

OPTIMALIZACE STRUKTURY NEURONOVÝCH SÍTÍ APLIKOVANÝCH V REÁLNÝCH TECHNICKÝCH PROSTŘEDÍCH

Michal MUSIL

Katedra provozní spolehlivosti, diagnostiky a mechaniky v dopravě

1. Úvod

Neuronové sítě představují v současné době jeden z hlavních prostředků systémů využívajících umělou inteligenci. Aplikační oblast neuronových sítí je značně rozsáhlá. Jedna z možných aplikací je využití neuronových sítí v diagnostických systémech, aplikace sítě pro detekci a klasifikaci poruch byla popsána v [13] a [14]. Při konkrétní aplikaci sítí je nutné zvolit vhodný typ a strukturu sítě. Pokud budeme pracovat z vícevrstevnými neuronovými sítěmi (MLP) je nutné stanovit optimální počet neuronů v jednotlivých vrstvách sítě (ve vstupní, skryté a výstupní vrstvě). V současné době neexistuje jednoznačné pravidlo pro stanovení optimálního počtu neuronů ve vrstvách sítě, v tomto příspěvku je popsána metoda stanovení počtu neuronů jednotlivých vrstev sítě využívané diagnostickým systémem s klasifikátorem chyb, ovšem je možné tuto metodu použít i v jiných aplikacích.

2. Vlastnosti neuronových sítí z hlediska počtu neuronů sítě

Velký počet neuronů ve vstupní a skryté vrstvě znesnadňuje jednak technickou realizaci sítě a jednak znesnadňuje proces učení se sítě (každý neuron navíc znamená zvýšení náročnosti algoritmu učení se sítě) a samozřejmě může docházet k jevům

jako je „přeučenosť neuronové sítě“, kdy lze výsledek učení se sítě vyjádřit tak že „sít si mnoho pamatuje, ale málo zobecňuje“, což v praxi znamená že sít je dokonale naučena na danou trénovací množinu, ovšem v režimu řešení úlohy již nedokáže aplikovat své naučené vlastnosti na konkrétní vstupní posloupnost a spolehlivě pracuje jen při řešení úlohy v níž se vyskytují téměř identické obrazy trénovací množiny. Což v případě klasifikátoru chyb znamená to, že sít je dokonale naučena na konkrétní vstupní diagnostický signál, ovšem v režimu řešení úlohy při vstupu signálu stejného typu (např. signál tenzometrů při měření) ovšem odlišného průběhu vykazuje sít již značnou výstupní chybu – jinak řečeno sít tento signál nedokáže spolehlivě rozpoznat.

Malý počet neuronů sítě sice urychlí proces učení se sítě, ale sít není schopna se naučit na danou vstupní trénovací posloupnost s požadovanou chybou. Sít je tedy na trénovací posloupnost naučena nepřesně, což se opět projeví při rozpoznávání konkrétního vstupního signálu v režimu řešení úlohy.

Realizovaný diagnostický systém – klasifikátor chyb je založen na tom, že sít je nejprve naučena na referenční bezchybný průběh signálu a v režimu řešení úlohy je na jeho vstup přiváděn signál obsahující rezidua odpovídající chybám diagnostického objektu. Pokud není neuronová sít naučena na referenční signál s požadovanou chybou, není poté sít schopna tyto odlišnosti od referenčního průběhu vyhodnotit. Je třeba podotknout, že správné nastavení neuronové sítě (počty neuronů v jednotlivých vrstvách a algoritmus učení se sítě) jsou základními předpoklady pro bezchybnou funkci neuronového klasifikátoru a je proto nutné tomuto nastavení věnovat dostatečnou pozornost.

Neuronové sítě použité v klasifikátoru chyb byly implementovány na PC prostřednictvím softwarového simulátoru neuronových sítí systémem Matlab/Simulink.

