

## Mašinsko učenje veštačke neuronske mreže sa radijalnim aktivacionim funkcijama Gausovog tipa na bazi Kalmanovog filtra – teorijske osnove

NAJDAN L. VUKOVIĆ, Univerzitet u Beogradu,

Inovacioni centar Mašinskog fakulteta, Beograd

ZORAN Đ. MILJKOVIĆ, Univerzitet u Beogradu,

Mašinski fakultet, Beograd

Originalni naučni rad

UDC: 004.85:004.032.26]:681.5.015.44

*U ovom radu se analizira problem mašinskog učenja veštačke neuronske mreže sa radijalnim aktivacionim funkcijama Gausovog tipa na bazi Kalmanovog filtra. Prikazana su tri nova sekvencijalna algoritma mašinskog učenja: prvi algoritam direktno primenjuje linearizovani Kalmanov filter kao algoritam mašinskog učenja, drugi algoritam primenjuje dual Kalmanovom filtru pod nazivom linearizovani informacioni filter, dok treći algoritam na poseban način aproksimira prvi i drugi moment Gausove raspodele. U radu se naglašavaju osnovne prednosti koje pomenuti algoritmi imaju u poređenju sa konvencionalnim vidovima mašinskog učenja. Za sva tri algoritma razvijen je odgovarajući matematički model veštačke neuronske mreže sa radijalnim aktivacionim funkcijama Gausovog tipa. Analizirane su osnovne postavke izvedenih algoritama u cilju njihove primene na složene probleme u inženjerskoj praksi.*

**Ključne reči:** mašinsko učenje, veštačke neuronske mreže, Kalmanov filter

### 1. UVOD

Veštačke neuronske mreže predstavljaju matematičko-softverski alat namenjen za modeliranje problema koje je teško ili nemoguće definisati matematički. Veštačke neuronske mreže su tokom niza godina razvoja uspešno primenjene na različite probleme iz inženjerske prakse [1], [2], [3], [4]. Prednosti razvoja modela problema primenom veštačkih neuronskih mreža ogledaju se u mogućnosti modeliranja kompleksnih problema bez potrebe za poznavanjem eksplicitnog matematičkog modela.

Razvoj modela problema primenom veštačkih neuronskih mreža podrazumeva sledeće korake:

- definisanje ulaznih i izlaznih veličina;
- izbor sistema za generisanje potrebnih podataka;
- akvizicija eksperimentalnih podataka;
- pretprocesiranje eksperimentalnih podataka;
- izbor veštačke neuronske mreže, arhitekture i tipa aktivacionih funkcija;
- izbor tipa mašinskog učenja (nadgledano ili ne-nadgledano) i odgovarajućeg algoritma učenja;

- optimizacija parametara veštačke neuronske mreže-mašinsko učenje;
- validacija kvaliteta obučenosti mreža;
- testiranje „performansi” dobijenih modela.

Korak broj 1 potpuno je definisan polaznim problemom. Poznavanje fizičkih osnova izučavanog problema predstavlja osnovu za primenu veštačkih neuronskih mreža. Na ovaj način se koraci br. 2, i br. 3 jednostavno sprovode poznavanjem fizičkog uticaja jednih veličina na druge. U koraku br. 4 se vrši skaliranje eksperimentalnih podataka koje za cilj ima preslikavanje polaznog eksperimentalnog skupa u neki od uobičajenih opsega promene. Nakon pretprocesiranja eksperimentalnih podataka potrebno je izabrati tip veštačke neuronske mreže, tip aktivacione funkcije i broj neurona, kao i broj slojeva ako je u pitanju višeslojna mreža [1], [4]. Navedeni parametri direktno utiču na performanse i kvalitet predikcije. U koracima broj 6 i broj 7 treba definisati vid mašinskog učenja koji direktno zavisi od polaznog problema, kao i algoritam učenja. Od sedmog koraka započinje iterativni postupak koji se sastoji od optimizacije parametara i provere „naučenog” (koraci broj 8 i broj 9). U ovom radu analizirana je primena tri algoritma u cilju optimizacije parametara veštačke neuronske mreže sa radijalnim aktivacionim funkcijama Gausovog tipa.

Adresa autora: Najdan Vuković, Univerzitet u Beogradu, Inovacioni centar Mašinskog fakulteta, Beograd, Kraljice Marije 16

Rad primljen: 07.05.2014.

Rad prihvaćen: 12.06.2014.

## 2. VEŠTAČKA NEURONSKA MREŽA SA RADIJALNIM AKTIVACIONIM FUNKCIJAMA GAUSOVOG TIPA

Veštačka neuronska mreža sa radijalnim aktivacionim funkcijama predstavljena je sledećim izrazom [1], [2], [5]:

$$\mathbf{y}_i = \sum_{j=1}^N \mathbf{w}_{ij} \mathbf{f}(\mathbf{x}_i, \boldsymbol{\mu}_j, \sigma_j) \quad (1)$$

Osnovni parametri su:

- vektor težinskih parametara veštačke neuronske mreže  $\mathbf{w}_{ij}$ ;
- vektor prototipova (repera)  $\boldsymbol{\mu}_j \in \mathbb{R}^{n_x}$ , gde  $n_x$  označava broj dimenzija ulaznog vektora;
- radijalne distance  $\sigma_j; (\sigma_j \in \mathbb{R}^+)$ ;
- $\mathbf{f}(\cdot, \cdot, \cdot)$  je aktivaciona funkcija veštačke neuronske mreže,

$\mathbf{x}_i \in \mathbb{R}^{n_x}$  i  $\mathbf{y}_i \in \mathbb{R}^{n_y}$  su ulazni i izlazni vektor (respektivno) koji formiraju skup za optimizaciju.

