

An Approach to Tacit Knowledge Classification in a Manufacturing Company

Justyna PATALAS-MALISZEWSKA, Małgorzata ŚLIWA

Abstract: This article attempts to classify the tacit knowledge acquired using the example of the research and development (R&D) department of a manufacturing company. Based on studies in the literature and direct interviews in the company analysed, the authors' own model for classifying the tacit knowledge in an R&D department was proposed. The description of this model has been divided into two parts. In the first part, viz., the classification of knowledge, through three planes: (1) selection of algorithm inputs and the grouping of knowledge accumulated in an enterprise; (2) algorithm activity, that is, the use of algorithms based on clustering them, for calculations; (3) interpretation of results. The Bayesian Network was used for this purpose, which was modelled on the defined relationships between representing of tacit knowledge. Then, on the basis of the case study, the classification of knowledge was prepared according to: (1) definition of knowledge in the R&D department and its modelling, (2) implementation of a suitable number of training sets, (3) verification of the knowledge base, that is, declaration of the value of the knowledge observed, followed by (4) assignment of the probability of returning to the node of network clusters containing interpretations of business benefits.

Keywords: Bayesian Network; case study; manufacturing company; tacit knowledge classification; web-application

1 INTRODUCTION

Managing knowledge generated by the organisation implies important challenges for any institution [1]. Knowledge in an enterprise may not only be explicit, that is, formalised and inclusive of the documentation, in such as reports, catalogues or patents but may also be tacit [2-4] in the natural surroundings and environment [5] and available in the minds of employees [6]. Tacit knowledge is perceived as a prime mover of explorations and creative activities [7] and is identified with research and development (R&D) [8].

Thus, there is a need to develop an approach to tacit knowledge (embedded in the minds of experts) acquiring and classification. Even more so because research on tacit knowledge aspects treats this resource from a behavioural perspective (involvement, trust, motivation), indicating the role of knowledge in the enterprise, but without ready-made commercial solutions [9, 10].

Organising knowledge and assigning it to appropriate categories is supported by simple techniques such as 'clustering', that is, the grouping together of similar categories [11], semantic networks and ontology. Creating logical links between resources is possible using network algorithms based on the probability of their relationships, such as decision trees, Bayesian Networks or the Monte Carlo method. The realisation of the knowledge classification can be also carried out using Petri Networks [12].

This article attempts to classify the tacit knowledge acquired using the example of the R&D department of a manufacturing company. For this purpose, a brief review of the literature on the subject of algorithms has been made, the application of which enables the classification of knowledge.

Based on studies in the literature and direct interviews in the company analysed, the authors' own model for classifying the tacit knowledge in an R&D department was proposed. The description of this model has been divided into two parts. In the first part, viz., the classification of knowledge, the authors suggest classifying three planes of knowledge: (1) selection of algorithm inputs and the grouping of knowledge accumulated in an enterprise; (2)

algorithm activity, that is, the use of algorithms based on clustering them, for calculations; (3) interpretation of results. Bayesian Networks, in conjunction with an internal knowledge base and containing formalised, tacit knowledge and explicit source knowledge were used to this end. The Network model was based on the defined relationships between nodes of tacit knowledge, such as key words, sub-domains and knowledge domains, departments to which knowledge was made accessible, activities performed for solving tasks, the documentation used, the documentation created, staff and nodes artificially created as a competent team, ready-made patterns and resultant nodes, along with the level of knowledge.

2 RELATED WORK

Many different measures are available for the purpose of this evaluation in the context of classification, as described [13]. In the case of clustering, a typical challenge for researchers is the quality of the results or objects which have been clustered but which still do not fit into any cluster. It is also recommended to use a multi-dimensional or criteria-based description of clusters such as colour \wedge shape [14]. Clustering is also one type of unsupervised learning where the goal is the division of objects into groups called clusters [15]. In order to classify knowledge, the expected maximisation algorithm can be used to determine the most likely estimators for models with incomplete data [16]. One good example of the application of this technique, with regard to the implementation of the knowledge classification, is its implementation, when analyzing pharmaco-kinetic/pharmaco-dynamic population modelling [17]. In the knowledge classification may be used also the Principal Component Analysis (PCA) a yield a linear function that reflects the dependency structure for the knowledge acquired. In the context of the knowledge classification, according to [18] also fuzzification permits the conversion of input data into a symbolic representation by means of a fuzzy set. The method for the knowledge clustering is also knowledge mapping, which is often used to quickly diagnose knowledge in organisations, for example when identifying sources of knowledge while

working in a department within the enterprise or, in the case of a process, flow tracking [19]. In the case of knowledge engineering, semantic web is used to defragment descriptions on visualisation. By analysing the image from the study, the words representing semantic terms are selected in appropriate clusters describing, for example, the shape of the object under investigation [20]. The assessment of knowledge contained in semantic networks is similar to that of resource mapping, namely, to measure oscillating knowledge according to its source, or to measure the extent of the hierarchy branches of the semantic network. The resultant modelling allows the use of classification, that is, the decision tree and regression trees. The Bayesian Network and the Bayesian Classifier are used for classifying objects, in the presentation of process dependencies, as well as for making decisions under the conditions of uncertainty [21] - with data and expert knowledge [22]. It is especially useful in the case of limited access to collections of historical data [23]. Petri nets (PNs) are used for the visual representation of knowledge and inferencing phenomenon in expert systems of knowledge modelling [24]. They are used in the analysis and modelling of discrete events which occur only partially [12]. During the statistical estimation of the values which consists directly of many variable factors, the universal Monte Carlo algorithm is very useful. It is used in the modelling of knowledge and is related to performance evaluation in complex systems under uncertainty [25] for financial analysis [26] or estimates with dynamically changing knowledge [27].