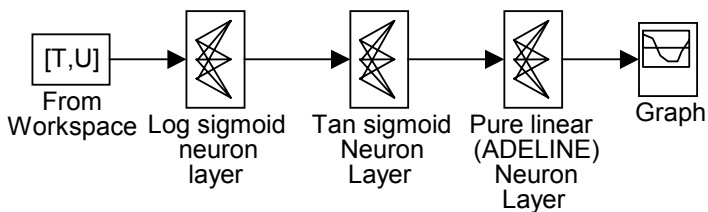
3. Proces učení se neuronové sítě a jeho průběh pro různé struktury sítí.

Učení se sítě algoritmem Back-Propagation v prostředí Matlab-Simulink

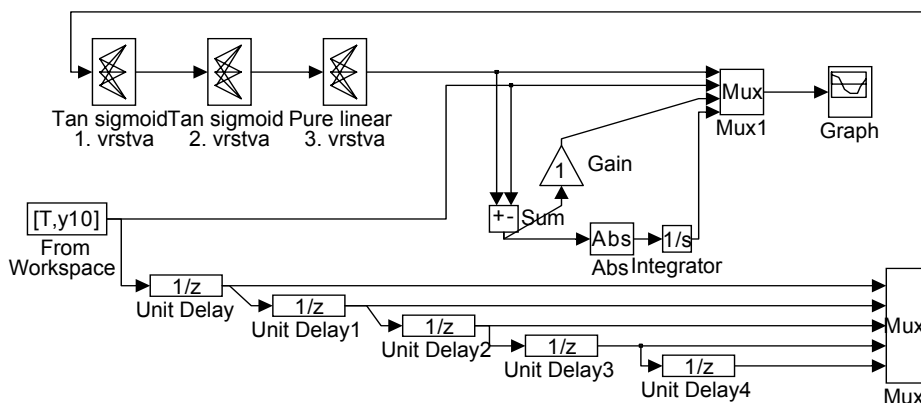
Na základě rozsáhlých experimentů byly jako nejvhodnější sítě vybrány pro klasifikátor chyb MLP sítě v třívrstevném uspořádání (vstupní, skrytá a výstupní vrstva), přičemž sít je zde využívána pro predikci průběhu signálu a jako nejvhodnější se ukázalo uspořádání sítě podle *obr. 1* kde je hledán optimální počet neuronů vstupní a skryté vrstvy, vrstva výstupní vzhledem k aplikaci sítě obsahuje jeden neuron, na jehož vstupy jsou přivedeny výstupy všech neuronů skryté vrstvy.

Z obrázku je zřejmá i volba vhodných přenosových funkcí v jednotlivých vrstvách sítě, zde opět praktické ověřování klasifikátoru ukázalo jako nejvhodnější využití ve výstupní vrstvě přenosové funkce čistě lineární a ve vrstvách vstupní a skryté přenosové funkce tangens-sigmoidální přenosovou funkcí. Na *obr. 2* je již uspořádání celé sítě pro predikci průběhu vstupního signálu, predikce je prováděna s pěti předchozími vzorky signálu (toto opět ukázalo praktické ověřování klasifikátoru,

kde by se na první pohled mohlo zdát, že z délkou „historie“ signálu kvalita predikce roste, ovšem není tomu tak a pro signály predikované klasifikátorem chyb se právě využití pěti předchozích vzorků signálu jeví jako nejvhodnější.



Obr. 1 Uspořádání neuronové sítě pro klasifikátor chyb
Fig. 1 The neural network ordering for error- classifier

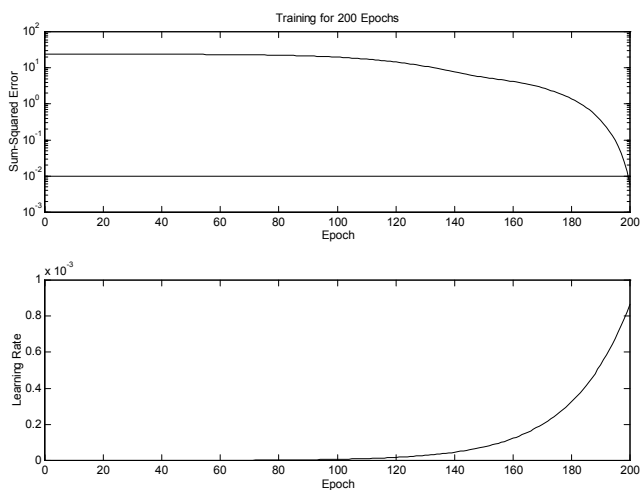


Obr. 2 Uspořádání neuronové sítě pro predikci signálu
Fig. 2 The neural network ordering for prediction of signal

Jako základní signál se kterým pracuje klasifikátor chyb byl zvolen sinusový signál, na níže uvedených grafech je zobrazen průběh učení se sítě na tento signál v různých uspořádáních sítí z hlediska počtu neuronů ve vstupní a skryté vrstvě.