Postoji više mogućnosti izbora aktivacione funkcije, a sam izbor zavisi od problema za koji se razvija ovaj tip veštačke neuronske mreže. Mogući tipovi aktivacionih funkcija su:

- Gausova aktivaciona funkcija:

$$\mathbf{f}(\mathbf{x}_i) = \exp\left(-\frac{1}{2\sigma_j^2} \boldsymbol{\xi}_{ij}^T \boldsymbol{\xi}_{ij}\right) \quad (2)$$

- Inverzna kvadratna aktivaciona funkcija:

$$\mathbf{f}(\mathbf{x}_i) = (\boldsymbol{\xi}_{ij}^T \boldsymbol{\xi}_{ij} + \sigma_j^2)^{-1/2} \quad (3)$$

- Kvadratna aktivaciona funkcija:

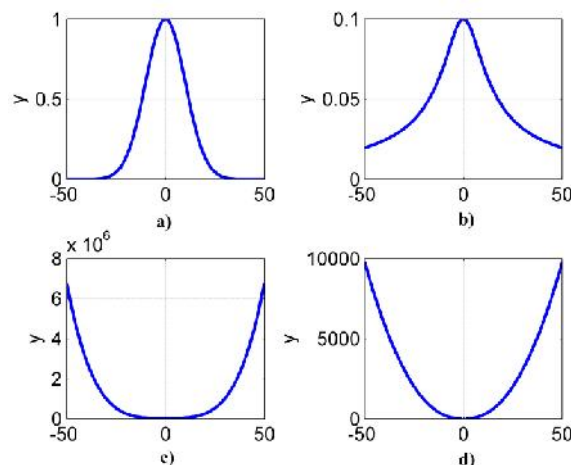
$$\mathbf{f}(\mathbf{x}_i) = (\boldsymbol{\xi}_{ij}^T \boldsymbol{\xi}_{ij} + \sigma_j^2)^{1/2} \quad (4)$$

- Logaritamska aktivaciona funkcija:

$$\mathbf{f}(\mathbf{x}_i) = \boldsymbol{\xi}_{ij}^T \boldsymbol{\xi}_{ij} \ln(\boldsymbol{\xi}_{ij}) \quad (5)$$

gde je  $\boldsymbol{\xi}_{ij} = \mathbf{x}_i - \boldsymbol{\mu}_j; (\boldsymbol{\xi}_{ij} \in \mathbb{R}^{n_x})$  razlika između  $i$ -tog ulaznog vektora  $\mathbf{x}_i$  i  $j$ -tog prototip vektora  $\boldsymbol{\mu}_j$ . Parametar  $\sigma_j$  određuje širinu aktivacione funkcije, odnosno graničnu vrednost pripadnosti  $i$ -tog ulaznog vektora  $\mathbf{x}_i$  skupu čiji je predstavnik  $j$ -ti prototip vektor  $\boldsymbol{\mu}_j$ . Na slici 1 su grafički prikazane četiri karakteristične aktivacione funkcije  $\mathbf{f}(\mathbf{x}_i)$ . Na osnovu ma-

tematičke formulacije aktivacionih funkcija, kao i na osnovu njihovog grafičkog prikaza, može se zaključiti da odziv neurona na ulazni vektor  $\mathbf{x}_i \in \mathbb{R}^{n_x}$  direktno zavisi od razlike  $\boldsymbol{\xi}_{ij}$  između ulaznog vektora  $\mathbf{x}_i \in \mathbb{R}^{n_x}$  i vektora prototipa  $\boldsymbol{\mu}_j \in \mathbb{R}^{n_x}$ .

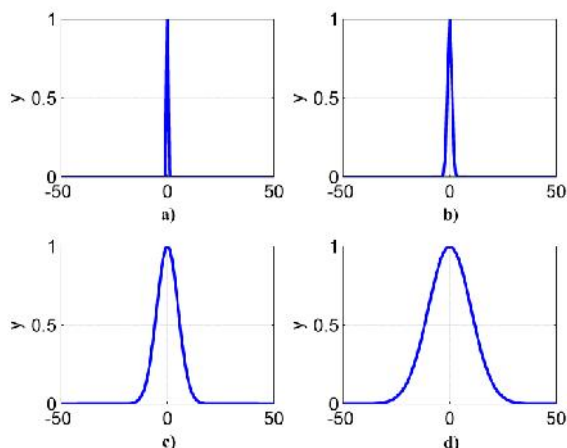


Slika 1 - Tipovi aktivacionih funkcija: a) Gausova aktivaciona funkcija; b) Inverzna kvadratna aktivaciona funkcija; c) Kvadratna aktivaciona funkcija; d) Logaritamska aktivaciona funkcija

U slučaju Gausove i inverzne kvadratne aktivacione funkcije, aktivacija neurona je veća ukoliko je  $\boldsymbol{\xi}_{ij}$  opada. U slučaju kvadratne aktivacione funkcije i logaritamske aktivacione funkcije aktivacija neurona se povećava sa porastom  $\boldsymbol{\xi}_{ij}$ . Moguće je uočiti i da je izlazna vrednost aktivacione funkcije (aktivacija neurona) uvek pozitivna veličina

$$\mathbf{f}(\mathbf{x}_i) > \mathbf{0}, \forall \mathbf{x}_i \in \mathbb{R}^{n_x}$$

U ovom istraživanju izabrana je veštačka neuronska mreža sa radijalnim aktivacionim funkcijama u vidu aktivacionih funkcija Gausovog tipa. Osnovni razlog za izbor Gausove aktivacione funkcije je pre svega u postavci, koja obezbeđuje veću aktivaciju neurona ukoliko  $\boldsymbol{\xi}_{ij}$  opada i manju aktivaciju ukoliko  $\boldsymbol{\xi}_{ij}$  raste. Potrebno je naglasiti i osnovnu razliku reprezentacije skupa podataka modeliranjem primenom višeslojne perceptron veštačke neuronske mreže i veštačke neuronske mreže sa radijalnim aktivacionim funkcijama. Naime, višeslojni perceptron uspostavlja klasifikacionu granicu između odgovarajućih elemenata skupa, dok radijalne aktivacione funkcije uspostavljaju sličnost između elemenata ulaznog skupa podataka [1]. U tom smislu, na slici 2 je dat prikaz Gausove aktivacione funkcije za različite vrednosti radijalne distance  $\sigma_i$



Slika 2 - Gausova aktivaciona funkcija za različite vrednosti parametra  $\sigma_i$ : a)  $\sigma_i = 0,1$ ; b)  $\sigma_i = 1$ ; c)  $\sigma_i = 5$ ; d)  $\sigma_i = 10$ . Sa porastom  $\sigma_i$  povećava se opseg aktivacione funkcije.