Based on the literature research, algorithm for the effective tacit knowledge clustering using the example of the research and development (R&D) department of a manufacturing company, was searched. The R&D department analysed, is engaged in the development of projects preliminarily characterised by the customer, which refer to previously executed orders, in their specificity, such as in their working conditions, loads, co-operating elements, quality and standard requirements. In the R&D department analysed, there is tacit knowledge on the projects carried out by the employees of the company. The activities of the R&D department considered include the design of the connecting elements that are part of the automotive pneumatic system and the devices supporting the production of these components. The sources of tacit knowledge, in the R&D department studied, were also located.

We analysed the presented classification algorithms using the following main criteria: simplicity, modelling with a small number of variables, decision support, including of knowledge a priori, a posteriori, model flexibility, possibility to declare exceptions, clear representation. Based on our analysis results it was decided to use the Bayesian Network to build a research model aimed at clustering knowledge.

3 AN APPROACH TO KNOWLEDGE CLUSTERING - A CASE STUDY

To carry out research based on tacit knowledge, and next – experiments, researchers conducted individual interviews with 2 employees of the company: with the leader of the R&D department and with the senior

constructor. At that time, the tested department numbers 4 workers. It suggests that the tacit knowledge sharing is easier in smaller group of employees trusting themselves [9]. The purpose of the interviews was defining the characteristics and work culture in the R&D department, affecting the knowledge sharing [20] and:

- defining expectations and necessary functionalities for a knowledge management tool supporting the R&D department,
- collecting information on ongoing projects: budget, implementation time,
- defining the sources, and then, methods to explicit and tacit knowledge obtaining,
- domain definition of knowledge, factors influencing the success of R&D projects implementation.

Thanks to the above findings, a questionnaire for acquiring tacit knowledge and groups of nodes comprising the Bayesian network scheme were built. In the next stage of the interviews, the "valve" as a group of products was chosen, based on which the research was carried out. Therefore, a closed set of nodes constituting alternatives within one group of nodes has been established, i.e. keywords (C_c), subdomains (B_{ij}), domains of knowledge (B_i) etc. (in Tab. 1). The last findings concerned the assessments of nodes and the rules construction (see p. 3.1).

In total, 7 meetings were held. The interviews were conducted over a time interval allowing for the analysis and optimization of the assumptions required to design a web-application that would: collate knowledge and interpret the estimated level of knowledge in the Bayesian network. The results were collected in the period from October 2016 to May 2017. The final stage was to obtain 5 completed questionnaires about the "valve" product.

According to [28] modelling with a simple, statistical algorithm appears to be more effective and reliable than do complex methods. This is confirmed even in the case of resistance to the impact of noise or imbalanced data. The concept of a Bayesian Network based on clustering, allows the simple representation and description of knowledge [29]. In the case of a Bayesian Network, designed and implemented for estimating knowledge in the enterprise, the clustering process can be presented in three stages.

Stage 1. Network construction

Pre-clustering: this involves designing network nodes. Their creation involves the initial separation of knowledge into elements influencing the final effect, that is the level of knowledge in an enterprise and in the R&D department. The creation of clusters and objects of knowledge according to their definition [14] refers to such criteria as knowledge localisation, its nature - for tacit knowledge it is node I, for explicit knowledge it is node F- and the subject matter of the knowledge. Regarding the other nodes those artificial nodes representing knowledge which had not been acquired through the form were created by logical implications. This stage is the longest and most difficult to implement. It is done by acquiring knowledge from experts - by discussion, observation, consultation and the use of metaphors [30].

Stage 2. Network performance

Proper clustering: this takes place while the network is running and relies on the statistical probability of the nodes

when declaring observed objects. It includes assignment to the cluster by determining the values for the dependent nodes and then specifying the probability of obtaining a given status for that node. This stage of the network is the shortest and the most automated.

Stage 3. Interpretation of probability

Final clustering: this is the assignment of the probability result, obtained for node Z, $P(Z=T)$ at the intervals specified, which indicate the potential for business benefits, in the context of the knowledge that is simulated in the network. The tie-in with this stage is again possible thanks to activities supporting the acquisition of tacit knowledge. As a result, the anticipated business benefits corresponding to the knowledge status in cluster Z, are declared. Modification of the compartments is coupled with statistical analysis.

A list of established Bayesian nodes, along with their designation, description and an indication of the number of selections, that is, their criteria for the initial clustering stage, is provided in Tab. 1.

Table 1 A list of nodes with their description, along with the number of selections and clustering criteria, based on [31]

| Symbol | Description of the node | Number of choices | Criteria for clustering |
|----------|--|---|---|
| B_i | Domain of knowledge | $\{b_{1-5}\}$ | according to the chosen domain |
| B_{ij} | Sub-domain of knowledge | $\{b_{11-13}\},$ $\{b_{41-47}\},$ $\{b_{51-56}\}$ | according to the chosen sub-domain |
| C_c | Keywords | $\{c_{1-98}\}$ | according to the keyword chosen |
| D_d | Who obtained access to knowledge (departments) | $\{d_{1-5}\}$ | according to the department in the enterprise |
| E_e | Actions performed for the solution | $\{e_{1-5}\}$ | according to the source of knowledge |
| F_f | Documents used for the solution | $\{f_{1-9}\}$ | |
| F_- | Groups of documents used to solve the problem | $\{f_{1-3}\}$ | - |
| G_g | Documents created after the solution | $\{g_{1-9}\}$ | according to the source of knowledge |
| I_i | Employee involved in the task | $\{i_{1-3}\}$ | |
| N | A competent team | $\{n_1, n_2\}$ | - |
| O | Ready-made templates | $\{o_1, o_2\}$ | - |
| Z | Level of knowledge | $\{z_1, z_2\}$ | - |

