

Tirpalų mišinių koncentracijų klasifikavimas*

Romas BARONAS (VU), Feliksas IVANAUSKAS (VU, MII),
Romualdas MASLOVSKIS (VU), Pranas VAITKUS (VU)

el. paštas: feliksas.ivanauskasmaf.vu.lt

Reziumė. Darbe aprašytas biojutiklio signalo klasifikatoriaus modelis. Sukonstruotas klasifikatorius pritaikytas duomenims, gautiems eksperimentų metu ir modeliuojant biojutiklio veiksmą su kompiuteriu.

Raktiniai žodžiai: neuroniniai tinklai, pagrindinių komponentų analizė, nepriklausomų komponentų analizė.

1. Įvadas

Biojutikliai – tai įrenginiai, kuriuos sudaro biologiškai aktyvi medžiaga, dažniausiai fermentas, ir elektroninis signalo keitiklis. Fermentui reaguojant su analizuojamu tirpalu, fizikiniai-cheminiai pasikeitimai yra paverčiami elektros signalu, kurio stiprumas priklauso nuo tirpalo koncentracijos.

Šio darbo tikslas sukonstruoti klasifikatorių, kuris klasifikuotų biojutiklio signalą. Darbe nagrinėjami duomenys gauti modeliuojant amperometriniu biojutiklio veiksmą [1] ir eksperimento metu. Nagrinėjant eksperimentinius duomenis buvo sukonstruoti klasifikatoriai, pagal biojutiklio signalą nustatantys tirpalų koncentracijų reikšmes ir galintys atskirti ar tirpalų koncentracijų reikšmės viršija leistiną ribą. Sumodeliuotų duomenų atveju sukonstruotas prediktorius, nustatantis tirpalų koncentracijų reikšmes.

Klasifikatoriams sukonstruoti buvo naudojami neuroniniai tinklai. Neuroninio tinklo įėjimams buvo pritaikytas pagrindinių komponentų metodas, o keliais atvejais ir nepriklausomų komponentų analizė. Sukonstruoti klasifikatoriai pasižymi aukštu tikslumu.

2. Matematinis modelis

Sumodeliuoti duomenys

Tegul $\vec{c} = (c_1, \dots, c_K)$ yra K tirpalų koncentracijų vektorius, $\vec{z} = \vec{z}(\vec{c}) = (z_1(\vec{c}), \dots, z_N(\vec{c}))$ yra biojutiklio signalas momentais t_1, \dots, t_N . \vec{z} yra biojutiklio atsakas į mišinį $\vec{c} = (c_1, \dots, c_K)$. Pažymėkime, $C = \{\vec{c}\}$ – aibė visų tirpalų galimų koncentracijų vektorių ir $Z = \{\vec{z}(\vec{c})\}$ – stebėtų biojutiklio signalų aibė. Aibė Z suskaidoma į dvi dalis: apmokymo ir testinę imtis. Naudojant apmokymo aibės elementus apskaičiuojamos l pagrindinių komponentų, atitinkančių l didžiausių tikrinių reikšmių.

*Darbą parėmė Lietuvos valstybinis mokslo ir studijų fondas, proektas Nr. C-03048

Pasirinktas prediktorius yra trijų sluoksnių neuroninis tinklas (įvesties sluoksnis, vienas paslėptasis sluoksnis ir išvesties sluoksnis):

$$c_k = \sum_{s=1}^p \alpha_{sk} \phi(\langle \vec{\beta}_s^T, \vec{X} \rangle + \gamma_s) + \epsilon_k, \quad k = 1, \dots, K, \quad (2.1)$$

kur c_k – tirpalų koncentracijų vektoriaus \vec{c} k -toji komponentė, $\alpha_{s,k}, \vec{\beta}_s$ – svoriai, γ_s, ϵ_k – koeficientai, \vec{X} – neuroninio tinklo įvestis (pagrindinių komponentių vektorius), ϕ – sigmoidinė aktyvacijos funkcija, t.y. $\phi(u) = \frac{1}{1+\exp(-u)}$, p -neuronų skaičius paslėptajame sluoksnyje.

