Acta Agraria Kaposváriensis (2007) Vol 11 No 2, 175-186
Kaposvári Egyetem, Állattudományi Kar, Kaposvár
University of Kaposvár, Faculty of Animal Science, Kaposvár

Applying decision trees to investigate the operating regimes of a production process

T. Varga, J. Abonyi, F. Szeifert
University of Pannonia, Faculty of Engineering, Department of Process Engineering, H-8200 Veszprém, Egyetem út 10.

ABSTRACT

Nowadays beside the improvement of the overall process performance, the maintenance of the safe operation conditions is the key element in the development of process control systems. To improve the quality of products, to reduce energy and materials waste, and to increase the flexibility of production the process operators require more insight in the behavior of the process. While the optimal operating conditions of production processes are getting closer and closer to the physical constrains, more and more important is the development of knowledge based expert systems for supporting the operators to keep the operation conditions in this narrow range. Next to this requirement an expert system has to be able to detect failures, discover the sources of failures and forecast the false operations to prevent from the development of production breakdowns. The aim of this work is to propose a novel approach based on process models and decision tree induction technique to discover and isolate the operating regimes of dynamic processes. The novelty of this approach is the application of a classical machine learning tool (decision tree induction) for the extraction of the hidden knowledge of process models into easily interpretable rule base that describes the operation regions of the process. In order to emphasize applicability of decision trees in extracting the relevant information from the model of a technology and how the rules represent operating regimes a detailed case study was performed based on a sophisticated model of an industrial heterocatalytic reactor.
(Keywords: decision tree, operator support system, heterocatalytic reactor)

ÖSSZEFOGLALÁS

Döntési fák alkalmazási lehetőségei technológiai rendszerek működtetési tartományainak diagnosztikai célú leírására
Varga T., Abonyi J., Szeifert F.
Pannon Egyetem, Mérnöki Kar, Folyamatmérnöki Intézeti Tanszék, 8200 Veszprém, Egyetem út 10.

Az ipari gyakorlatban használt folyamatirányító rendszerek fejlesztésében törekednek arra, hogy a lehető legjobban kihasználják a technológiai folyamat nyújtotta lehetőségeket. Az optimális üzemeltetési körülmények már a fizikai és kémiai törvények által meghatározott korlátokat közelítik, így egyre fontosabbá válik olyan szakértői rendszerek kidolgozása, amelyek segítik az operátorokat a termelő technológiák eme szűk üzemeltetési tartományban történő működtetésében, a rendszerben előforduló hibák detektálásában, a hibaforrások feltárásában, illetve előrejelzésében, megelőzve esetleg azok kialakulását. A tanulmány célja egy olyan új megközelítésmód bemutatása, amely alkalmas a technológia modelljén alapulva, döntési fák alkalmazásával az üzemeltetés *szempontjából lényeges hibaforrások kialakulási okainak feltárására, izolálására. Az így megszerzett szabály alapú ismeretek rendszerezésével egy az operátor döntését segítő szakértői rendszer kidolgozására nyílik lehetőség. A javasolt megközelítésmódot egy heterokatalitikus reaktor üzemeltetését támogatandó információk kezelésével kapcsolatos vizsgálatok sorozatán keresztül mutatjuk be, rávilágítva arra, hogy a gépi tanulás eszközei, pl. döntési fák, miként alkalmazhatók a különböző szimulátorokból kinyerhető információk összesítésében, és könnyen értelmezhető, illetve kezelhető formában történő reprezentálásában.*

(Kulcsszavak: döntési fa, szakértői rendszer, heterokatalitikus reaktor)