### 3. OPTIMIZACIJA PARAMETARA VEŠTAČKE NEURONSKE MREŽE SA RADIJALNIM AKTIVACIONIM FUNKCIJAMA GAUSOVOG TIPA – MAŠINSKO UČENJE

Osnovna prednost veštačke neuronske mreže sa radijalnim aktivacionim funkcijama je u mogućnosti razdvajanja optimizacije parametara aktivacionih funkcija i težinskih parametara same veštačke neuronske mreže. Npr. moguće je primeniti nenadgledani vid mašinskog učenja [6], [7] za optimizaciju centara i radijalnih distanci na osnovu isključivo ulaznih podataka  $\mathbf{x}_i \in \mathbb{R}^n$  iz celokupnog skupa za optimizaciju. Nakon ovog postupka, primenjuje se nadgledani vid mašinskog učenja u kome se uzimaju u obzir i ulazni i izlazni parovi iz skupa za optimizaciju. Proces optimizacije centara i radijalnih distanci aktivacionih funkcija se može nastaviti i u drugom koraku ali ih možemo i smatrati za konstantne parametre. Primeri ovakvog kombinovanog mašinskog učenja se mogu videti u [7], [8]. Optimizacija parametara veštačke neuronske mreže može biti sprovedena primenom metode najmanjeg gradijenta [9], [10], ili rekurzivne metode najmanjih kvadrata [4]. Moguće je sprovesti i optimizaciju parametara i simultano menjati broj neurona uzimajući u obzir grešku učenja [2], [11], [12], [13]. Primena ovog pristupa rezultira bržom optimizacijom i za rezultat ima optimalne arhitekture u pogledu broja neurona, sa poboljšanom tačnošću i preciznošću [2], [11].

U ovom radu su razvijena tri sekvencijalna algoritma nadgledanog tipa mašinskog učenja. Za potrebe optimizacije parametara veštačke neuronske mreže analizirani su sledeći algoritmi:

- Linearizovani Kalmanov filter (LKF);
- Linearizovani informacioni filter (LIF);
- Algoritam specifične aproksimacije momenata (ASAM).

Primena LKF algoritma za potrebe optimizacije odgovarajućih parametara veštačke neuronske mreže je sprovedena u [10] i predstavlja (skoro) „standardni” postupak. Međutim, osnovna razlika pomenutih rezultata u odnosu na algoritam prikazan u ovom radu ogleda se u sekvencijalnoj obradi podataka koji je primenjen u ovom radu (*on line* režim), kao i različitim tipu aktivacionih funkcija. Po saznanjima autora, primena linearizovanog informacionog filtra za ove potrebe nije uobičajeni postupak i pretraživanjem referentnih izvora nisu pronađeni odgovarajući rezultati. Algoritam specifične aproksimacije momenata primenjen je u nekoliko referentnih izvora u cilju optimizacije odgovarajućih parametara, međutim s obzirom na to da ASAM predstavlja relativno nov koncept u teoriji ocenjivanja (potpuno zaokružen na teoretskom nivou početkom 2000. godine) njegova primena za različite inženjerske probleme nije dovoljno analizirana i istražena. Prethodna činjenica predstavlja osnovnu ideju zbog koje je izabran razvoj algoritma obučavanja na bazi specifične aproksimacije momenata.

Pre nego što uvedemo izabrane algoritme neophodno je formirati vektor stanja s obzirom na to da sva tri algoritma počivaju na ovakvoj formulaciji. Na osnovu jednačine (1) veštačka neuronska mreža sa radijalnim aktivacionim funkcijama je data kao:

$$\mathbf{y}_i = \begin{bmatrix} w_{i0} & w_{i1} & \dots & w_{iJ} \\ w_{20} & w_{21} & \dots & w_{2J} \\ \vdots & \vdots & \vdots & \vdots \\ w_{J0} & w_{J1} & \dots & w_{Jj} \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{f}_1(\mathbf{x}_i, \mu_1, \sigma_1) \\ \vdots \\ \mathbf{f}_J(\mathbf{x}_i, \mu_J, \sigma_J) \end{bmatrix} \quad (6)$$

odnosno, zapisano u matricnom obliku

$$\mathbf{y}_i = \mathbf{W}_{i \times J} \mathbf{G}_{J \times 1} \quad (7)$$

gde  $\mathbf{W}_{i \times J}$  predstavlja matricu svih težinskih parametara, a  $\mathbf{G}_{J \times 1}$  vektor aktivacije koga čine sve aktivacione funkcije. Vektor stanja sistema čine svi parametri veštačke neuronske mreže koje je potrebno optimizovati u odnosu na zadati kriterijum. Znajući da je svaka vrsta matrice težinskih parametara  $\mathbf{W}_{i \times J}$  poseban vektor, polazna matrica se može napisati i u sledećem obliku:

$$\mathbf{W}_{ij} = \begin{bmatrix} w_{i0} & w_{i1} & \dots & w_{iJ} \\ w_{20} & w_{21} & \dots & w_{2J} \\ \vdots & \vdots & \vdots & \vdots \\ w_{J0} & w_{J1} & \dots & w_{Jj} \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \mathbf{w}_i^T \end{bmatrix} \quad (8)$$

Da bi se formirao vektor stanja neophodno je napisati težinske parametre u vektorskom obliku:

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_1^T & \mathbf{w}_2^T & \dots & \mathbf{w}_i^T \end{bmatrix}^T \quad (9)$$

Preostali elementi vektora stanja čine prototip vektori  $\boldsymbol{\mu}_j$  i sve radijalne distance  $\sigma_j$  koje određuju pripadnost  $j$ -tom prototip vektoru, odnosno:

$$\mathbf{t} = \begin{bmatrix} \boldsymbol{\mu}_1^T & \boldsymbol{\mu}_2^T & \dots & \boldsymbol{\mu}_J^T & \sigma_1^T & \sigma_2^T & \dots & \sigma_J^T \end{bmatrix}^T \quad (10)$$

Dakle, vektor stanja čine svi težinski parametri veštačke neuronske mreže  $\mathbf{W}_{i \times j}$  i parametri svih aktivacionih funkcija  $\mathbf{t}$

$$\begin{aligned} \boldsymbol{\lambda} &= \begin{bmatrix} \mathbf{w}^T & \mathbf{t}^T \end{bmatrix}^T = \\ &= \begin{bmatrix} \mathbf{w}_1^T & \mathbf{w}_2^T & \dots & \mathbf{w}_i^T & \boldsymbol{\mu}_1^T & \boldsymbol{\mu}_2^T & \dots & \boldsymbol{\mu}_J^T & \sigma_1^T & \sigma_2^T & \dots & \sigma_J^T \end{bmatrix}^T \end{aligned} \quad (11)$$

Ukupan broj parametara  $N_p$  veštačke neuronske mreže, a time i broj dimenzija vektora stanja sistema  $\boldsymbol{\lambda}$  je:

$$\begin{aligned} N_p &= \dim(\boldsymbol{\lambda}) = \\ &= n_{\text{neurona}}(\dim(\mathbf{x}) + \dim(\mathbf{y}) + 1) \end{aligned} \quad (12)$$

Funkcija  $\dim(\cdot)$  određuje broj dimenzija odgovarajućeg argumenta,  $\mathbf{x}$  i  $\mathbf{y}$  predstavljaju ulazni i izlazni vektor veštačke neuronske mreže respektivno, a  $n_{\text{neurona}}$  je ukupan broj neurona.