3.1 The Bayesian Network

The Bayesian Network, using the "a priori" type of probability, was used, which is the value of the constants assigned to the model. In addition, networks combine "a posteriori" probability, taking into account the knowledge gained from the experts and contained in the databases. As a result of consultations with the leader of the R&D department, 5 supplementary forms on R&D projects were obtained and implemented as a training set.

The values of the node state were determined based on the actual number of elements, in relation to their maximum, potential number; the states of the nodes, meanwhile, were artificially defined, with the lack of knowledge representing a given node in the knowledge

base; these were determined according to the rules defined in Eq. (1), (2), (3).

The rule for node F:

$$P(F_{-}=1 (\text{True})) = \sum_1^3 (S_{f'} \cdot p_{f'}) \tag{1}$$

where: $S_{f'}$ - the alternate state in node F_f (1 True/0 False), $p_{f'}$ - the probability of the occurrence of F_f node in the base. f - the number of alternatives from 1 to 3; $f \in N$.

The rule for node O:

$$P(O=1 (\text{True})) = \left(0,65 \sum_1^9 \frac{S_{g'} \cdot w_g}{\sum_1^9 w_g} \right) + \left(0,35 \sum_1^3 \frac{S_{f_-} \cdot w_{f_-}}{\sum_1^3 w_{f_-}} \right) \tag{2}$$

where: $S_{g'}$ - the alternative state in node G_g (1 True/0 False), S_{f_-} - the alternative state in node F_f (1 True/0 False), w_g - the level of significance in the resulting document, determined by the AHP method, w_{f_-} - the weight of the group of documents used (w_f takes on a scale 1-2-3), p_g - the probability of occurrence of G_g node in the base, g - the number of alternatives from 1 to 9; $g \in N$, f_- - the number of alternatives from 1 to 3; $f \in N$.

The rule for node N:

$$P(N=1 (\text{True})) = \left(0,25 \sum_1^5 S_{e'} \cdot p_e \right) + \left(0,15 \sum_1^5 S_{d'} \cdot p_d \right) + \left(0,6 \sum_1^3 S_{i'} \cdot p_i \right) \tag{3}$$

where: $S_{e'}$ - the alternative state in node E_e (1 True/0 False), $S_{d'}$ - the alternative state in node D_d (1 True/0 False), $S_{i'}$ - the alternative state in node I_i (1 True/0 False), p_e - the probability of the occurrence of E_e node in the base, p_d - the probability of the occurrence of D_d node in the base, p_i - the probability of the occurrence of I_i node in the base, e - the number of alternatives from 1 to 5; $e \in N$, d - the number of alternatives from 1 to 5; $d \in N$, i - the number of alternatives from 1 to 3; $i \in N$.

The rule for node Z

Based on consultation with the company's representative, specific values were devised for the states of node Z relative to node O and N, which had only two states: 0 - False and 1 - True. For node Z, the intervals of expected benefits from the process for the enterprise, managed with knowledge, were declared and comprised faster completion of the project, a reduction in financial inputs and internal training, the selection of employees and innovation. Benefit intervals indicated low level – $0.00 \leq P(Z=T) < 0.35$; average level – $0.35 \leq P(Z=T) < 0.70$ and high level – $0.70 \leq P(Z=T) \leq 1,00$.

The research results are based on the received five knowledge questionnaires from the leader of the R&D department. The values in Tab. 2 include knowledge that State1 $\neq 0$. The value of the State 1 for the knowledge not included in the table is 0, the states in other nodes are defined by the rules (see rules for node F_- , O, N).

Meters allowing verification of network operation are identical with its calibration and learning. In this case,

learning is done by manually updating the network, which suggests 100% compatibility of the probability presented in the JPT tables, what is reflected in the respondents' answers. The approach with goal measures ("Z" node) is based mainly on the statistics module in the KnowledgeNets web-application.

Table 2 Designated elements of knowledge

| Symbol | No of choices | Symbol | Cardinality | State 1 | Symbol | Cardinality | State 1 | Symbol | Cardinality | State 1 |
|----------|---------------|--------|-------------|---------|--------|-------------|---------|--------|-------------|---------|
| B_i | 1 | B1 | 1 | 0.20 | B4 | 2 | 0,40 | B5 | 2 | 0.40 |
| B_{ij} | 3 | B11 | 1 | 0.07 | B48 | 1 | 0,07 | B53 | 1 | 0.07 |
| | | B42 | 1 | 0.07 | B51 | 1 | 0,07 | - | - | - |
| C_c | 10 | C1 | 1 | 0.02 | C30 | 1 | 0,02 | C73 | 1 | 0.02 |
| | | C2 | 1 | 0.02 | C63 | 1 | 0,02 | C74 | 1 | 0.02 |
| | | C37 | 1 | 0.02 | C64 | 1 | 0,02 | C75 | 1 | 0.02 |
| | | C61 | 1 | 0.02 | C72 | 1 | 0,02 | C76 | 1 | 0.02 |
| | | - | - | - | - | - | - | - | - | - |
| D_d | 5 | D1 | 1 | 0.04 | D4 | 2 | 0,08 | - | - | - |
| E_e | 5 | E1 | 1 | 0.04 | E2 | 2 | 0,08 | E3 | 3 | 0.12 |
| F_f | 9 | F1 | 1 | 0.02 | F3 | 1 | 0,02 | F7 | 1 | 0.02 |
| | | F2 | 2 | 0.04 | F6 | 1 | 0,02 | - | - | - |
| G_g | 9 | G1 | 3 | 0.07 | G3 | 1 | 0,02 | G7 | 2 | 0.04 |
| | | G2 | 2 | 0.04 | G6 | 1 | 0,02 | - | - | - |
| I_i | 1 | I1 | 5 | 1 | - | - | - | - | - | |

