Knowledge Discovery with Qualitative Influences and Synergies

Jesús Cerquides, Ramon López de Màntaras

Artificial Intelligence Research Institute, IIIA
Spanish Council for Scientific Research, CSIC
08193, Bellaterra, Barcelona, Spain
{cerquide,mantaras}@iiia.csic.es

Abstract We review some approaches to qualitative uncertainty and propose a new one based on the idea of Absolute Order of Magnitude. We show that our ideas can be useful for Knowledge Discovery by introducing a derivation of the Naive-Bayes classifier based on them: the Qualitative Bayes Classifier. This classification method keeps Naive-Bayes accuracy while gaining interpretability, so we think it can be useful for the Data Mining step of the Knowledge Discovery process.

1 Introduction

Comprehensibility is a key characteristic for algorithm results to be useful in Knowledge Discovery in Databases tasks.

Bayesian reasoning has been usually criticized as hard to explain and understand, but achieves high performance rates with simple constructs, as happens for instance with the Naive-Bayes classifier[5].

Some approaches to increasing Bayesian reasoning comprehensibility appear in [3, 6, 12, 14, 15]. The main idea in all these approaches was to attach linguistic labels as "probable" or "very unlikely" to numerical probabilities, that is to absolutes degrees of belief. Bayesian reasoning works primarily with changes in probability values, and these approaches do not seem to give any interpretation of such changes, giving as result hardly understandable explanations. It has been accepted that, unlike physical parameters, absolute probabilities do not seem to have values (except the endpoints) that are universally interesting [13].

This problem was noticed also by Elsaesser, that in [1] proposed the use of a version of Polya's "shaded inductive patterns" [10] for linguistic explanation of Bayesian inference. Elsaesser uses Oden's model [8] to create the linguistic labels related to changes in probability. Elsaesser explanations are comprehensible, but we have no security that reasoning with the information given by these explanations really bring us to coherent conclusions, this is because explanation and reasoning are performed at different levels, and we are not allowed to use a previous explanation in a future case.

Another approach is the one followed by Neufeld [7], Wellman [13] and Parsons [9], using ideas from the field of qualitative uncertainty. The idea behind their work is finding whether a fact A is favoured, unfavoured or not altered by another fact B. Quoting Parsons:

> *Whereas in probabilistic networks the main goal is to establish probabilities of hypotheses when particular observations are made in qualitative systems the main aim is to establish how values change.*

Our approach can be viewed as a refinement of qualitative probabilistic networks (QPNs) showing that slightly modified, Elsaesser explanations can be used not just for explanatory purposes but also for reasoning and prediction achieving results similar to those of non-qualitative probabilistic reasoning, while keeping intact its interpretability.

Next section briefly introduces qualitative probabilistic networks, concretely Wellman and Neufeld approaches. Section 3 introduces our qualitative approach to influences and synergies, making use of the absolute orders of magnitude model. Section 4 describes our proposal to use the qualitative influences and synergies in a Qualitative Bayesian classifier. Finally, Section 5 describes an empirical comparative study based on 15 datasets and analyses the results obtained with the aim of showing the good performance of our approach in terms of classification accuracy.

2   Introduction to Qualitative Probabilistic Networks

Two main approaches have been done to the concept of QPN. We will shortly review both here.

2.1   Wellman approach

For Wellman, a QPN is a pair $G = (V, Q)$, where V is the set of variables or vertices of the graph and Q is a set of qualitative relationships among the variables. He introduces two main concepts for modelling QPNs as are *qualitative influences* and *qualitative synergies*.

Wellman qualitative influences. Qualitative influences can be thought of as qualitative relations describing the sign (direction) of the relationship between a pair of variables. A variable can influence another positively (+), negatively (-), or in no way (0). We should also consider the possibility that the sign of the influence is unknown to us (?). If we use δ to denote one of $\{+,-,0,?\}$ we say a qualitative influence of a on b in direction δ holds in the graph $G = (V, Q)$ if $(a, b, \delta) \in Q$. For formally introducing the probabilistic semanticof this concept the way Wellman does, we need to previously define the set of predecessors that influence a variable in a network.

$$pred_G(b) = \{a | (a, b, \delta) \in Q, for some \; \delta \in \{+, -, ?\}\} \tag{1}$$

Now we can assign probabilistic meaning to influences. We say that an influence edge $(a, b, +) \in Q$ is satisfied in a concrete domain if for all $x \in pred_G(b) - \{a\}$ such that x is consistent with both a and $\neg a$, we have

$$Pr(b|a, x) \geq Pr(b|\neg a, x) \tag{2}$$

The meaning of this expression can be stated as: under any circumstances (x) that are known to affect b, the presence of a makes b more likely than its absence.

