HAL
open science

Customer relationship management and small data - application of bayesian network elicitation techniques for building a lead scoring model

Youssef Benhaddou, Philippe Leray

To cite this version:

Youssef Benhaddou, Philippe Leray. Customer relationship management and small data - application of bayesian network elicitation techniques for building a lead scoring model. 14th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA 2017), Oct 2017, Hammamet, Tunisia. pp.251-255, 10.1109/AICCSA.2017.51 . hal-01619307

HAL Id: hal-01619307

https://hal.science/hal-01619307v1

Submitted on 15 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Customer Relationship Management and Small Data - Application of Bayesian Network Elicitation techniques for building a Lead scoring model

Youssef Benhaddou, Philippe Leray
LS2N UMR 6004, DUKe research group, University of Nantes, France
Email: youssef.benhaddou@etu.univ-nantes.fr, philippe.leray@univ-nantes.fr

***Abstract*—Customer Relationship Management is an every day task for companies, even the ones dealing with Small Data. We are more interested here by Lead Scoring that refers to the practice of calculating and assigning a score to leads (business contacts or qualified prospects) of the company.**

In this paper, we present one way of building a Lead scoring model with a Bayesian network using a small amount of data. In addition to its ability of handling uncertainty, Bayesian networks are knowledge representation models that can be built from expert knowledge. In our specific context, we then propose to build our Lead scoring model from expertise and apply usual heuristics to decrease the complexity of our model (parent divorcing, NoisyOr). We specifically propose three ways of estimating the parameters of our NoisyOr submodels. The only data available is used to validate our approach, with good precision and recall results on a small set of 23 examples.

***Index Terms*—Customer Relationship Management, Lead Scoring, Bayesian Networks, Expert Knowledge, Preferences elicitation, Probability elicitation.**

I. Introduction

Customer Relationship Management (CRM) includes devices, marketing operations or support, which aim to optimize the customer relationship's quality, retain and maximize the turnover or the margin by customer. CRM brings together the collect and analysis techniques of customer data, marketing operations and operations support [Chorianopoulos, 2016]. The concept of CRM has sometimes tended to widen toward the prospect and is more and more complementary to a device for the optimization of the digital channels integration both for data collection and for campaigns targeting and customizing.

CRM aims to identify, understand and fit the customers in order to improve their satisfaction and maximize profits. Different data mining techniques may be used to understand, anticipate and respond to customer needs [Ngai et al., 2008], [Soltani and Navimipour, 2016].

Lead scoring[1] is a CRM subtask that refers to the practice of calculating and assigning a score to leads (business contacts or qualified prospects) of the company. The score can be calculated from the characteristics of the lead (sector, company size, responsibility) or from its behavior (contact history, type of request, visit behavior on a website, etc.). According to its components, the score is supposed to reflect the potential of the prospect, its degree of appetite for the product / service or its position in the purchase cycle. The score allows you to select targets, establish contact priorities and personalize marketing action.

Some of the existant CRM tools use probabilistic graphical models like Bayesian Networks for tasks like churn prediction [Lee and Jo, 2010], [Sun et al., 2013], [Verbraken et al., 2014]. In these papers, the model is learnt from existing data.

Small data is an emerging trend in reaction to Big Data. By reviewing [Lindstrom, 2016]'s book about Small Data on Forbes blog, R.Dooley[2] explains that *Marketers can obtain significant market insights from gathering Small Data by engaging with and closely observing real people in their own environments.*

In this paper, we present one way of building a Lead scoring model with a Bayesian network.

In addition to its ability of handling uncertainty, Bayesian networks are knowledge representation models that can be built from expert knowledge. In our specific context, we then propose to build our Lead scoring model from expertise. We apply usual heuristics to decrease the complexity of our model (parent divorcing, NoisyOr) and three ways of estimating the parameters of our NoisyOr submodels. The only data available is used to validate our approach, with good precision and recall results.

