Marcin MATUSZNY, MSc* ORCID: 0000-0002-4124-7355

DOI: 10.15199/180.2021.1.4

* University of Bielsko-Biała, Faculty of Mechanical Engineering and Computer Science,
speciality: Mechanical Engineering
e-mail: marcinmatuszny@wp.pl

INDUCTION OF DECISION TREES FOR BUILDING KNOWLEDGE BASES OF PRODUCTION PROCESSES

INDUKCJA DRZEW DECYZYJNYCH DLA BUDOWY BAZ WIEDZY PROCESÓW PRODUKCYJNYCH

Summary: *The article presents the process of acquiring the knowledge based on the induction of decision trees, graphically illustrating the differences between acquiring the knowledge in a traditional way, from the expert, and the process of acquiring the knowledge supported by the machine learning methods. The methods of acquiring the knowledge are discussed and specified. The practical part represents the use of De Treex 4.0 software dedicated to the induction of decision trees, which is a part of the Sphinx 4.0 artificial intelligence package.*

Keywords: *decision trees, production processes, knowledge bases, knowledge acquisition, production engineering*

Streszczenie: *W artykule przedstawiono proces pozyskiwania wiedzy w oparciu o indukcję drzew decyzyjnych, w sposób graficzny zilustrowano różnice pomiędzy pozyskiwaniem wiedzy w sposób tradycyjny, od eksperta, a także procesem pozyskiwania wiedzy wspomaganym metodami uczenia maszynowego. Omówiono i wyszczególniono metody pozyskiwania wiedzy. W części praktycznej przedstawiono wykorzystanie oprogramowania DeTreex 4.0 dedykowanego do indukcji drzew decyzyjnych wchodzącego w skład pakietu sztucznej inteligencji Sphinx 4.0.*

Słowa kluczowe: *drzewa decyzyjne, procesy produkcyjne, bazy wiedzy, pozyskiwanie wiedzy, inżynieria produkcji*

1. Acquiring the knowledge from production processes for construction of knowledge bases

Construction of knowledge base in expert system, supporting the design of technological treatment processes requires, in many cases, development of the methods for acquiring technological knowledge. The process of acquiring the technological knowledge is aimed, first of all, at obtaining the knowledge and experience in a strictly defined range of the tasks in the field of designing the technological processes from the identifiable knowledge sources as well as recording of the acquired knowledge in the way enabling its application in the process of supporting the decision-making during solving the tasks concerning design of technological processes [1].

The knowledge acquisition from a viewpoint of building the knowledge bases is strongly connected with the conception of machine learning. Speaking in general, the knowledge acquisition may be defined as learning, i.e. obtaining a symbolic knowledge, connected with acquiring the capabilities of utilizing the mentioned knowledge in the effective way [2].

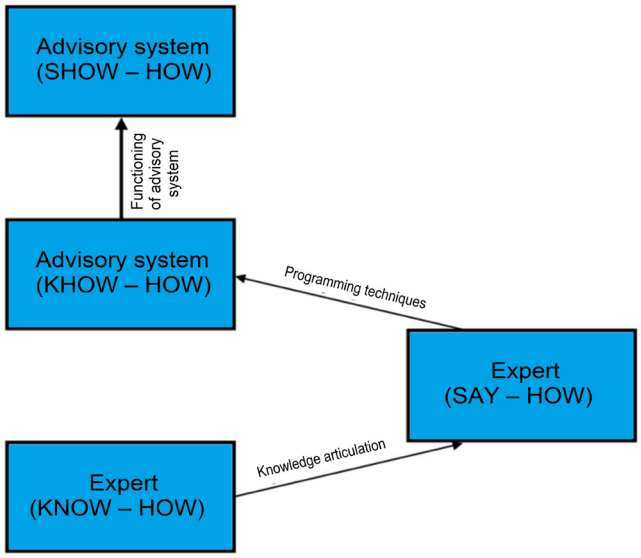
Together with the development of the studies in the field of constructing expert systems, the obstacles appeared; they resulted mainly from the necessity of possessing more and more advanced knowledge bases. It was necessary to develop such methods for knowledge acquiring which would be effective due to the abbreviation of the time of constructing and verifying the knowledge bases in respect of their non-contradiction, completeness and elimination of the excess of information. The most frequently employed methods for knowledge acquiring directly from the expert are as follows:

- Direct consultations with the expert,
- Analysis and observation of the work, performed by the expert,
- The knowledge records, made by the expert in the intentionally developed electronic or paper document.