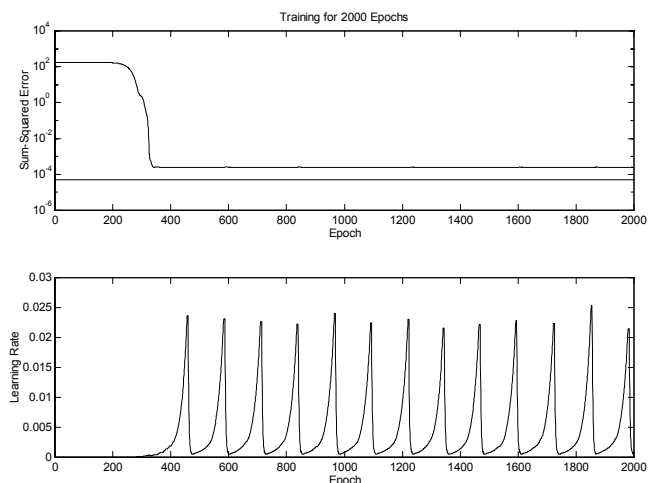
Síť NS1 zde byla v následujícím uspořádání : vstupní vrstva (FL) - 5 tansig neuronů, skrytá vrstva (SL) - 8 tansig neuronů, výstupní vrstva (OL) - 1 purelin neuron, počet trénovacích vektorů (TV) - 1000, velikost povolené chyby (Err) - 0,01 rychlost učení se sítě (LR) - 0,0005, počet epoch (E) – 5000

Jako hlavní kritérium pro hodnocení úspěšnosti učení se sítě byla zvolena metoda SSE což je součet kvadrátů chyb učení se sítě přepočtených na počet kroků. Chybou učení se sítě je myšlen rozdíl mezi okamžitou hodnotou vstupního a predikovaného signálu. Z obr. 3 je zřejmé že pro požadovanou chybu a v požadovaném uspořádání dosáhle síť po 200 krocích požadovaných parametrů.



Obr. 3 Průběh procesu učení se neuronové sítě na sinusový vstupní signál
Fig. 3 Learning the neural network on sine wave input signal

Dále tedy bude probíhat proces učení se sítě s přísnějším požadavkem na chybu SSE. Chyba zde byla nastavena v inicializačním souboru na hodnotu $8 \cdot 10^{-5}$, v tomto případě ovšem nastává druhý limitní případ, kdy síť v dané konfiguraci není schopna „naučit se“ na průběh daného signálu s požadovanou přesností. Zde přichází v úvahu dvě možnosti řešení : jednak je možno zmírnit požadavek na chybu SSE, ovšem to zaleží na požadované citlivosti klasifikátoru chyb a jednak je možno změnit strukturu sítě, tedy zejména počty neuronů ve vstupní a skryté vrstvě.



Obr. 4 Výsledek průběhu učení se sítě na vstupní sinusový signál, proces není úspěšně dokončen

Fig. 4 Learning the neural network on sine wave input signal, action isn't successfully finalization

4. Možnosti řešení stanovení optimálního počtu neuronů ve vstupní a skryté vrstvě sítě

Na předcházejících dvou příkladech bylo ukázáno, jak se prakticky projevuje nevhodně zvolený počet neuronů sítě nebo nevhodně zvolená velikost chyby SSE, pokud nechceme měnit požadovanou velikost chyby SSE je nutno pro její dosažení změnit strukturu sítě. Zde se tedy nabízí aplikace optimalizační metody Hidden Unit Search (HUS) - je to iterační metoda hledající optimální počet neuronů ve skryté vrstvě neuronové sítě. Síť učíme na určitou funkci s tím, že ve skryté vrstvě je minimální počet neuronů, tento počet postupně zvyšujeme a porovnáváme výsledky „učení se“ neuronové sítě na danou funkci (v našem případě daný vstupní např. sinusový signál). Porovnávání výsledků provádíme pomocí chyby SSE.