INTRODUCTION

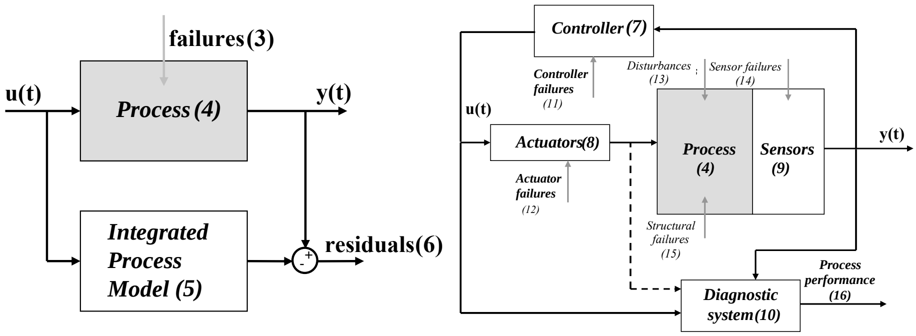
Nowadays beside the improvement of the overall process performance, the maintenance of the safe operation conditions is the key element in the development of process control systems. In the process industries the increasingly complex technologies effect considerable challenges in their design, analysis, manufacturing and management for successful operation. For many applications, it is necessary that maintain the process variables within strict limits. Due to complexity of production systems the fault prevention, diagnoses and the control of abnormal events became more and more complicated for the process operators. Many thousands of process variables are observed and stored every second in a large process plant. To support the operators in improving the quality of products, reducing the energy and materials waste, and increasing the flexibility of production it is necessary to increase their insight in the behavior of the process. Due to this demand the number of claims against for developing diagnostic systems is continuously increasing.

On-line diagnostic and fault detection are based on comparison between measured process data and the values of state variables calculated by the integrated process model. The scheme of such on-line diagnostic can be seen in *Figure 1a* (*Venkatasubramanian*, 2005; *Németh et al.*, 2004). Usually the comparison is a simple subtraction of the measured and simulated variables. These residuals are called symptoms, ie. reactor temperature is higher or the particle size distribution changes in wider limits than expected. The first example is important from the view point of safety aspect, while the second one is necessary to maintain the product quality. The obtained symptoms can be served as the inputs of a knowledge-based diagnostic method which is based on the heuristic information of the observed symptoms and process system. In the case of lack of heuristic knowledge about the interactions between the faults and symptoms statistical and data mining tools can be applied for fault diagnostic. Otherwise, the interactions can be described by if-then rules and a deduction method is applied for diagnostic. All the failures and malfunctions in a controlled system can be distributed in three classes (*Figure 1b*) (*Venkatasubramanian*, 2005):

- *gross parameter changes*: the change in parameter value occurs when one or more exogenous variables are varied due to the disturbance entering the process from the environment, e.g. catalyst ageing.
- *structural changes*: occurrence due to the hard malfunctions, e.g. controller malfunction.
- *sensors and actuators malfunctioning*: actuators and sensors cause gross errors due to a fixed failure or an out-of-range failure.

Figure 1

The scheme of on-line fault detection and diagnosis



a) Principle of fault diagnosis (1) *b) The failure possibilities (2)*

1. ábra: On-line hibajelzés és diagnózis

A hibadiagnosztika alapstruktúrája(1), Szabályozókörben előforduló meghibásodások(2), Hibák(3), Technológiai folyamat(4), Integrált folyamat modell(5), Eltérés/szimptóma(6), Szabályozó(7), Beavatkozó(8), Érzékelő(9), Diagnosztikai rendszer(10), Szabályozási hibák(11), Beavatkozó hibái(12), Zavarok(13), Érzékelő hibák(14), Strukturális meghibásodások(15), A folyamat teljesítőképessége(16)

Every process model or controller, even models applied in on-line diagnostic systems have a limited range of validity. This range may be restricted by several factors such as validity of linearization, modeling assumptions or stability properties. A model that is not able to describe the state of system at every operating condition is called local model and it is not valid in each operating regime, as opposed to a global model that is valid in the full range of operation (*Figure 2*.). Of course, the final goal is the creation of a global model in every modeling task. The fundamental consideration is the application of unique local model or controller in each unique operating regime (*Johansen at al.*, 1997; *Rodriguez et al*., 2003). These local models are then combined in such a way to yield a global model. Hence, the model development within this framework typically consists of the following steps:

