Content Determination for Natural Language Descriptions of Predictive Bayesian Networks

M. Pereira-Fariña[a] **and Alberto Bugarín**[b]

[a]Departamento de Filosofía e Antropoloxía Social,
Universidade de Santiago de Compostela, martin.pereira@usc.es
[b]Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS),
Universidade de Santiago de Compostela, alberto.bugarin.diz@usc.es

Abstract

The dramatic success of Artificial Intelligence and its applications has been accompanied by an increasing complexity, which makes its comprehension for final users more difficult and damages their trustworthiness. Within this context, the emergence of Explainable AI aims to make intelligent systems decisions more transparent and understandable for human users. In this paper, we propose a framework for the explanation of predictive inference in Bayesian Networks (BN) in natural language to non-specialized users. The model represents the embedded information in the BN by means of (fuzzy) quantified statements and reasons using the a fuzzy syllogism. The framework provides how this can be used for the content determination stage in Natural Language Generation explanation systems for BNs. Through a number of realistic scenarios of use examples, we show how the generated explanations allows the user to trace the inference steps in the approximate reasoning process in predictive BNs.

Keywords: Content Determination in natural language generation, Linguistic descriptions, Fuzzy syllogism.

1 Introduction

The dramatic success of applications in Artificial Intelligence (AI) has highlighted the importance of making the decisions and actions of intelligent systems more understandable for their users. This is precisely the main goal of the Explainable AI (XAI) paradigm, which has been catalysed by workshops and conferences [2, 1]. The two main goals of XAI are [3]:

- To develop more transparent models without affecting their effectiveness and accuracy;

- To deliver understandable and trustworthy explanations.

Furthermore, the European Union's General Data Protection Regulation [11] actually has created what many authors consider a "de-facto right to explanation". From April 2018, whereby a user can ask for an explanation of an algorithmic decision that was made about him/her. This legislation is starting to fuel the explanation perspective in AI-based systems and applications, which also will contribute to clarify the open debate about possible biases in these systems within the general framework of Fairness, Accountability, Transparency and Ethics (FATE) perspective in AI in general.

In this paper, we focus on automatic linguistic explanation in Bayesian Networks [7] (BN), since understanding of the underlying approximate knowledge representation and reasoning mechanisms (e.g. Conditional Probability Tables, belief updating, message passing algorithms, etc.) is not straightforward for inexperienced users. Upon BNs model, a number of software applications have been proposed for (mostly in a graphic way) helping users and designers to build BN-based models for different application areas. For instance, Promedas [17] is a probabilistic decision support system for medical diagnosis with two goals: i) to offer patients specific diagnosis advice and, ii) educational purposes. In both cases, for a regular patient is not enough to see the knowledge coded in a graph but he/she also needs to understand the reasons that support the output of the system in order to be responsive. On the other hand, several fallacies and misunderstandings that inexperienced users commit when using or applying BNs [6, 5] are described in the literature; therefore, explainability is still being an open question in this type of formalisms.

Other software tools for BNs [9] help to visualise the knowledge represented within a BN and to perform the calculations associated with the approximate reasoning process (belief propagation and updating), but they do not provide meaningful information about the logic underlying the model. In other words, graphical tools support the elaboration of the BNs but they do not address interpretability.

The challenge is to design a computational system that will be able to generate an explanatory narrative of the reasoning process in a register of language which is adequate for non-experienced users.

The focus of this paper is on one of the key stages of Natural Language Generation (NLG) systems, which is *Content Determination* [14]. Content determination in this context consists in extracting the most relevant information to be included in the final explanations in natural language which are conveyed to the users. The content determined in this stage is usually represented in terms of an intermediate language which is subsequently realised in plain natural language in other stages (planning, micro-planning and linguistic realisation). Template-based NLG approaches [16] usually merge these three subsequent stages into a single one (linguistic realisation), since templates take care of both planning and micro-planning simultaneously in a straightforward manner. Our content determination approach is based on the fuzzy syllogistic theory [13], which allows us to build a Knowledge Base (KB), made up of binary Quantified Statements (QS), that comprises the relevant information which supports the BN reasoning. Using the BN topology, we determine the key contents for a template-based short narrative based on the semantic concept of "causal chaining" in different levels of depth, where the user can navigate from the query variable to the root nodes in order to know how each step is involved in the belief updating process. The obtained QSs in the content determination stage can be integrated in a template-based NLG system with the linguistic realisation stage, in order to obtain an automatic system for the generation of explanations of BN. However, this task is out of the aim of this paper and will be addressed as future work.