Apmokymo metu, naudojant Levenberg–Marquardt algoritmą, yra įvertinami parametrai $\alpha_{s,k}, \vec{\beta}_s, \gamma_s, \epsilon_k$.

Neuroninio tinklo klasifikavimo tikslumui įvertinti pasirinktas dydis

$$Q_k = \frac{1}{T} \sum_{i=1}^T \sum_y \text{Ind}(O_{i,k} \in \Delta y) \cdot \text{Ind}(c_{i,k} = y), \quad k = 1, \dots, K, \quad (2.2)$$

kur $\text{Ind}(\cdot)$ – indikatorius, Δy – intervalas $(y - \delta_{1,y}, y + \delta_{2,y})$, y – koncentracijos reikšmė, T – apmokymo arba testinės imties stebėjimų skaičius, O – neuroninio tinklo išvestis.

Eksperimentiniai duomenys

Tarkime $\vec{z}_{ji}(\vec{c})$, $j = 1, \dots, J$, $i = 1, \dots, I$, j -tojo biojutiklio signalo i -tasis matavimas. Kiekvienam tirpalų koncentracijos vektoriui \vec{c} ir kiekvienam biojutiklio numeriui j atsitiktinai generuojamas skaičius i^* iš aibės $\{1, \dots, I\}$. Vektoriai $\vec{z}_{ji^*}(\vec{c})$ sudaro testinę aibę. Naudojant apmokymo aibę apskaičiuojamos l pagrindinių komponentių. Kaip ir sumodeliuotų duomenų atveju, šią užduotį spręsti pasirinktas trijų sluoksnių neuroninis tinklas

$$c_k = \sum_{s=1}^p \alpha_{sk} \phi(\langle \vec{\beta}_s^T, \vec{X}^* \rangle + \gamma_s) + \epsilon_k, \quad k = 1, \dots, K, \quad (2.3)$$

kur $\vec{X}^* = (X_1, \dots, X_l, j)$ – neuroninio tinklo įvestis, susidedantis iš l pagrindinių komponentių ir biojutiklio numerio, c_k – tirpalų koncentracijų vektoriaus \vec{c} k -toji komponentė. Klasifikavimo tikslumas matuojamas naudojant dydį (2.2).

Sudarome klasifikatorių galintį atskirti du atvejus:

- 1) bent viena koncentracijos reikšmė viršija nustatytą kritinę reikšmę;
- 2) visos koncentracijų reikšmės mažesnes už nustatytas kritines reikšmes.

Duomenys skirstomi į dvi aibes naudojant taisyklę:

- 1) parenkamas kritinių reikšmių vektorius $\vec{c}^* = (c_1^*, \dots, c_K^*)$;
- 2) jei bent vienam k $c_k \geq c_k^*$, tai $\vec{z}_{ji}(\vec{c})$ priskiriamas aibei A , $j = 1, \dots, J$, $i = 1, \dots, I$;
- 3) jei visiems k $c_k < c_k^*$, tai $\vec{z}_{ji}(\vec{c})$ priskiriamas aibei B , $j = 1, \dots, J$, $i = 1, \dots, I$.

Apmokymo ir testinė imtis sudarinėjamos kaip ir konstruojant tirpalų koncentracijų atpažinimo klasifikatorių. Klasifikatoriaus modelis yra

$$D = \sum_{s=1}^p \alpha_{sk} \phi(\langle \vec{\beta}_s^T, \vec{X}^* \rangle + \gamma_s) + \epsilon_k, \quad k = 1, \dots, K,$$

kur $\vec{X}^* = (X_1, \dots, X_l, j)$ – neuroninio tinklo įėjis, susidedantis iš l pagrindinių komponentių ir biojutiklio numerio, D – įgyja reikšmes iš aibės $\{-1, 1\}$ (-1 kai $\vec{z}(\vec{c})$ priklauso aibei A ir 1 kai $\vec{z}(\vec{c})$ priklauso aibei B).