The presented below figure shows a scheme of the traditional process of knowledge acquisition from the expert what makes the verification, construction and control of the complicated knowledge bases difficult due to the frequent problems in the knowledge articulation by the mentioned expert and the errors in the knowledge records.

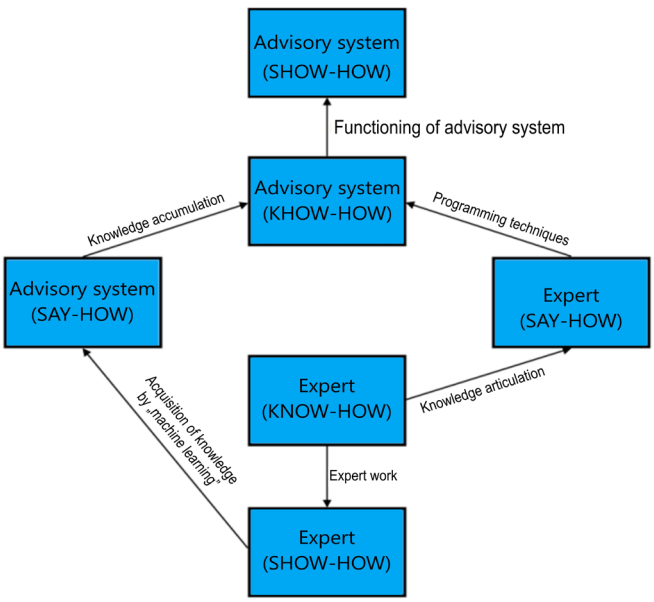
To support the partly characterized above process of acquiring the knowledge from expert, many methods have been developed which facilitate the computerised automated acquiring the knowledge. When employing the mentioned methods, we do not require a direct participation of the expert

Fig. 1. The knowledge acquiring in the "traditional" way; own development on the grounds of [3]



in the process of the knowledge acquisition. The knowledge which was acquired by the above described methods is obtained basing on the results of the expert work and the earlier collected data. The scheme as shown below (Fig. 2) illustrates acquisition of the knowledge from the expert in the process supported by the methods of machine learning.

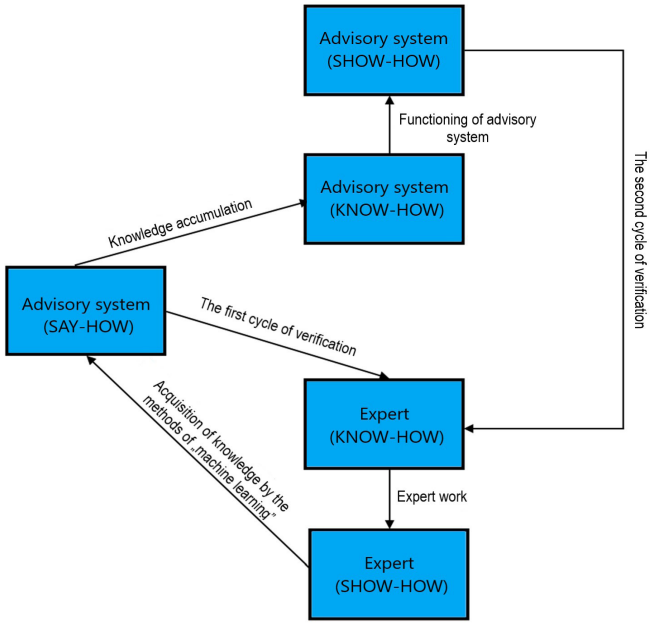
Fig. 2. Acquisition of the knowledge directly from the expert s a process which has been supported by the methods of machine learning; own elaboration on the grounds of [3]



