

METODY PRO GENEROVÁNÍ FUZZY PRAVIDEL Z DAT

R. Zitta, M. Palatová

Vysoká škola chemiko-technologická Praha, Ústav počítačové a řídicí techniky

Abstract

This contribution shortly introduces two methods from literature and one own method, that were used for processing of measured data. Three methods for fuzzy rules generation were tested: Fuzzy-ROSA-method based on statistical analysis, ANFIS method (Adaptive Neuro-Fuzzy Inference System) and an own method newly created working in Matlab software and its Fuzzy Toolbox.

1 Úvod

Jedna z možností jak zpracovat naměřená data je vytvořit s jejich pomocí pravidla (např. fuzzy pravidla), která umožní pro další podobný proces předpovídat stav, do kterého se systém dostane.

V příspěvku stručně seznamujeme se dvěma metodami z literatury a jednou vlastní, které jsme použili při zpracovávání naměřených dat. Ta jsme získali během experimentů v laboratoři Ústavu počítačové a řídicí techniky při studiu řízení produkce xylitolu z xylózy.

Generovali jsme fuzzy pravidla typu:

$$\text{If } A \text{ then } Y = B_k,$$

kde antecedent A odpovídá jednomu stavu vstupních proměnných

$$A = (X_1 = A_{1k_1}) \wedge (X_2 = A_{2k_2}) \wedge \dots \wedge (X_n = A_{nk_n})$$

X_i je i -tá vstupní proměnná, A_{iki} je k_i -tá lingvistická hodnota i -té vstupní proměnné, Y je výstupní proměnná, B_k její k -tá lingvistická hodnota.

Jako vstupní proměnné pravidlového systému byly zvoleny některé z veličin: koncentrace rozpuštěného kyslíku, dávkování substrátu, objem báze dodané do systému, respirační kvocient, koncentrace xylózy, změna koncentrace xylózy, koncentrace biomasy, změna koncentrace biomasy. Výstupní veličinou byla koncentrace xylitolu nebo její změna. Vstupní proměnné figurují v antecedentu pravidla a výstupní v konsekventu. Pro každou vstupní a výstupní proměnnou bylo třeba definovat jejich lingvistické hodnoty a jejich funkce příslušnosti.

Testovali jsme metodu Fuzzy-ROSA-Verfahren, založenou na principu statistické analýzy, metodu ANFIS (Adaptive Neuro-Fuzzy Inference System) a vlastní algoritmus vytvořený v prostředí Matlab s využitím fuzzy toolboxu.

V příspěvku jsou uvedeny převážně grafické výsledky použití jednotlivých metod. Vzhledem k značné různosti charakterů jednotlivých dat (experimentů) není přenositelnost jednotlivých souborů pravidel na jiná data příliš uspokojivá.

2 Metody

2.1 WINROSA

Program WINROSA slouží k automatickému generování fuzzy pravidel z databáze naměřených dat. Je založen na metodě Fuzzy-ROSA-Verfahren [3], která představuje jedno z rozšíření metody ROSA-Verfahren [4, 5]. Program je složen z několika komponent. Každá komponenta nabízí více výpočetních metod, díky čemuž je možné přizpůsobit proces generování pravidel požadavkům konkrétní úlohy.

Je to postupně komponenta pro

- definici projektu
- generování relevantních fuzzy pravidel
- redukci počtu pravidel
- analýzu pravidel

- export pravidel

Kompletní soubor pravidel je často v praktických aplikacích příliš obsáhlý. Proto vytváří WINROSA také generalizovaná pravidla, která pokrývají více než jen jeden stav vstupních proměnných. V těchto pravidlech chybí v antecedentu jedna nebo více vstupních proměnných. V mnoha aplikacích se totiž ukázalo, že vysoký počet proměnných v antecedentu nevede ke zlepšení výsledku. Ve WINROSE má uživatel možnost nastavit maximální počet proměnných v antecedentu.