The organization of this paper is as follows: some recalls about Bayesian networks and their building from expertise are presented in Section II. Section III presents our specific problem, and the several steps performed to build our model. We show in Section IV some examples of the application of our methodology for building our Lead scoring model, and some preliminary results. Finally, Section V presents conclusion and future work.

II. Background

A. Bayesian Network

A Bayesian network [Pearl, 1988] is an annotated directed acyclic graph that encodes a joint probability distribution over a set of random variables U.

[1] http://www.definitions-marketing.com/definition/lead-scoring/

[2] https://www.forbes.com/sites/rogerdooley/2016/02/16/small-data-lindstrom/

Formally, a Bayesian network for U is a pair: $B = \langle G, \Theta \rangle$. The first component, G, is a directed acyclic graph (DAG) whose vertices correspond to the random variables $X_1, ..., X_n$, and whose edges represent direct dependencies between the variables.

The graph G encodes independence assumptions: each variable X_i is independent of its non descendants given its parents in G.

The second component of the pair, namely Θ, represents the set of parameters that quantifies the network. It contains a parameter $\Theta_{x_i|\Pi x_i} = P_B(x_i|\Pi x_i)$ for each possible value x_i of X_i, and Πx_i of ΠX_i , where ΠX_i denotes the set of parents of X_i in G.

A Bayesian network B defines a unique joint probability distribution over U given by:

$$P_B(X_1, ..., X_n) = \prod_{i=1}^{n} P_B(X_i|\Pi X_i) = \prod_{i=1}^{n} \Theta_{X_i|\Pi_{X_i}}. \quad (1)$$

The complexity of a Bayesian network is defined as its dimension, i.e. the number of independent parameters used to describe its conditional probability distributions.

$$Dim(B) = \sum_{i=1}^{n} (r_i - 1)q_i \quad (2)$$

where r_i is the cardinality of X_i and $q_i = \prod_{X_j \in \Pi_{X_i}} r_j$ is the number of configuration of the parents of X_i.

For instance, if a boolean variable has 10 boolean parents, the corresponding contribution of its conditional probability distribution to the model complexity is $(2-1) * 2^{10} = 1024$.

In order to simplify this complexity, some approximate distributions have been proposed, such as the NoisyOr model, proposed by [Henrion, 1989], aims at simplifying one conditional probability distribution $P(Y|X_1...X_p)$ by determining only one subset of conditional probabilities :

$$p_i = P(Y = true|X_1 = false, ...X_i = true, ...X_p = false) \quad (3)$$

With a NoisyOr model, the contribution of our previous boolean variable with 10 parents to the model complexity would be reduced from 1024 to 10.

This model works with boolean variables (but has been extended to multivalued variables with the NoisyMax model [Díez and Druzdzel, 2000]) and makes the assumption that parents effects are independent from each others.

This model implies that the target value Y is false when all its parent are false. Thus Henrion [Henrion, 1989] proposed an extension of the Noisy-OR structure called Leaky Noisy-OR by introducing a new parameter called leak probability. The leak probability corresponds to the fact that it may exist other parent variables to define more precisely the child variable. This leak probability can be modeled by using another virtual parent variable L with a leak probability p_L.

$$P(Y = true|\Pi_Y = \pi_y) = 1 - (1 - p_L) \times \prod_{X_i \in X_T} (1 - p_i) \quad (4)$$

where X_T is the subset of Π_Y which are observed at their true value.

B. Building Bayesian networks from expertise

In this paper, we are interested by constructing a Bayesian network from expertise. [Kjaerulff and Madsen, 2008] proposes a state of the art for this goal that we can divide into two tasks, (a) structure determination, and (b) parameter elicitation.

As described by [Kjaerulff and Madsen, 2008] (chapter 6), manually building a Bayesian network can be a labor-intensive task, and requires skill, creativity as well as close communication with problem domain experts. Two main tasks need to be addressed in this process: identification of the relevant variables and identification of the links between the variables.

Parameter elicitation aims at eliciting subjective conditional probabilities from expertise, and had been studied by [Druzdel et al., 2000], [van der Gaag et al., 2002], [Renooij, 2001]. One of these solutions proposes a mapping of verbal statements of probability to probabilities with an "elicitation scale" as described in Figure 2.