The network modelled, along with probability result Z, in the resting state of the network, is shown in Fig. 1. The occurrence of event Z, also referred to as the level of knowledge, is estimated at 38%.

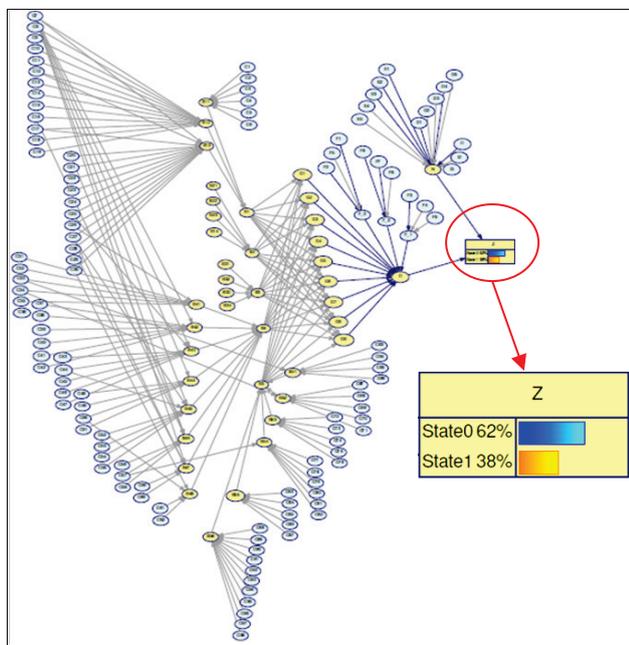


Figure 1 The Bayesian Network at rest; made in GeNie

Based on the data from Tab. 2, three selected simulations of network operation were performed; these comprised the clustering of the knowledge studied on the available training set, under given observation conditions:

Simulation A - performed for an artificially modelled, intermediate node; conditions of observation A: Target = N, Z; Set evidence = D1, E2 - State 1 (True)

The probability of the occurrence of an event is $P(N=T) = 0.63$.

Simulation B and C - performed for the artificially modelled end node; conditions of observation B: Target = Z; Set evidence = D1, D4, E1, E2, I1 - State 1 (True)

The probability of the occurrence of an event is $P(N=T) = 0.39$; conditions of observation C: Target = Z; Set evidence = D1, D4, E1, E2, I1, O - State 1 (True)

The probability of the occurrence of an event is $P(N=T) = 0.90$.

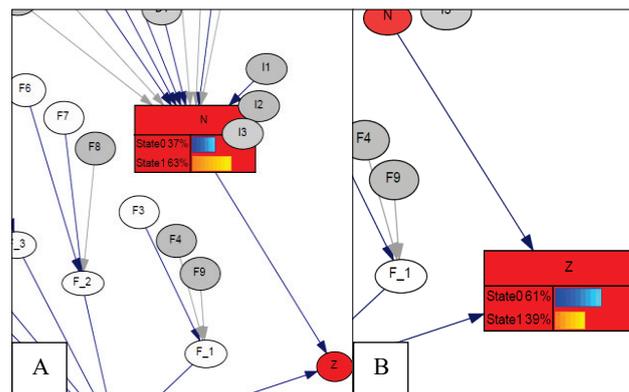


Figure 2 Example simulations carried out in the network modelled and named; made in GeNie, own study

Verification of the interpretation of the final node Z is possible thanks to the introduction and observation of data concerning the project, implemented in the R&D department. This, in turn, is transformed into statistics, backed up by visualisation.

The results obtained from the simulation (in Fig. 2), have been interpreted as follows:

- simulation A (in Fig. 2A): $P(N=T) = 0.63$ belongs to the average range: $0.35 \leq P(X=T) < 0.70$.
Tip: It is recommended to supplement the knowledge base of the knowledge areas observed.
- simulation B (in Fig. 2B): result $P(N=T) = 0.39$ belongs to average range: $0.35 \leq P(X=T) < 0.70$.
Tip: A quicker completion of the project may be effected using available resources: knowledge, human resources and tools from the knowledge area observed (assumption: 5-10% time reduction).

Reduction of financial expenditures on new projects utilising knowledge areas observed (assumption: reduction by 4%).

Reduction of internal employees training- or their faster engagement- from knowledge areas observed (assumption: reduction of the implementation time by 10%).

Selection of employees: this assumes the moderate competence of the particular internal worker delegated to the project who uses the knowledge areas observed.

The likelihood for the creation of innovation when designing a project using observed knowledge areas is average.

Simulation C: the result of $P(N=T) = 0.90$ belongs to the high range: $0.70 \leq P(X=T) \leq 1$.

Tip: Faster project completion (assumption: 15% reduction of time).