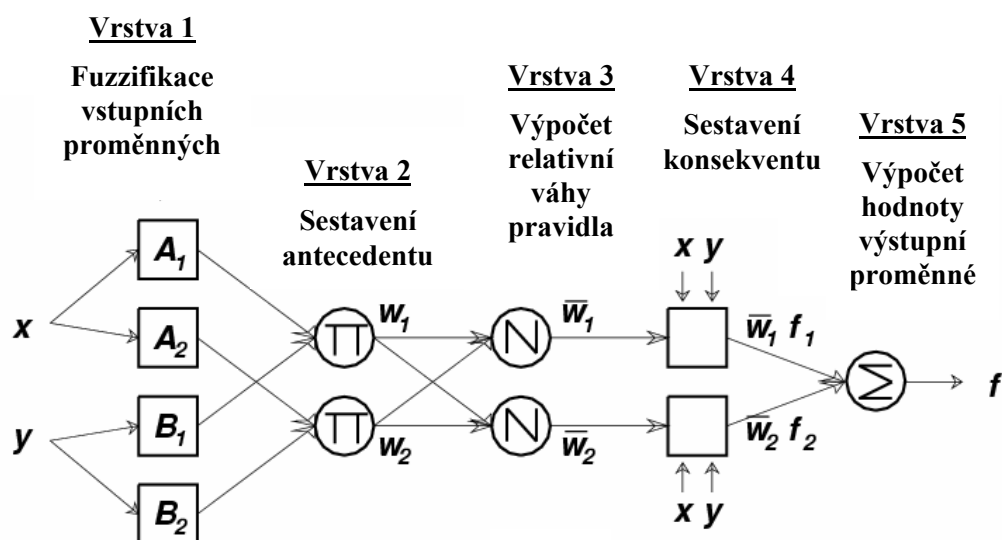
Analýza souboru pravidel se provádí na základě ohodnocení pravidel a slouží k tomu, aby bylo možno si utvořit úsudek o vygenerovaných pravidlech. Současně se tím naskýtá možnost zhodnotit důležitost vstupních proměnných a definovaných lingvistických hodnot.

2.2 ANFIS

ANFIS (Adaptive Neuro-Fuzzy Inference System) [2] je adaptivní dopředná neuronová síť, která je funkčně ekvivalentní fuzzy inferenčnímu systému typu Sugeno (Takagi-Sugeno), to znamená, že generuje pravidla typu:

$$\text{If } (X_1 = A_{1k_1}) \wedge (X_2 = A_{2k_2}) \wedge \dots \text{ then } Y = f(X_1, X_2, \dots),$$

kde X_i jsou vstupní proměnné, A_{iki} , jejich lingvistické hodnoty a Y je výstupní proměnná. Funkce $f(X_1, X_2, \dots)$ v konsekventu pravidla Sugenoova typu může být obecně jakákoliv, ANFIS však umožňuje zvolit pouze mezi konstantní funkcí a lineární funkcí vstupních proměnných. Architektura a funkce jednotlivých vrstev sítě jsou znázorněny na obr. 1, více viz např. [7].



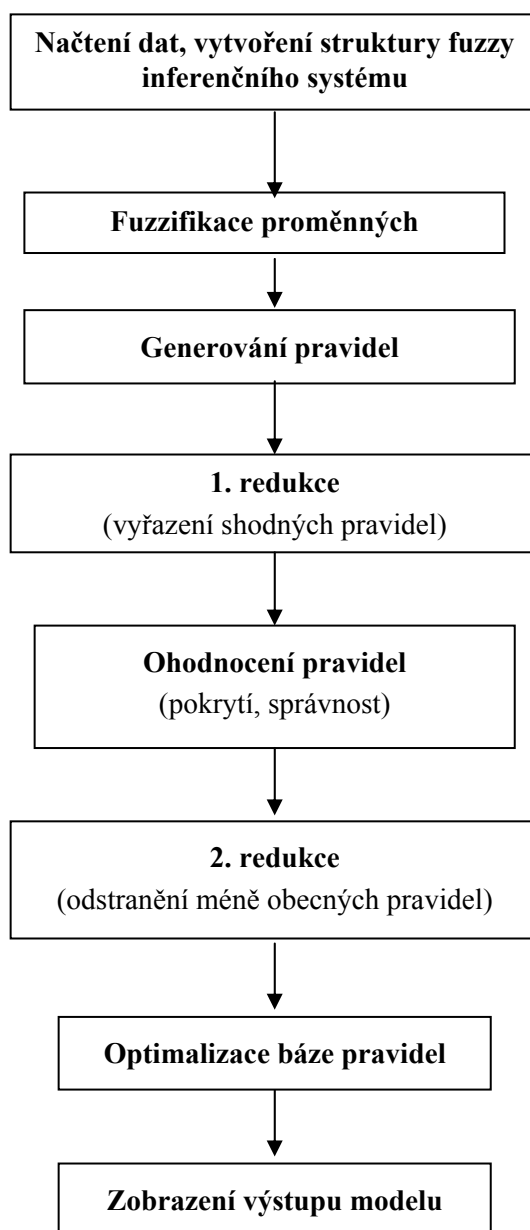
Obr. 1: Architektura sítě ANFIS, čtverce označují adaptivní uzly, kruhy neadaptivní uzly.