Významnou metodou pro optimalizaci struktury NS je využití tzv. Genetických algoritmů. Jedná se o metodu založenou na principech přirozené selekce a genetiky, metoda vychází z Darwinovy teorie evoluce. Tato optimalizační metoda využívá tři základní operace - reprodukce (reproduction), křížení (crossover), přeměna (mutation).

Typický genetický algoritmus pracuje v cyklu:

1. Vygenerování základní struktury neuronové sítě,
2. Vyhodnocení vlastností neuronové sítě,
3. Vytváření nové neuronové sítě s využitím základních operací (reprodukce, křížení, přeměna).

Tato metoda se používá v modifikaci Multilevel distributed Genetic Algritm.

Což znamená, že několik výchozích vzájemně odlišných struktur NS se zpracovává paralelně na několika počítačích a podle dosažených výsledků se sítě upravují podle zásad GA.

Pro stanovení optimálního počtu neuronů jsme zvolili metodu Hidden Unit Search, přičemž při sestavování výsledné struktury jsme využili i vlastností metody GA.

Vždy bylo sestaveno paralelně několik sítí, přičemž poměr neuronů ve vstupní vrstvě a skryté vrstvě byl 1:1.5, postupným testováním sítí daným signálem (např. základní sinusový signál), přičemž vždy sledujeme dosaženou velikost chyby SSE, strukturu kde je velikost této chyby nejmenší použijeme jako nejvhodnější.

V prostředí Matlabu je možno uložit vypočtené váhové koeficienty sítě pro každou strukturu zvlášť a následně provádět testování sítě s různými nastaveními pro různé signály a vybrat nejvhodnější konfiguraci.

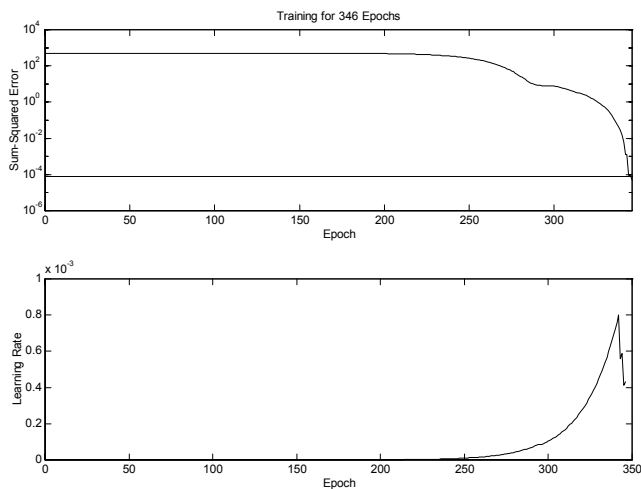
Na základě provedené optimalizace byla stanovena následující struktura neuronové sítě :

Síť NS2: vstupní vrstva (FL) - 10 tansig neuronů, skrytá vrstva (SL) - 16 tansig neuronů, výstupní vrstva (OL) - 1 purelin neuron, počet trénovacích vektorů (TV) - 1000,

velikost povolené chyby (Err) - 0,0001 rychlost učení se sítě (LR) - 0,0005, počet epoch (E) – 500.

Na obr. 5 je zobrazen průběh učení se sítě ve výše uvedené konfiguraci, z grafu je zřejmé, že síť dosáhne požadovaných parametrů chyby SSE je po 400 cyklech.

