FCT/UNL Quinta da Torre - 2829 -516 CAPARICA, PORTUGAL.
Telephone: +351- 212948536 (E. 10714)
Fax: 212948541
email: nmm@di.fct.unl.pt

Abstract: In this paper we present a theoretical model based on Bayesian Networks for evaluating agent's belief. We model the agent's belief as a threshold function of agent's certainty. The main problem is that agent's certainty is unobservable. This model is based on results from different tests. A propagation algorithm is presented. We also discuss an example for determining statistical agent's belief in a probability model. Finally, a proposal for application of our model to natural language subcategorization purposes is made.

Keywords: Agent's Belief, Statistical Tests, Weak Monotonous Bayesian Rule, Bayesian Network, Propagation Algorithm.

Topics: Belief Revision, Reasoning under Uncertaninty, Natural Language Processing, Multi-Agent Systems.

Agent's Belief: A Stochastic Approach

Veska Noncheva[1,2] Nuno Cavalheiro Marques[2]

[1] University of Plovdiv, Faculty of Mathematics and Informatics;
nonchev@plovdiv.techno-link.com

[2] CENTRIA - DI- Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa;
nmm@di.fct.unl.pt

Abstract. In this paper we present a theoretical model based on Bayesian Networks for evaluating agent's belief. We model the agent's belief as a threshold function of agent's certainty. The main problem is that agent's certainty is unobservable. This model is based on results from different tests. A propagation algorithm is presented. We also discuss an example for determining statistical agent's belief in a probability model. Finally, a proposal for application of our model to natural language subcategorization purposes is made.

1 Introduction

One of the main goals of Artificial Intelligence is to create agents. These agents embody expertise and intelligent behaviour. The states of the agents consist of components such as knowledge, belief, desire, intention, and obligation. Until now, most well known agent's belief work is based on logical approaches [2].

In this paper a general stochastic model for determining the agent's belief is proposed. This model has the following features:

- It is necessary to choose one of several alternative agent's belief states determined in advance.
- The agent's belief problem is formulated through the stochastic terms. One of the random variables, being of particular interest, is unobservable.
- A numeric measure called utility, measuring the profit of every agent's belief state, is given and the aim is the expected utility to be maximized.

One particular case of the problem under discussion is considered in [6]. Both Bayesian and utility Networks are used to present agent's belief. A possible application of this work to statistical subcategorization [3,4] learning is discussed.

The agent's belief can be presented through a threshold function of the degree of the agent's certainty (see [5]). However, the agent's certainty is unobservable, and because of that testing is needed. The test results are a kind of estimators of the degree of agent's certainty. The information from the tests is used to make decision about the agent's belief. Furthermore, the higher result values received from the first tests will result in lower requirements for the next test results to be obtained.

We are to discuss the decision-making rules having a monotonous form. For instance, the agent is certain that a statement is true, if the value of the variable, representing the certainty, is bigger than the threshold value preliminary determined. The agent, on the other hand, rejects the statement, if the value of the variable is smaller than the threshold value.

2 General Stochastic Model of the Agent's Belief

The general formulation of the problem for determining the agent's belief state is following:

1. Let X_i **,** $i=1,2,\ldots n$ be continuous random variables defined on sample spaces $\Omega_i = [0,1], i = 1,\ldots,n$, which random variables can be observed. We interpret X_i as a test result and $X_1, X_2, \ldots, X_n$ as a sequence of the results from the tests.
2. Let T be a continuous random variable defined on $\Omega_t = [0,1]$, which cannot be observed and it is being interpretted as *the agent's certainty*.
3. The Bayesian model of the probability structure is known; consequently, the joint probability distribution $f(x_1, \ldots, x_n, t)$ of the random variables $X_1,\ldots, X_n, T$ is known, as well.
4. The finite set of the possible agent's belief states D is known, too.
5. The utility function $U(t,d)$: $\Omega_t \times D \rightarrow [0,1]$ is also known.

A decision-making rule is the $\delta(x_1,\ldots,x_n)$ rule, which for each possible realization $(x_1,\ldots,x_n)$ of the random vector $(X_1,\ldots,X_n)$ determines which state $a_j \in D$, $j=1,\ldots,k$, will be acquired by the agent's belief. That is, the decision-making rule is a function of random variables $X_1,\ldots,X_n$ defined on $\Omega_1 \times \ldots \times \Omega_n$ and with range space D. The goal is to find a decision-making rule, which is to maximize the expected utility.