Vrstva 1 je tvořena adaptivními uzly, jejichž přenosové funkce odpovídají funkcím příslušnosti daných lingvistických hodnot, tedy každý uzel představuje lingv. hodnotu vstupní proměnné. Vrstva 2 je tvořena neadaptivními uzly, které pouze násobí příchozí signály a na výstupu tak poskytují váhu w pravidel, jejichž antecedent je dán kombinací lingv. hodnot jednotlivých proměnných. Vrstva 3 je tvořena neadaptivními uzly, které na výstupu poskytují poměr váhy jednotlivých pravidel ku součtu vah všech pravidel \bar{w} . Vrstva 4 je tvořena adaptivními uzly, jejichž přenosová funkce je dána požadovaným tvarem konsekventu. Může být tedy konstantní nebo lineární. Vrstvu 5 tvoří jediný neadaptivní uzel, jenž počítá celkový výstup jako součet všech vstupních signálů.

Pro trénování sítě používá ANFIS hybridní učicí metodu. Parametry v antecedentu pravidla (určující funkce příslušnosti vstupních proměnných) jsou optimalizovány gradientní metodou, parametry konsekventu (konstanty nebo koeficienty lineární funkce) jsou počítány metodou nejmenších čtverců.

2.3 Vlastní algoritmus

Algoritmus pro generování fuzzy pravidel je vytvořen v prostředí MATLAB s využitím fuzzy toolboxu. Schéma jednotlivých kroků algoritmu je uvedeno v následujícím schématu:



Při generování pravidel byl vzat v úvahu poznatek z předchozích generování pomocí programu WINROSA, ve kterých bylo zjištěno, že majoritní část báze pravidel tvoří pravidla s maximálně dvěma proměnnými v antecedentu. Generování pravidel probíhá následujícím způsobem:

Pro každý sloupec matice fuzzifikovaných vstupních proměnných, představující jednotlivá měření, se postupně pro každou proměnnou naleznou nenulové hodnoty, určující příslušnost k nějaké lingvistické hodnotě určité proměnné a vytvoří se antecedent s touto hodnotou dané proměnné. K němu se přiřadí konsekvent z matice fuzzifikované výstupní proměnné z odpovídajícího měření. Takto se vytvoří báze pravidel s jednou proměnnou v antecedentu. Z této báze se pak vytvoří pravidla s dvěma proměnnými v antecedentu prostým zkombinováním pravidel, které vznikly ze stejného měření.

Pokud se veličiny mění vzhledem k periodě vzorkování pomalu, obsahuje takto vytvořená báze velký počet shodných pravidel, které se musí z báze odstranit. Následně se provádí ohodnocení zbylých pravidel pro získání informace o míře užitečnosti pravidla. V tomto algoritmu se používá

pokrytí pravidla a správnost pravidla. Míra pokrytí vyjadřuje, jak velkou částí prostoru vstupních stavů je pravidlo aktivováno, tato míra je tedy ovlivněna pouze antecedentem pravidla. Pro *míru správnosti* je použito podmíněné pravděpodobnosti $P(K|A)$, která udává, s jakou pravděpodobností nastane konsekvant K , je-li splněn antecedent A pravidla. Podrobněji viz [7].

Dále je nutné odstranit z báze pravidel ta pravidla, která pokrývají menší prostor vstupních stavů než jiné pravidlo, které má s prvním shodný konsekvant, má vyšší správnost a přitom je jeho antecedent obecnější. Například z následujících pravidel (v závorce je uvedena správnost)

If X je střední a RQ je nízké Then P je střední (0,425)

If X je střední a DO je střední Then P je střední (0,388)