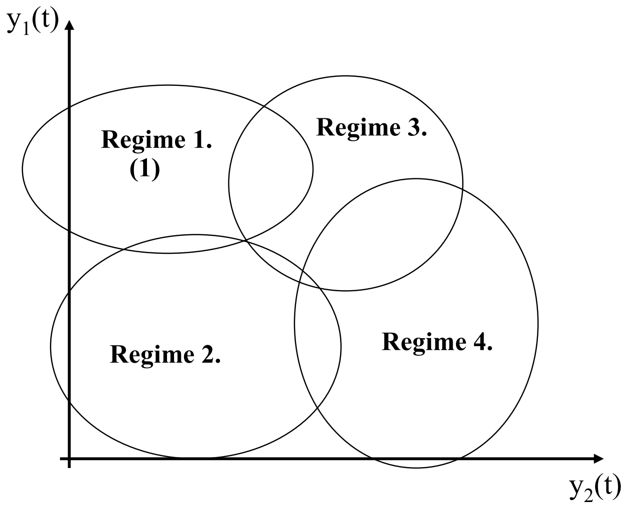
- decompose full operation range of the system into operating regimes;
- select local model/controller structures within each operating regime;
- the local model or controller structures are usually parameterized by certain variables that must be determined, e.g. using nonlinear parameter identification.

A lot of possible methods exist to determine the limits of each operating regime based on experience of the operators or applying different kind of machine learning algorithms (*Gugaliya et al.*, 2005). Learning from examples, i.e. concepts acquisition, is one of the most important branches of machine learning that has been generally regarded as the bottle-neck of expert system development. For this purpose a wide range of models and identification algorithms have been developed.

Through this paper one among the wider range of possible approaches the binary decision trees are applied to create rule-based of the classifier. Decision trees are widely used in pattern recognition, machine learning and data mining applications due to the interpretable representation of the detected information.

Figure 2

Decomposition of the global operating regime



2. ábra: A teljes üzemelési tartomány felosztása

Tartomány(1)

A binary decision tree consists of two types of nodes: (i) internal nodes having two children, and (ii) terminal nodes without children. Each internal node is associated with a decision function to indicate which node to visit next. Each terminal node represents the output for a given input that leads to this node, i.e. in classification problems each terminal node contains the label of the predicted class (*Abonyi*, 2005; *Abonyi et al.*, 2003; *Han*, 2000). The resulting knowledge, in the form of decision trees is easily comprehensible. This is attractive for a wide range of users who are interested in domain understanding, classification capabilities, or the symbolic rules that may be extracted from the tree and subsequently used in a rule-based decision system. An illustrative example for a decision tree is given in *Figure 3*. As this figure illustrates such a model is easily interpretable, so it can be easily integrated into an operator support system.

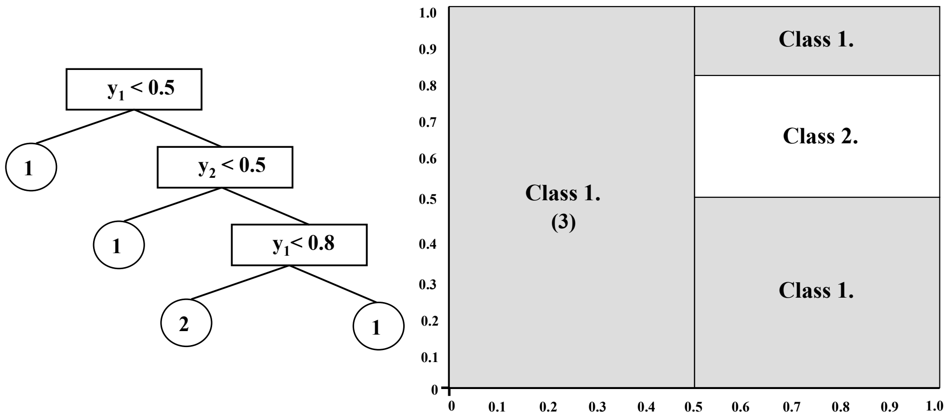
This work shows the applicability of decision trees for mining information from a one-dimensional steady-state model of an industrial heterocatalytic reactor. Two phenomena are investigated closely:

- reactor runaway: it means a sudden and considerable change in process variables, e.g. a highly exothermic reaction takes place in any kind of reactor, in case the generated heat can't be removed, the reaction rate increases due to an increase in temperature, causing a further increase in temperature and hence a further increase in the reaction rate.
- catalyst ageing: due to this phenomenon the catalyst activity is decreasing during the production. It is obvious that a lower catalyst activity results lower process potential.

First we introduce a one dimensional steady-state model of the reactor followed by applied algorithm for decision tree generating. Finally by solving the two above mentioned problems the way of generating learning samples and using the decision trees to detect and follow these phenomena will be shown.

