Using a Bayesian Network in the ProdFLOW™ Approach

Stefan Wagner
Technische Universität München
Garching b. München, Germany
wagnerst@in.tum.de

Melanie Ruhe
Siemens AG
Munich, Germany
melanie.ruhe@siemens.com

Abstract

ProdFLOW™[1] is a new approach for the productivity analysis and management of research & development organisations created by the research department of the Siemens AG. Its core are organisation-specific models based on the respective substantial levers of productivity. Levers that are both influenceable and measurable are compiled together with the experts of the organisation. This paper proposes to use Bayesian networks for building such models. It is shown how the networks are structured, how they are parameterised and used to analyse different improvement scenarios. Experiences from a case study suggest that Bayesian networks are a suitable technique for the organisation-specific models in ProdFLOW™.

1. Introduction

Research and Development (R&D) activities are strongly affected by the human factor and dominated by cognitive activities and knowledge work, no matter whether its software, hardware or system development. This is a major difference to the manufacturing business. Usually the input and output of the R&D process differ from one R&D project to another and the R&D process is unique, i.e. not directly and completely repeatable. That means existing approaches of measuring and improving productivity in manufacturing need to be adapted to the characteristics of R&D. In the research department of the Siemens AG a new approach for the management of productivity in the area of R&D is being created. Former studies often start with fixed models for the productivity and try to calculate quantitatively the relation between productivity and its influencing factors by analysing regression models. This procedure presupposes that productivity is significantly determined by these influencing factors and that these indicators can easily be changed by the organisation. However, we have made the experience that the so-called typical productivity factors are not that typical within Siemens. Therefore, we depart from the fixed model approach, which might not fit to the conditions of the organisation, and develop a new approach called ProdFLOW™.

Problem To build a new and specific model for each organisation is potentially a very elaborate and difficult task. Also the high uncertainty in the factors and their interrelationships render their results potentially unreliable. The problem we address in this paper is how to build organisation-specific productivity models in a way that the effort is justifiable and their results are still useful.

Contribution We describe how Bayesian networks can be used as a modelling technique for organisation-specific productivity models. This technique fits well because of its suitability for modelling influences between levers as well as modelling uncertainties. The latter are mainly introduced by the questionnaires that are used to fill the model with values. We show the validity of the approach by a case study at Siemens.

Organisation We start by discussing generally what productivity in R&D businesses means in Sec. 2 and summarise the Siemens ProdFLOW™ approach in Sec. 3. The Bayesian network approach for productivity models is proposed in Sec. 4 and the corresponding case study is described in Sec. 5. Conclusions are given in Sec. 6. Related work is cited where appropriate.

2. Productivity in R&D businesses

The well-known definition of productivity (relation of output quantity to the input quantity) is often disputed in the world of R&D. Due to the diversity of disciplines that use the term productivity, there is no clear cut definition of productivity and related terms. This lack of common agreement on what constitutes productivity is perceived as a major obstacle for a substantiated discussion of productivity. Product management aims to release as early as possible, in order to maximise the relative market value, whereas the product development team wants to maximise the creation

[1] ProdFLOW is a registred trademark of the Siemens AG.

of value in the sense of the fulfilment of all customer requirements [1, 3]. The motto *deliver value in time* motivates effective (value) and efficient (in time) product development. Thus, we define productivity in R&D as the relation between value creation for the customer (output) and the effective budget for research and development (input):

$$\text{Productivity} = \frac{\text{R\&D value creation}}{\text{Effective R\&D budget}} \quad (1)$$

Productivity increases if more or better products are developed from the same resources. Better products may be products of higher quality, higher reliability or flexible products (=higher value for the customer) and thus the created value (for the customer) increases. If we develop the same products with fewer resources, productivity may also increase.