In this paper, we describe our framework for the case of predictive reasoning in BNs, both without fixed evidences and with them. For the sake of simplicity, we illustrate the behaviour of this model by means of a well-known example in the BN literature [8].

The paper is organised as follows: Section 2 describes the main features of the syllogistic theory that supports how can probabilities can be worded as binary quantified statements; Section 3 describes our concept of explainability in BNs and how the text with the explanation can be generated, both in predictive inference without evidence and with it; and, finally, Section 4, summarises our main conclusions and propose some tasks for future work.

2 Syllogistic Expression of Conditional Probabilities

Probability and uncertainty are not just mathematical concepts, but they are very common in natural language. Thus, qualitative expressions such as *very high probability* or *unlikely* or quantitative ones such as *100%* or *between 30%*

and 35% are frequently used in everyday language and reasoning for dealing with situations involving uncertainty.

This idea applies to BNs as well. Schwartz [15] and Oaksford and Charter [10] defined the equivalence between the mathematical notation of probability ($P(X) = Z$), conditional probability ($P(X|Y) = Z$) and binary QS. QSs are of the form Q ST are PT, where Q stands for a probability Z, which can be quantitative (such as the crisp "5%", the fuzzy "around 20%", etc.) or qualitative ("a few", "most", etc.); ST is the the subject-term of the statement (a single term or a Boolean combination) and stands for the universe of reference in the case of *a priori* probabilities ($P(X) = Z$) or Y in the case of conditional probabilities ($P(X|Y) = Z$); and, PT is the predicate-term (which can also be a single one or a Boolean combination) and stands for X. For instance, the a priori probability $P(Cancer) = 1.16\%$ can be expressed as "1.16% of people have cancer" whilst the Conditional Probability $P(Cancer \mid Smoker) = 5\%$ can be expressed as "5% of smokers have cancer" or "5% of people who are smokers have cancer".

Quantified statements employ a set-based semantics instead of a Bayesian one. Therefore, each statement is denoting a particular set or subset of the universe and the quantifier heading the sentence its cardinality. From the point of view of human understanding of reasoning, its interpretation is not a trivial step but it affects the persuasiveness of the argument for the user [4]; i.e., how the information is presented has cognitive effects on the addressee.

Based on these assumptions, Pereira-Fariña et al. [13] enhanced the theory of syllogistic logic to handle arguments with more than two fuzzy premises and three terms, which opens the possibility of using fuzzy syllogistics for developing a linguistic version of BNs [12]. The general inference pattern for the generalised fuzzy syllogism is described as follows:

$$
\begin{array}{ll}
PR1: & Q_1\ L_{1,1}\ \text{are}\ L_{1,2} \\
PR2: & Q_2\ L_{2,1}\ \text{are}\ L_{2,2} \\
\ldots & \\
\underline{PRN:} & \underline{Q_N\ L_{N,1}\ \text{are}\ L_{N,2}} \\
C: & Q_C\ L_{C,1}\ \text{are}\ L_{C,2}
\end{array} \qquad (1)
$$

where PRn, $n = 1,\ldots,N$ denote the premises and C the conclusion. In any premise, PRn, Q_n denotes are its (fuzzy) quantifier, $L_{n,j}$, $j = 1, 2$ denote an arbitrary Boolean combination among the terms considered in the syllogism (subject-term *ST* for *j*=1, predicate-term *PT* for *j*=2). Similarly, for the Conclusion C, Q_C stands for its quantifier (which is the value to be calculated) and $L_{C,1}$ and $L_{C,2}$ stand respectively for its subject-term and predicate-term.