Figure 3

Example of a binary decision tree



a) Binary decision tree (1) *b) The decomposed features space (2)*

3. ábra: Bináris döntési fa példa

Bináris döntési fa(1), Felosztott változó mező(2), Osztály(3)

THE STUDIED REACTOR AND ITS MODEL

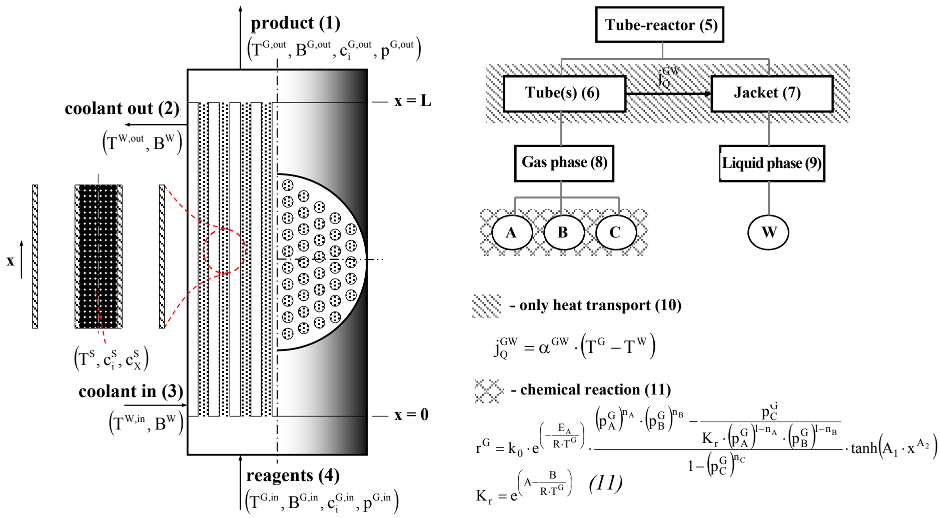
The studied vertically arranged reactor contains a great number of tubes with catalyst as shown in *Figure 4a*. The second order reaction occurs as the reactants moving upwards through the fixed bed of catalyst particles and the heat generated by the reaction removed by the cooling water. Due to the exothermic reaction it has a great chance for development of hot spots somewhere in reactor which increase the rate of catalyst ageing. The selection of operation conditions is important to avoid the development of reactor runaway and to increase the lifetime of catalyst at same time. It is obvious that the two investigated phenomena is in close relationship, the runaway decreases the lifetime of catalyst while as the activity of catalyst is decreasing the reaction rate is being changed in the same way.

Before the introduction of details of the reactor model, the modeling assumptions are summarized and the resulted structure are presented (*Figure 4b*):

- the reaction takes place in the gas phase;
- to calculate the rate of reaction the Langmuir-Hinselwood kinetic is modified by a tanh-term (as you can see in *Figure 4b*), this term makes possible to simulate how the catalyst activity profile modifies the reaction rate in each spot of reactor;.
- the temperature of the gas and solid phase are equal;
- to calculate the pressure drop in the reactor a modified Ergun-equation is used.

Figure 4

The scheme of the industrial heterocatalytic reactor (a) and the structure of steady-state model (b)



a) *b)*

4. ábra: Az ipari reaktor sémája(a) és a kidolgozott modell struktúrája(b)

Termék(1), Hűtőközeg ki(2), Hűtőközeg be(3), Reagensek(4), Csőreaktor(5), Cső(vek)(7), Köpenytér(8), Gázfázis(8), Folyadék fázis(9), Csak hőtranszport(10), Kémiai reakció(11)

The mean of model parameters and variables are summarized in the notation. Based on the assumptions just shown the balance of each component flux is the following:

$$\frac{d\left(B^G \cdot c_i^G\right)}{dx} = \upsilon_i \cdot V^S \cdot r^G \,, \qquad (1)$$

where i = {A; B; C}. The temperature of reactor and jacket are calculated by the next equations:

$$B^G \cdot \rho^G \cdot c_p^G \cdot \frac{dT^G}{dx} = V^S \cdot r^G \cdot \left(-\Delta H_r\right) - A^{GW} \cdot j_Q^{GW} \qquad (2)$$

$$B^W \cdot \rho^W \cdot c_p^W \cdot \frac{dT^W}{dx} = A^{GW} \cdot j_Q^{GW} \,. \qquad (3)$$