3. ProdFLOW$^{\text{TM}}$

ProdFLOW$^{\text{TM}}$ stands for "Productivity in R&D with FLOW". Especially in the context of knowledge work the status of flow [3] should increase productivity. Within ProdFLOW$^{\text{TM}}$ we focus on specific levers when improving productivity in terms of increased value creation and emphasise this using the abbreviation FLOW, which stands for "Focus on Levers to optimise your Work". More details on the approach in general can be found in [9].

The procedure can be applied both to small and large R&D organisations and is also scalable to single phases of the development process. Important is, that the results are developed individually for the evaluated organisation and that it is thus not transferable to other organisations. The aspect of comparability is excluded from this approach because we consider it more important to improve individual productivity than to compare it. We split up the approach into four working steps and the preparation step 0 "Customise Analysis". Those steps, if necessary, can be iterated several times. The main objective is to improve the identified, major/top levers and to lift these into a balanced condition. In the following these individual steps of our approach are described with the respective activities, tasks and results.

3.1. Step 0 – Customise analysis

Step 0 of ProdFLOW$^{\text{TM}}$ has the objective to prepare and plan the subsequent steps. Therefore, the basic goals and characteristics, economic data and future strategic plans of the organisation are analysed. A stakeholder diagram is elaborated to get indications for potential areas of major productivity levers by understanding the network of internal and external stakeholders and their expectations, impacts, cooperation as well as priorities.

3.2. Step 1 – Identify productivity levers

Step 1 has the objective to identify and define the substantial levers with influence on productivity. For the collection of data we employ individual interviews. We make use of an interview guideline for the preparation of the interviews and as a starting point into the productivity analysis. The interview itself is an open but guided conversation with the focus on how value is created for the customer of the analysed organisation. Important is also to query facts and no opinions or rumours. In the follow-up to the interview, minutes with the major information of the interview are developed, i.e. the logging represents an interpretation. Afterwards the logging minutes are provided to the interview partner for authorisation. That gives the evaluation the guarantee that every analysis is based on the right facts. This procedure is based on the work of [1].

Finally a list with levers and their definition based on the aggregated results of the analysis is created. This final activity is critical based on our experience, since the definition of the levers must be formulated objectively and be generally understandable.

3.3. Step 2 – Rank and filter levers

Step 2 has the objective to rank the identified levers regarding the criteria *importance* and *improvement potential* as well as to filter the levers according to the criteria *measurability* and *influenceability*. The results are collected, cumulated and evaluated in the sense of an average ranking, i.e. average rank and the standard deviation for each lever are calculated. In the case of high standard deviation, i.e. very different opinions, the results must be discussed in the organisation to be clarified.

The results are visualised in a so-called prioritisation matrix to present the ranking. The two dimensions of the matrix represent (a) the importance of the lever and (b) its improvement potential. Levers, which are regarded as important as well as having high potential for improvement, are located in the matrix in the upper right quadrant.

In a further step the levers with high priority are filtered regarding to the criteria measurability and influenceability. Levers, which cannot be influenced by the organisation or cannot be measured, will be marked and not further considered in the next steps of the approach.

3.4. Step 3 – Define lever indicators and initiate improvement

Step 3 has the objective, to evaluate and measure the identified top levers and to initiate a coordinated productivity improvement project. This is focused on the top levers that considerably affect productivity. To determine suitable measurement or evaluation instruments measurement expertise is required. The identified levers usually are not standard factors that can be looked up in some reference textbooks. In parallel to the definition of the measurement and evaluation instruments, the organisation has to define measures to improve the levers. This can include a first estimation of the current status of the lever based on the defined measurement instruments.

3.5. Step 4 – Measure levers and determine balance

Step 4 has the objective to track the progress of the productivity improvement project as well as to analyse the balance of levers. The balance is the precondition for the increased value creation at the customer and the related increased productivity. The idea of balance in the context of ProdFLOW$^{\text{TM}}$ is a new concept which tries to analyse and understand the dependencies and influences of the productivity levers on each other as well as on productivity itself. The idea is in line with the concepts of Pareto optimality or Pareto efficiency: An economic system that is Pareto efficient implies that no individual (lever) can be made better off without another being made worse off [6]. Within our approach it means, that there is no reason to improve one single lever if another important lever may deteriorate. Due to the fact, that the levers may have a mutual influence the direction (positive, negative) as well as the intensity (high, medium and low) of the influence has to be analysed.