This schema has the same structure of any propositional logic argument, where the premises convey the available

information and, the conclusion, the variable of interest (query). A critical step for defining the equivalent syllogism to a BN is how the information conveyed in a BN has to be expressed. To the best of our knowledge, this task has not been addressed yet in the literature.

In Table 1 we define the five steps procedure to be followed for converting a BN into an equivalent KB made up of binary QSs. Once the universe of reference U is set (step 1), we distinguish two well-differentiated parts in a BN that must be independently analysed: root nodes (steps 2 and 3) and non-root nodes (step 4).

Root-nodes do not have parents and, consequently, the corresponding variables are not conditioned by others. The values of these variables are associated with *a priori* probabilities. As described in Table 1-step 2, the *ST* of the QSs is U and their *PT* are the pairs variable-value. The values of the crisp quantifiers are directly extracted from the *a priori* probabilities. The independence relationship between the variables of the root-nodes is implicit in the BN, but it has to be made explicit in order to obtain a full representation of it, and a consistent explanation for the reasoning process. This is addressed in Table 1-step 3. The quantifier is calculated here using the product rule because of the independence condition.

Table 1-step 4 deals with non-root nodes. Here, the *ST* of the *QS* are the variables in the parent nodes of the *PT*. The values of the quantifiers are now taken from the Conditional Probability Tables (CPT) in the BN.

The last point of the procedure addresses the wording of the QS in the KB; i.e., the intermediate language of the content determination step. Each one of the QS follows the form *Q ST are PT*, where Q stands for the corresponding quantifier and *ST* and *PT* stand for Subject-term and Predicate-term correspondingly.

For illustrative purposes, we consider in Fig. 1 a handbook example of BN about the medical diagnosis of cancer [8, p. 29-36]. Table 2 contains the equivalent KB made up of binary QSs obtained as a result of applying the procedure described in Table 1. We assume the fuzzy protoform "*Q X is A and Y is B and...*" which is represented in Table 2) with the following intermediate language n-tuple notation: $(Q, X \text{ is } A, Y \text{ is } B, \ldots)$. Three different blocks or types of QSs are generated, which correspond to: i) variables related to root-nodes *pollution* (QS1-QS2) and *smoker* (QS3-QS4), ii) independence condition between these variables (QS5-QS8) and iii) non-root nodes variables *cancer* (QS9-QS16), *XRay* (QS17-QS20) and *dyspnoea* (QS21-QS24). The KB in Table 2 describes in an intermediate language all the relevant information that is explicitly or implicitly included in the BN of the example of Fig. 1. Chaining among QSs provide explanation paths for the approximate reasoning in the BN, as we show in the
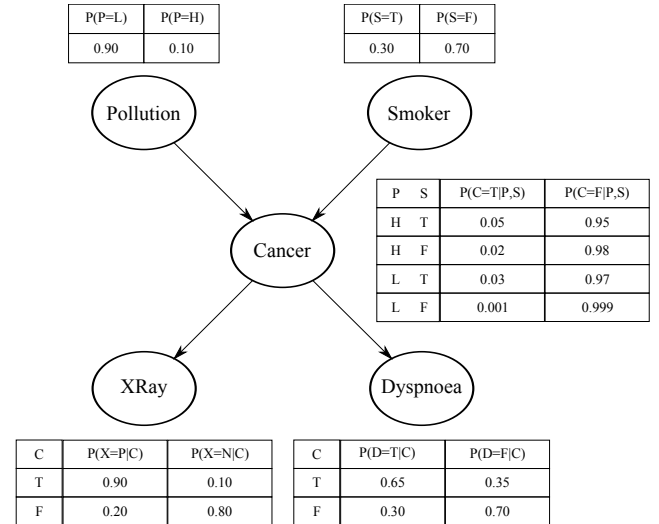
P(P=L)	P(P=H)
0.90	0.10