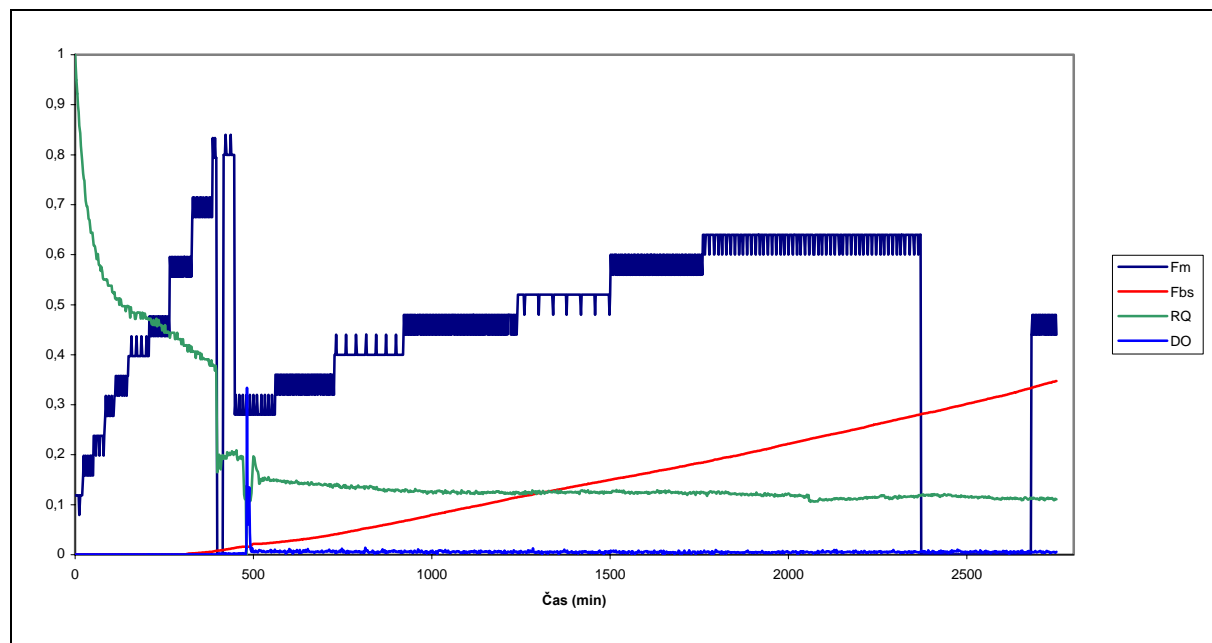
If X je střední Then P je střední (0,496)

jsou první dvě nadbytečná.