Reduction in financial outlay (assumption: reduction by 8%).

Reduction of internal employee training (assumption: reduction of implementation time by 20%). Selection of employees - the company has internal experts dealing with subjects from the area of knowledge observed.

It is unlikely that innovation will be forthcoming.

3.2 The KnowledgeNets Web Application

A web application named KnowledgeNets (KN) supports the modelled Bayesian network. It is based on acquiring tacit knowledge in the R&D department using a questionnaire with a closed list of responses (in Tab. 1, form is only for nodes B to F), and on indexing this knowledge in the database (in Fig. 3). The role of the application is to increase the know-how, but also the ability to verify the expected benefits for the company.

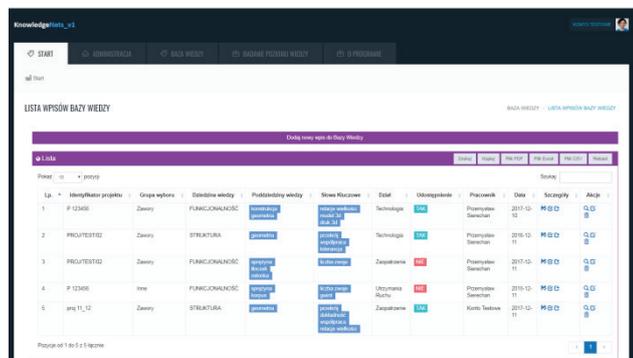


Figure 3 View of tab with knowledge base records, own study

Collecting data necessary to keep statistics on the benefits for an enterprise described in the work, allows the LEVEL of KNOWLEDGE tab → ADD DATA TO STATISTICS → Add data for analysis (in Fig. 4). Two forms to collect data were distinguished – describing internal training (in Fig. 5), and there:

- time for internal training (for R&D employees), here: 200, realised in [month, year], here: 12 2017,
- number of R&D employees in a given month, here: 6.

Form regarding completed projects, consists of five fields:

- project identifier, here: PROJ/TEST/45,
- start date of the project - in YYYY-MM-DD format,
- end date of the project - in YYYY-MM-DD format,
- real realised time of R&D project and assumed time,
- the used budget and the planned budget of R&D project in thou. PLN.

The LEVEL of KNOWLEDGE TEST tab → LEVEL of KNOWLEDGE is used to implement activities related to the observation of the level of knowledge in the enterprise. Previously, it is necessary to make the analysis of the observed knowledge in the external GeNIE program, with the taught Bayesian network (pp. 3.1, in Fig. 2). Next, its user should enter the results from GeNIE to the KN tab: in the field "Result from the end node, State 1 =", the field "Observed nodes" where should be added information

about the analysis; here are the chosen nodes: C2, C8, C10, C17, B5, B12, I2. After clicking on Analyze button, the contents of the interpretation are displayed, which are automatically archived. This provides help in analyzing and reinterpreting the level of knowledge for further measurements.

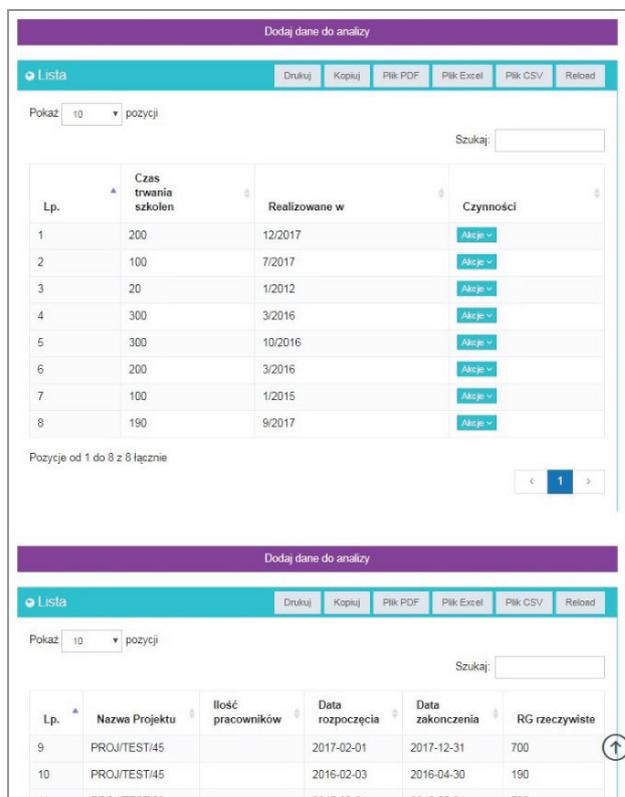


Figure 4 View of tab with statistical data records, own study

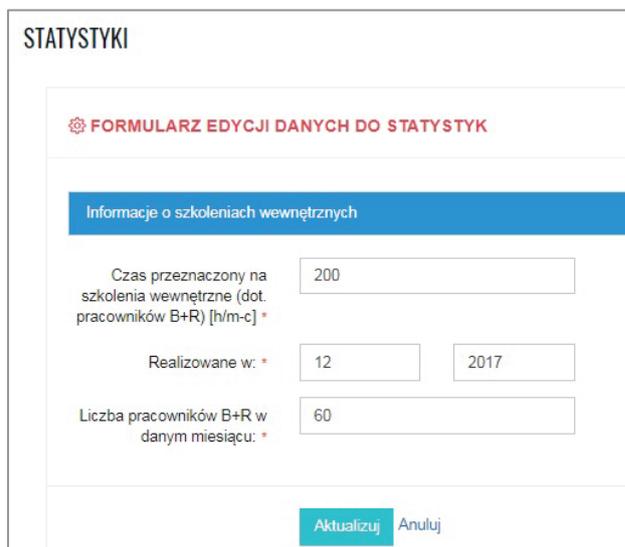


Figure 5 View of tab for entering data into statistics: internal trainings, own study