P(S=T)	P(S=F)
0.30	0.70

P	S	P(C=T|P,S)	P(C=F|P,S)
H	T	0.05	0.95
H	F	0.02	0.98
L	T	0.03	0.97
L	F	0.001	0.999

C	P(X=P|C)	P(X=N|C)
T	0.90	0.10
F	0.20	0.80

C	P(D=T|C)	P(D=F|C)
T	0.65	0.35
F	0.30	0.70

Figure 1: A BN for the lung cancer problem described in [8].

example scenarios in Section 3.2.[1]

3 Explainability of Bayesian Predictive Inference

Forward or predictive reasoning can be executed with or without evidence. Without evidence, it is the immediate belief propagation through the network once the values of the CPTs are defined, and, with evidence, a particular variable in one of the nodes takes a concrete value and this triggers a new belief updating process [8]. From the point of view XAI and a final user, inference with evidence is most interesting because it allows the user to define specific conditions and see the updated probability explicable to them.

In this section, we will describe the general procedure for the generation of the linguistic explanation and then illustrate it using the BN in Fig. 1. This network is composed of five nodes, with two root nodes, all containing bi-valued variables ($Pollution = \{Low, High\}$, $Smoker = \{True, False\}$, $Cancer = \{True, False\}$, $XRay = \{Positive, Negative\}$ and $Dyspnoea = \{True, False\}$).

3.1 General Procedure

Predictive inference in BNs can involve evidence or not. In the former case, and without loss of generality, let us consider a node N in a BN, associated to variable X and let x be one of its values. The procedure, in this case, will obtain the QSs associated with the BN which explain the inferred

---

[1]Furthermore, performing syllogistic reasoning in the KB in Table 2, is equivalent to predictive reasoning in BNs. For instance, one of the QS obtained as a conclusion is "1.163% *of people have cancer*."

Let U be the universe of reference, X_i the i-th probabilistic variable (associated to the corresponding node of the BN), $VAL_{X_i} = \{x_{i,n_i}; n_i = 1, \ldots, N_i\}$ the set of its linguistic values. Each linguistic value is denoted x_{i,n_i} and the number of these linguistic values is N_i.

1. BUILDING THE STATEMENTS CORRESPONDING TO **ROOT NODES** (A PRIORI PROBABILITIES).
   For each variable X_i associated to a root node and $\forall x_{i,n_i} \in VAL_{X_i}$, a binary quantified statement is added to the KB, with

   - *Quantifier* $= P(X_i = x_{i,n_i})$
   - *Subject-term = noun corresponding to the U*
   - *Predicate-term = noun phrase corresponding to the variable and the value that it takes; i.e., "X_i is x_{i,n_i}"*

   Therefore, the total number of statements to be generated in this step is $\sum_{i=1}^{I_{RN}} N_i$, being I_{RN} the number of root nodes.

2. BUILDING THE STATEMENTS CORRESPONDING TO **ROOT NODES** (INDEPENDENCE CONDITIONS).
   For each different combination of all the values of the I variables X_i associated to root nodes, a binary quantified statement is added to the KB, with

   - *Quantifier* $= \prod_{i=1}^{I} P(X_i = x_{i,n_i})$
   - *Subject-term = noun corresponding to the U*
   - *Predicate-term= conjunction using and with all the corresponding values; i.e., "X_1 is x_{1,n_1} and $\ldots$ and X_I is x_{I,n_I}"*

   Therefore, the total number of statements to be added in this step is $\prod_{i=1}^{I_{RN}} N_i$, being I_{RN} the number of root nodes.