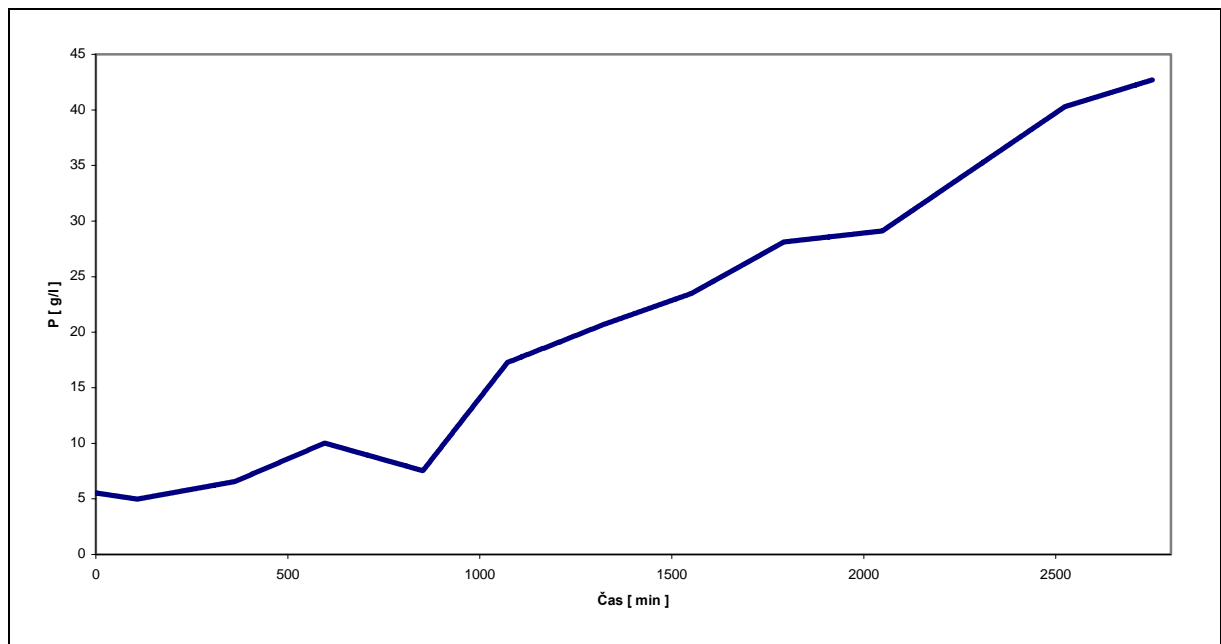
3 Výsledky

Pravidla jsme generovali na základě dat získaných při pokusech, viz [6]. Pro tvorbu modelu byly vybrány jedna až čtyři vstupní veličiny z následující množiny: koncentrace rozpuštěného kyslíku DO , dávkování substrátu Fm , objem báze dodané do systému Fbs , respirační kvocient RQ , koncentrace xylózy S , změna koncentrace xylózy dS , koncentrace biomasy X , změna koncentrace biomasy dX . Výstupní veličinou byla koncentrace xylitolu P nebo její změna dP . Vstupní veličiny byly vzorkovány s periodou 2 minuty, koncentrace xylózy a biomasy s periodou 2 hodiny a jejich hodnoty mezi měřeními aproximovány lineární závislostí. Výstupní veličina byla též vzorkována s periodou 2 hodiny a její hodnoty mezi měřeními aproximovány lineární závislostí.

Data z jednoho pokusu jsou uvedena na obr. 2 a obr. 3.



Obr. 2: Průběh vstupních proměnných (normalizovaných na interval $\langle 0;1 \rangle$) v čase. Fm dávkování substrátu, Fbs objem dodané báze, RQ respirační kvocient, DO koncentrace rozpuštěného kyslíku.

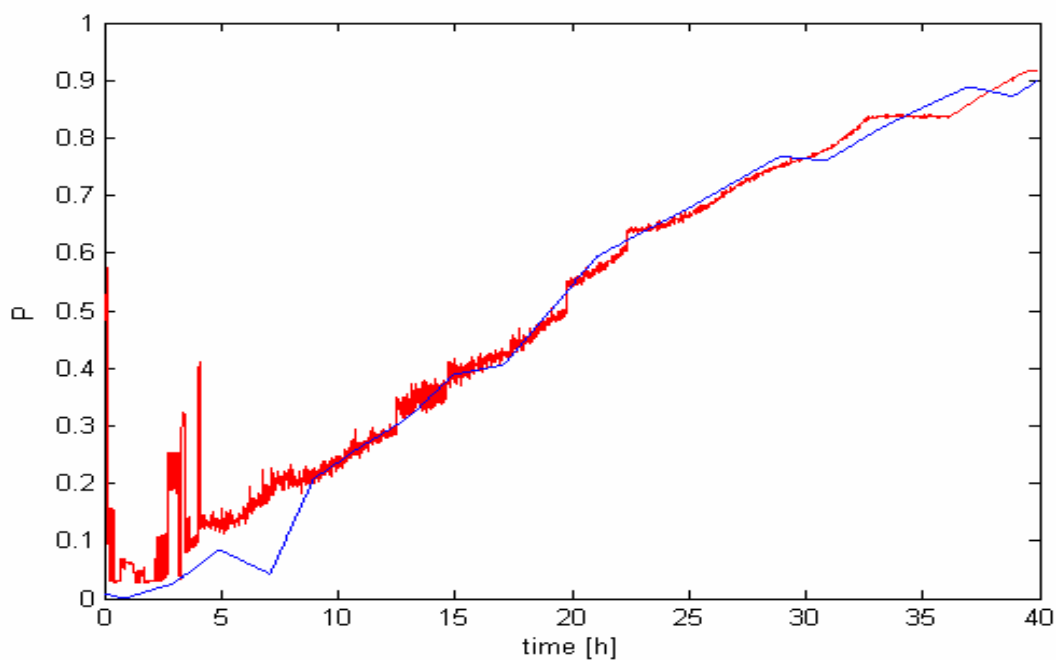


Obr. 3: Průběh výstupní proměnné (koncentrace xylitolu P) v čase.

3.1 Výsledky WINROSA

Pravidla byla generována v programu WINROSA metodou kompletního prohledávání. Antecedent pravidel byl omezen na maximálně tři lingvistické proměnné. Soubor vygenerovaných pravidel byl podroben situation-based redukcí, po které soubor obsahoval 82 pravidel s relevancí od 0,19 do 0,75.

Na obr. 4 je zobrazen průběh koncentrace xylitolu v čase získaný z naměřených dat i z modelu. Z grafů na obr. 2 a obr. 3 je zřejmé, že velké odchylky modelovaných hodnot koncentrace xylitolu od měřených (patrných na počátku procesu), způsobuje vstupní proměnná Fm (dávkování substrátu), jejíž prudká změna během procesu je způsobena přechodem na jinou řídicí strategii.