The successive diagram (Fig. 3) illustrates the successive stage in the processes of knowledge acquisition. In the discussed process, the expert does not transfer his knowledge but performs its precise verification. The knowledge was acquired by the methods of machine learning and the task of the expert is also to verify the correctness of expert system functioning as well as to check the conclusions and explanations, obtained from the system. In the case of stating the errors and other irregularities, the expert should recognize the corrections in the knowledge basis as being necessary or even repeated acquisition of the discussed knowledge, based on the supplemented and amended data. The presented below process is more efficient in comparison to two previous ones due to the following reasons:

- Acquisition of the knowledge with the application of the methods based on machine learning is quicker,
- The knowledge bases obtained in the automated way contain lower number of errors or, at optimum situation, do not contain them at all,
- Verification of the knowledge acquired in the described way by the expert is a task considerably simpler and quicker for him than the articulation of the mentioned knowledge.

Fig. 3. The process of acquiring the knowledge supported by the machine learning methods; own elaboration based on [3]



1.1. The methods of knowledge acquisition

The basic idea of learning includes acquiring the knowledge with the application of few methods of reasoning – induction, deduction or analogy. In the specific case, learning may require only duplication of information, supplied by the environment or transformation of the mentioned information and separation only certain part of it which is relevant for us. The process of acquiring the knowledge is classified in dependence on many criteria. We may distinguish the following methods of the knowledge acquisition:

- Direct knowledge acquisition – it does not require concluding and knowledge transformation from the system subjected to learning; it is implemented, as an example, via direct programming; the method is applied in relation to the simple bases of knowledge;

- Acquisition of the knowledge on the grounds of instructions – it requires the necessity of cooperation between the teacher and the learner; it is implemented via the application of the appropriate knowledge sources, indicated by the teacher and their transformation into the language, being acceptable by the learner;
- Acquisition of the knowledge based on the analogy – it consists in such transformation of the existing information that it could be used for the description of facts, similar to those ones, being contained in the knowledge base; they are implemented, for example, by the modification of the computer program;
- Acquisition of the knowledge based on the examples – the method is very often employed when constructing the knowledge bases; it consists in generation of a general description of the classes on the grounds of the collection of examples and counter-examples, representing the mentioned classes; general description is obtained on the basis of induction principle.
- Acquisition of the knowledge based upon the observation – the mentioned methods requires greater participation of the learner during the learning process; the learner may make passive and active observations

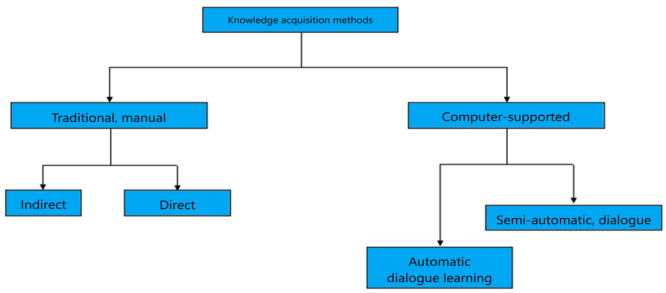
From among many induction methods, we may distinguish the following methods (on the grounds of examples):

- Induction of the rules, using generation of coverage [4],
- Induction of decision trees [5],
- Induction of the rules, with the application of approximated sets [6].

When reassuming the above, the methods of the knowledge acquisition may be classified in a following way (due to the involvement of the software in the process of knowledge acquisition (Fig. 4) [7] :

- Manual methods
- Semi-automatic methods
- Automatic methods

Fig. 4. Classification of the knowledge acquisition methods on the ground of [7]



To employ the mentioned methods, it is necessary to use the so-called attribute model of the description of the area concerned, i.e. determination of the so-called Domain. To make the acquisition of the knowledge possible based upon the examples, it is necessary to specify:

- Objects – phenomena, problems which are to be classified;
- Attributes, describing the given objects and
- Values which will be adopted by particular attributes. Values of the attributes for the selected group of the objects constitute the teaching set and the examples, describing the objects, are called learning examples.