Parallel definitions can be done for $(a, b, -)$, $(a, b, 0)$ and $(a, b, ?)$, replacing $\geq$ by $\leq$, $=$, and "no condition at all" ($(a, b, ?)$ always holds) respectively.

Wellman qualitative synergies. Qualitative synergies describe the qualitative interaction among influences. The idea behind them is that two variables synergically influence a third if their joint influence is greater than separate, statistically independent, influences. The formalisation of this idea can be seen in [13], and will be skipped here.

2.2 Neufeld approach

Neufeld formalises the idea of qualitative influence by means of the concept of favouring. He says a favours b if $Prob(b|a) > Prob(a)$. He includes four types of edges in what he calls inference graphs:

- *Defeasible links.* Given a, b is more likely to happen.
$$a \rightarrow b \text{ if } 1 > Prob(b|a) > Prob(b) \tag{3}$$

- *Logical links.* Given a, b will surely happen.
$$a \Rightarrow b \text{ if } 1 = Prob(b|a) > Prob(b) \tag{4}$$

- *Negative defeasible links.* Given a, b is less likely to happen.
$$a \nrightarrow b \text{ if } 1 > Prob(\neg b|a) > Prob(\neg b) \tag{5}$$

- *Negative logical links.* Given a, b will not happen.
$$a \nRightarrow b \text{ if } 1 = Prob(\neg b|a) > Prob(\neg b) \tag{6}$$

Once introduced these concepts Neufeld uses them to do common sense reasoning. For more details on his approach to qualitative uncertainty see [7].

3 Influences and synergies revisited

Neufeld and Wellman ideas are useful for common sense reasoning, planning under uncertainty and when qualitative differential equations are not applicable. Our idea is to adapt them in order to make them useful for classification and characterization of sets.

The qualitative model used by both approaches is the signs model, composed of three categories +,-,0 and ? for representing the unknown. More sophisticated models have risen from the field of qualitative reasoning. One of these models is the absolute orders of magnitude model [11], that considers a finer partition of the real line than the one given by the signs, allowing also distinctions in quantities of the same sign. This model qualifies quantities into seven classes, from *Negative Large* to *Positive Large*, including *Zero*. Quantities of the same sign are divided into three classes(*Large, Medium* and *Small*) that are very natural in human reasoning. We have discretized influences into this new model, in a way
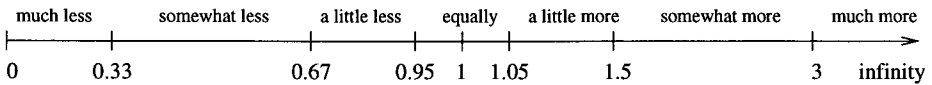
coherent with Neufeld works. Neufeld states a *favours* b if $Prob(b|a) > Prob(b)$ that is, if $\frac{Prob(b|a)}{Prob(b)} > 1$. This quotient was also used by Elaesser, in his work trying to explain bayesian reasoning, to denote the shift in belief that a produces in b. We will define influence of a in b as:

$$Influence(a, b) = \frac{Prob(b|a)}{Prob(b)} \tag{7}$$

We note that:

$$Influence(a, b) = Influence(b, a) \tag{8}$$

Once we have a definition for influence, we make use of the absolut order of magnitudes model to make influences comprehensible. By discretizing influences into the seven classes seen in Figure 1, we perform a process similar to that of Elsaesser assigning linguistic labels as "much more likely", "a little less likely", and so on. The boundary values established in Figure 1 were selected over a set of alternatives because they performed better than the rest in the classification experiments described in Section 5.

much less	somewhat less	a little less	equally	a little more	somewhat more	much more
0 0.33		0.67	0.95 1 1.05	1.5		3 infinity

Figure1. Influence discretization scale

We have given an expression for influences and a scale for their discretization. What about synergies?. Synergies can be seen as the difference in influence between two facts that happen together with respect to these two facts happening separately. We can give the following expression for synergies of two variables:

$$Synergy(\{a_1, a_2\}, b) = \frac{Influence(a_1 \cap a_2, b)}{Influence(a_1, b) * Influence(a_2, b)} \tag{9}$$

4 An application of qualitative influences and synergies: The Qualitative Bayesian Classifier

In this section we will show that qualitative influences and synergies can be used for reasoning and, more concretely, for classification tasks, giving high classification rates. We will use them to get a qualitative version of a well known classification method, the Bayesian classifier.