If there are several decisions, resulting in one and the same maximal expected utility, then we can consider each of these decisions as optimal. In this case the randomised decision-making rules are acceptable, but they have no priorities.

It is intuitively obvious that the high value of the result from test i will result in low requirements towards the result from test j, when $j>i$. That is, the preliminary obtained information influences the decision-making rules. The decision-making rules in which the decisions from the test j are functions of the obtained result from the test i, $i<j$, are called *weak rules*.

It is natural to discuss the decision-making rules, having a *monotonous form*, i.e. the rules with threshold points $x_i^c, i = 1, 2, \ldots, n$, forming partitions of the sample spaces $\Omega_i = [0,1]$ in the following manner $\Omega_i = A_i + \overline{A_i}, i = 1,..., n$, where $A_i = \{x_i : x_i < x_i^c\}$ and $\overline{A_i} = \{x_i : x_i \geq x_i^c\}$. Therefore, the problem for determining the agent's belief state means that we are to find n threshold points $x_i^c, i = 1,2,..., n$, for each $\boldsymbol{X}_i$, $i = 1,2,..., n$, which points are optimal in accordance with the Bayesian approach.

Hence the purpose is to find a *weak monotonous Bayesian rule* for determining the agent's belief state.

3. Example

Statistical inferences are conclusions for different characteristics of the population, being made on the basis of observations and assumptions for the population. The most popular and correct form of statistical inferences is the statistical hypothesis.

The assumptions may be for independence, symmetry, normality, stationarity, etc. The assumptions, being strict mathematical notions, are either incredibly close to our intuitive concepts for independence, stationarity, or have a proper interpretation in the real world.

However, as all mathematical notions have shortcomings, the assumptions have theirs as well. Strictness is probably the main one. This means that the requirements are so strict, that practically the assumptions cannot be verified.

To put it simply, when we say that two events are independent or the population distribution is normal, we put much more *belief* in these assumptions, than we could mathematically verify.

Consider statistical agent's belief in the probability model i.e. in the type of population distribution.

In order to form its belief in the type of population distribution, the agent can start with a statistical test for symmetry. When the hypothesis for symmetry cannot be rejected, the agent has to continue with tests for normality. Usually a Goodness-of-fit test is first used and after that the user opinion is asked. The results from these three tests are respectively 1-p_1, 1-p_2, where p_1 and p_2 are the p-values of the statistics of the two statistical tests, and the degree of the user certainty of normality, represented as numbers in the interval [0,1].

Designate with X_1 the observed value 1-p_1, where p_1 is p-value of the statistics of the test for symmetry. Designate with X_2 the observed value 1-p_2, where p_2 is p-value of the statistics of the test for normality. Designate with X_3 the degree of the user's certainty of normality. Designate with T the agent's certainty of normality, which cannot be observed.

Assume that X_1, X_2, X_3 and T are continuous random variables with a joint probability density function $f(x_1,x_2,x_3,t)$.

The decision rule $\delta(x_1, x_2, x_3)$ determines the state a_j, $j=0,1,2,3$, of the agent's belief for each possible realization (x_1, x_2, x_3) of the random vector (X_1, X_2, X_3).

The weak decision rule δ in this case has the form:

$$\{(x_1,x_2,x_3):\delta(x_1,x_2,x_3)=a_0\}=A_1\times[0,1]\times[0,1]$$
$$\{(x_1,x_2,x_3):\delta(x_1,x_2,x_3)=a_1\}=\overline{A_1}\times A_2(x_1)\times[0,1]$$
$$\{(x_1,x_2,x_3):\delta(x_1,x_2,x_3)=a_2\}=\overline{A_1}\times\overline{A_2}(x_1)\times A_3(x_1,x_2)$$
$$\{(x_1,x_2,x_3):\delta(x_1,x_2,x_3)=a_3\}=\overline{A_1}\times\overline{A_2}(x_1)\times\overline{A_3}(x_1,x_2)\,,$$

where A_1 and $\overline{A_1}$ are the sets of values of the test leading respectively to the rejection and to the acceptance of the statement for symmetry, $A_2(x_1)$ and $\overline{A_2}(x_1)$ are sets of values of the pre-test leading respectively to the rejection and to the acceptance of the statement for normality, $A_3(x_1,x_2)$ and $\overline{A_3}(x_1,x_2)$ are sets of values of the post-test leading respectively to the rejection and to the acceptance of the statement for normality.