3.6. Model Requirements

Step 4 of ProdFLOW$^{\text{TM}}$ is the one we concentrate on in the following. So far no explicit models of productivity have been created. The models in this step need to allow an analysis, simulation, and balancing of the productivity levers. From these goals, we can derive a set of requirements on the modelling technique that we use for modelling and analysing the productivity factors.

First, we believe that there are significant influences in between levers that are partly contrary. Hence, it is necessary that the modelling technique allows to model those influences. Second, the productivity factors are diverse and can be measured in various kinds of ways in different scales and units. Therefore, the modelling technique needs to be able to handle those different kinds of data and their combination. Third, for several of the identified productivity factors, it will be necessary to determine their value by expert opinion. This and other measurement methods introduce uncertainty in the data. The modelling technique is required to be able to represent that uncertainty. Fourth, a graphical notation is seen as helpful in order to handle the complex interrelationships. Finally, the effort for developing the model itself needs to be small so that the benefits outweigh the costs.

4. Bayesian network model

We first introduce briefly what Bayesian networks are and how they can be used to model productivity. We explicitly address the difficult issue of probability elicitation for the model and provide the assumptions made in the model.

4.1. Bayesian networks

Bayesian networks, also known as Bayesian belief nets or belief networks, are a modelling technique for causal relationship based on Bayesian inference. It is represented as a directed acyclic graph (DAC) with nodes for uncertain variables and edges for directed relationships between the variables. This graph models all the relationship abstractly. For each node or variable there is a corresponding *node probability table* (NPT). These tables define the relationships and the uncertainty of these variables. The variables are usually discrete with a fixed number of states. For each state, the probability that the variable is in this state, is given. If there are parent nodes, i.e. a node that influences the current node, these probabilities are defined in dependence on the states of these parents. An example is shown in Tab. 1. There the variable is for example with a probability of 70% in the state *low* if both parents are in the state *low*, and with 55% in *low* if the first parent is in *high* and the second is in *low*.

Table 1. An example NPT for a variable with two states and two parents

	low			high		
	low	med	high	low	med	high
low	0.7	0.65	0.4	0.45	0.23	0.07
high	0.3	0.35	0.6	0.55	0.77	0.93

With respect to the above defined requirements on the modelling technique for productivity analysis, we can conclude that Bayesian networks are able to model influences between levers. However, these influences can only exist in one direction. Mutual influences are not possible. Different scales and units can be used as long as there is a way to discretise them. Finally, uncertainty is a first-level citizen in Bayesian networks. Hence, we believe that they are a suitable basis for the models used in our methodology.

Moreover, Stamelos et al. [10] have already used Bayesian belief networks for productivity prediction. They employed the COCOMO I factors and combined them in such a network. They conclude that this is a suitable approach especially when expert judgement has to be included.

4.2. Building a Network

We assume that we have from the first two steps the 3–5 identified top levers from an organisation as input. We only consider those, and not all identified levers, in the network. The reason is the effort for thoroughly modelling these factors and elicit the probabilities. As is stated in [7] "In order to reduce the size of the elicitation task, one should seek to consider only those variables and their values that are absolutely necessary for the specific problem." For these levers, the influences between them need to be included.

From each of these levers, there is an edge to *Productivity* which is measured in *low*, *medium* and *high*. Furthermore, it is common to employ *node divorcing* in order to have not more than two parents per node. If there are more parents, one or more additional nodes can be included in between to reduce the number of parents. Especially for the relation to *Productivity*, this can be employed. However, in our case, usually it is not necessary because we use a simple combination of the input nodes. We also introduce the additional node *Other influences* that accounts for all the other productivity levers not explicitly modelled. This "trick" avoids that a single or very few levers can have a too huge influence on the overall productivity. The complete standard pattern for the used models is shown in Fig. 1.