Nakon definisanja vektora stanja sistema moguće je nastaviti uvođenje linearizovanog Kalmanovog filtra, linearizovanog informacionog filtra i algoritma specifične aproksimacije momenata, kao izabranih algoritama za sekvencijalnu optimizaciju parametara veštačke neuronske mreže sa radijalnim aktivacionim funkcijama Gausovog tipa.

Linearni dinamički sistem u diskretnom obliku dat je sledećim jednačinama:

$$\begin{aligned} \boldsymbol{\lambda}_k &= \boldsymbol{\lambda}_{k-1|k-1} + \boldsymbol{\varepsilon} \\ \mathbf{y}_k &= \mathbf{g}(\boldsymbol{\lambda}_{k|k-1}, \mathbf{x}_k) + \boldsymbol{\omega} \end{aligned} \quad (13)$$

$\mathbf{g}(\boldsymbol{\lambda}_{k|k-1}, \mathbf{x}_k)$  je veštačka neuronska mreža.  $\boldsymbol{\varepsilon}$  i  $\boldsymbol{\omega}$  su odgovarajući beli šumovi ( $\boldsymbol{\varepsilon} \sim N(\boldsymbol{\varepsilon}, \mathbf{Q})$  i  $\boldsymbol{\omega} \sim N(\boldsymbol{\omega}, \mathbf{R})$ );  $\mathbf{y}_k$  je željeni izlazni vektor. Kao što se može videti, optimizacija parametara podrazumeva linearan sistem, što je predstavljeno identičnim preslikavanjem  $\boldsymbol{\lambda}_{k-1|k-1} \rightarrow \boldsymbol{\lambda}_k$ . Uvode se sledeće pretpostavke:

$$E[\boldsymbol{\lambda}_{0|0}] = \hat{\boldsymbol{\lambda}}_{0|0}; \quad E\left[\left(\boldsymbol{\lambda}_{0|0} - \hat{\boldsymbol{\lambda}}_{0|0}\right)\left(\boldsymbol{\lambda}_{0|0} - \hat{\boldsymbol{\lambda}}_{0|0}\right)^T\right] = \mathbf{P}_{0|0}$$

$$E[\boldsymbol{\varepsilon}_k] = \mathbf{0}; \quad E\left[\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_{k+1}^T\right] = \mathbf{Q} \delta_{k,k+1}$$

$$E[\boldsymbol{\omega}_k] = \mathbf{0}; \quad E\left[\boldsymbol{\omega}_k \boldsymbol{\omega}_{k+1}^T\right] = \mathbf{R} \delta_{k,k+1}$$

gde je  $\delta_{k,k+1}$  Kronekerov delta simbol. U nastavku su detaljno predstavljeni razvijeni sekvencijalni algoritmi mašinskog učenja bazirani na Kalmanovom filtru.

### 3.1 Mašinsko učenje primenom linearizovanog Kalmanovog filtra

Algoritam optimizacije parametara na bazi LKF algoritma prikazan je u Algoritmu 1.

Algoritam 1 - Pseudokod linearizovanog Kalmanovog filtra za optimizaciju parametara veštačke neuronske mreže sa radijalnim aktivacionim funkcijama Gausovog tipa.

1. **Algoritam LKF\_RGAF** ( $\hat{\boldsymbol{\lambda}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}, \mathbf{x}(k), \mathbf{y}(k)$ )

2.  $E[\boldsymbol{\lambda}_{0|0}] = \hat{\boldsymbol{\lambda}}_{0|0}; \quad E\left[\left(\boldsymbol{\lambda}_{0|0} - \hat{\boldsymbol{\lambda}}_{0|0}\right)\left(\boldsymbol{\lambda}_{0|0} - \hat{\boldsymbol{\lambda}}_{0|0}\right)^T\right] = \mathbf{P}_{0|0}$

3.  $\mathbf{Q} = \text{diag}(\dim(\boldsymbol{\lambda})); \quad \mathbf{R} = \text{diag}(\dim(\mathbf{y}_k))$

4. *while*  $Er > Er_{\min}$

5.  $\hat{\boldsymbol{\lambda}}_{k|k-1} = \hat{\boldsymbol{\lambda}}_{k-1|k-1}$

6.  $\mathbf{P}_{k|k-1} = \mathbf{P}_{k-1|k-1} + \mathbf{Q}$

7.  $\hat{\mathbf{y}}_k = \mathbf{g}(\hat{\boldsymbol{\lambda}}_{k|k-1}, \mathbf{x}(k))$

8.  $\mathbf{H}_k = \nabla_{\boldsymbol{\lambda}} \mathbf{y}_k^T = \left[ \nabla_{\mathbf{w}} \mathbf{g}(\cdot, \cdot), \nabla_{\boldsymbol{\mu}} \mathbf{g}(\cdot, \cdot), \nabla_{\sigma} \mathbf{g}(\cdot, \cdot) \right]$

9.  $\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R})^{-1}$

10.  $\hat{\boldsymbol{\lambda}}_{k|k} = \hat{\boldsymbol{\lambda}}_{k|k-1} + \mathbf{K}_k (\mathbf{y}(k) - \hat{\mathbf{y}}_k)$

11.  $\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1}$

12.  $Er = \frac{1}{2} (\mathbf{y} - \mathbf{g}(\hat{\boldsymbol{\lambda}}_{k|k}, \mathbf{x}))^T (\mathbf{y} - \mathbf{g}(\hat{\boldsymbol{\lambda}}_{k|k}, \mathbf{x}))$

13. *endwhile*

14. *return*  $\hat{\boldsymbol{\lambda}}_{k|k}, \mathbf{P}_{k|k}$

Na osnovu poznatih karakteristika analiziranog problema, potrebno je usvojiti početne vrednosti sledećih veličina: vektora stanja sistema  $\hat{\boldsymbol{\lambda}}_{0|0}$ , matrice kovarijansi  $\mathbf{P}_{0|0}$ , kovarijanse šumova sistema  $\mathbf{Q}$  i merenja  $\mathbf{R}$ . Ulazne veličine u algoritam su ocena vektora stanja sistema  $\hat{\boldsymbol{\lambda}}_{k-1|k-1}$ , matrica kovarijansi vektora stanja  $\mathbf{P}_{k-1|k-1}$  u trenutku  $k-1|k-1$  i željena izlazna vrednost kao primer iz skupa za optimizaciju  $\mathbf{y}(k)$ . Može se primetiti da u ovakvoj postavci Kalmanovog filtra željena vrednost  $\mathbf{y}(k)$  iz skupa za optimizaciju predstavlja „merenje”. Dakle, za zadate početne uslove i usvojenu očekivanu grešku optimizacije  $Er_{\min}$  algoritam rekurzivno ocenjuje vektor