3. BUILDING THE STATEMENTS CORRESPONDING TO **NON-ROOT NODES**.
   Let X_i be a variable associated to an intermediate node and $P_{X_i} = \{X_1, \ldots, X_K\}$ the set of variables associated to its K parents. The corresponding values for each $X_k, k = 1, \ldots, K$ is denoted as $VAL_{X_k} = \{x_{k,n_k}; n_k = 1, \ldots, N_k\}$. For all $n_i \in \{1, ..., N_i\}$, for each different combination of the values of the variables X_k associated to the parent nodes, a binary quantified statement is added to the KB, with

   - *Quantifier* $= P(X_i = x_{i,n_i} \mid X_1 = x_{1,n_1}, \ldots, X_K = x_{K,n_K})$
   - *Subject-term= the variables corresponding to the parent nodes: i.e., "X_1 is x_{1,n_1} and $\ldots$ and X_K is x_{K,n_K}"*
   - *Predicate-term= the variables corresponding to the child node; i.e., "X_i is $x_{i,ni}$"*

   These QS make explicit the independence conditions between variables associated to root nodes. Therefore, the total number of statements to be added in this step is $\sum_{k=1}^{K} N_k$, being K the number of parent nodes.

4. WORDING OF THE QS IN THE KB.
   Each of the binary quantified statements added in the previous steps are worded in the usual form:

   $$Q \ ST \ are \ PT,$$

   where Q, ST and PT are, respectively, the corresponding previously defined *Quantifier*, *Subject-term* and *Predicate-term*

Table 1: Procedure for the Description of a Bayesian Network in terms of Binary Quantified Statements (Content determination stage).

Procedure of Table 1	Statement	Content determination (Q X is A and Y is B and ...)
Step 1: Statements for root-nodes	QS1	(90%, *people, Pollution is Low*)
	QS2	(10%, *people, Pollution is High*)
	QS3	(30%, *people, Smoker is True*)
	QS4	(70%, *people, Smoker is False*)
Step 2: Statements for independence of root-nodes	QS5	(27%, *people, Pollution is Low, Smoker is True*)
	QS6	(63%, *people, Pollution is Low, Smoker is False*)
	QS7	(3%, *people, Pollution is High, Smoker is True*)
	QS8	(7%, *people, Pollution is High, Smoker is False*)
Step 3: Statements corresponding to non-root nodes	QS9	(5%, *people, Pollution is High, Smoker is True, Cancer is True*)
	QS10	(95%, *people, Pollution is High, Smoker is True, Cancer is False*)
	QS11	(2%, *people, Pollution is High, Smoker is False, Cancer is True*)
	QS12	(98%, *people, Pollution is High, Smoker is False, Cancer is False*)
	QS13	(3%, *people, Pollution is Low, Smoker is True, Cancer is True*)
	QS14	(97%, *people, Pollution is Low, Smoker is True, Cancer is False*)
	QS15	(0.1%, *people, Pollution is Low, Smoker is False, Cancer is True*)
	QS16	(99.9%, *people, Pollution is Low, Smoker is False, Cancer is False*)
	QS17	(90%, *people, Cancer is True, X Ray is Positive*)
	QS18	(10%, *people, Cancer is True, X Ray is Negative*)
	QS19	(20%, *people, Cancer is False, X Ray is Positive*)
	QS20	(80%, *people, Cancer is False, X Ray is Negative*)
	QS21	(65%, *people, Cancer is True, Dyspnoea is Positive*)
	QS22	(35%, *people, Cancer is True, Dyspnoea is Negative*)
	QS23	(30%, *people, Cancer is False, Dyspnoea is Positive*)
	QS24	(70%, *people, Cancer is False, Dyspnoea is Negative*)

Table 2: Intermediate language expression with crisp quantifiers obtained applying the procedure described in Table 1 to the example of Figure 1

conclusion for the query pair variable-value $P(X = x)$, which corresponds to QS: *Q Us are x*, being U the universe of X.

In Algorithm 1 we describe the general procedure for the extraction of the relevant QS (SKB) in the KB for the content determination of the linguistic explanation of the inferred conclusion (query pair variable-value $P(X = x)$). In the first place, the BN is converted into a KB made up of QS, following the procedure in Table 1. Secondly, the iterative search of the QS is performed. In the first place, all the QS involving value x in their PT are selected (to this aim, the values list is initialised with x). After that, the corresponding ST are considered as the new values to be iteratively searched for in the KB and therefore added to the value list. Value x is no longer considered. This is repeated until the values list is empty. At this point, the procedure ends and all the selected QS are returned. These are the QSs which actually correspond to the variables in the ancestors of node N in the BN.