3. Skaičiavimų rezultatai

Sumodeliuotų duomenų analizė

Naudojant modelį pateiktą [1], buvo modeliuojamas biojutiklio atsakas į keturių ($K = 4$) tirpalų mišinį, vonios (BA) ir apipurškimo (FIA) režimais. Vonios režimu biojutiklio veikimas modeliuotas dviem membranos storiams: $d = 0,02$ ir $d = 0,05$ cm. Apipurškimo režime membranos storis buvo $d = 0,02$ cm. Naudotos aštuonios kiekvieno tirpalo koncentracijos: 1, 2, 4, 8, 12, 16, 32, 64. Iš viso buvo gauta 4096 skirtingų tirpalų koncentracijų vektorių. Biojutiklio signalo modeliavimo laikas: BA atveju ($d = 0,02$) 300 sekundžių ($N = 300$), BA atveju ($d = 0,05$) 500 sekundžių ($N = 500$) ir FIA atveju 150 sekundžių ($N = 150$).

Sudarant testinę aibę atsitiktinai be grąžinimo išrinkti 2000 tirpalų koncentracijų vektorių, likusieji 2096 vektoriai sudarė apmokymo aibę. Bandymų būdu buvo sukonstruoti neuroniniai tinklai: BA ($d = 0,02$) atveju su dvylika neuronų paslėptajame sluoksnyje, FIA ir BA ($d = 0,05$) naudojami aštuonių neuronų paslėpti sluoksniai. Visais atvejais neuroninių tinklų įvesties sluoksnius sudarė šeši neuronai (šešios ($l = 6$) pagrindinės komponentės). Įvertinus neuroninių tinklų svorius ir pasirinkus tikslumo intervalus Δ (pateikti 1 lentelėje) buvo apskaičiuotas klasifikavimo tikslumas naudojant dydį (2.2). Klasifikavimo tikslumas pateiktas 2 lentelėje.

Pritaikius neuroninio tinklo įėjimams nepriklausomų komponentių analizę [2], BA ($d = 0,02$) atveju sumažėjo tikslumas klasifikuojant trečią tirpalą, bet daugiau nei 4 procentais padidėjo tikslumas klasifikuojant apmokymo aibės ketvirtąjį tirpalą ir daugiau nei 6 procentais padidėjo tikslumas klasifikuojant testinės aibės ketvirtą tirpalą. Tikslumui vertinti naudojant Δ_2 intervalus, žymiai padidėjo ketvirtojo tirpalo klasifikavimo tikslumas. Rezultatai pateikti 3 lentelėje.

1 lentelė. Tikslumo intervalai Δ

y	1	2	4	8	12	16	32	64
Δ_1	$< 1,5$	[1,5;3)	[3;6)	[6;10)	[10;14)	[14;24)	[24;48)	≥ 48
Δ_2	[0;1,5)	[1,5;2,9)	[3,1;5)	[7;9)	[11;13)	[15;17)	[31;33)	[63;65)

2 lentelė. Klasifikavimo tikslumas vonios ir apipurškimo režimais, naudojant intervalus Δ_1 ir Δ_2

Tirpalas	BA, $\Delta_1, d = 0,02$		BA, $\Delta_2, d = 0,02$		BA, $\Delta_2, d = 0,02$		FIA, $\Delta_2, d = 0,02$	
	Apm. aibė	Test. aibė	Apm. aibė	Test. aibė	Apm. aibė	Test. aibė	Apm. aibė	Test. aibė
1	100	100	100	100	99,9	100	100	100
2	100	100	100	100	99,8	99,8	100	100
3	99,76	99,6	92,56	96	99,9	100	100	100
4	87,97	86,95	61,16	57,4	99,95	99,9	99,85	99,75

3 lentelė. Klasifikavimo tikslumas vonios ($d = 0,02$) režimu, naudojant intervalus Δ_1 ir Δ_2 , pritaikius neuroninio tinklo įėjimams nepriklausomų komponentių analizę

Tirpalas	BA, $\Delta_1, d = 0,02$		BA, $\Delta_2, d = 0,05$	
	Apm. aibė	Test. aibė	Apm. aibė	Test. aibė
1	100	100	100	100
2	100	100	100	100
3	97,23	98,05	95,61	96,65
4	92,94	93,45	80,25	79,55