stanja shodno informaciji  $\mathbf{y}(k)$ . U redu broj 5 i redu broj 6 se vrši predikcija vektora stanja  $\hat{\boldsymbol{\lambda}}_{k|k-1}$  i matrice kovarijansi  $\mathbf{P}_{k|k-1}$ . Red br. 7 definiše predviđanje merenja, što predstavlja izlazni vektor veštačke neuronske mreže. S obzirom na nelinearnost senzorskog modela, neophodno je izvršiti linearizaciju razvojem u Tejlorov red. Jakobijan veštačke neuronske mreže  $\hat{\mathbf{y}}_k = \mathbf{g}(\hat{\boldsymbol{\lambda}}_{k|k-1}, \mathbf{x}(k))$  u odnosu na elemente vektora stanja  $\hat{\boldsymbol{\lambda}}$  je dat u redu br. 8. Matrica pojačanja filtra je data u redu br. 9, dok se korak korekcije sprovodi u redu br. 10 za vektor stanja sistema i u redu br.11 za matricu kovarijansi. Vektor inovacije  $(\mathbf{y}(k) - \hat{\mathbf{y}}_k)$  predstavlja razliku između stvarne izlazne vrednosti  $\mathbf{y}(k)$  i predviđanja veštačke neuronske mreže  $\hat{\mathbf{y}}_k = \mathbf{g}(\hat{\boldsymbol{\lambda}}_{k|k-1}, \mathbf{x}(k))$ , koje je funkcija ocene vektora stanja sistema  $\hat{\boldsymbol{\lambda}}_{k|k-1}$  i ulaznog vektora  $\mathbf{x}(k)$ . Vremenski indeks  $k$  označava da se u jednoj iteraciji koristi jedan i samo jedan element iz skupa ulazno-izlaznih parova koji se koriste za optimizaciju parametara. Nakon sprovedene korekcije, izračunava se nova vrednost greške optimizacije za novu ocenu vektora stanja  $\mathbf{g}(\hat{\boldsymbol{\lambda}}_{k|k}, \mathbf{x})$  i svih ulaznih vektora iz skupa za optimizaciju. Obratiti pažnju da vrednost greške  $Er$  predstavlja kvadratnu formu formiranu od svih ulaznih primera i svih odziva mreže.

### 3.2. Mašinsko učenje primenom linearizovanog informacionog filtra

Za razliku od Kalmanovog filtra koji Gausovu raspodelu modelira prvim i drugim momentom, informacioni filter u rekurzivnu estimacionu formu uvodi inverziju matrice kovarijansi  $\mathbf{P}$  koja se naziva informaciona matrica i definiše se kao:

$$\mathbf{I} = \mathbf{P}^{-1} \quad (14)$$

Matrica  $\mathbf{P}$  predstavlja meru neodređenosti stanja sistema, a velike vrednosti nam ukazuju na to da postoji značajna neodređenost u pogledu ocene vektora stanja. Ukoliko  $\mathbf{P} \rightarrow \infty$  to nam govori da nemamo nikakvo predznanje o sistemu. Naravno, kada  $\mathbf{P} \rightarrow \mathbf{0}$  imamo „idealnu” situaciju, znamo sve potrebne informacije o sistemu. Problem koji se javlja prilikom ocenjivanja u slučaju kada nemamo dovoljno informacija o sistemu je kada  $\mathbf{P} \rightarrow \infty$ . U praktičnim primenama Kalmanovog filtra nemoguće je računaru predočiti ovu situaciju na optimalan način. Zbog toga se uvodi ocenjivanje u informacionom prostoru, koji podrazumeva informacioni vektor i informacionu matricu. Inverzijom matrice kovarijansi moguće je

računaru „predstaviti” količinu našeg znanja/neznanja o sistemu na sledeći način:

$$\begin{aligned} \mathbf{P} = \mathbf{0} &\Rightarrow \mathbf{I} = \infty \\ \mathbf{P} = \infty &\Rightarrow \mathbf{I} = \mathbf{0} \end{aligned} \quad (15)$$

Dakle, imajući u vidu (15), slučaj kada je naše neznanje o sistemu veliko ( $\mathbf{P} \rightarrow \infty$ ) prelaskom u informacioni prostor se jednostavno algebarski definiše kao  $\mathbf{I} = \mathbf{0}$ . Linearni dinamički sistem u diskretnom obliku je formulisan jednačinom (13). Osnovne pretpostavke koje su usvojene tom prilikom važe i u ovom slučaju. Pseudokod linearizovanog informacionog filtra je dat u Algoritmu 2. Inicijalizacija algoritma započinje u redu br. 2 i redu br. 3, gde se određuju početne vrednosti vektora stanja  $\hat{\boldsymbol{\lambda}}_{0|0}$  i početne matrice kovarijansi  $\mathbf{P}_{0|0}$ , kao i kovarijansne šuma sistema  $\mathbf{Q}$  i šuma merenja  $\mathbf{R}$ . Predikcija stanja sistema se obavlja u redovima br. 5 i br. 6. Predikcija vektora stanja je identična kao i prilikom primene algoritma linearizovanog Kalmanovog filtra. Prva razlika se može videti u redu broj 6, gde se predviđa nova vrednost informacione matrice. Treba obratiti pažnju da ova operacija podrazumeva dvostruku inverziju odgovarajućih matrica. Jakobijan veštačke neuronske mreže je dat u redu broj 8. Korak korekcije algoritma započinje određivanjem informacione matrice  $\mathbf{I}_{k|k}$  u trenutku  $k|k$  i pojačanja filtra  $\mathbf{K}_k$ . Na kraju, u redu br. 11 se vrši modifikacija vektora stanja sistema na osnovu vektora inovacije  $(\mathbf{y}(k) - \hat{\mathbf{y}}_k)$ . Određivanje trenutne greške učenja je dato u redu broj 12.