3.2 Three Scenarios of the Explanation Algorithm Applied to Different Cases of Predictive Reasoning

Let us consider a computational system that implements the BN of Fig. 1 and a user that interacts with it by means of a natural language. We define three possible scenarios for the interaction between the user and the system that in-clude predictive reasoning with and without evidence. As universe of reference, let us suppose the population of a city, which is labelled as "people".

In the first scenario, the user is just interested in knowing the probability of having cancer, without any other additional information:

> User: "What is the chance that I will have cancer?"
> System: "Without any other additional information about where do you live or whether or not you are a smoker, you have 1.163% of chance of having cancer."
> User: "Why?"

This question triggers the explainability module. According to Algorithm 1, KB is initialised following the procedure in Table 2 and LIST with "*Cancer is True*". The y value is "*Cancer is True*" (query variable and value) and all those statements in KB that have "*Cancer is True*" in the PT are extracted, obtaining $SKB = \{QS9, QS11, QS13, QS15\}$ after the first iteration of the external loop. In the second and third iterations, the values of y are all the values of variables *Pollution* and *Smoker*; therefore, all those statements with these variables in the ST must be added to SKB, obtaining as final result *SKB={QS1, Q2S, Q3S, QS4, QS5, QS6, QS7, QS8, QS9, QS11, QS13, QS15}*.

Data: X (query variable), x (value of X)
Result: SKB, the selected QS
KB = Knowledge Base of QS obtained using the
 procedure in Table 1;
N = | KB |, the number of QS in KB;
SKB = $\emptyset$;
LIST = { "X = x"};
while *LIST* $<>$ $\emptyset$ **do**   /* still values to look
 for in the PTs of the KB */
   SELECT y $\in$ LIST /* any of the remaining
    PTs */;
   for *k=1 to N* **do**   /* looking for y in the
    PT of the KB */
      if *y is a PT of QS_k* **then**
         $SKB = SKB \bigcup \{QS_k\}$ /* QS_k added
          to the KB */;
         LIST = LIST $\bigcup Y$, being Y all the values in
          the ST of QS_k /* All the values in
          the ST to be considered */;
      end
   end
   LIST = LIST $\setminus y$;
end
Algorithm 1: Algorithm in pseudo-code for the selection
of the relevant set of QS for the explanation of a predictive
inference without evidence (query variable $X = x$).

Once the content determination step is completed (i.e.,
SKB), a template-based linguistic realisation, following
the NLG pipeline, is generated for answering the question
"Why?". In order to make clear the causal relation support-
ing the different QS involved in the explanation, we use the
linguistic causality marker "because":

```
 There is a chance of 1.163% for
people having cancer because cancer
is related to pollution and smoking.
Since people can be in high or low
polluted areas or can be smokers
or not, this chance arises from
considering all the combinations:
```

- 0.15% is the chance for people
  to have cancer when they are
  in high polluted areas and are
  smokers. This is because they
  are 3% of total people and there
  is a 5% chance for them to have
  cancer.
- 0.14% is the chance for people
  to have cancer when they are in
  high polluted areas and are not
  smokers. This is because they
  are 7% of total people and there
  is a 2% of chance for them to
  have cancer.
- 0.81% is the chance for people

- to have cancer when they are
  in low polluted areas and are
  smokers. This is because they
  are 27% of total people and
  there is a 3% of chance for them
  to have cancer.
- 0.063% is the chance for people
  to have cancer when they are
  in low polluted areas and are
  not smokers. This is because
  they are 63% of total people and
  there is a 0.1% of chance for
  them to have cancer.

In a second scenario, we address reasoning with evidence
when it is in the parent node of the query variable. Thus,
without loss of generality, let us consider a node N in a
BN, associated to variable X and let x be one of its values.
Let Z be the variable in one of the ancestors of N and z
one of its values, for which there is some fixed evidence
$P(Z = z) = 1$. The procedure in this case will obtain
the QSs associated to the BN which explain the inferred
conclusion $P(X = x \mid Z = z)$, which corresponds to QS:
Q zs are x.