III. Contribution

A. Problem

We work with *Uneek*, a company that develops Customer Relationship Management (CRM) tools and want to implement a Lead Scoring model build on its expertise.

Our main source of data is different companies News available on the internet. Since the large amount of data is not available to define the structure (DAG) and the parameters of the scoring model we had to use knowledge based methods to come across this problem.

The main objective is to calculate a score that defines if the company present in the news text is interesting for *Uneek* and shall be contacted or not.

B. Variables determination

All the factors used in this article are boolean. if there is a present keyword in the news annotating the factor DF_i then the value of this factor goes to true (i.e. false in the other case). Table I give us examples of such factors.

Our target variable is denoted $Kept$. $P(Kept = true|evidence)$ will give us the lead scoring of the candidate prospect.

We identify two types of decision factors DF_i. First, **Mandatory decision factors** must be positive in order to keep the company.If one of the mandatory factors $MF_i = false$ then $Kept = false$. For instance, we can define that the headquarter location of the candidate prospect has to be the same than our company otherwise, the corresponding scoring is equal to zero.

Then, **Secondary decision factors** correspond to usual inputs of a scoring model. These factors don't have the same impact on the target variable, but the more the factors are positive, the more interesting is the prospect.

These decision factors are not directly observed. They are related to some keywords that need to be present in the input

TABLE I
DECISION FACTORS DF_i

Mandatory	Secondary	Target
Headquarter Location	Interest	Kept (Final scoring)
	Activity Sector	
	Company Type	
	Growing Company	
	Reorganization	

TABLE II
KEYWORDS EXAMPLE FOR "REORGANIZATION" FACTOR.

Reorganization
Modernization
Moving
New headquarter
Fusion
Repurchase
Absorption

News text. For instance, Table II describes some keywords associated to the decision factor *Reorganization*.

C. Structure

Fig. 1 depicts the structure of our Bayesian network. Starting from the right side, our target variable is named *Kept* and expresses the binary result of contacting a company or not. The distinction between mandatory factors MF and secondary ones DF is modelled by a logical AND in order to ensure that $Kept = false$ if $MF = false$.

The dependency between the final scoring and the DF can lead to a complex conditional probability distribution. For this reason, we proposed to use one heuristic to reduce this complexity. First, we decomposed it into a hierarchical structure. This modeling technique, referred to as parents divorcing [Olesen et al., 1989], is used for reducing the complexity of the model by adding some intermediate decision factors. In such a hierarchy, we consider that the observed decision factors are the leaves of the decomposition.

Finally, we consider that each possible keyword can be generated differently with respect to a specific decision factor.

D. Parameters

The probability of each keyword given its associated factor is for now defined as a constant distribution, with $P(keyword = true|factor = true) = 0.8$ and $P(keyword = true|factor = false) = 0.3$.

To capture expert's preferences about decision factors in the recommendation model, we propose using another approximation method to decrease the complexity of the model. We model each conditional probability distribution associated to a factor DF_i by a leaky Noisy-Or model. We estimate $P(Y = DF_i|X_1, X_2...X_n)$ where $(Y, X_1...X_n)$ are boolean.

As described in equation 4, if one factor has p parents, we need to identify $p + 1$ parameters (including the leak probability) where each parameter p_i defines the interest of the news when the factor X_i is true and all the others are false.

Inspired from [Chulyadyo and Leray, 2014], we propose several ways to obtain these parameters:

Elicitation method : we propose here to estimate each parameter by asking questions to the experts, as described in section II-B. The approach is based on the use of verbal statements like very unlikely or almost certain that are then mapped to probabilities (cf. Figure 2). The use of such a limited set of verbal statements often makes it quite a lot easier for the domain expert to provide assessments of conditional probabilities.

With this solution, we need to ask $p+1$ questions, without directly asking expert' preferences.