Algoritam 2 - Pseudokod linearizovanog informacionog filtra za optimizaciju parametara veštačke neuronske mreže sa radijalnim aktivacionim funkcijama Gausovog tipa.

```

1. Algoritam LIF_RGAF ( $\hat{\boldsymbol{\lambda}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}, \mathbf{x}(k), \mathbf{y}(k)$ )
2.  $E[\boldsymbol{\lambda}_{0|0}] = \hat{\boldsymbol{\lambda}}_{0|0}; \left\{ E[(\boldsymbol{\lambda}_{0|0} - \hat{\boldsymbol{\lambda}}_{0|0})(\boldsymbol{\lambda}_{0|0} - \hat{\boldsymbol{\lambda}}_{0|0})^T] \right\}^{-1} = \mathbf{I}_{0|0}$ 
3.  $\mathbf{Q} = \text{diag}(\text{dim}(\boldsymbol{\lambda}))$ ;  $\mathbf{R} = \text{diag}(\text{dim}(\mathbf{y}_k))$ 
4. while  $Er > Er_{\min}$ 
5.  $\hat{\boldsymbol{\lambda}}_{k|k-1} = \hat{\boldsymbol{\lambda}}_{k-1|k-1}$ 
6.  $\mathbf{I}_{k|k-1} = ((\mathbf{I}_{k-1|k-1})^{-1} + \mathbf{Q})^{-1}$ 
7.  $\hat{\mathbf{y}}_k = \mathbf{g}(\hat{\boldsymbol{\lambda}}_{k|k-1}, \mathbf{x}(k))$ 
8.  $\mathbf{H}_k = \nabla_{\boldsymbol{\lambda}} \hat{\mathbf{y}}_k = [\nabla_{\mathbf{v}} \mathbf{g}(\cdot) \quad \nabla_{\boldsymbol{\mu}} \mathbf{g}(\cdot) \quad \nabla_{\boldsymbol{\sigma}} \mathbf{g}(\cdot)]$ 
9.  $\mathbf{I}_{k|k} = \mathbf{I}_{k|k-1} + \mathbf{H}_k^T \mathbf{R}^{-1} \mathbf{H}_k$ 
10.  $\mathbf{K}_k = (\mathbf{I}_{k|k})^{-1} \mathbf{H}_k^T \mathbf{R}^{-1}$ 
11.  $\hat{\boldsymbol{\lambda}}_{k|k} = \hat{\boldsymbol{\lambda}}_{k|k-1} + \mathbf{K}_k (\mathbf{y}(k) - \hat{\mathbf{y}}_k)$ 
12.  $Er = \frac{1}{2} (\mathbf{y} - \mathbf{g}(\hat{\boldsymbol{\lambda}}_{k|k}, \mathbf{x}))^T (\mathbf{y} - \mathbf{g}(\hat{\boldsymbol{\lambda}}_{k|k}, \mathbf{x}))$ 
13. endwhile
14. return  $\hat{\boldsymbol{\lambda}}_{k|k}, \mathbf{I}_{k|k}$ 

```

### 3.3. Mašinsko učenje primenom algoritma specifične aproksimacije momenata

Linearizacija nelinearnih jednačina modela sistema i senzorskog modela razvijanjem u Tejlorov red rezultira uvođenjem grešaka u proces ocenjivanja, s obzirom na to da odgovarajući Jakobijani ne mogu sa dovoljnom tačnošću da modeliraju promenu stanja nelinearnog sistema. Ove greške dovode do narušavanja optimalnosti (time i tačnosti) i mogu izazvati divergenciju filtra. Algoritam specifične aproksimacije momenata (ASAM) ne vrši linearizaciju nelinearnog sistema razvijanjem jednačina modela sistema i senzorskog modela u Tejlorov red. Osnovna ideja ASAM-a počiva na intuiciji da je jednostavnije aproksimirati poznatu gustinu raspodele stohastičke promenljive nego polazne nelinearne modele [1]. Algoritam počiva na specifičnoj transformaciji nelinearne funkcije stohastičke promenljive.

Neka je data proizvoljna nelinearna funkcija  $\mathbf{y}=\mathbf{f}(\mathbf{x})$  i neka argument funkcije  $\mathbf{x}$  podleže Gausovoj raspodeli sa poznatim statistikama  $\bar{\mathbf{x}}$  i  $\mathbf{P}$ . Potrebno je izračunati nepoznate statistike promenljive  $\mathbf{y}$  nakon što je primenjena nelinearna funkcija  $\mathbf{f}(\cdot)$  na argument  $\mathbf{x}$ . Prvo je potrebno formirati matricu sigma tačaka  $\Sigma$  na sledeći način:

$$\begin{aligned}\Sigma_0 &= \bar{\mathbf{x}} \\ \Sigma_i &= \bar{\mathbf{x}} + \left( \sqrt{(\dim(\mathbf{x}) + \gamma) \mathbf{P}} \right)_i, \\ &\quad i = 1, \dots, \dim(\mathbf{x}) \\ \Sigma_i &= \bar{\mathbf{x}} - \left( \sqrt{(\dim(\mathbf{x}) + \gamma) \mathbf{P}} \right)_{i - \dim(\mathbf{x})}, \\ &\quad i = \dim(\mathbf{x}) + 1, \dots, 2 \cdot \dim(\mathbf{x})\end{aligned}\quad (16)$$

$\gamma$  je skalarni parametar koji određuje širinu opsega tačaka  $\Sigma_i$  i izračunava se kao

$$\gamma = \alpha^2 (\dim(\mathbf{x}) + \kappa) - \dim(\mathbf{x}) \quad (17)$$

Ostali skalarni parametri su  $\alpha \in [10^{-4}, 1]$ ;  $\kappa \in \{0, \dim(\mathbf{x})\}$ . Vektori kolone pomoću kojih se vrši aproksimacija nelinearne funkcije nazivaju se sigma tačke. Dakle, matricu  $\Sigma$  čini  $2 \dim(\mathbf{x}) + 1$  sigma tačaka. Nakon formiranja matrice sigma tačaka  $\Sigma_i$ , na njene vektor kolone se primenjuje nelinearna funkcija  $\mathbf{f}(\cdot)$ , odnosno:

$$\Psi_i = \mathbf{f}(\Sigma_i), \quad i = 0, \dots, 2 \dim(\mathbf{x}) \quad (18)$$

Matrica  $\Psi_i$  je matrica svih mogućih vrednosti koje promenljiva  $\mathbf{y}=\mathbf{f}(\mathbf{x})$  može da ima na osnovu

prethodnog znanja da promenljiva  $\mathbf{x}$  podleže Gausovoj raspodeli sa očekivanom vrednošću  $\bar{\mathbf{x}}$  i kovarijansom  $\mathbf{P}$ .