As it was assumed in the assumptions the pressure drop on catalyst bed is calculated by a modified Ergun-equation:

$$\frac{dp^G}{dx} = -2 \cdot f_c \cdot \frac{\rho^G \cdot \left(B^G\right)^2}{d_p \cdot (A)^2} \cdot \frac{1-\varepsilon}{\varepsilon^3} \cdot \left(1.75 + 150 \cdot \frac{1-\varepsilon}{Re}\right) \,. \qquad (4)$$

The model given by the equation (1-4) was implemented in MATLAB. The developed process simulator was used to generate learning samples for the induction of the decision trees trained to extract information related to the two earlier mentioned problems.

RESULTS AND DISCUSSION

Determining the instability regime

To separate the operating regimes of the reactor and avoid the development of reactor runaway analytical and data mining techniques are introduced. First let's consider the analytical solution. The very first step is the calculation of equilibrium reactor temperature ($T^{G,eq}$) in the full scale of conversion. Another possibility to calculate the reaction equilibra making use of the partial pressures of each component:

$$K = \exp\left(A - \frac{B}{R \cdot T^G}\right) = \frac{p_C^G}{p_A^G \cdot p_B^G} \quad , \tag{5}$$

then

$$T^{G,eq} = -\frac{\ln\left(\frac{p_C^G}{p_A^G \cdot p_B^G}\right) - A}{B} \cdot R \quad . \tag{6}$$

The optimal reactor temperature ($T^{G,opt}$) can be easily determined based on the following equation:

$$\frac{d\, r\left(T^G, \text{conversion}\right)}{d\, T^G} = 0 \quad . \tag{7}$$

To investigate the model stability Ljapunov indirect stability analysis method was applied, which is useful in reactor stability analysis (*Varga et al.*, 2006). The results of calculations in case of this reactor are shown in *Figure 5*.

Decision tree for determining the regime of instability in the reactor

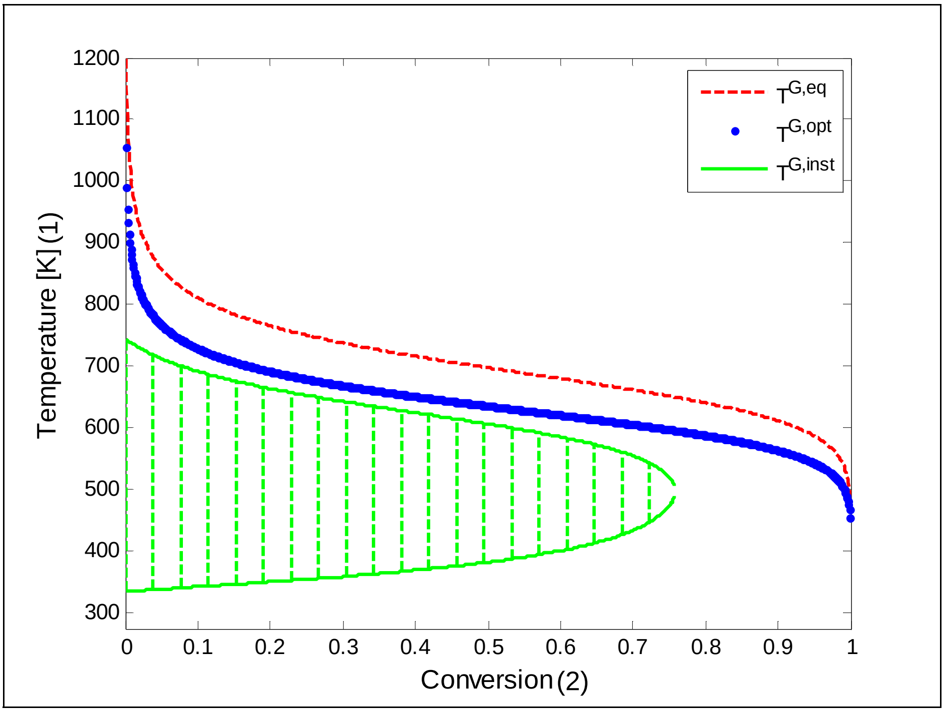
Sometimes it is too complicated to analytically determine the boundaries of stability. In such case instead of any analytical calculations a data mining method can be used to hedge in the instability. The main purpose of this paper is to show how the decision tree technique can be applied in the solution of this problem. The results of the Ljapunov indirect stability analysis performed in case of a great amount of randomly generated inlet conditions are proper for gathering learning samples to obtain a decision tree that is suitable for determining the boundaries of instability. In our previous work (*Varga et al.*, 2006) it has been presented that the cooling fluid inlet temperature and the reagents inlet pressure have the main effect on development of runaway in this reactor. That is why in this case only two from the five inlet parameters are used to generate the decision tree and detect the development of reactor runaway. The learning samples are plotted in *Figure 6a* where the stars present the inlet conditions when runaway occurs, and the circles when doesn't. It is easy to draw a line separating the two regimes but this example is just for presenting the applicability of decision trees in operation regime determination. The generated tree can be seen in *Figure 6b*. The direction of the tree is left to right. In a decision tree the leaves contain the answer of the investigation. In this case 1 means runaway doesn't occur while 2 means the opposite. Based on this tree the instability regime can be hedged in as shown in *Figure 6c*.