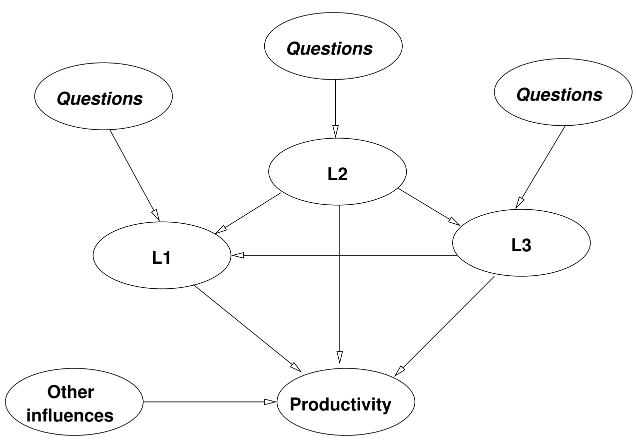

Figure 1. General model structure

4.3. Probability Elicitation

It is an acknowledged problem in Bayesian networks that the elicitation of the probabilities for the node probability tables (NPT) is difficult and elaborate [4]. This is especially the case when the figures need to be estimated by an expert. First, the number of estimates increases exponentially with the states of the predecessor nodes. Second, people are bad at estimating probabilities in general [2]. As in our method probably most of the identified levers need to be estimated with expert opinion, we need clear guidelines for the probability elicitation.

There are several general guidelines w.r.t. eliciting expert judgement [2]. The elicitation questions must be clearly expressed and validated with test runs in order to minimise misunderstandings. Furthermore, it turned out that the expert should know how the responses will be processed. This helps them to give a more accurate estimation.

Various methods have been proposed to mitigate these problems. However, as is discussed for example in [11], they all have drawbacks. A graphical *probability scale* allows the expert to mark they estimated probability on a horizontal or vertical line, usually marked with 0, 50 and 100. The support from the method is conceived as rather low [11] and the so-called *spacing effect* occurs, i.e., the experts tend to organise the probabilities in a way to increase visual attractiveness. It is also often acknowledged [2] that it is easier to think in *frequencies* instead of probabilities. This means it is cognitively easier to imagine that something holds for 5 out of 100 people than for 5%. Yet, in [11] it is reported that the experts had difficulties using frequencies for rare combinations. Finally, *gambles* are also proposed for a more accurate judgement. However, gambles are demanding for the experts and take significant time. What seems to be practically useable is a probability scale with both numbers and words [8].

Fenton et al. [5] employ a truncated normal (TNorm) distribution as basis to generate the corresponding probabilities in an NPT. The experts describe their intuition first for extreme points in a kind of "truth table" and then calibrate the corresponding distribution. This is the state of the art in defining probabilities for Bayesian networks. We will use this technique as implemented in the AgenaRisk tool. However, this still leaves us with the task of defining the mean and variation of the TNorm distribution TNorm(μ, σ^2). For this, we will build on the influences. For all values to be determined, we use expert opinion asked in the form of a questionnaire. The complete elicitation is structured in five parts.

4.3.1 Part I – Lever influences

First we need to determine which levers are influenced by others. This is done by asking for each pair of levers whether they influence each other. If an influence given in both directions, the stronger influence is included in the model. In Fig. 1 *L2* influences *L1* and *L3*, and *L3* also influences *L1*.

4.3.2 Part II – Independent levers

Next, we look at all lever nodes that do not depend on another lever. In the network from Fig. 1, this is the case for *L2*. As in most cases, these nodes will not have a natural distribution, we assume a uniform distribution. This is the standard procedure if no further knowledge is available.