Očekivana vrednost (prvi moment) i kovarijansa (drugi moment) promenljive  $\mathbf{y}$  su:

$$\begin{aligned}\bar{\mathbf{y}} &\approx \sum_{i=0}^{2 \dim(\mathbf{x})} W_i^m \Psi_i \\ \mathbf{P} &\approx \sum_{i=0}^{2 \dim(\mathbf{x})} W_i^k (\Psi_i - \bar{\mathbf{y}})(\Psi_i - \bar{\mathbf{y}})^T\end{aligned}\quad (19)$$

a konstantni težinski koeficijenti su

$$\begin{aligned}W_0^m &= \gamma / (\dim(\mathbf{x}) + \gamma) \\ W_0^k &= \gamma / (\dim(\mathbf{x}) + \gamma) + (1 - \alpha^2 + \beta) \\ W_i^m &= W_i^k = 1 / (2 (\dim(\mathbf{x}) + \gamma)), \\ &\quad i = 1, \dots, 2 \dim(\mathbf{x})\end{aligned}\quad (20)$$

Optimalna vrednost konstantnog parametra  $\beta$  za Gausovu raspodelu je  $\beta = 2$  [1]. Prikazana aproksimacija momenata promenljive  $\mathbf{y}$ , koja je nelinearna funkcija nekih drugih stohastičkih promenljivih, višeg je reda tačnosti od linearizacije ostvarene razvijanjem u Tejlorov red.

Nakon uvođenja osnovne transformacije na kojoj ASAM počiva, moguće je pokazati kako se ova ideja može primeniti na problem mašinskog učenja. Algoritam optimizacije baziran je na osnovnoj rekurzivnoj formulaciji linearizovanog Kalmanovog filtra. Međutim, u ASAM algoritmu nema potrebe za izračunavanjem Jakobijana. Algoritam je prikazan u Algoritmu 3.

Algoritam 3 - Pseudokod algoritma ocenjivanja parametara veštačke neuronske mreže sa radijalnim aktivacionim funkcijama Gausovog tipa baziran na Kalmanovom filtru i specifičnoj aproksimaciji momenata.

Algoritam započinje inicijalizacijom u redu br. 2 i redu br. 3, gde se definišu početne vrednosti vektora stanja  $\hat{\lambda}_{0|0}$ , početne matrice kovarijansi  $\mathbf{P}_{0|0}$ , kovarijanse šuma sistema  $\mathbf{Q}$  i šuma merenja  $\mathbf{R}$ . U redovima broj 5 i broj 6 vrši se predikcija ocene vektora stanja  $\hat{\lambda}_{k|k-1}$  i matrice kovarijansi  $\mathbf{P}_{k|k-1}$ .

Prva razlika u odnosu na LKF može se videti u redu broj 7 gde se određuje matrica  $\Sigma_i$ , čije su kolone sigma tačke. Kao što se može videti, u redu br. 8 se određuje izlazna vrednost veštačke neuronske mreže, na osnovu mogućih vrednosti koje mogu imati parametri mreže (vrednosti određene sigma tačkama) i poslednjeg ulaznog vektora  $\mathbf{x}_k$ . U redu br. 9 se od-

ređuje (uslovno) matematičko očekivanje izlazne vrednosti veštačke neuronske mreže na osnovu izračunatih svih izlaznih vrednosti uvođenjem sigma tačaka. Ocenjivanje parametara se nastavlja određivanjem kovarijanse  $\mathbf{P}_{\hat{\mathbf{y}}_k \hat{\mathbf{y}}_k}$  u redu br. 10 i kroskovarijanse  $\mathbf{P}_{\hat{\mathbf{y}}_k \hat{\mathbf{x}}_k}$  u redu broj 11, kao i izračunavanjem Kalmanovog pojačanja  $\mathbf{K}_k$  u redu broj 12.

```

1. Algoritam ASAM_RGAF ( $\hat{\boldsymbol{\lambda}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}, \mathbf{x}(k), \mathbf{y}(k)$ )
2.  $E[\boldsymbol{\lambda}_{0|0}] = \hat{\boldsymbol{\lambda}}_{0|0}; \quad E\left[\left(\boldsymbol{\lambda}_{0|0} - \hat{\boldsymbol{\lambda}}_{0|0}\right)\left(\boldsymbol{\lambda}_{0|0} - \hat{\boldsymbol{\lambda}}_{0|0}\right)^T\right] = \mathbf{P}_{0|0}$ 
3.  $\mathbf{Q} = \text{diag}(\text{dim}(\boldsymbol{\lambda})); \quad \mathbf{R} = \text{diag}(\text{dim}(\mathbf{y}_k))$ 
4. while  $Er > Er_{\min}$ 
5.  $\hat{\boldsymbol{\lambda}}_{k|k-1} = \hat{\boldsymbol{\lambda}}_{k-1|k-1}$ 
6.  $\mathbf{P}_{k|k-1} = \mathbf{P}_{k-1|k-1} + \mathbf{Q}$ 
7.  $\boldsymbol{\Sigma}_i = \begin{bmatrix} \hat{\boldsymbol{\lambda}}_{k|k-1} & \hat{\boldsymbol{\lambda}}_{k|k-1} + \gamma\sqrt{\mathbf{P}} & \hat{\boldsymbol{\lambda}}_{k|k-1} - \gamma\sqrt{\mathbf{P}} \end{bmatrix}$ 
8.  $\boldsymbol{\Psi}_{i,k|k-1} = \mathbf{g}(\boldsymbol{\Sigma}_i, \mathbf{x}(k))$ 
9.  $\hat{\mathbf{y}}_k = \sum_{i=1}^{2\text{dim}(\mathbf{x})} W_i^m \boldsymbol{\Psi}_{i,k|k-1} \approx E\left[\mathbf{g}(\hat{\boldsymbol{\lambda}}_{k|k-1}, \mathbf{x}_k)\right]$ 
10.  $\mathbf{P}_{\hat{\mathbf{y}}_k \hat{\mathbf{y}}_k} = \sum_{i=1}^{2\text{dim}(\mathbf{x})} W_i^k \left(\boldsymbol{\Psi}_{i,k|k-1} - \hat{\mathbf{y}}_k\right)\left(\boldsymbol{\Psi}_{i,k|k-1} - \hat{\mathbf{y}}_k\right)^T + \mathbf{R}$ 
11.  $\mathbf{P}_{\hat{\mathbf{y}}_k \hat{\mathbf{x}}_k} = \sum_{i=1}^{2\text{dim}(\mathbf{x})} W_i^k \left(\boldsymbol{\Sigma}_i - \hat{\boldsymbol{\lambda}}_k\right)\left(\boldsymbol{\Psi}_{i,k|k-1} - \hat{\mathbf{y}}_k\right)^T$ 
12.  $\mathbf{K}_k = \mathbf{P}_{\hat{\mathbf{y}}_k \hat{\mathbf{x}}_k} \left(\mathbf{P}_{\hat{\mathbf{y}}_k \hat{\mathbf{y}}_k}\right)^{-1}$ 
13.  $\hat{\boldsymbol{\lambda}}_{k|k} = \hat{\boldsymbol{\lambda}}_{k|k-1} + \mathbf{K}_k \left(\mathbf{y}(k) - \hat{\mathbf{y}}_k\right)$ 
14.  $\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{P}_{\hat{\mathbf{y}}_k \hat{\mathbf{y}}_k} \mathbf{K}_k^T$ 
15.  $Er = \frac{1}{2} \left(\mathbf{y} - \mathbf{g}(\hat{\boldsymbol{\lambda}}_{k|k}, \mathbf{x})\right)^T \left(\mathbf{y} - \mathbf{g}(\hat{\boldsymbol{\lambda}}_{k|k}, \mathbf{x})\right)$ 
16. endwhile
17. return  $\hat{\boldsymbol{\lambda}}_{k|k}, \mathbf{P}_{k|k}$ 

```

Nakon svih sprovedenih koraka, korekcija vektora stanja i greške ocene je data u redu broj 13 i redu broj 14. Određivanje trenutne vrednosti greške optimizacije je dato u poslednjem redu.