Ranking method: In the opposite here, we propose to simply obtain expert' preferences by asking him to sort the parent factors according to their importance. We then assign weights $W_i^{Ranking}$ to each parent (1 for the worst parent, and $p + 1$ to the best one).

Normalizing these weights (by dividing by the max. value)) give us values between $\frac{1}{p}$ and 1. Using these values as probabilities can lead to bias our model by giving too much importance to the leak when the number of parents is small, and a huge importance to the best factor. For this reason, we propose to re-calibrate these weights, by asking two questions to the expert, corresponding to the probabilities P_{min} of the worst factor (the one with $W_i^{Ranking} = 1$ and P_{max} for the best factor (the one with $W_i^{Ranking} = p + 1$).

$$P_i^{Ranking} = P_{min} + \frac{W_i^{Ranking} - 1}{p}(P_{max} - P_{min}) \quad (5)$$

Analytic hierarchy process method: We apply Analytic Hierarchy Process (AHP) [Saaty, 1980] to rank the parent factors and find their weights. For this, we ask the expert for his relative importance for every pair of factors. We can apply the scale shown in Figure 3, proposed in [Saaty, 2008], to express the relative importance of the decision factors. These pairwise comparisons can be expressed as a matrix $[a_{il}]$, referred to as pairwise comparison matrix, where each row and column represents a decision factor such that each cell represents the relative importance between a pair of decision factors.

First we normalize by the max of the column:

$$\overline{a_{ij}} = \frac{a_{ij}}{\max_i(a_{ij})}. \quad (6)$$

Then, the weight W_i^{AHP} of each factor is defined as the mean of all its $\overline{a_{ij}}$:

$$W_i^{AHP} = \frac{\sum_j \overline{a_{ij}}}{p+1}. \quad (7)$$

Unlike the previous method, using the numerical scale helps us to provide a small weight to the worst factor, but using this weights as a probability will lead to overestimate the probability of the best factor.

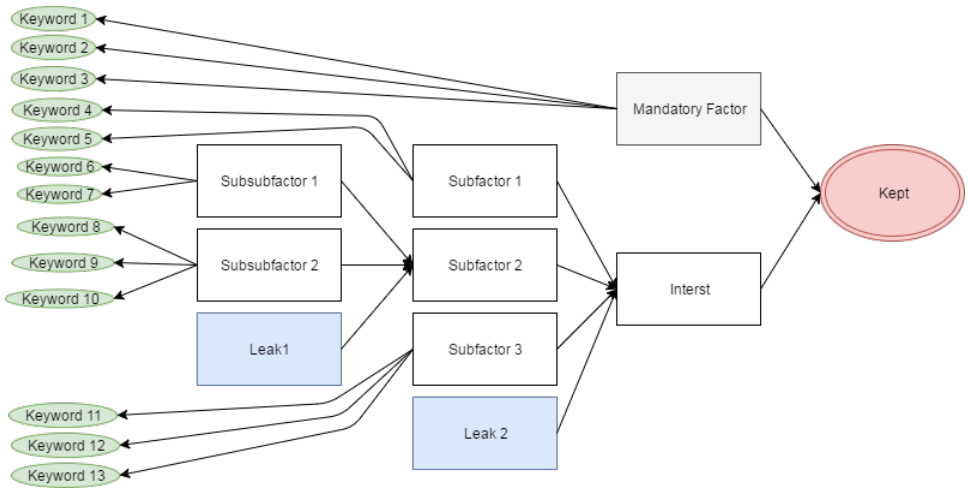

Fig. 1. Bayesian network structure constructed for Lead Scoring

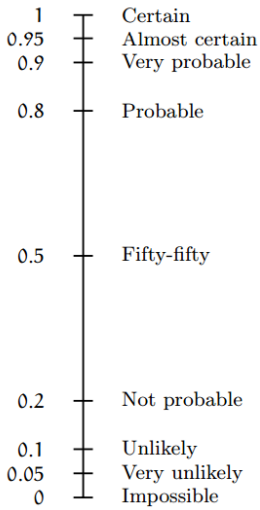

Fig. 2. Scale for the experts' answers [Kjaerulff and Madsen, 2008].