4.3.3 Part III – Dependent levers

The third part contains the determination of the NPTs for the levers that are influenced by other levers. First it has to be determined if the influence is positive or negative, then the strength of the influence needs to be set. We do this in the questionnaire already when asking which levers do influence each other by asking for each pair if there is a light, medium or strong influence. From this the mean μ of the TNorm distribution is derived. This could also be enriched with weights as described in part IV below. For the variance, we need an additional question: How confident are you about the influence from Lever 1 on Lever 2 on a scale from 0 to 10? This question could be aided by a scale with numbers and words describing what that valuation is supposed to mean, for example from *neglectable* to *directly determined*.

4.3.4 Part IV – Productivity

The NPT of the *Productivity* node is determined accordingly. We again use the mean of all levers (using the $1 - l$ normalisation if necessary) in order to determine productivity. The variation is similar to above determined. In this part, it is especially important to add weights to the mean calculation. It is obvious that not all levers will contribute with the same intensity to productivity. This needs to be reflected in weights. We ask the experts to assign 100 points on the influencing factors depending on the strength of the influence. These points can then be used as weights and the Bayesian network is fully completed.

4.3.5 Part V – Current state

For each lever, a set of questions is derived that determines its current state. For example, for a lever called *storing and finding knowledge* we used as one question "How satisfied are you with the current knowledge management system?". The answers of these questions are all on a scale from 1 to 6. In the Bayesian network (Fig. 1 each of the questions is modelled as a separate node as indicated by three example nodes. Their joint influence on each lever determines the current state of the lever in the analysed organisation.

4.4. Assumptions

For the above sketched approach, we use several assumptions to make it practical. We briefly discuss these assumptions in the following.

1. The first assumption is that there is always a dominate influence in one direction from one lever to the other. Bayesian networks do not allow cycles in the graph. Hence, it is not possible to model influences in both directions between two levers. However, from the experiences with the levers we found in the interviews, we are confident that it is sufficient to have only an influence in one direction.

2. The second assumption is that we use a uniform distribution for the node states of nodes that have no parent (i.e. no levers that influence them). A prior distribution does only make sense if there is a natural distribution available, possibly determined empirically. For most levers, this distribution will not be available because of the lack of information for that specific department. A standard statistics procedure is then to use the uniform distribution.

3. Using the *TNorm* distribution for the probabilities in a node with parents. Fenton, Neil and Galan Caballero [5] describe this approach in detail. The idea is that having ranked nodes (i.e. nodes in which the states have a ranking order), the experts find it easier to give the central tendency of the node based on the value of the influencing node. However, this relationship is not completely certain and hence needs an uncertainty distribution around it. This is similar to linear regression where a normal distribution is used to model the uncertainty. For the ranked nodes, this is only changed to the doubly truncated Normal distribution that is only defined in the $[0, 1]$ region. Hence, "This enables us to model a variety of shapes, including a uniform distribution, achieved when the variance $\sigma^2 \rightarrow \infty$, and highly skewed distributions, achieved when $\sigma^2 \rightarrow 0$." [5]

5. Case Study

The first steps of the ProdFLOW$^{\text{TM}}$ approach have been used in several case studies now. We went back to one of case studies done with a Siemens department and added the next step of model building. A questionnaire was prepared with the questions to determine the NPTs and the current state as described above.

In order to then validate whether the model actually describes reality in an acceptable way, the model derived from the questionnaire was presented to 5 volunteers that provided their expert opinion. It is analysed whether the output of the model corresponds to the expectation of the experts. For this, we used 7 scenarios in the model that were rated by the experts on 6-point scale from “Does barely meet expectation (1)” to “Does completely meet expectation (6)”.

The results are shown in Tab. 2. It shows that for nearly all of the scenarios the evaluation lies between 5 and 6 which represents a high correspondence with the expectations. Only The scenario in which one lever was strongly improved, the result is poorer. For this scenario, we repeatedly got the comment that the effect of a single lever should not be that strong.