#### 4. DISKUSIJA

Na kraju ove analize bi trebalo izvršiti poređenje parametara razvijenih algoritama mašinskog učenja sa poznatim algoritimima mašinskog učenja [4]. U tim algoritimima se uvode konstantni parametri poznatiji kao parametri učenja pomoću kojih se kontroliše brzina optimizacije. Analogno ovim parametrima, odgovarajuće matrice Kalmanovog filtra imaju sličnu ulogu. Matrica šuma sistema  $\mathbf{Q}$  direktno utiče na ocenu kovarijanse sistema. Povećavanjem matrice šuma sistema možemo kontrolisati dužinu inkrementalnih koraka, odnosno brzinu konvergencije filtra. Pove-

ćanje matrice šuma merenja  $\mathbf{R}$  direktno utiče na „poverenje” filtra u merenje. Ukoliko je  $\mathbf{R}$  veliko, filter će više verovati koraku predikcije; ukoliko je  $\mathbf{R}$  malo, onda će filter „imati poverenja” u merenje  $\mathbf{y}(k)$ .

Kao i svi iterativni postupci određivanja minimuma funkcije cilja, i Kalmanov filter može odrediti jedan od mogućih višestrukih lokalnih minimuma umesto globalnog minimuma. Međutim, Kalmanov filter je algoritam drugog reda tačnosti, s obzirom na to da uzima u obzir i lokalnu zakrivljenost funkcije cilja. Ova činjenica je od važnosti za izbegavanje lokalnog minimuma.

U drugom delu ovog rada prikazani su eksperimentalni rezultati ostvareni primenom veštačke neuronske mreže sa radijalnim aktivacionim funkcijama Gausovog tipa, čiji parametri su određeni pomoću uvedenih sekvencijalnih algoritama mašinskog učenja na bazi Kalmanovog filtra, za modeliranje četiri realna problema iz inženjerske prakse.

#### ZAHVALNOST

Rad je nastao u okviru istraživanja na projektu „Inovativni pristup u primeni inteligentnih tehnoloških sistema za proizvodnju delova od lima zasnovan na ekološkim principima“ (TR-35004), koji je podržan od strane Ministarstva prosvete, nauke i tehnološkog razvoja Vlade Republike Srbije.

#### LITERATURA

- [1] Vuković, N., Razvoj mašinskog učenja inteligentnog mobilnog robota baziran na sistemu veštačkih neuronskih mreža, doktorska disertacija, Univerzitet u Beogradu - Mašinski fakultet, 2012. (DOI: 10.2298/BG20120928VUKOVIC)
- [2] Vuković, N., Miljković, Z., A Growing and Pruning Sequential Learning Algorithm of Hyper Basis Function Neural Network for Function Approximation, Neural Networks, 46C, pp. 210-226, 2013. (DOI: 10.1016/j.neunet.2013.06.004)
- [3] Miljković, Z., Sistemi veštačkih neuronskih mreža u proizvodnim tehnologijama, Univerzitet u Beogradu - Mašinski fakultet, Beograd 2003.
- [4] Miljković, Z., Aleksendrić, D., Veštačke neuronske mreže - zbirka rešenih zadataka sa izvodima iz teorije, Univerzitet u Beogradu - Mašinski fakultet, Beograd 2009.
- [5] Vuković, N., Miljković, Z., Babić, B., Bojović, B., Training of Radial Basis Function Networks with  $H_{\infty}$  Filter-Initial Simulation Results, Proceedings of the 6th International Working Conference "Total Quality Management - Advanced and Intelligent Approaches", pp. 163-168, Belgrade, Serbia, 2011.

- [6] Lazaro, M., Santamari I., Pantaleo C., A new EM-based training algorithm for RBF networks, *Neural Networks*, 16, pp. 69-77, 2003.
- [7] Schwenker, F., Kestler, H.A., Palm, G., Three learning phases for radial-basis-function networks. *Neural Networks*, 1(4-5), pp. 439-58, 2001.
- [8] Vuković, N., Miljković, Z., Babić, B., Čović, N., Towards implementation of intelligent mobile robots in a manufacturing environment, *Proceedings of the 4th International Conference on Manufacturing Engineering (ICMEN 2011)*, pp 367-376, Thessaloniki, Greece, 2011.
- [9] Karayiannis, N. B., Reformulated Radial Basis Neural Networks Trained by Gradient Descent, *IEEE Transactions on Neural Networks*, 10(3), pp. 657-671, 1999.
- [10] Simon, D, Training Radial Basis Function Neural Networks with the Extended Kalman Filter, *Neurocomputing*, 48, pp. 455-475, 2001.
- [11] Huang, G. B., Saratchandran, P., Sundararajan, N., A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation. *IEEE Transactions on Neural Networks*, 16(1), pp. 57-67, 2005.
- [12] Bortman, M., Aladjem M., A Growing and Pruning Method for Radial Basis Function Networks, *IEEE Transactions on Neural Networks*, 20(6), pp. 1039-1045, 2009.
- [13] Han, H. G., Chen, Q. L., Qiao, J. F., An efficient self-organizing RBF neural network for water quality prediction. *Neural Networks*, 24(7), pp. 717-725, 2011.

## SUMMARY

### MACHINE LEARNING OF RADIAL BASIS FUNCTION NEURAL NETWORK BASED ON KALMAN FILTER - INTRODUCTION

*This paper analyzes machine learning of radial basis function neural network based on Kalman filtering. Three algorithms are derived: linearized Kalman filter, linearized information filter and unscented Kalman filter. We emphasize basic properties of these estimation algorithms, demonstrate how their advantages can be used for optimization of network parameters, derive mathematical models and show how they can be applied to model problems in engineering practice.*

**Key words:** machine learning, artificial neural network, Kalman filter