Thus, the user says to the system that s/he is a smoker (the
evidence *Smoker is True*), obtaining the answer "there is
a chance of 3.2% for smokers to have cancer". This ev-
idence, in addition to be in the parent node of the query
variable, it is also in a root node of the network and this
simplifies the explanation. Both the QSs of KB about in-
dependence and all QSs involving values different of z are
also discarded, since the corresponding quantifiers become
null. Thus, following the Algorithm 1, the y value is again
$Cancer\ is\ True$ but the ST is $Smoker\ is\ True$, therefore,
we have to select those QSs in the KB satisfying this con-
straint; obtaining in the first iteration of the external loop
SKB={*QS9, QS13*}. In the next iteration, since the inde-
pendence QSs are not required, we just have to add the QS
corresponding to the other term in the ST of *QS9* and *QS13*;
i.e., $Pollution\ is\ High$ and $Pollution\ is\ Low$. As a re-
sult, *SKB*={*QS9, QS13, QS1, QS2*}.

The generated linguistic summary is the following one:

```
 There is a chance of 3.2% for
smokers to have cancer because once
you are a smoker, cancer is related
to pollution.  Since people can be
in high or low polluted areas, this
chance arises from considering all
the combinations:
```

- 2.7% is the chance for smokers
  to have cancer when they are
  in low polluted areas. This is
  because they are 90% of total
  people and there is a 3% chance
  for them to have cancer;

```
• 0.5% is the chance for smokers
  to have cancer when they are in
  high polluted areas.  This is
  because they are 10% of total
  people and there is a 5% chance
  for them to have cancer.
```

In the third scenario, the only difference with respect to the previous one is that Z is an ancestor which is not a parent of X. This case can be solved applying the same procedure. Once again, QSs related to the separation or involving values different of z will be discarded.

Thus, the user asks the system the chances of having dyspnoea given he is a smoker. The system replies "there is a chance of 31.12% for smokers to have dyspnoea, because dyspnoea is related to cancer". For building the explanation, according to Algorithm 1, the y value in this case is "*Dyspnoea is Positive*" and the first set of statements are those where y is the PT; i.e., *SKB={QS21, QS23}*. In the next iteration of the external loop, we add all the QS where the STs of the QS of SKB are PTs taking into the evidence; i.e., "*Cancer is True*" and "*Cancer is False*" given "*Smoker is True*"; i.e., *SKB={QS14, QS13, QS10, QS09}*. Finally, *SKB={QS1, QS2}* are selected.

The resulting linguistic explanation is as follows:

```
 There is a chance of 31.12%
for smokers to have dyspnoea,
because dyspnoea is related to
cancer.  Since smokers can be either
have cancer or not, this chance
arises from considering all the
combinations:
    • 2.08% is the chance for smokers
      to have dyspnoea when they have
      cancer because they are 3.2% of
      total people and 65% of people
      with cancer have dyspnoea;
    • 29.040% is the chance for
      smokers to have dyspnoea when
      they do not have cancer because
      they are 96.8% of total people
      and 30% of people without cancer
      have dyspnoea.
```

The remaining part of the explanation can be generated in the same way of scenario 2.

4 Conclusions

We have presented a novel method for the explanation of predictive Bayesian Networks (BNs) using a linguistic approach which makes use of the equivalence of BNs and binary Quantified Statements. The method focuses on the content determination stage of Natural Language Generation systems, by generating Quantified Statements (of the form *Q ST are PT*) which facilitates the generation of explanations of the reasoning process in BNs in a template-based NLG system. Particularly, these Quantified Statements make explicit some issues that are implicit in the BNs approximate reasoning mechanism, such are the independence condition for root node variables as well as D-separation when evidence for input variables is fixed. Since we just focus on the relevant information for the variable in the query, we do not need to explain how all the values of the BN are updated, which simplifies its interpretability for non-expert users.