Table 2. The results from the validation

Expert	1	2	3	4	5	Med.
Scenario “Baseline”	6	5	6	6	6	6.0
Scenario “Very good”	6	5	5	6	5	5.0
Scenario “Very bad”	6	6	3	6	5	6.0
Scenario “Improving L1”	3	4	6	6	6	6.0
Scenario “Improving L2”	5	5	5	6	4	5.0
Scenario “Improving L3”	5	5	5	6	4	5.0
Scenario “Strong L3”	4	2	5		5	4.5

Another frequent comment was that the lever *L2* should have a stronger effect than the lever *L3*. In the model it is the other way round as derived from the questionnaire.

The validation showed mainly three lessons:

- The model in general meets the expectations of the experts.
- There should be an additional factor *Other influences* that has an effect on productivity so that a single lever has less influence. This is already incorporated in the description in section 4.
- The way the questionnaire asked for the influence weights was not optimal as the experts agreed on a different ranking that the questionnaire average.

Possible alternatives to the simple questionnaire would either be a graphical representation that fosters intuitive understanding or a coached, workshop-like interview.

6. Conclusions

ProdFLOW^{TM} is a new approach to productivity analysis for R&D organisations based on the assumption that there cannot be a fixed model of productivity factors valid for all. Therefore, the approach contains organisation-specific models that contain only the most relevant productivity levers. It is challenging to find a suitable modelling technique that provides the necessary mechanisms and most importantly allows an efficient and effective creation of such models.

We propose to use Bayesian networks in such organisation-specific networks because they are able to handle influences between levers, they can work with different scales and units, and they directly support to model the uncertainty in the data. We employed the truncated normal distribution approach [5] for an efficient determination of the needed data that was elicited by a questionnaire. We showed in a case study with a department of Siemens that it is practically possible to build such a model and that it fits well to the experts expectations.

References

[1] P. Aisenbrey. *Deductive Empirical Social Research Based on the Non-Classic Theory of Self-Monitoring*. PhD thesis, FU Berlin, 2007.

[2] R. M. Cooke. *Experts in Uncertainty. Opinion and Subjective Probability in Science*. Environmental Ethics and Science Policy Series. Oxford University Press, 1991.

[3] T. DeMarco and T. Lister. *Peopleware. Productive Projects and Teams*. Dorset House Publishing, 1987.

[4] N. Fenton and M. Neil. Managing risk in the modern world. Applications of Bayesian networks. Knowledge transfer report, London Mathematical Society, 2007.

[5] N. E. Fenton, M. Neil, and J. G. Caballero. Using ranked nodes to model qualitative judgments in Bayesian networks. *IEEE Transactions on Knowledge and Data Engineering*, 19(10):1420–1432, 2007.

[6] T. Little. Value creation and capture: a model of the software development process. *IEEE Software*, 21(3):48–53, 2004.

[7] E. Rajabally, P. Sen, S. Whittle, and J. Dalton. Aids to Bayesian belief network construction. In *Proc. Second IEEE International Conference on Intelligent Systems*, volume 2, pages 457–461. IEEE, 2004.

[8] S. Renooij and C. Witteman. Talking probabilities: Communicating probabilistic information with words and numbers. *International Journal of Approximate Reasoning*, 22:169–194, 1999.

[9] M. Ruhe and S. Wagner. Using the ProdFLOW approach to address the myth of productivity in R&D organizations. In *Proc. 2nd International Conference on Empirical Software Engineering and Measurement (ESEM ’08)*. ACM Press, 2008.

[10] I. Stamelos, L. Angelis, P. Dimou, and E. Sakellaris. On the use of Bayesian belief networks for the prediction of software productivity. *Information and Software Technology*, 45(1):51–60, 2003.

[11] L. C. van der Gaag, S. Renooij, C. L. M. Witteman, B. M. P. Aleman, and B. G. Taal. Probabilities for a probabilistic network: A case-study in Oesophageal Carcinoma. Technical Report UU-CS-2001-01, Institute of Information and Computing Sciences, Utrecht University, 2001.