The weak monotonous rule for making a decision about the agent's belief is defined as follows:

$$\delta(x_1,x_2,x_3) = \begin{cases} a_0, & \text{if } X_1 < x_1^c, X_2 \in [0,1], X_3 \in [0,1]) \\ a_1, & \text{if } X_1 \ge x_1^c, X_2 < x_2^c(x_1), X_3 \in [0,1] \\ a_2, & \text{if } X_1 \ge x_1^c, X_2 \ge x_2^c(x_1), X_3 < x_3^c(x_1,x_2) \\ a_3, & \text{if } X_1 \ge x_1^c, X_2 \ge x_2^c(x_1), X_3 \ge x_3^c(x_1,x_2)\,, \end{cases}$$

where x_1^c, x_2^c, x_3^c are the threshold values for X_1, X_2, X_3, and a_j , *j=0,1,2,3* are the following agent's belief states:

- a_0 - the agent rejects the assumption for the symmetry of the distribution, describing the population under investigation. In the process of the statistical analysis the agent will use the median as the best estimate of the "center" of the distribution since the mean is strongly influenced by outliers in the data.
- a_1 - the agent rejects the assumption for normality of the distribution. It will make only use of the assumption for symmetry in the statistical analysis.
- a_2 - the agent supposes (suspects) that the distribution of the population being investigated is normal. In the process of the statistical analysis it will make use only of tests which are not sensitive to moderate deviations from the assumption for normality. An example of such a robust test is the *t*-test.
- a_3 – the agent convinced that the distribution describing the population is normal. It will also use statistical tests, which are sensitive to deviations from the assumption for normality. Such tests are, for example, Pearson's, Fisher's and Bartlett's tests.

Assume that utility structure has the following form:

$$u(t) = \begin{cases} u_1(\mathrm{t}), & \text{if } a_0 \\ u_2(\mathrm{t}), & \text{if } a_1 \\ u_3(\mathrm{t}), & \text{if } a_2 \\ u_4(\mathrm{t}), & \text{if } a_3 \end{cases}$$

where $u_i(t)$, $i=1,2,3,4$ are continuous, monotonic and bounded functions.

It is well known that if the data is lognormal, that means that it will be normally distributed after a log transformation. Therefore, this model could be improved by adding possibility for data transformation, thus trying to obtain normally distributed responses.

The aim is to find a weak monotonous Bayesian rule for determining the statistical agent's belief in the probability model. That means to find the threshold points x_1^c, x_2^c, x_3^c, which are optimal in accordance with the Bayesian approach.

4. Data Structures

Let us assume that $\boldsymbol{X}=\{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\}$ is a finite set of continuous random variables. The *event tree* is a binary treelike structure having the following properties:

- the nodes and the leaves are mapped events.
- the sample space Ω is mapped in the root.
- each node has 0 or 2 children.
- If nodes A_l and A_r are respectively left and right child of A_i node, then A_l maps the event $\{X_l < x_l^c\}$, whereas A_r maps the complementary event $\{X_l \geq x_l^c\}$. Thus, we may designate the following equation: $\overline{A_l} = A_r = \{X_l \geq x_l^c\}$.

The event tree is in *canonical form* if the indices of the random variables - associated with the nodes - aligned from the tree root to the leaves and from left to right, coincide with the first *n* natural numbers. From now on we are to consider event trees in canonical form only.

The event tree, presenting the conditions of the decision rule from *Example* is represented in *Figure 1*.

The path that goes from the first level to a leaf in the event tree is called a *factor*.

We must bear in mind that we are to interpret the factor as events simultaneously occurring, i.e. as an intersection of the factor's events.

Let $F=\{F_i, i=1,2,\ldots,n+1\}$ be the set of the factors in the event tree. It presents the decision rule conditions. For convenience's sake we are to number the factors in event tree from left to right.