Intensity of importance	Definition
1	Equal importance
2	Weak or slight importance
3	Moderate importance
4	Moderate plus
5	Strong importance
6	Strong plus
7	Very strong or demonstrated importance
8	Very, very strong
9	Extreme importance
Reciprocals of above	If DF_i has one of the above non-zero numbers assigned to it when compared with DF_j, then DF_j has the reciprocal value when compared with DF_i

Fig. 3. The scale of absolute numbers for expressing the relative importance of a pair of decision factors [Saaty, 2008].

We propose then to re-calibrate these weights by asking only the expert (with the previous elicitation method) the probability of this best factor P_{max} in order to estimate :

$$P_i^{AHP} = W_i^{AHP} \times P_{max} \tag{8}$$

.

IV. Case Study

Uneek is a company who developed a CRM tool named Kosmopolead. In order to sell this tool, their commercials browse the internet searching for news about companies near to Uneek's headquarter and which may be interested by their product and try to contact them. The proposed method aims to model the commercials work and predict the companies that eventually will be interested by buying the tool.

Starting from news about companies extracted from newspaper websites, words are extracted and compared to a list of significant words (one list for each decision factor that models the experts criteria).

As seen on the Figure 1 the model is composed from a multiple hierarchy Noisy-OR models. Each of the methods presented in Parameters section have been presented to the expert with examples that shows him the way that it would succeed.

Finally,our expert decided to use the parameter elicitation method derived from AHP since he was comfortable with it.

A. Example

For the sub model (Subsubfactor1, Subsubfactor2, Subfactor2), the first step is to complete the AHP matrix (Table III) by letting the experts giving a comparison scores using the scale Figure 3.

Once the experts complete the comparison, a question concerning the best rated factor is asked to him in order to define the probabilities:

TABLE III
AHP EXAMPLE - COMPARISON OF PARENTS OF SUBFACTOR2.

	Subsubfactor1	SubsubFactor2	Leak
Subsubfactor1	1	3	9
SubsubFactor2	1/3	1	9
Leak	1/9	1/9	1

TABLE IV
LEAKY NOISY-OR SUBFACTOR2 PARAMETERS FOR PMAX=0.5 .

Factors	Probabilities
Subsubfactor1	0.5
SubsubFactor2	0.28
Leak	0.43

TABLE V
THE MODEL RECALL AND PRECISION FOR N=23 EXAMPLES

Precision	0.80
Recall	0.88
Accuracy	0.75

- **Question** : *If the SubsubFactor1 exists alone and no other factor is existing in the news, how much you think this would affect the result of the SubFactor2 ?*

The answer is checked on the scale Figure2. in this example the experts answered with $P_{max} = 0.5$, then the parameters of the Leaky Noisy-Or model are calculated (from equation 8) in Table IV.

B. Validation

Uneek provides us 17 positive news and 6 negative news. Then the model calculates a score for each news based on the news's text. The scores are presented to the expert in order to evaluate them. The experts validation will allow us to update the model in order to make the results better. Until now this task is still open and the work still going on this part since the amount of available data is limited.

The very first version of our model has given good results until now and described in Table V in term of Precision, Recall and Accuracy. One problem that we still face is actually for the negative examples since we have only few examples.

V. CONCLUSION

In this paper, we presented one way of building a Lead scoring model with a Bayesian network.

In addition to its ability of handling uncertainty, Bayesian networks are knowledge representation models that can be built from expertise.

In our specific context, we then proposed to build our Lead scoring model from expertise. We applied usual heuristics to decrease the complexity of our model (parent divorcing, NoisyOr), and proposed three ways of estimating the parameters of our NoisyOr submodels.

The only data available has been used to validate our approach, with good precision and recall results.

Our model has yet to be improved by interacting with the expert, as proposed by [Amershi et al., 2014], in order to provide rapid, focused and incremental learning cycles resulting in a tight coupling between the expert and the model, where the two influence one another.

REFERENCES

[Amershi et al., 2014] Amershi, S., Cakmak, M., Knox, W. B., and Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*.