The Naive-Bayes classifier [5] is a classification method based on Bayesian reasoning. Given a test example e from a probabilistic viewpoint we must choose the class i that maximizes $P(C = i|\mathbf{E} = \mathbf{e})$. Developing this conditional probability according to the Bayes rule we have:

$$P(C = i|\mathbf{E} = \mathbf{e}) = P(C = i) * \frac{P(\mathbf{E} = \mathbf{e}|C = i)}{P(\mathbf{E} = \mathbf{e})} \tag{10}$$

If the attributes are independent given the class, it holds:

$$P(\mathbf{E} = \mathbf{e}|C = i) = \prod_{j=1}^{N} P(E_j = e_j|C = i) \tag{11}$$

where N is the number of features of the example. Since $P(\mathbf{E} = \mathbf{e})$ is independent of the class, Naive-Bayes tells us to choose the class which maximizes:

$$P(C = i) * \prod_{j=1}^{N} P(E_j = e_j|C = i) \tag{12}$$

An interpretation of this formula can be that $P(C = i)$ is our initial belief in the fact that i is the class of our example, and each one of the factors $P(E_j = e_j|C = i)$ can be seen as shifts that modify this belief. We will adapt these shifts in belief to coincide with our previously defined influences.

Returning to Equation 10, and assuming statistical independence between the attributes, we are allowed to develop the denominator the same way we have done with the numerator:

$$P(\mathbf{E} = \mathbf{e}) = \prod_{j=1}^{N} P(E_j = e_j) \tag{13}$$

and substituting in 10 we have:

$$P(C = i|\mathbf{E} = \mathbf{e}) = P(C = i) * \frac{\prod_{j=1}^{N} P(E_j = e_j|C = i)}{\prod_{j=1}^{N} P(E_j = e_j)} \tag{14}$$

that in terms of influences can be expressed as

$$P(C = i|\mathbf{E} = \mathbf{e}) = P(C = i) * \prod_{j=1}^{N} Influence(E_j = e_j, C = i) \tag{15}$$

Now we can apply this rule with qualitative influences and analyze the differences in accuracy between applying the Naive-Bayes classifier where shifts in belief vary continuously from 0 to 1 and our qualitative influences framework, where shifts only can have the seven values we have previously seen.

Before doing this, we want to introduce the idea of synergy in our classifier, because we have made two independence assumptions and synergy can improve the performance when the classification problem at hand does not fulfill this assumptions, because synergies precisely try to express interattribute dependencies with respect to the class. Our first idea was to calculate all the synergies between all pairs of variables and apply them. The problem with this approach is that it is not an approximation of the Bayes formula, and hence is not theoretically well founded and empirically does not perform correctly. In fact, applying synergies that way made accuracy get worse. The reason is that synergies can be seen as corrections of the approximation to the probabilities given by influences. It is not correct to apply a synergy correction for two variables E_i and E_j and also apply it to E_j and E_k, because we are correcting E_j influence twice. That is why we will follow the next schema:

> We first classify the set of synergies that affect to our example into Large, Medium and Small synergies, no matter if they are positive or negative. Then we try to apply as many Large synergies as possible. Once this has been done we repeat the same process for medium and finally for small synergies.

Algorithmically, it can be expressed as shown in Figure 2. The operations in the algorithm are calculated by using a representative for each interval. We tested two approaches, one taking as representative a value in the center of each interval and the other taking the value of the class that is nearer to equality. Using the discretization values of 1, our first method will give as representative of "much less" 0.165, as representative of "somewhat less" 0.5 and so on, while for the same values, the second method will choose 0.33 as representative of "much less", 0.67 for "somewhat less" and so on. The second approach performed better empirically, consequently it is the one we will use from now on.

When our Qualitative Bayesian Classifier (QBC) is restricted to influences we call it First Order QBC (FOQBC), when synergies of two variables are applied we call it Second Order QBC (SOQBC). The development for order greater than second is not trivial because different developments of the greater order approximations are possible.

```
program SOQBC;
    foreach class i

        ClassProb[i] = P(C=i) * ∏_{j=1}^{N} Influence(E_j = e_j, C = i)

        InfluencesCorrected[i] = ∅;
        ApplySynergies(LargeSynergies(class i));
        ApplySynergies(MediumSynergies(class i));
        ApplySynergies(SmallSynergies(class i));
    end
    Select the class i with highest ClassProb[i];
end
procedure ApplySynergies(SynergySet)
    While SynergySet ≠ ∅
            Randomly choose a Synergy (namely {E_j = e_j, E_k = e_k})
                    and delete it from SynergySet
            if ({E_j = e_j, E_k = e_k} ∩ InfluencesCorrected) = ∅
                ClassProb[i] = ClassProb[i] * Synergy({E_j = e_j, E_k = e_k}, C=i)
                InflCorrected[i] = InflCorrected[i] ∪ {E_j = e_j, E_k = e_k}
```