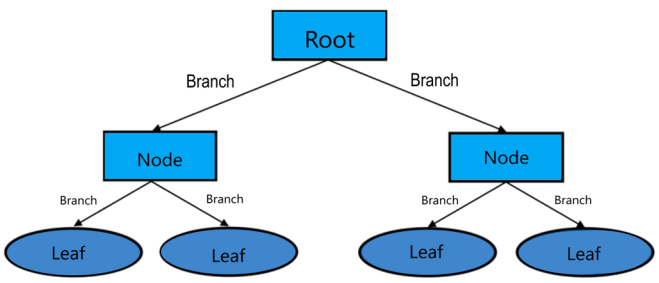
When possessing the above data, we may commence the work on the induction of decision trees from which the rules are created later on. Quinlan algorithm, which is very helpful in this procedure, is employed when we have the data on the grounds of which it is possible to generate the rules. Construction of the decision trees based upon the data may lead to extremely developed trees and, consequently, complicated, developed rules, giving no effects in the concluding process; when basing on the data, we cannot say that the presence of a certain trait or its absence implies the immediate conclusion. Owing to the application of software based upon the Quinlan algorithm, it is possible to specify the sequence of attributes, chosen when building the decision tree [8].

2. The structure of decision tree

To have a correct induction of decision tree, it is indispensable to get familiarized with the structure of decision tree (Fig. 5). The formal definition of the decision free is as follows: the decision tree is a tree-like structure, each node of which corresponds to conducting of a certain test of value of one attribute and each leaf contains decision on classification of the example. From particular nodes as many branches are led out as many possible results of the test, corresponding to the mentioned nodes, are obtained. Each of the branches leads to sub-tree (node) serving for classification of the discussed objects for which the mentioned test has a defined result.

The decision trees are constructed by the recurrent division of the training data into subsets, based on the fact how the mentioned division affects the states of the output variable [9].

Fig. 5. The structure of the decision tree shown in a form of graphical presentation with nomenclature of the particular componential elements of the tree, own development based on [10]



The nodes of the decision trees, as described above, represent the attributes of the considered problem. The edges of the tree are connected with the finite sets of the values of attributes. The decision tree should be constructed with the commencement from the most significant attribute. Then, the remaining attributes on the lower levels of the tree should be utilized. The choice of the attribute is also based on its abilities

of classifying. The measure of the significance of the attribute has a nature of entropy. After building the tree, each new object may be classified by passing through the tree down – from the root (top) to leaf (final node) [11].

3. Utilization of the software dedicated to the induction of decision trees

In the present part of the study, practical utilization of software DeTreex 4.0, dedicated to the induction of the decision trees, entering the composition of the artificial intelligence package Sphinx 4.0, has been presented. The application of De Treex is a tool serving, first of all, for the support of the knowledge acquisition process. Owing to the employed induction method of machine learning, it is possible to construct the decision trees and recording the mentioned trees in a form of rules; it is worth pointing out that the rules are the most frequently applied method of knowledge representation in the knowledge bases of expert systems.

DeTree 4.0 is independent in respect of domains and may be utilized not only in the construction of decision trees in the production processes but also in other domains, inter alia, in economy or medicine. The discussed application may be used where there is a problem of decision-making, classification of the objects, quick verification of the acquired rules or quick acquisition of the decision rules from the set of teaching examples.

In order to make the induction of decision tree with the software utilization possible, we have based on the teaching example which refers to the process of glass production; the output variable in the described example was a type of the produced glass while the input variables included: refraction coefficient and the content of the particular elements in the produced glass: sodium, magnesium, aluminium, silicone, potassium, calcium, barium and iron.

Based upon the teaching set consisting of 214 examples, the domain was generated and the attributes were presented below together with their values given in a form of intervals (Fig. 6).

Then, using the option of decision tree generation, the minimum number of examples, creating the leaf of the tree was adopted as 5 and the option of cutting the tree by 25% (with the program assumption that such possibility is available) was chosen. After setting and confirming the mentioned above parameters, the tree was generated. The generated tree is legible (readable) and the effect of the work, in a form of graphical and text tree has been presented below in Fig. 7 and 8, respectively.