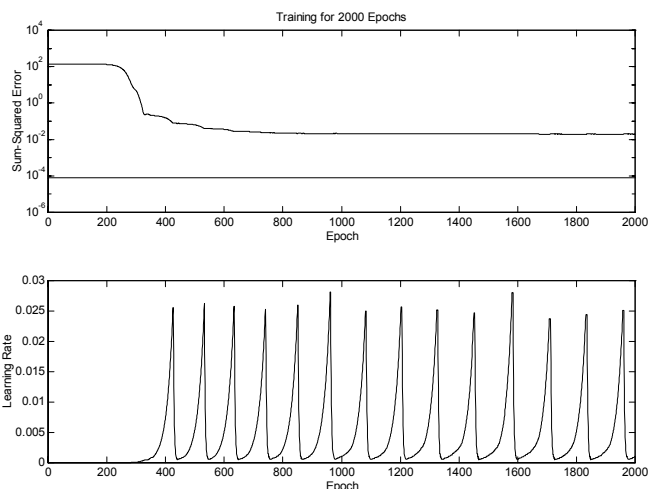
Stejným postupem je možno provést optimalizaci struktury pro další uvažované signály, v našem případě pracujeme kromě základního sinusového signálu se signály z tenzometrických měření a signály zvukovými (jedná se o akustické signály převedené do digitální formy). Pro tyto signály je již proces učení se sítě náročnější a zejména pro komplikované akustické signály se nepodařilo dosáhnout zcela přesné naučení se sítě na daný průběh vstupního signálu, ovšem praktickým ověřováním bylo zjištěno že pro potřebnou detekci chyb v signálech stupeň naučení se sítě (resp. velikost chyby SSE) je postačující.



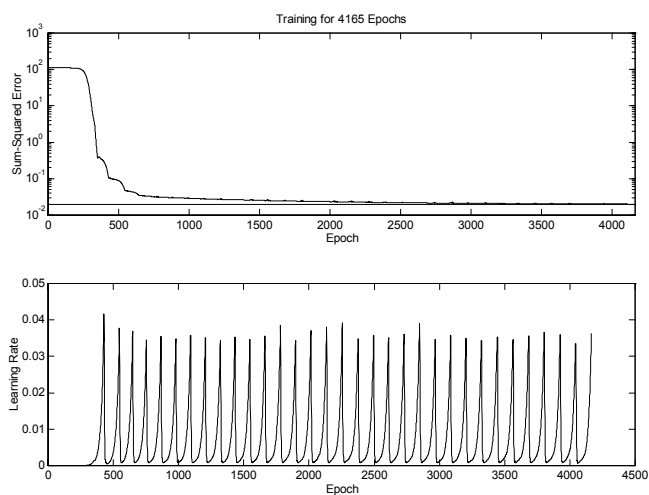
Obr. 5 Průběh procesu učení se neuronové sítě NS2 na sinusový vstupní signál
Fig. 5 Learning the neural network NS2 on sine wave input signal

Na obr. 6 je zobrazen průběh učení se sítě NS2 na signál z tenzometrických měření, tento obrázek jasně dokumentuje to, že pro jiný typ signálu je nutno sestavit jinou neuronovou síť, síť NS2 je optimalizována pro sinusový signál. Pokud na vstup takto optimalizované sítě NS2 přivedeme signál jiný síť není schopna se v dané konfiguraci na signál naučit. Je tedy nutné síť patřičně upravit a výše popsáním způsobem provést optimalizaci pro konkrétní signál – např. signál z tenzometrických měření. Na obr. 7 je průběh učení se sítě NS3 na tento signál, optimalizovaná síť má následující parametry :

Síť NS3: vstupní vrstva (FL) - 13 tansig neuronů, skrytá vrstva (SL) - 20 tansig neuronů, výstupní vrstva (OL) - 1 purelin neuron, počet trénovacích vektorů (TV) - 1000, velikost povolené chyby (Err) - 0,02 rychlost učení se sítě (LR) - 0,004, počet epoch (E) – 5000.