Figure2. Synergies application strategy

5   Empirical comparison

We have evaluated the classification accuracy for First and Second Order QBC and compared it with the Naive-Bayes classifier, as well as with other widely used machine learning algorithms. Our experiment consists in evaluating the average accuracy of each classifier, as well as its standard deviation for 15 datasets from the Irvine repository. Some information regarding these datasets can be seen in 1 For each dataset and classification method we performed 50 runs, keeping

the 70% of the dataset as training set and the remaining 30% as test set. We included our algorithms in the $\mathcal{MLC}++$ [4] library, and used the facilities this library provides for machine learning experimentation. We compared the First and Second Order QBC with the well known machine learning algorithms CN2, Naive-Bayes classifier, IBL, and ID3. The results are summarized in Table 2.

Furthermore, in table 3 we show how many times each classification algorithm ranked in the position indicated by the column identifier. In this table "1" means the algorithm ranked once in the position specified by the column under which it appears, "22" that it ranked twice , "333" that it ranked three times and so on.

Dataset	Attributes	Instances	Classes	Missing
BREAST-CANCER	9	286	2	none
BREAST	10	699	2	16
CRX	15	690	2	few
GLASS2	10	214	2	none
HEART	13	270	2	none
HYPOTHYROID	25	3162	2	some
IRIS	4	150	3	none
MONK1	6	432	2	none
MONK2	6	432	2	none
MONK3	6	432	2	5%
PARITY 5+5	10	100	2	none
SOYBEAN-LARGE	35	316	19	some
SOYBEAN-SMALL	35	47	4	none
VOTES	16	435	2	few
WAVEFORM-21	21	5000	3	none

Table1. Datasets information

Dataset	CN2	Bayes	IBL	ID3	FOQBC	SOQBC
BREAST-CANCER	73.82 ± 0.39	69.56 ± 0.5	**77.01 ± 0.54**	74.04 ± 0.52	69.98 ± 0.52	71.69 ± 0.49
BREAST	95.88 ± 0.15	**97.5 ± 0.11**	97.09 ± 0.11	96.2 ± 0.15	97.42 ± 0.11	97.35 ± 0.11
CRX	86.86 ± 0.29	86.51 ± 0.21	87.17 ± 0.23	86.97 ± 0.23	86.22 ± 0.2	**87.75 ± 0.19**
GLASS2	78.92 ± 0.79	77.89 ± 0.75	**83.28 ± 0.59**	80.89 ± 0.83	76.72 ± 0.76	76.96 ± 0.73
HEART	79.70 ± 0.62	82.11 ± 0.42	79.90 ± 0.49	80.76 ± 0.52	81.34 ± 0.48	**82.25 ± 0.5**
HYPOTHYROID	98.9 ± 0.04	98.43 ± 0.05	97.7 ± 0.05	**98.94 ± 0.03**	98.56 ± 0.05	98.44 ± 0.05
IRIS	96.24 ± 0.42	94.99 ± 0.35	96.52 ± 0.33	**96.80 ± 0.37**	94.69 ± 0.4	94.92 ± 0.36
MONK1	**91.63 ± 0.95**	74.67 ± 0.61	81.68 ± 0.47	83.76 ± 0.98	71.66 ± 0.66	88.73 ± 0.62
MONK2	**72.77 ± 0.59**	57.41 ± 0.53	70.01 ± 0.51	69.2 ± 0.73	60.94 ± 0.47	61.37 ± 0.54
MONK3	90.04 ± 0.63	**92.25 ± 0.36**	85.83 ± 0.53	91.60 ± 0.39	91.82 ± 0.51	91.46 ± 0.4
PARITY 5+5	64.26 ± 0.71	50.45 ± 0.76	**65.43 ± 0.7**	64.14 ± 0.75	51.71 ± 0.64	50.67 ± 0.66
SOYBEAN-LARGE	89.52 ± 0.28	90.66 ± 0.22	**92.59 ± 0.2**	91.44 ± 0.25	88.04 ± 0.2	86.33 ± 0.39
SOYBEAN-SMALL	90.93 ± 1.36	**99.14 ± 0.77**	98.55 ± 0.81	93.76 ± 1.11	99.01 ± 0.78	99.02 ± 0.77
VOTES	94.89 ± 0.2	89.74 ± 0.25	93.99 ± 0.23	**95.09 ± 0.18**	88.17 ± 0.28	91.61 ± 0.26
WAVEFORM-21	74.33 ± 0.43	82.81 ± 0.35	82.16 ± 0.41	78.31 ± 0.39	79.98 ± 0.4	**83.07 ± 0.4**