For future work, we will address the extension of this framework for dealing with other types of reasoning in BNs, such as inter-causal or diagnostic ones. In the linguistic module, we will address how the use of fuzzy quantifiers can improve the overall quality of the explanations for the users. Finally, we will integrate this framework in a template-based NLG system for automatic generation of explanations in natural language for BN knowledge representation, including an evaluation phase in order to quantify how useful are these explanations for the user and how they can be improved.

We have laid the groundwork for building explanations for BN using a NLG approach, which makes them legible for non-experienced users. We adopted a new viewpoint about the pair node-variable, with the use of binary Quantified Statements, and the reasoning structure, with the chaining between these statements.

Acknowledgement

References

[1] D. V. Aha, T. Darrell, P. Doherty, D. Magazzeni, IJCAI-18 Workshop on Explainable AI (XAI), `http://home.earthlink.net/~dwaha/research/meetings/faim18-xai`, last accessed May 15, 2019. (2018).

[2] D. V. Aha, T. Darrell, M. Pazzani, D. Reid, C. Sammut, P. Stone, IJCAI-17 Workshop on Explainable AI (XAI), `http://www.intelligentrobots.org/files/IJCAI2017/IJCAI-17_XAI_WS_Proceedings.pdf`, last accessed May 15, 2019. (2017).

[3] DARPA, Explainable Artificial Intelligence (XAI), `https://www.darpa.mil/program/explainable-artificial-intelligence`, last accessed May 15, 2019. (2016).

[4] P. Johnson-Laird, B. G. Bara, Syllogistic inference, Cognition 16 (1984) 1 – 61.

[5] D. Kahneman, Thinking, Fast and Slow, Farrar, Straus and Giroux, 2011.

[6] K. Korb, Bayesian informal logic and fallacy, Informal Logic 23 (2) (2004) 41–70.

[7] K. Korb, R. McConachy, I. Zukerman, A cognitive model of argumentation, in: Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society, 1997, pp. 400–405.

[8] K. Korb, A. Nicholson, Bayesian Artificial Intelligence, Chapman & Hall/CRC, 2004.

[9] Norsys, Netica application, `http://www.norsys.com/netica.html`, last accessed May 15, 2019. (2017).

[10] M. Oaksford, N. Chater, Bayesian Rationality. The probabilistic approach to human reasoning, Oxford University Press, Oxford, 2007.

[11] Parliament and Council of the European Union, General data protection regulation (GDPR), `http://data.europa.eu/eli/reg/2016/679/oj`, last accessed May 15, 2019. (2016).

[12] M. Pereira-Fariña, A. Bugarín, Approximate syllogism as argumentative expression for knowledge representation and reasoning with generalized Bayes' theorem, in: D. Mohammed, M. Lewinski (Eds.), Argumentation and Reasoned Action: Proceedings of the First European Conference on Argumentation, Studies in Logic and Argumentation. College Publications, 2016, pp. 817–830.

[13] M. Pereira-Fariña, J. C. Vidal, F. Díaz-Hermida, A. Bugarín, A fuzzy syllogistic reasoning schema for generalized quantifiers, Fuzzy Sets and Systems 234 (2014) 79 – 96.

[14] E. Reiter, R. Dale, Building Natural Language Generation Systems, Cambridge University Press, 2000.

[15] D. G. Schwartz, Dynamic reasoning with qualified syllogisms, Artificial Intelligence 93 (1997) 103–167.

[16] K. van Deemter, E. Krahmer, M. Theune, Squibs and discussions: Real versus template-based natural language generation: A false opposition?, Computational Linguistics 31, 1 (2005) 15 – 23.

[17] W. Wiegerinck, M. Nijman, W. ter Burg, E. ter Braak, Y. O, J. Neijt, H. Kappen, Inference and advisory system for medical diagnosis second report STW-NGN55. 3614, `http://www.snn.ru.nl/v2/serve.php?doc=eind1.pdf` (1999).