Fig. 6. Domain of the problem presented in a form of screenshot from software DeTreex 4.0

Dziedzina problemu

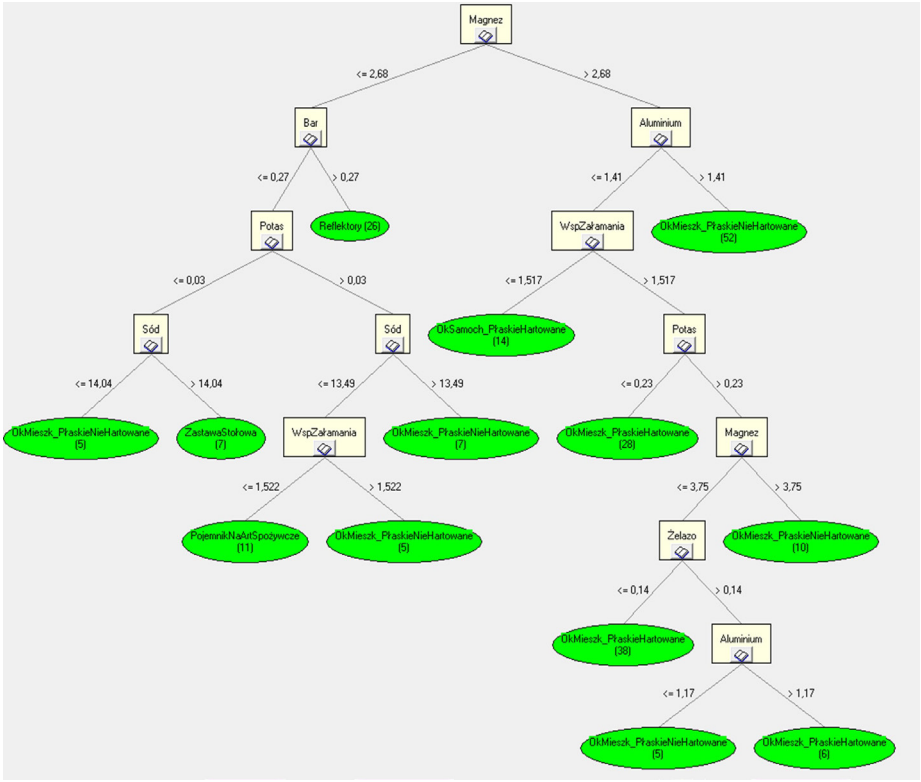
WspZałamania	Sód	Magnez	Aluminium	Krzem	Potas	Wapń	Bar	Żelazo	RodzajSzkła
min. 1.51115	min. 10.73	min. 0.00	min. 0.29	min. 69.81	min. 0.00	min. 5.43	min. 0.00	min. 0.00	OkMieszk_PłaskieHartowane
max. 1.53393	max. 17.38	max. 4.49	max. 3.50	max. 75.41	max. 6.21	max. 16.19	max. 3.15	max. 0.51	OkMieszk_PłaskieNieHartowane
									OkSamoch_PłaskieHartowane
									PojemnikNaArtSpożywcze
									ZastawaStołowa
									Reflektory

Fig. 7. The decision tree, generated with the application of software Detreex 4.0, as presented in a text form

RodzajSzkła - drzewo decyzyjne

```
RodzajSzkła
  Magnez <= 2.68 :
    Bar > 0.27 : Reflektory (26.0)
    Bar <= 0.27 :
      Potas <= 0.03 :
        Sód <= 14.04 : OkMieszk_PłaskieNieHartowane (5.0)
        Sód > 14.04 : ZastawaStołowa (7.0)
      Potas > 0.03 :
        Sód > 13.49 : OkMieszk_PłaskieNieHartowane (7.0)
        Sód <= 13.49 :
          WspZałamania <= 1.52152 : PojemnikNaArtSpożywcze (11.0)
          WspZałamania > 1.52152 : OkMieszk_PłaskieNieHartowane (5.0)
  Magnez > 2.68 :
    Aluminium > 1.41 : OkMieszk_PłaskieNieHartowane (52.0)
    Aluminium <= 1.41 :
      WspZałamania <= 1.51707 : OkSamoch_PłaskieHartowane (14.0)
      WspZałamania > 1.51707 :
        Potas <= 0.23 : OkMieszk_PłaskieHartowane (28.0)
        Potas > 0.23 :
          Magnez > 3.75 : OkMieszk_PłaskieNieHartowane (10.0)
          Magnez <= 3.75 :
            Żelazo <= 0.14 : OkMieszk_PłaskieHartowane (38.0)
            Żelazo > 0.14 :
              Aluminium <= 1.17 : OkMieszk_PłaskieNieHartowane (5.0)
              Aluminium > 1.17 : OkMieszk_PłaskieHartowane (6.0)
```