Obr. 6 Průběh procesu učení se neuronové sítě NS2 na signál z tenzometrických měření
Fig. 6 Learning the neural network NS2 on signals from tensometric measurements



Obr. 7 Průběh procesu učení se neuronové sítě NS3 na signál z tenzometrických měření
Fig. 7 Learning the neural network NS2 on signals from tensometric measurements

5. Závěr

Praktickým ověřováním známých metod optimalizace sítě byla jako nejvhodnější pro danou aplikaci zvolena metoda hledající optimální počet neuronů skryté vrstvy sítě. Poměr neuronů vstupní a skryté vrstvy byl stanoven na 1:1,5. Je třeba podotknout, že práce s neuronovými sítěmi představuje především učení se sítí na dané trénovací

množiny (v našem případě vstupní diagnostické signály) a pro každou aplikaci resp. vstupní signál, je třeba najít optimální strukturu sítě. Neexistuje tedy jednoznačné pravidlo pro stanovení počtu neuronů sítě a počtu vrstev sítě. Také při návrhu sítě pro nejkomplicovanější typ signálu – akustický digitalizovaný signál, bylo po optimalizaci struktury dosaženo lepších výsledků při učení se sítě na tento signál než v původně realizovaném klasifikátoru (naučená neuronová síť vykazovala menší chybu SSE než původní síť klasifikátoru). Menší počty neuronů v síti také částečně eliminují negativní vlastnost sítí s velkým počtem neuronů a tou je tzv. přeúčenost neuronové sítě (tzn. síť málo zobecňuje).

Lektoroval: Doc. Ing. Spalek, Doc. Ing. Muzikářová, Ing. Franeková

Předloženo: 14.5.2003

Literatura

1. BÍLA, J. Umělá inteligence a neuronové sítě v aplikacích. (2. přepracované vydání). Vydavatelství ČVUT, Praha, (1998).
2. ŠNOREK, M., JIŘINA, M. Neuronové sítě a neuropočítače. ČVUT, Praha, (1996).
3. BÍLA, J., KOŘAN, L. Application of Matroids in Detection. In Proc. Of 12th Int. Conference „Process Control 99“, High Tatras,) pp. 78-82 S.R., Vol. 1., (1999).
4. BÍLA, J., KREILD, M., HOUFEK, P., KOŘAN, L. Neural Networks in Ultrasonic Testing of Thin Walled Welded Constructions used in Aerospace Industry. In Proc. Of 5th Int. Conference on Sof Computing – MENDEL 99pp. 325-328, Brno, C.R., (1999).
5. KREIDL, M. Diagnostické systémy. Vydavatelství ČVUT, Praha, (1995).
6. VITKAJ, J. Analysis of Chaotic Signals by means of Nonlinear Methods. In Proc. Of 4th Int. Conf. on Soft Computing, pp. 296-299, Brno, C.R., (1998).
7. LÁNSKÝ, M. Teorie automatizovaných diagnostických systémů v dopravě a spojkách. Nadas, Praha (1990).
8. LÁNSKÝ, M. Teorie automatizovaných diagnostických systémů v dopravě a spojkách, Nadas, Praha (1990).
9. PATTON, R. J. Robust model - based fault diagnosis The state of the art, Model - Based Fault Diagnosis, pages 40 – 60, Zakopane (1994).
10. SEDLÁČEK, M. Zpracování signálů v měřící technice. Skriptum ČVUT, Praha, (1993).
11. DALMI, I., KOVÁCS, L., TERSTVÁNSZKY, G. Diagnosing Priori Unknown Faults by Radial Basis Function Neural Network. Proc. of ECSAP-97 The First European Conference on Signal Analysis and Prediction, Prague, Czech Republik, s.228-231, June (1997).
12. MUSIL, M. Diagnostika analogových a číslicových přenosových kanálů. Sborník prací konference "Současné problémy v kolejových vozidlech" (J. Machalíková edit.), Univerzita Pardubice, (1994).
13. MUSIL, M. Aplikace neuronových sítí pro detekci poruch signálů. Scientific papers of the University Pardubice, Series B, The Jan Perner Transport Faculty, Pardubice, str. 165-174, ISBN - 80 -7194 -283-9, ISSN - 1211 – 6610, (1999).
14. MUSIL, M. Návrh a realizace klasifikátoru chyb diagnostického systému s využitím neuronových sítí. Scientific papers of the University Pardubice, Series B, The Jan Perner Transport Faculty, Pardubice str. 5-25, ISBN - 80 -7194 -548-0, (2001).