Entering and observing project data enables verification of the assumed benefits for enterprise in the final node (Z). These are transformed into statistics, supported by graphic visualization. As to the interpretation of benefits, corresponding to the content assigned to individual probability intervals, management have the discretion to change it. It is possible in the module LEVEL OF KNOWLEDGE tab → STATISTICS. Generating a

view of a selected chart may reflect the annual average of observed values or all points separately. In the KN application also proposed an optional declaration of the analysis period by indicating the date range. Monitoring concerns:

- the number of completed projects, including: all projects [pcs], number of completed projects in the declared period [pcs/time],
- duration of projects, including: real realised time of R&D project [man hour (m-h)], assumed realised time of R&D project [m-h] for the duration of projects in the declared period [m-h/period],
- project budget, including: the total budget used for the realised projects [thou. PLN], used budget in the period declared [thou. PLN / period], the ratio of the real to assumed budget, within the declared period [-/period],
- number of internal trainings in a given year, including: total training time per employee [h/man (h/m)], total training time per employee in the declared period [h/m/period],
- R&D department employee activity (entries to the knowledge base), including: total employee activity [pcs. entries], employee activity in the declared period [pcs. entries/period].

In the case of the analysis of the duration of the project and its budget, an additional classification was introduced, classifying records into 4 groups: A, B, C and D, and in them to separate data series, as described below:

- A: duration up to 3 months and: up to 20 thou. PLN; over 20 to 100,000 thou. PLN
 B: duration up to 6 months and: up to 20 thou. PLN; over 20 to 100 thou. PLN; over 100 thou. PLN
 C: duration up to 1 year and: up to 100 thou. PLN; over 100 thou. PLN
 D: duration over 1 year and: up to 100 thou. PLN; over 100 thou. PLN.

This means that the introduction, in addition to the actual time of realisation of the R&D project and budget, contains the third dimension – passing time on the abscissa. Such action will allow for a better validity and confrontation with the interpretation of benefits for the enterprise in the case of observation of the trend line. This means that, for example, an analysis of the budget reduction of the XYZ project with an assumed duration of 2.5 months and a budget of 45 thou. PLN, should be carried out only into category A projects (time up to 3 months), and more precisely – in relation to the data from the series "above 20 to 100 thou. PLN".

For presenting example in the STATISTICS tab, the budget of the project was selected for observation with the presentation of all values option (from the period from 2015-01-01 to 2018-01-01). The generated effect is mainly charts of all categories and series, assuming that there is at least one record that suits it. Additionally, after hovering over any point on the straight line or on the graph, the description is displayed, ex.: $y = 0.0028x + 1110.604$, $y = 63.71$ thou. PLN, or the selected value along with the series description (ex. Budget > 20 thou. PLN, this point is for 120 thou. PLN), as shown in Fig. 6.

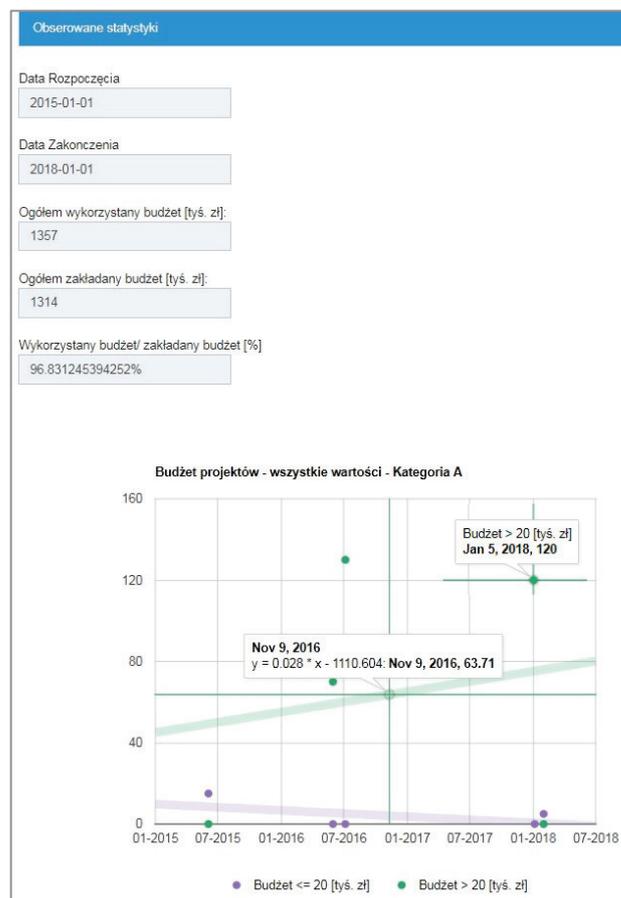


Figure 6 View of the tab with statistics for the budget of R&D project (category A) with the trend line description (left), the point's description (right), own study

The work with the KN application helps to verify the approach - interpretation of the estimated knowledge. Based on the knowledge research methods described above, it is possible to edit the assumptions as well as the range of probability intervals. As a result, it allows for a proper selection of instructions related to the level of knowledge.