It is with each factor F_i that one of the agent's belief states is associated. In such case we say that the set of agent's belief states is associated with the set of factors from the event tree. The decision-making rule for the agent's belief state can be presented by the above mentioned sets.

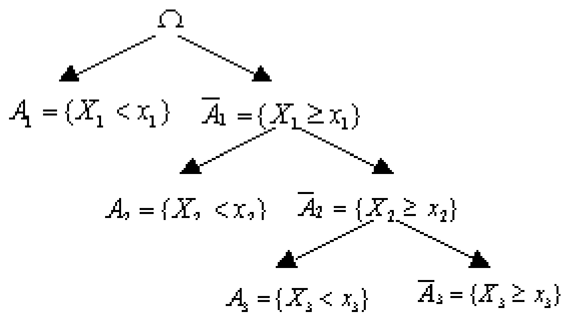

Fig. 1. Event tree from *Example*.

Further on, we will associate a utility function u_i with each factor F_i, $i=1,2,\ldots,n+1$, which is to say that each leaf from the event tree is associated with a utility node. Therefore, a set of utility nodes is associated with the set of factors in the event tree. Consequently, a set of factors, as well as, a set of utility functions is associated with the event tree.

The pair (F, U), where F is the set of factors and U is the set of utility nodes - both associated with event tree - is called a *utility network*.

5. Propagation Algorithm

We are now ready to present a propagation algorithm for generation of equations, whose roots are the optimal values $x_i^c, i = 1,2,\ldots, n$. The theoretical backgrounds of this algorithm are given in [5].

Algorithm: An algorithm for generating symbolic integral equations.

Input:

- A utility network, associated with the event tree.
- A Bayesian network

Output: n symbolic integral equations.

Initialization: Each A_i leaf in the event tree is associated with two symbol variables as follows:

A_i.*head*:=$U(A_i)$, where $U(A_i)$ is the utility function, associated with A_i leaf,

A_i.*tail*:=ϕ, where ϕ is the empty set.

Steps:

1. Find $(A_i, \overline{A_i})$ element with the greatest index ***i*** from the set of paired leaves, i.e. the leaves having the maximum level number in the event tree.

2. Find $\psi(A_i)$ - the list of the random variables associated with A_i leaf, then find i – the index of the element with the greatest index from $\psi(A_i)$.

3. Find the left hand of the equation:

$\boldsymbol{left}$:=E{[$\overline{A_i}$.*head* (T)-A_i.*head* (T)]/$\psi(A_i)$ }+ $\overline{A_i}$.*tail* - A_i.*tail*

4. Generate the equation:

$\boldsymbol{left}$ =0

5. Find A_k=π(A_i) , which is the paired leaves parent $(A_i, \overline{A_i})$. If it turns to be the tree's root go to *step* **9**. Otherwise designate $\overline{A_i^{\,0}} = [x_i^{\,c};1]$ and find the random variables list $\psi(A_k) = \psi(A_i) \setminus X_i$.

6. If π(A_i)= $\overline{A_k}$, where π(A_i) is A_i node's parent, then accept $\overline{A_k}$.*head*:= A_i.*head*,

$$\overline{A_k}\ .tail := \int\limits_{\overline{A_i^{\,0}}} \{\boldsymbol{left}\}.f_i(\boldsymbol{x}_i / \psi(A_k))\mathrm{d}\,\boldsymbol{x}_i + A_i\ .tail,$$

else accept $\boldsymbol{A}_k$.*head*:= $\boldsymbol{A}_i$.*head*,

$$A_k\ .tail := \int\limits_{\overline{A_i^{\,0}}} \{\boldsymbol{left}\}.f_i(\boldsymbol{x}_i / \psi(A_k))\mathrm{d}\,\boldsymbol{x}_i + A_i\ .tail$$

7. Remove A_i and $\overline{A_i}$ leaves.

Hence, we have articulated an equation for x_i^c and a new utility network.