Table2. Average accuracies and their standard deviations

	First	Second	Third	Fourth	Fifth	Sixth
ID3	333	333	55555	1	333	
IBL	4444	333	22	333	1	22
SOQBC	333	22	1	55555	333	1
CN2	22	333	333	22	1	4444
BAYES	333	22	1	22	4444	333
FOQBC		22	333	22	333	55555

Table3. Ranking table

5.1 Result analysis and justification

Table 3 shows that the SOQBC has an accuracy in the level of the best classifiers being the best one in 3 out of 15 times, so it can be considered as a valuable alternative to these methods. Our method has an accuracy at least as good as the one provided by Naive-Bayes and offers the advantage of a common classification-explanation space. That is, we are not keeping two separate representations, one for reasoning and the other for explanation. That allows the user to actually *apply* the discovered knowledge in his own decisions with a greater confidence, because we have shown that reasoning with it gives acceptable accuracy results.

On the other hand, we consider the FOQBC results good enough considering the simplicity of the classifier induced. Its results are not too far from the ones given by continuous Bayes (surprisingly four times FOQBC outperforms Bayes as shown in Table 2). This can be seen as a confirmation of the idea exposed by Friedman in [2]:

> *Good probability estimates are not necessary for good classification; similarly, low classification rates does not imply that the corresponding class probabilities are being estimated (even remotely) accurately.*

FOQBC makes a extremely inexact estimation of the probabilities, but has classification results only slightly worse than the ones given by more complex classifiers. Our intuition is that the difference in performance between FOQBC and Naive-Bayes will increase when the number of examples in the dataset increases but we have not tested it yet.

6 Conclusions

We have introduced qualitative influences and synergies based on the absolute orders of magnitude model. We have developed a competitive learning algorithm (SOQBC) based on these ideas, that offers a good balance between the accuracy of its predictions and its understandability. These facts make us believe that both the ideas of qualitative influences and synergies and the classifier can be useful to the Knowledge Discovery community.

7 Acknowledgements

References

1. Christopher Elsaesser. Explanation of probabilistic inference. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*, pages 387–400. Elsevier Science Publishers B.V. (North-Holland), 1989.
2. Jerome H. Friedman. On Bias, Variance, 0/1-Loss, and the Curse-of-Dimensionality. *Data Mining and Knowledge Discovery*, 1:55–77, 1997.
3. Edgard M. Johnson. Numerical encoding of qualitative expressions of uncertainty. Technical Report 250, Army Research Institute for the Behavioural and Social Sciences, Arlington, Virginia, 1973.
4. R. Kohavi, G. John, R. Long, D. Manley, and K. Pfleger. MLC++: A machine learning library in C++. In *Tools with Artificial Intelligence*, pages 740–743. IEEE Computer Society Press, 1994.
5. Pat Langley, Wayne Iba, and Kevin Thompson. An Analysis of Bayesian Classifiers. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 223–228. AAAI Press and MIT Press, 1992.
6. S. Lichtenstein and J. R. Newman. Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychon. Sci.*, 9(10), 1967.
7. Eric Neufeld. A probabilistic commonsense reasoner. *International Journal of Intelligent Systems*, 5:565–594, 1990.
8. G. C. Oden. Integration of fuzzy logical information. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4):565–575, 1977.
9. Simon Parsons. Further results in qualitative uncertainty. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 3(2):187–210, 1995.
10. George Polya. *Mathemathics and Plausible Reasoning, Vol II: Patterns of Plausible Inference*. Princenton, New Jersey: Princenton University Press, 1954.
11. L. Trave and N. Piera. The orders of magnitude models as qualitative algebras. In *11th IJCAI*, Detroit, 1989.
12. T. S. Wallsten, D. V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth. Measuring the vague meanings of probability terms. Technical Report 173, The L. L. Thurstone Psychometrich Laboratory, Chapel Hill, N.C., 1985.
13. Michael P. Wellman. Fundamental concepts of qualitative networks. *Artificial Intelligence*, 44:257–303, 1990.
14. Alf C. Zimmer. Verbal vs. numerical processing of subjective probabilities. In R. W. Scholz, editor, *Decision Making Under Uncertainty*, pages 159–182. Elsevier Science Publishers B.V. (North-Holland), 1983.
15. Alf C. Zimmer. The estimation of subjective probabilities via categorical judgments of uncertainty. In *Proceedings of the Workshop on Uncertainty and Probability in Artificial Intelligence*, pages 217–224. UCLA, 1985.