Decision trees for monitoring catalyst activity

In every catalytic process the activity of applied catalyst decreases in time but the rate of catalyst ageing is not the same. To generate a decision tree for detecting the catalyst activity in the reactor the necessary learning samples must be collected. In order to collect learning samples two steady-state models, one of them including a tanh-term are used. The calculated outlet values by these models are compared as shown in *Figure 7a.* Varying the catalyst activity and inlet conditions a great amount of learning samples are generated. The generated tree (*Figure 7b*) is able to sign only if the catalyst activity is lower than a fixed value. In case some trees which sign in different values are systemized a decision forest can be obtained as it can be seen in Figure 7c. This decision forest is able to measure the catalyst activity in reactor based on the scheme can be seen in *Figure 1a* where the integrated process model is the presented steady-state model without tanh-term.

Figure 5

The analytic investigation of reactor in phase-space

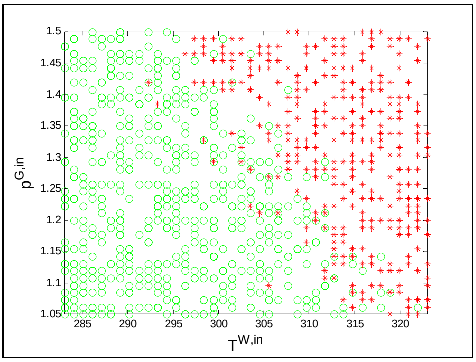


5. ábra: A reaktor analitikus vizsgálata a fázistérben

Hőmérséklet [K](1), Konverzió(2)

Figure 6

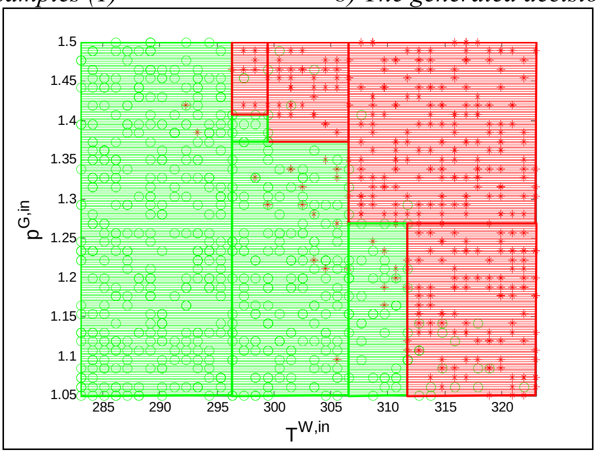
The boundaries of instability



a) The learning samples (1)

$T^{W,in} \le 306.6$:
$\vert\;\; T^{W,in} \le 296.3 : 1\ (384.0/8.2)$
$\vert\;\; T^{W,in} > 296.3$:
$\vert\;\; \vert\;\; p^{G,in} \le 1.37 : 1\ (181.0/26.3)$
$\vert\;\; \vert\;\; p^{G,in} > 1.37$:
$\vert\;\; \vert\;\; \vert\;\; T^{W,in} > 299.4 : 2\ (42.0/7.6)$
$\vert\;\; \vert\;\; \vert\;\; T^{W,in} \le 299.4$:
$\vert\;\; \vert\;\; \vert\;\; \vert\;\; p^{G,in} \le 1.41 : 1\ (5.0/2.3)$
$\vert\;\; \vert\;\; \vert\;\; \vert\;\; p^{G,in} > 1.41 : 2\ (13.0/6.5)$
$T^{W,in} > 306.6$:
$\vert\;\; p^{G,in} > 1.27 : 2\ (192.0/6.7)$
$\vert\;\; p^{G,in} \le 1.27$:
$\vert\;\; \vert\;\; T^{W,in} \le 311.7 : 1\ (50.0/19.2)$
$\vert\;\; \vert\;\; T^{W,in} > 311.7 : 2\ (133.0/23.8)$