The approach proposed, with regard to the classification of knowledge, enables knowledge regarding the research and development projects implemented within an enterprise, to be organised. The result of the proposed model is the obtainment of the values of node Z states, relative to nodes O and N, which makes it possible to evaluate the expected benefits for the enterprise, vis-à-vis the implementation of the knowledge management process. The present knowledge classification model is tailored to the needs and requirements of the R&D department of the enterprise in question. However, construction of this model assumes the usefulness of its use in the knowledge management process for other companies. At the design stage of the model, it is important to define:

- alternatives assigned to a given knowledge cluster – possible extension and reduction,
- alternative attributes - possible expansion and reduction,
- dependencies between nodes,
- rules of evaluation for artificially defined nodes which are key for the separation of component rules,
- supporting methods for simple Bayesian inference, viz., the linking of algorithms,

- benefits,
- probability ranges for benefits – possible expansion and reduction,
- value of benefits and boundaries of probability intervals.

The decisive advantage of the approach is a quantitative character to the issue of tacit knowledge acquiring and classification in the R&D department and inference when evaluating: (1) available in the knowledge department and (2) project management (mainly in the initial scheduling).

5 CONCLUSION

The article describes the techniques selected, viz., EMT, PCA, Fuzzy logic, knowledge mapping, semantic net, classification, regression trees, Bayes, Petri net and the Monte Carlo Algorithm, which support the knowledge classification. The selection was made and the Bayesian Network was chosen and implemented for its immunity to the impact of noise or imbalanced data and the ability to describe knowledge [29]. The design of the classification model by means of the Bayesian Network, without supportive rules/algorithms, is only possible if the number of knowledge elements and states is not fully developed. According to the present authors, a maximum of 13 inputs per single intermediate node should be aimed for, assuming two states that is T/F . The number of state combinations in a given node is defined as the number of states (s) in the input node in relation to the power of the sum of the inputs (w) of all nodes: sw . To avoid excessive expansion of the structure, node F was introduced. Then, on the basis of the case study, the classification of knowledge was prepared according to: (1) definition of knowledge in the R&D department and its modelling, (2) implementation of a suitable number of training sets, (3) verification of the knowledge base, that is, declaration of the value of the knowledge observed, followed by (4) assignment of the probability of returning to the node of network clusters containing interpretations of business benefits.

The developed approach is based on universal stages involved in our research model and can be used in another industry. As the example, our model was verified in the medium size pneumatic enterprise from the automotive sector. To adapt our approach to tacit knowledge classification of another company the two main activities should be implemented: adjusting the dictionary related to the knowledge within a company and verification of the computational efficiency of the Bayes network structure for a larger number of entry nodes.

5 REFERENCES

- [1] Cantú, F. J. & Ceballos, H. G. (2010). A multiagent knowledge and information network approach for managing research assets. *Expert Systems with Applications*, 37(7), 5272-5284. <https://doi.org/10.1016/j.eswa.2010.01.012>
- [2] Nonaka, I. & Takeuchi, H. (1995). *The knowledge-creating company. How Japanese Companies Create the Dynamic of Innovation*. New York, NY: Oxford University Press.
- [3] Nonaka, I., Toyama, R., & Konno, T. (2000). SECI, Ba and Leadership: A Unified Model of Dynamic Knowledge Creation. *Long Range Planning*, 33(1), 5-34. [https://doi.org/10.1016/S0024-6301\(99\)00115-6](https://doi.org/10.1016/S0024-6301(99)00115-6)
- [4] Nonaka, I. & Von Krogh, G. (2009). Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organisational Knowledge Creation Theory. *Organisation Science*, 20(3), 635-652. <https://doi.org/10.1287/orsc.1080.0412>
- [5] Holste, J. S. & Fields, D. (2010). Trust and Knowledge Sharing and Use. *Journal of Knowledge Management*, 14(1), 128-140. <https://doi.org/10.1108/13673271011015615>
- [6] Jashapara, A. (2006). *Zarządzanie wiedzą*. Warszawa: Polskie Wydawnictwo Ekonomiczne.
- [7] Yang, S. C. & Farn, C. K. (2009). Social capital, behavioural control and tacit knowledge sharing – A multi-informant design. *International Journal of Information Management*, 29(3), 210-218.
- [8] Song, J., Asakawa, K., & Chu, Y. (2011). What determines knowledge sourcing from host locations of overseas R&D operations? A study of global R&D activities of Japanese multi-nationals. *Research Policy*, 40(3), 380-390. <https://doi.org/10.1016/j.respol.2011.01.002>
- [9] Shao, Z., Wang, T., & Feng, Y. (2015). Impact of organizational culture and computer self-efficacy on knowledge sharing. *Industrial Management & Data Systems*, 115(4), 590-611. <https://doi.org/10.1108/IMDS-12-2014-0377>
- [10] Liebowitz, J., Ayyavoo, N., Nguyen, H., Carran, D., & Simien, J. (2007). Cross-generational knowledge flows in edge organizations. *Industrial Management & Data Systems*, 107(8), 1123-1153. <https://doi.org/10.1108/02635570710822787>
- [11] Ghattas, B., Michel, P., & Boyer, L. (2017). Clustering nominal data using unsupervised binary decision trees: Comparisons with 'state of the art' methods. *Pattern Recognition*, 67, 177-185. <https://doi.org/10.1016/j.patcog.2017.01.031>
- [12] Cassandras, C. G. & La Fortune, S. (2008). *Introduction to Discrete Event Systems*. New York, NY: Springer-Verlag.
- [13] Pfitzner, D., Leibbrandt, R., & Powers, D. (2009). Characterisation and evaluation of similarity measures for pairs of clusterings. *Knowledge and Information Systems*, 19(3), 361-394. <https://doi.org/10.1007/s10115-008-0150-6>
- [14] Färber, I., Günemann, S., Kriegel, H., Müller, E., Schubert, E., Seidl, T., & Zimek, A. (2010). On Using Class-Labels in the Evaluation of Clusterings, *16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, USA, July 25-28, Retrieved from <http://www.imada.sdu.dk/~zimek/publications/MultiClustAtKDD2010/Faerberetal.pdf>
- [15] Cornuéjols, A., Wemmert, C., Gańczarski, P., & Bennani, Y. (2018). Collaborative clustering: Why, when, what and how. *Information Fusion*, 39, 81-95. <https://doi.org/10.1016/j.inffus.2017.04.008>
- [16] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1-38. Retrieved from <http://www.jstor.org/stable/2984875>
- [17] Yates, J. W. T. (2009). An implementation of the Expectation-Maximisation (EM) algorithm for population pharmacokinetic – pharmacodynamic modelling in ACSLXTREME. *Computer Methods and Programs in Biomedicine*, 96(1), 49-62. <https://doi.org/10.1016/j.cmpb.2009.03.011>
- [18] Nait-Said, R., Zidani, F., & Ouzraoui, N. (2008). Fuzzy risk graph model for determining safety integrity level. *International Journal of Quality, Statistics, and Reliability*, 2008, 1-12. <http://dx.doi.org/10.1155/2008/263895>
- [19] Śliwa, M. & Patalas-Maliszewska, J. (2016). A strategic knowledge map for the research and development department in a manufacturing company. *Foundations of Management*, 8(1), 151-166. <https://doi.org/10.23743/acs-2017-20>