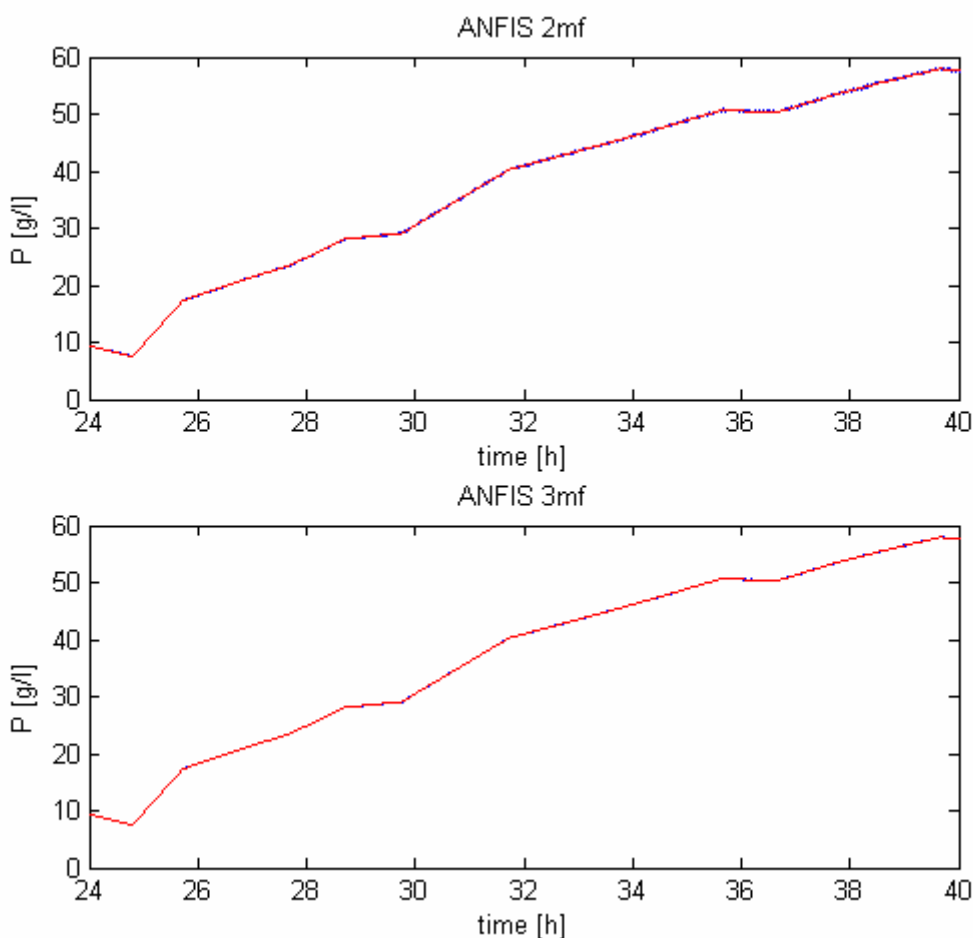


Obr. 4: Průběh koncentrace xylitolu P normalizované na interval $\langle 0,1 \rangle$ v čase. Výstup modelu je znázorněn červenou křivkou, naměřené hodnoty modrou.

Při pokusech o využití stávajícího pravidlového systému na další tři dostupné databáze experimentálních dat byla zjištěna nízká přenositelnost vytvořené báze pravidel. Tento výsledek je způsoben značným rozdílem mezi rozsahy měřených veličin v jednotlivých pokusech. Tento nedostatek nebylo možno odstranit ani zúžením intervalů použitých k modelování procesu.

3.2 Výsledky ANFIS

Model, který měl systém ANFIS vytvořit, byl navrhován za použití stejných veličin jako model vytvořený předchozím algoritmem, tzn. jako vstupní proměnné byly použity veličiny: množství dodané báze Fbs , respirační kvocient RQ , změna koncentrace xylózy dS a změna koncentrace biomasy dX . Výstupní proměnnou byla změna koncentrace xylitolu dP . Volba počtu lingv. hodnot těchto vstupních proměnných byla omezena počtem použitých měření na dvě až tři. Výsledky použití modelů vytvořených systémem ANFIS jsou na obr. 5. Vytvořená pravidla dokáží přesně popsat výstupní veličinu a ani nezáleží zda užijeme dvě nebo tři lingvistické hodnoty pro jednotlivé proměnné.