b) The generated decision tree (2)



c) The operating regimes (3)

6. ábra: A reaktor instabil tartományának meghatározása döntési fa technikával

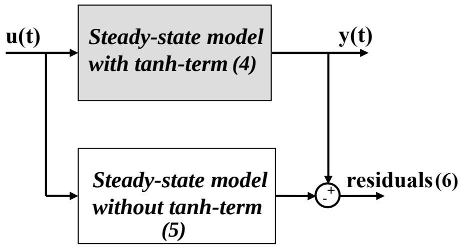
A tanulási minták(1), A generált döntési fa(2), A működési tartományok(3)

CONCLUSIONS

This work demonstrated how decision trees can be used to determine the operating regimes of complex processes, e.g. stability-instability regions of a heterocatalytic reactor. The results show that the proposed decision tree based approach based on a set of linguistic rules extracted from data obtained by the analysis of the steady-state model of the process is able to distinguish between runaway and non-runaway situations. It is shown that decision tree technique is able to extract information from steady-state models to simulate a dynamic phenomenon. The generated decision forest can be applied for the monitoring of catalyst activity in the reactor and the regeneration time can be scheduled by using the proposed tool.

Figure 7

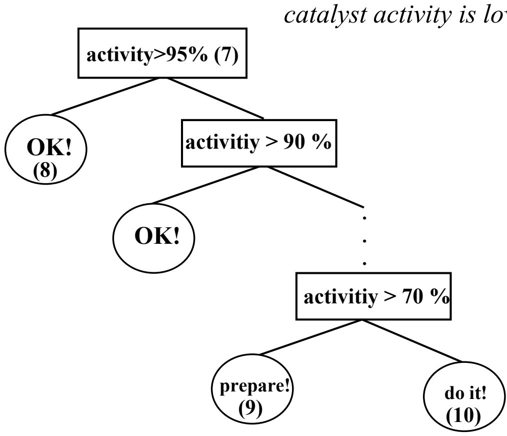
The detection of catalyst activity using decision trees



a) Generating the learning samples (1)

$\hat{p}^{G,out} - p^{G,out}$ > -0.00464 : 1 (236.0/28.8)
$\hat{p}^{G,out} - p^{G,out}$ <= -0.00464 :
| $\hat{T}^{G,out} - T^{G,out}$ > 3.254 : 2 (448.0/3.0)
| $\hat{T}^{G,out} - T^{G,out}$ <= 3.254 :
| | $\hat{p}^{G,out} - p^{G,out}$ <= -0.0119 : 2 (142.0/12.0)
| | $\hat{p}^{G,out} - p^{G,out}$ > -0.0119 :
| | | $\hat{n}_A^{G,out} - n_A^{G,out}$ <= 0.0307 :
| | | | $\hat{n}_C^{G,out} - n_C^{G,out}$ > -0.0146 : 2 (125.0/44.3)
| | | | $\hat{n}_C^{G,out} - n_C^{G,out}$ <= -0.0146 :
| | | | | $\hat{T}^{G,out} - T^{G,out}$ <= 2.737 : 1 (5.0/2.3)
| | | | | $\hat{T}^{G,out} - T^{G,out}$ > 2.737 : 2 (4.0/2.1)
| | | $\hat{n}_A^{G,out} - n_A^{G,out}$ > 0.0307 :
| | | | $\hat{n}_C^{G,out} - n_C^{G,out}$ <= -0.379 : 2 (5.0/2.3)
| | | | $\hat{n}_C^{G,out} - n_C^{G,out}$ > -0.379 : 1 (35.0/2.9)

b) Decision tree to detect when the catalyst activity is lower than 70% (2)



c) The worked out decision forest (3)