Eksperimentinių duomenų analizė

Pirmo eksperimento metu buvo dirbama su vienu ($K = 1$) tirpalu: katecholiu. Galimos katecholio koncentracijos: 0,2, 0,5, 1, 5, 10, 20, 40, 60, 100. Eksperimente buvo naudojami septyni ($J = 7$) biojutikliai, kurie veikė vonios režime. Dalyje bandymų papildomai buvo naudojama gliukozė. Kiekvienai katecholio reikšmei buvo atlikti trys ($I = 3$) pakartotiniai matavimai. Sudarant testinę aibę, kiekvienam biojutikliui buvo atsitiktinai išrinktas vienas iš trijų pakartotinių matavimų. Likusieji matavimai sudarė apmokymo aibę. Naudojant apmokymo aibę buvo apskaičiuotos šešios pagrindinės komponentės. Bandymų būdu buvo parinktas neuroninis tinklas su architektūra 8–6–1 (aštuoni neuronai įvesties sluoksnyje, šeši neuronai paslėptajame sluoksnyje ir vienas neuronas išvesties sluoksnyje). Įvestį sudarė: šešios pagrindinės komponentės, biojutiklio numeris ir gliukozės reikšmė (1 jei buvo naudojama gliukozė, 0 jei ne). Neuroninio tinklo išvestis: katecholio koncentracija. Vertinant klasifikavimo tikslumą buvo naudojamas (2.2) dydis. Intervalai pateikti 4 lentelėje.

4 lentelė. Tikslumo intervalai Δ

y	0,2	0,5	1	5	10	20	40	60	100
Δ	< 0,3	[0,3; 0,75)	[0,75; 2,5)	[2,5; 7,5)	[7,5; 15)	[15; 30)	[30; 50)	[50; 80)	≥ 80

Vertinant klasifikavimo tikslumą, naudojant (2.2) parametą, gauti tokie rezultatai: apmokymo aibė – $Q = 91,09$, testinė aibė – $Q = 91,07$.

Antro eksperimento metu šeši biojutikliai ($J = 6$) dirbo vonios režime. Buvo naudojami du tirpalai ($K = 2$): katecholis ir p-kresolis. Galimos tirpalų koncentracijos: 0, 10, 20, 30, 40, 50. Pasirinkus koncentracijų kritines reikšmes $\vec{c}^* = (30, 30)$ duomenys buvo suskirstyti į dvi aibes A ir B. Naudojant tris pagrindines komponentes ir biojutiklio numerį buvo sukonstruotas modelis, galintis klasifikuoti šias dvi aibes. Klasifikavimo tikslumui vertinti buvo naudojamas sąlyginės tikimybės empirinis analogas. Gauti tokie rezultatai: apmokymo aibė – $\hat{P}(O \in B | \vec{z}(\vec{c}) \in B) = 0,979$, $\hat{P}(O \in A | \vec{z}(\vec{c}) \in A) = 0,916$, testinė aibė – $\hat{P}(O \in B | \vec{z}(\vec{c}) \in B) = 0,948$, $\hat{P}(O \in A | \vec{z}(\vec{c}) \in A) = 0,888$, kur O – neuroninio tinklo išvestis.

4. Išvados

Sukonstruotą klasifikatorių galima sėkmingai taikyti klasifikuojant biojutiklio signalą net ir tada, kai eksperimente naudojami keli biojutikliai. Jis leidžia nustatyti tirpalų koncentracijų reikšmes ir gali būti taikomas nustatant, ar tirpalų koncentracijų reikšmės viršija leistiną ribą.

Literatūra

1. R. Baronas, J. Christensen, F. Ivanauskas, J. Kulys, Computer simulation of amperometric biosensor response to mixtures of compounds, *Nonlinear Analysis: Modelling and Control*, **7**(2), 3–14 (2002).
2. E.G. Learned-Miller, J.W. Fisher III, ICA using spacings estimates of entropy, *Journal of Machine Learning Research*, **4**, 1271–1295 (2003).

SUMMARY

R. Baronas, F. Ivanauskas, R. Maslovskis, P. Vaitkus. The classification of concentration of mixture of analytes

This paper presents a system which is used for the classification of biosensor signals. The proposed system is applied to the synthesized and experimental data. The developed system showed good prediction performance.

Keywords: neural networks, principal components analysis, independent components analysis.