8. Repeat the steps above beginning with *Step* **1**.

9. End.

The problem for decision making about the agent belief state, presented in *Example*, may be represented by both Bayesian network and the utility network in *Figure* 2. Then the equations, generated by means of Propagation algorithm are:

$$E[u_2(T) - u_1(T)/x_1] + \int\limits_{\overline{A_2^0}} \{E[u_3(T) - u_2(T)/x_1, x_2] +$$

$$+\int\limits_{\overline{A_3^0}} E[u_4(T) - u_3(T)/x_1, x_2, x_3] f(x_3/x_1, x_2)dx_3\} f(x_2/x_1)dx_2 = 0$$

$$E[u_3(T) - u_2(T)/x_1, x_2]$$

$$+\int\limits_{\overline{A_3^0}} E[u_4(T) - u_3(T)/x_1, x_2, x_3] f(x_3/x_1, x_2)dx_3 = 0$$

$$E[u_4(T) - u_3(T)/x_1, x_2, x_3] = 0$$

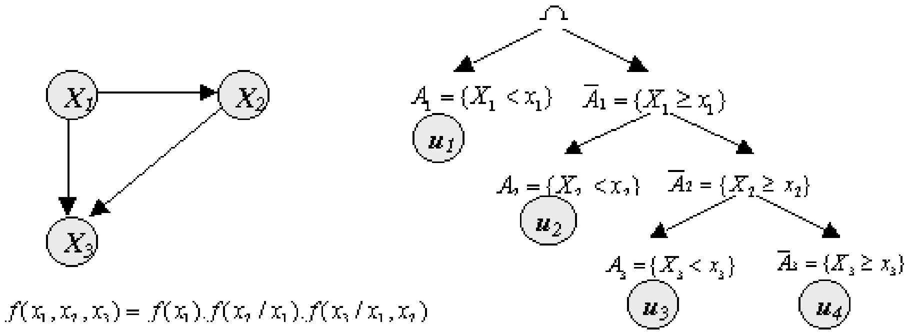

Fig. 2 Bayesian network and the utility network from *Example*.

6 Conclusions and Future work

The conception of agent's belief is a powerful abstraction and we are going to apply it in order to model the behaviour of an agent dealing with natural language processes. Statistical verb subcategorization is addressed in [3,4] by using loglinear modelling. One important problem in loglinear modelling is to find the most suitable model. We think that this problem can be stated in the framework of the agent's belief theory here proposed. In future work we will try to present the choice of the best loglinear model as an agent's belief in the goodness-of-fit of a loglinear model.

[1] uses undirected graph theory to represent loglinear models that contain only single-factor and two-factor terms. Representation of loglinear models based on Bayesian networks will be the topic of another paper and it will allow automatically selection among several possible loglinear models for dictionary building [3,4]. This application will help us to devise both a practical application to our model and a new way to assign good estimates for the probabilities. In this way, statistical subcategorization will be improved by the agent's belief model here proposed.

References

[1] J. N. Darroch, S. L. Lauritzen, and T. P. Speed: Markov fields and log-linear models. Ann. Statist. 8: 522-539. 1980.

[2] M. Georgeff, B. Pell, M. Pollack, M. Tambe, M. Wooldridge: The Belief-Desire-Intention Model of Agency. Jörg P. Müller, M. P. Singh, A. S. Rao (eds.), Proceedings 5th International Workshop (ATAL'98): Intelligent Agents V: Agent Theories, Architectures, and Languages. *Lecture Notes in Artificial Intelligence* 1555, pp. 1-10. Springer Verlag. 1998.

[3] N. Marques, J. Gabriel Lopes, and C. Agra Coelho. Learning Verbal Transitivity Using LogLinear Models. In Claire Nédellec and Céline Rouveirol, editores, Proceedings of the 10th European Conference on Machine Learning, *Lecture Notes in Artificial Intelligence* 1398, pp. 19-24, Chemnitz, Germany. Springer Verlag. 1998.

[4] N. Marques, J. Gabriel Lopes, and C. Agra Coelho. Using Loglinear Clustering for Subcategorization Identification. In Jan M. Zytkow and Mohamed Quafafou, editores, Proceedings of the Second European Conference on Principles of Data Mining and Knowledge Discovery, *Lecture Notes in Artificial Intelligence* 1510, pp. 379-387, Nantes, France. Springer Verlag. 1998.

[5] V. Noncheva. Agent Belief: Presentation, Propagation, and Optimization. *INFORMATICA*, 10(4): 403-426, 1999.

[6] V. Noncheva: A Statistical Agent's Belief in Gaussian Model. *Mathematica Balkanica*, 14(3-4), 2000.