7. ábra: A katalizátor aktivitásának meghatározása döntési fa technikával

A tanulási minták generálása(1), Döntési fa a 70% alatti katalizátor aktivitás jelzésére(2), A döntési fákból összeállított döntési erdő(3), Stacionárius modell a tanh-taggal(4), Stacionárius modell a tanh-tag nélkül(5), Eltérés/szimptóma(6), A katalizátor aktivitása(7), Nincs szükség a katalizátorágy regenerálására(8), Készülj a katalizátor cserére(9), A katalizátorágy regenerálására van szükség(10)

ACKNOWLODGEMENTS

The authors would like to acknowledge the support of the Cooperative Research Centre (VIKKK) (project III/2) and Hungarian Research Found (OTKA T049534).

János Abonyi is grateful for the support of the Bolyai Research Fellowship of the Hungarian Academy of Sciences and the Öveges Fellowship.

NOTATION

Notation	Meaning	Unit
α^{GW}	The heat transfer coefficient between the catalyst bed and the jacket	$\frac{W}{m^2 \cdot K}$
A^{GW}	The contact are between the catalyst bed and the jacket	m^2
A	The cross-section area of the catalyst bed	m^2
B^G	Volume velocity	$\frac{m^3}{s}$
$c_p^G ; c_p^W$	The heat capacity of the gas phase and the jacket	$\frac{J}{kg \cdot K}$
ΔH_r	Heat of the reaction	$\frac{J}{mol}$
d_p	The diameter of the catalyst particle	m
ε	The ratio of the volume of solid phase and the volume of catalyst bed	-
K_r	The reaction equilibrium	-
υ_i	Stoichometric coefficient	-
p^G	Pressure	Pa
p_i^G	Partial pressure of components	Pa
r^G	Rate of the reaction	$\frac{mol}{m^3 \cdot s}$
R	The ideal gas constant	$\frac{J}{mol \cdot K}$
Re	Reynold’s number	-
$\rho^G ; \rho^W$	The density of the gas phase and the jacket	$\frac{kg}{m^3}$
$T^G ; T^W$	The temperature of the gas phase and the jacket	K
V^S	The volume of the solid phase	m^3
x	Reactor length without dimension, $x \in [0, 1]$	-

REFERENCES

Abonyi, J. (2005). Adatbányászat – a hatékonysága eszköze. ComputerBooks kiadó, 12-34.

Abonyi, J., Roubos, J.A., Szeifert, F. (2003). Data-driven generation of compact, accurate, and linguistically sound fuzzy classifier based on decision-tree initialization. International Journal of Approximate Reasoning, 32. 1-21.

Gugliya, J.K., Gudi, R.D., Lakshminarayanan, S. (2005). Multi-model decomposition of nonlinear dynamics using a fuzzy-CART approach. Journal of Process Control, 15. 417-434.

Han, J., Kimber, M. (2000). Data Mining: concepts and techniques. Chapter 7, Morgan Kaufman, 279-334.

Johansen, T.A., Foss, B.A. (1997). Operating regime based process modeling and identification. Computers and Chemical Engineering, 21. 159-176.

Németh, E., Cameron, I.T., Hangos, K.M. (2004). Diagnostic goal driven modelling and simulation of multiscale process systems. Computers and Chemical Engineering, 29. 783-796.
Rodriguez, J.A., Romagnoli, J.A., Goodwin, G.C. (2003). Supervisory multiple regime control. Journal of Process Control, 13. 177-191.
Varga, T., Abonyi, J., Szeifert, F. (2006). Heterokatalitikus reaktorok vizsgálata. Acta Agraria Kaposváriensis, 10. 3. 121-135.
Venkatasubramanian, V. (2005). Prognostic and diagnostic monitoring of complex systems for product lifecycle management: Challenges and opportunities, Computers and Chemical Engineering, 29. 1253-1263.

Levelezési cím *(corresponding author):*

Varga Tamás
Pannon Egyetem, Folyamatmérnöki Intézeti Tanszék
8200 Veszprém, Egyetem út 10.
University of Pannonia, Department of Process Engineering
H-8200 Veszprém, Egyetem út 10.
Tel.: 36-88-624-447, Fax: 36-88-624-171
e-mail: vargat@fmt.uni-pannon.hu