Michal Musil:

15. KOTEK, Z., BRŮHA, I., CHALUPA, V., JELÍNEK, J. Adaptivní a učící se systémy. Praha, SNTL, (1980).

Resumé

OPTIMALIZACE STRUKTURY NEURONOVÝCH SÍTÍ APLIKOVANÝCH V REÁLNÝCH TECHNICKÝCH PROSTŘEDÍCH

Michal MUSIL

Příspěvek se zabývá optimalizací struktury neuronové sítě aplikované v diagnostickém systému jako klasifikátor chyb. Síť je tedy optimalizována pro tuto aplikaci, ovšem optimalizační metody lze využít obecně. Optimalizované sítě jsou typu MLP a jako učící algoritmus je využíván algoritmus Back-propagation. Vhodně zvolená struktura neuronové sítě má podstatný vliv na náročnost realizace a vlastnosti použité sítě.

Summary

OPTIMIZATION STRUCTURES THE NEURAL NETWORKS APPLY IN REAL TECHNICAL WORLD

Michal MUSIL

This contribution deals with optimization structure neural network applied in diagnostic system like error-classifier. Network is then optimization for this application, indeed optimization method it is possible use generally. Optimized sieve they are type MLP and like algorithm is exploitation algorithm Back-propagation. Becomingly election structure neural network has substantial influence over heftiness realization and characteristics networks.

The neural networks are the main means for realisation of the classifier. Representative types of the neural networks and their usability for solution of a project of the classifier are mentioned in the introductory part of the paper. Deterministic methods for the classification of errors in the system as well as their mathematical description are given in the next part. Methods of discriminated functions and methods of minimum distance from a standard are presented here.

Diagnostics of an error is divided into the three phases - detection, localisation and classification of the error. First of all, the described system realised by means of the neural networks, performs a detection and localisation of the error - the neural networks are learnt to non-defective course of diagnostic signals and, each defect in the system is displayed as a symptom in the diagnostic signal, the symptom is detected and localised by means of the neural networks. Further more, the part of classification of the error follows, when the separated part of a signal containing this error (symptom of the error) is analysed by means of the error-classifier which is realised with parallel-working neural networks, when each network is learnt to a particular type of the error. The double-learning principle of the neural networks is applied here - first, to the non-defective course of the signal (detection, localisation) and, subsequently to the particular symptoms of errors in the signal (classification).

The neural networks for realisation of the error-classifier are constituted by means of the software products Matlab/Simulink and its Neural Network Toolboxes.

The fundamental part of the contribution is aimed at the project, realisation and verification for the operation of the diagnostic system with use of the neural networks for the detection of errors in signals. The survey of the considered errors of a signal and their classification into the particular categories is presented here. Detection, localisation and classification of errors in the signals, is described in the next part. The principle of generation of the differential vector and classification of errors with use of the symptom analysis method in the differential vector, as well as with use of

the method of analysis of a separated error part of a signal is explained here. This principle of localisation and classification of errors is elucidated here with use of the examples of diagnostics of errors in representative types of signals. As a practical application, the diagnostics of errors in signals from tensometric measurements as well as from real audio signals is given in the paper. Results relating to the detection and localisation of an error as well as to the classification of the error by means of the neural-classifier are referred here.

The evaluation of an efficiency of the described error-classifier by means of the defined criteria as well as the possibility of the next application of the error-classifier, is mentioned in conclusion.

Zusammenfassung

DIE OPTIMISIERUNG AUFBAUS NEURAL NETZE APPLIZIEREN IN REAL TECHNISCH WELT

Michal MUSIL

Beitrag sich betreibt Optimisierung Aufbaus Neuron Netze angewandt in Diagnosesystem wie Klassierer Fehler. Netzwerk ist so Abstimmung fürs folgende Applikation, allerdings optimalizacni Methode kann man nutzen allgemein. Getunt Netze sind Modell MLP und wie Algorithmus ist ausgenutzt Algorithmus Back-propagation. Entsprechend gewählte Struktur Neuron Netze hat wesentlich Einfluß an Ansprüche Realisation und Eigenschaften Genuß Netze.