Obr. 5: Modelování koncentrace xylitolu P systémem ANFIS při nastavení dvou lingv. hodnot pro každou proměnnou (horní obrázek) a pro nastavení tří lingv. hodnot pro každou proměnnou (dolní obrázek). Výstup modelu je znázorněn modrou čarou, naměřené hodnoty červenou (v obou obrázcích se křivky překrývají).

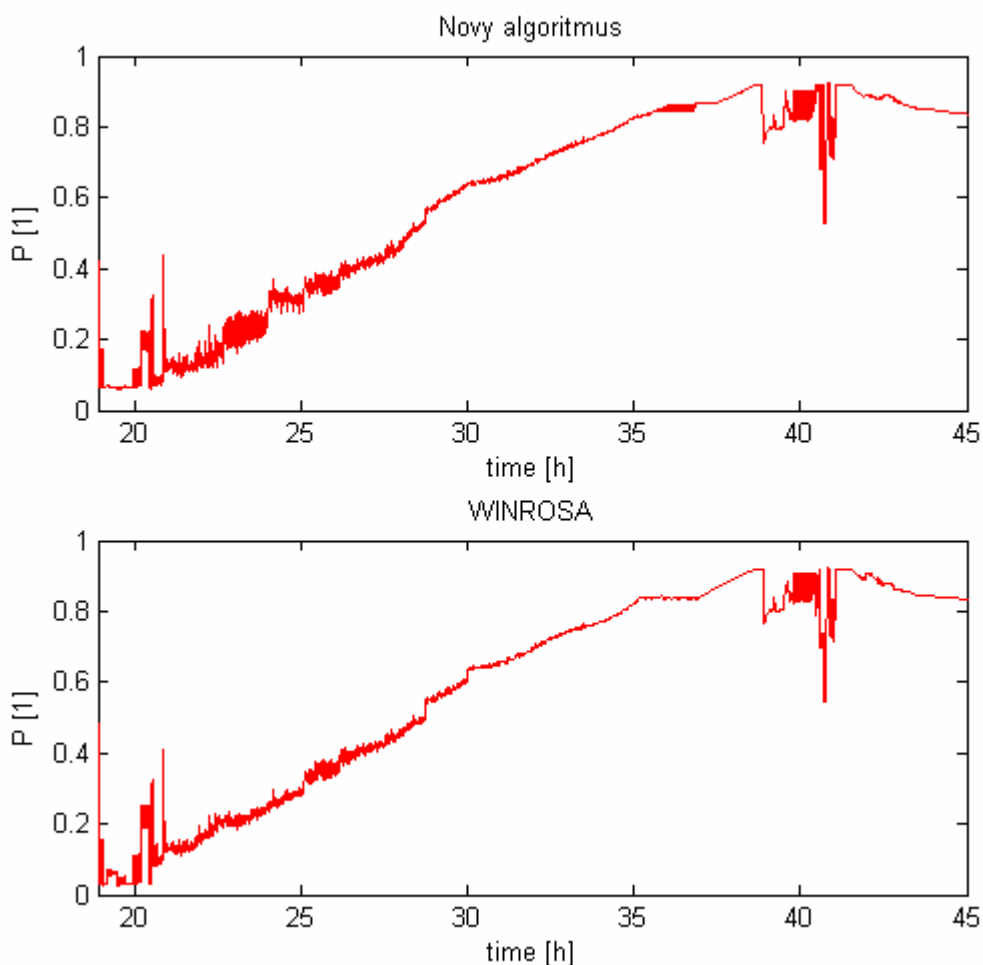
Při pokusech o využití stávajícího pravidlového systému na další tři dostupné databáze experimentálních dat byla zjištěna prakticky nepoužitelnost přenést vytvořené báze pravidel. Důvody jsou podobné jako u Winrosy. Neznamená to, že metody jsou špatné, ale charakter pokusů právě tak jako rozsahy vstupních a výstupních veličin se měnily tak značně, že pravidla nezabírala.

3.3 Výsledky vlastní algoritmus

Algoritmem popsaným v kap. 2.3 byla z každé databáze experimentálních dat generována pravidla modelující bioproces. Počet použitých měření v jednotlivých pokusech se pohyboval mezi dvěma až třemi tisíci, počet vygenerovaných pravidel pro jednotlivé modely se pohyboval v řádech desetitisíců až statisíců, po redukci mezi třemi až čtyřmi sty. Optimalizací byly vytvořeny báze pravidel čítající 30 až 150 pravidel.