[Chorianopoulos, 2016] Chorianopoulos, A. (2016). *Effective CRM Using Predictive Analytics*. Wiley.

[Chulyadyo and Leray, 2014] Chulyadyo, R. and Leray, P. (2014). A personalized recommender system from probabilistic relational model and users' preferences. *Procedia Computer Science*, 35(0):1063 – 1072. Knowledge-Based and Intelligent Information & Engineering Systems 18th Annual Conference, KES-2014 Gdynia, Poland, September 2014 Proceedings.

[Díez and Druzdzel, 2000] Díez, F. and Druzdzel, M. (2000). Canonical probabilistic models for knowledge engineering. Technical report, Research Centre on Intelligent Decision-Support Systems.

[Druzdel et al., 2000] Druzdel, M., Van der Gaag, L., Henrion, M., and Jensen, F. (2000). Building probabilistic networks: "where do the numbers come from?" guest editors introduction. *IEEE Transactions on Knowledge and Data Engineering*, 12:481–486.

[Henrion, 1989] Henrion, M. (1989). Some practical issues in constructing belief networks. In Kanal, L. N., Levitt, T. S., and Lemmer, J. F., editors, *Proceedings of the Third Annual Conference on Uncertainty in Artificial Intelligence (UAI '87)*, volume 8 of *Machine Intelligence and Pattern Recognition*, pages 161–174, Amsterdam. North-Holland.

[Kjaerulff and Madsen, 2008] Kjaerulff, U. B. and Madsen, A. L. (2008). *Bayesian Networks and Influence Diagrams A Guide to Construction and Analysis*. Springer.

[Lee and Jo, 2010] Lee, K. C. and Jo, N. Y. (2010). *Bayesian Network Approach to Predict Mobile Churn Motivations: Emphasis on General Bayesian Network, Markov Blanket, and What-If Simulation*, pages 304–313. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Lindstrom, 2016] Lindstrom, M. (2016). *Small Data: The Tiny Clues That Uncover Huge Trends: New York Times Bestseller*. Hachette UK.

[Ngai et al., 2008] Ngai, E., Xiu, L., and Chau, D. (2008). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2):2592–2602.

[Olesen et al., 1989] Olesen, K. G., Kjaerulff, U., Jensen, F., Jensen, F. V., Falck, B., Andreassen, S., and Andersen, S. K. (1989). A munin network for the median nerve-a case study on loops. *Appl. Artif. Intell.*, 3(2-3):385–403.

[Pearl, 1988] Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann.

[Renooij, 2001] Renooij, S. (2001). Probability elicitation for belief networks: Issues to consider. *Knowledge Engineering Review*, 16(3):255–269.

[Saaty, 1980] Saaty, T. (1980). *The Analytic Hierarchy Process, Planning, Piority Setting, Resource Allocation*. McGraw-Hill, New york.

[Saaty, 2008] Saaty, T. (2008). Decision making with the analytic hierarchy process. *International journal of services sciences*, 1(1):83–98.

[Soltani and Navimipour, 2016] Soltani, Z. and Navimipour, N. J. (2016). Customer relationship management mechanisms: A systematic review of the state of the art literature and recommendations for future research. *Computers in Human Behavior*, 61:667 – 688.

[Sun et al., 2013] Sun, P., Guo, X., Zhang, Y., and Wu, Z. (2013). Analytical model of customer churn based on bayesian network. In *2013 Ninth International Conference on Computational Intelligence and Security*, pages 269–271.

[van der Gaag et al., 2002] van der Gaag, L., Renooij, S., Witteman, C., Aleman, B., and Taal, B. (2002). Probabilities for a probabilistic network: a case study in oesophageal cancer. *Artificial Intelligence in Medicine*, 25(2):123–148.

[Verbraken et al., 2014] Verbraken, T., Verbeke, W., and Baesens, B. (2014). Profit optimizing customer churn prediction with bayesian network classifiers. *Intell. Data Anal.*, 18(1):3–24.