- [20] Kesti, M. & Syväjärvi, A. (2010), Human tacit signals at organization performance development. *Industrial Management & Data Systems*, 110(2), 211-229. <https://doi.org/10.1108/02635571011020313>
- [21] Fenton, N. & Neil, M. (2014). Decision Support Software for Probabilistic Risk Assessment Using Bayesian Networks. *IEEE Software*, 31(2), 21-26. <https://doi.org/10.1109/MS.2014.32>
- [22] Fenton, N. & Neil, M. (2012). *Risk assessment and decision analysis with Bayesian Networks*. Boca Raton, FL: Crc Press.
- [23] Constantinou, A. C., Fenton, N., & Neil, M. (2016). Integrating expert knowledge with data in Bayesian Networks: Preserving data-driven expectations when the expert variables remain unobserved. *Expert Systems with Applications*, 56, 197-208. <https://doi.org/10.1016/j.eswa.2016.02.050>
- [24] Yanrong, H. & Yang, S. X. (2004). A knowledge based genetic algorithm for path planning of a mobile robot, Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference, IEEE, 5, 4350-4355. <https://doi.org/10.1109/ROBOT.2004.1302402>
- [25] Torres-Jiménez, M., García-Alonso, C. R., Salvador-Carulla, L., & Fernández-Rodríguez, V. (2015). Evaluation of system efficiency using the Monte Carlo DEA: The case of small health areas. *European Journal of Operational Research*, 242(1), 525-535. <https://doi.org/10.1016/j.ejor.2014.10.019>
- [26] Glasserman, P. (2004). *Monte Carlo methods in financial engineering*. New York, NY: Springer-Verlag.
- [27] Chen, K. H. (2012). Dynamic randomization and domain knowledge in Monte-Carlo Tree Search for Go knowledge-based systems, *Knowledge-Based Systems*, 34, 21-25. <https://doi.org/10.1016/j.knosys.2011.08.007>
- [28] Van Hulse, J. & Khoshgoftaar, T. (2009). Knowledge discovery from imbalanced and noisy data. *Data & Knowledge Engineering*, 68, 1513-1542. <https://doi.org/10.1016/j.datak.2009.08.005>
- [29] Xing, P., Jun, W., & Lu, L. (2007). Knowledge Sharing Model based on Concept Clustering. *Systems Engineering – Theory & Practice*, 27(2), 126-132. [https://doi.org/10.1016/S1874-8651\(08\)60014-9](https://doi.org/10.1016/S1874-8651(08)60014-9)
- [30] Śliwa, M. (2016). Uwarunkowania procesu modelowania i oceny efektywności systemu wspomagającego zarządzanie wiedzą ukrytą w przedsiębiorstwach produkcyjnych. *Przedsiębiorczość i Zarządzanie*, 17(7/2), 264-278. Retrieved from <http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.ekon-element-000171488667>
- [31] Śliwa, M. (2017). Koncepcja oceny poziomu wiedzy technicznej w dziale badawczo-rozwojowym: studium przypadku. *Zeszyty Naukowe Wydziału Elektroniki i Informatyki*, 11, 87-108. Retrieved from <http://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-818ee9c2-599f-4b2e-bb9d-f0fe7a3b6c31>

Contact information:

Justyna PATALAS-MALISZEWSKA, Prof. D.Sc. Eng.
University of Zielona Góra, Mechanical Department,
ul. prof. Z. Szafrana 4,
65-516 Zielona Góra, Poland
J.Patalas@iizp.uz.zgora.pl

Małgorzata ŚLIWA, M.Sc. Eng.
University of Zielona Góra, Mechanical Department,
ul. prof. Z. Szafrana 4,
65-516 Zielona Góra, Poland
M.sliwa@iizp.uz.zgora.pl