Jednoduchost vlastního algoritmu nám dovolila vytvářet model z mnoha různých vstupních veličin, tzv. pohrát si s ním. Výsledky byly srovnatelné s metodou WINROSA.

Pro ověření správné činnosti algoritmu byly jeho výsledky srovnávány s výsledky generování pomocí programu WINROSA. Na obr. 6 je srovnání výstupu modelů popisujících přítokovanou fázi jednoho pokusu. Pro generování pravidel byly v obou případech nastaveny shodné parametry tj. stejný počet a typ funkcí příslušnosti lingv. hodnot.



Obr. 6: Porovnání výstupů modelů (normalizované hodnoty koncentrace xylozolu P) vytvořených programem WINROSA a vlastním algoritmem.

Co se týče použitelnosti (aplikovatelnosti) pravidlového systému na jiná data, je hodnocení zhruba analogické předchozím.

4 Závěr

Pro účely modelování bioproců jsou vhodné umělé neuronové sítě i techniky odvozování pravidel. V této práci byla studována možnost modelování bioproců fuzzy-pravidlovým systémem.

Jako vstupní proměnné pravidlového systému byly zvoleny veličiny, které mají podstatný vliv pro určení metabolického stavu kvasinek (např. koncentrace rozpuštěného kyslíku, respirační

kvocient), dále přítok živin, který má přímý vztah k množství vytvořeného produktu a celkový objem dodané báze, který dobře popisuje průběh růstu kvasinek v čase.

Pro odvozování fuzzy pravidel z databáze experimentálních dat byl nejprve použit program WINROSA. Tento program nevyžaduje pro generování báze pravidel pravidlového systému žádné apriorní znalosti o modelovaném systému.

Z výsledků je zřejmé, že model je schopen správně určovat trendy koncentrace xylytolu, ale modelovaná hodnota koncentrace xylytolu se liší od měřené až o 6% (nepočítáme-li chyby na počátku přítokované fáze zapříčiněné nastavením na maximální živení). Zároveň je patrné kmitání výstupu modelu, které je způsobeno šumem v měřených hodnotách vstupních veličin.

Aby bylo možno vyzkoušet více modelů a pokusit se tak o nalezení pravidlového systému, který by byl přenositelný z jedné databáze na druhou, byl vytvořen v prostředí programu MATLAB algoritmus schopný generovat fuzzy pravidla s výsledky porovnatelnými s výsledky programu WINROSA. Přes množství provedených experimentů s tímto algoritmem se nepodařilo nalézt kombinaci vstupních proměnných, která by přenositelnost systému zajistila.

Systém ANFIS vykazoval nejlepší výsledky, co se modelování učícího procesu týče, ovšem při použití na data z jiných experimentů vykazoval největší odchylky od měřené hodnoty výstupní proměnné.

Tato práce byla vypracována za podpory programu č. MSM 6046137306 MŠMT ČR.

Literatura

- [1] Hand D., Mannila H., Smyth P.: *Principles of data mining*. The MIT Press, Cambridge, 2001.
- [2] Jang, J.-S. R.: *ANFIS: Adaptive-Network-based Fuzzy Inference Systems*. IEEE Transactions on Systems, Man and Cybernetics, 23(1993), No. 3, pp. 665-685.
- [3] Kiendl H.: *Fuzzy control methodenorientiert*. Oldenbourg Verlag, München, 1997.
- [4] Krabs M., Kiendl H.: *Anwendungsfelder der automatischen Regelgenerierung mit dem ROSA – Verfahren*. Automatisierungstechnik 43(1995), No. 6, pp. 269-276.
- [5] Krone A., Frencik Ch., Russak O.: *Design of a fuzzy controller for an alkoxylation process using the ROSA Method for automatic rule generation*. EUFIT'95, Aachen, pp.760-764.
- [6] Náhlík J., Palatová M., Girio F., Roseiro C.: *Model identification and physiological control of xylytol production using Debaryomyces hansenii*. Process Biochemistry, 38(2003), pp.1695-1705.
- [7] Zitta R.: *Generování fuzzy pravidel z experimentálních dat a jejich aplikace při modelování a řízení bioprocesů*, diplomová práce, 2004, VŠCHT Praha.

Ing. Roman Zitta, RNDr. Marta Palatová, CSc.
Vysoká škola chemicko-technologická
Ústav počítačové a řídicí techniky
Technická 5, 166 28 Praha 6
tel. +420 220 444 295, fax. +420 220 445 053
e-mail: Marta.Palatova@vscht.cz