Fig. 8. The decision tree, generated with the application of software Detreex 4.0, as presented in a graphical form



4. Conclusions

In the paper, the process of knowledge acquisition was described and the methods of acquiring the knowledge with the consideration of the method of induction of decision trees were presented. The mentioned method was also presented on the practical example with the utilization of the software dedicated to the mentioned procedures. The software Sphinx 4.0 allows generating the decision tree in relatively simple way and, also, based upon the readable interface of use, obtaining the searched knowledge in a form of the rules, on the grounds of the teaching set. The acquisition of the knowledge using the decision trees is extremely important for the decision-making problems in the production enterprises which have the heuristic nature of the solutions. In the case of such problems, the application of expert systems in the selected areas should be purposeful.

The basis for the development of the method for construction of knowledge base for production processes, especially for the field of technological production includes a symbolic and object representation of data and their mutual relationships which may be recorded in a form of the rules, with the application of the appropriate methods of knowledge acquisition and the correctly conducted induction of the decision trees.

It should be also mentioned that the correctly developed schemes of processing the technological knowledge, beginning from its acquisition, via the presentation and ending at construction of the knowledge base, facilitate the preparation and recording of the indispensable knowledge in expert system what has a direct effect on optimization of the costs and optimization of the production processes in the enterprise.

References

[1] Knosala R. i zespół: Zastosowania metod sztucznej inteligencji w inżynierii produkcji. Wydawnictwo Naukowo-Techniczne, Warszawa 2002.
[2] Cholewa W., Pedrycz W.: Systemy doradcze, skrypt uczelniany Politechniki Śląskiej nr 1447, Gliwice 1987.
[3] Michie D.: Current Developments in Experts Systems. Applications of Expert Systems, Turing Institute Press in association with Addison-Wesley Publishing Company, 1987.
[4] Michalski R. S.: A Theory and Methodology of Inductive Learning, Artificial Intelligence 20, pp. 111-161, 1983.
[5] Quinlan J. R.: C4.5 Program for Machine Learning, Morgan Kaufmann, San Mateo, CA, 1993.
[6] Ohm A., Komorowski J., Skowron A., Synak P.: The Design and Implementation of a Knowledge Discovery Toolkit Based on Rough Sets – The Rosseta System. Rough Sets in Knowledge Discovery, Physica Verlag, 1998.
[7] Jagielski J.: Inżynieria wiedzy. Oficyna Wydawnicza Uniwersytetu Zielonogórskiego, Zielona Góra 2005.
[8] Kwiatkowska A. M.: Systemy wspomagania decyzji. Jak korzystać z wiedzy i informacji w praktyce. Wydawnictwo Naukowe PWN SA, Warszawa 2007.
[9] Szeliga M.: Data science I uczenie maszynowe. Wydawnictwo Naukowe PWN SA, Warszawa 2017.
[10] Popova O., Popov B., Karandey V., Gerashchenko A.: Entropy and Algorithm of Obtaining Decision Trees in a Way Approximated to the Natural Intelligence. International Journal of Cognitive Informatics and Natural Intelligence (IJCINI) 13(3), 2019.
[11] Pedrycz W., Sosnowski Z. A.: Drzewa decyzyjne z rozmytą granulacją wiedzy. Inżynieria Wiedzy i Systemy Ekspertowe. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2000.

Article reviewed
Received: 15.01.2021 r./Accepted: 13.02.2021 r.