

UOT 004.93

İsmayilov E.Ə.

Azərbaycan Dövlət Neft və Sənaye Universiteti, Bakı, Azərbaycan
elviz.ismayilov@gmail.com

AZƏRBAYCAN DİLİNDƏ ÇAP ƏLYAZMA SİMVOLLARININ TANINMASINA YENİ ƏLAMƏTLƏR SİNFİ VƏ DAYAQ VEKTORLAR ÜSULUNUN TƏTBİQİ

Əlifbamızda bütün dünyada geniş yayılmış latın hərflərindən daha çox istifadə olunsada xüsusi simvollar və dilimizin morfoloji tərkibi digər dillər üçün mövcud tanıma sistemlərinin tətbiqini qeyri-mümkün edir. Başqa sözlə, Azərbaycan dilində çap əlyazma simvollarının tanınması fərdi yanaşma tələb edir. Məqalədə Azərbaycan dilində çap əlyazma simvollarının tanınmasında “soft” əlamətlər (insanın daha asan anlaya biləcəyi, əlifbanı öyrənərkən istifadə etdiyimiz xarakteristikalar əsasında tərtib olunmuş) sinfi istifadə olunmuş, simvolların klasifikasiyası üçün dayaq vektorlar üsulunun bootstrap resampling üsulundan istifadə olunmuşdur.

Açar sözlər: “soft” əlamətlər, Dayaq Vektorlar Üsulu, çap-əlyazma simvolları, tanıma sistemi.

Giriş

Süni intellektin əsas istiqamətlərindən biri olan süni neyron şəbəkələr klasifikasiya üsulu kimi tanıma sistemlərinin qurulmasında çox geniş istifadə olunur. Lakin müxtəlif xarakterli tanıma məsələlərinin həlli və sistemlərin istismarı zamanı dəqiqliyi artırmaq üçün neyron şəbəkənin strukturunu və hesablama alqoritmlərini dəyişdirmək tələb olunur ki, bu da sistemin qurulmasını kifayət qədər mürəkkəbləşdirir. Bundan əlavə neyron şəbəkələrin tətbiq olunmasının əsas problemi öyrətmə zamanı arzu olunan nəticənin alındığı halda, sistemin istismarı zamanı eyni nəticələrin əldə edilməməsidir. Qeyd edilən problemi həll etmək məqsədi ilə V.N. Vapnik tərəfindən struktur risk minimallaşdırması nəzəriyyəsi əsasında hazırlanmış, öyrənmə və tanımanın uzlaşması üçün riyazi həll yollarını özündə cəmləşdirən Dayaq Vektorları Üsulu - DVÜ (Support Vector Machine - SVM) təklif edildi [1].

Respublikamızda müxtəlif alimlər tərəfindən Azərbaycan dilində çap, əlyazma və çap əlyazma mətnlərinin tanınması üçün müxtəlif üsullardan (süni neyron şəbəkələr, qeyri-səlis çoxluqlar nəzəriyyəsi, sadə, DVÜ və s.) istifadə olunmuşdur [2-4]. Bütün bu üsulların tətbiqi zamanı fərqli əlamət sinflərindən istifadə edilmiş, yeni yaradılan hər bir sistemin digərindən üstünlüyü eksperimentlər vasitəsilə əsaslandırılmışdır. Lakin buna baxmayaraq Azərbaycan dilində çap əlyazma simvollarının tanınması məsələsini tam olaraq həll olunmuş hesab etmək olmaz. Müxtəlif üsulların və ya yeni əlamətlər sinflərinin tətbiq olunması ilə tanıma sistemlərinin dəqiqliyini artırmaq cəhdi böyük əhəmiyyətə malikdir. Bu məqsədlə məqalədə Azərbaycan dilində çap əlyazma simvollarının tanınması üçün struktur əlamətlər və bootstrap nüvəli DVÜ əsasında tərtib olunmuş sistem təsvir olunmuş, nəticələrin əsaslandırılması məqsədilə eksperimentlər aparılmışdır.

Müxtəlif əlifbaların (xüsusilə azsaylı xalqların istifadə etdiyi əlifbalar) simvollarının tanınması üçün tədqiqatçılar tərəfindən çoxlu sayda əlamətlər sinfi təklif olunmuşdur [5-8]. Lakin bu əlamətlərin əksəriyyəti insan üçün anlaşılmayan, daha çox kompüterlərin rahat emal etməsi və hesablama üçün nəzərdə tutulmuş əlamətlərdir. Məhz buna görə də bu əlamətlərin həm sayı çox olur, həm də hesablanması xeyli vaxt və daha güclü emal vasitəsi tələb edir. Lakin bütün bunlara baxmayaraq tanımanın keyfiyyəti arzuolunan səviyyədə olmur, sistemin yol verdiyi səhvlərin anlaşılması çətinləşir. Bu məqalədə Azərbaycan dilində çap əlyazma simvollarının tanınması üçün təklif olunan əlamətlər insan tərəfindən asanlıqla başa düşülür. Belə ki, bu əlamətlər bizim hərfləri yadımızda saxlamaq üçün istifadə etdiyimiz xarakteristikalar əsasında hazırlanmışdır. Buna görə də şərti olaraq bu əlamətləri “soft” əlamətlər adlandıracağıq. Etiraf edək ki, bu əlamətlərin təyin olunması nisbətən çətinidir, yəni fərqli yanaşmalar, alqoritmlər tələb edir. Lakin buna baxmayaraq

nəticədə həm tanımanın keyfiyyəti yüksəlir, həm də əlamətlərin hesablanmasına tələb olunan vaxt xeyli azalır.

Beləliklə, təklif olunan yanaşmanın əsas üstünlüyü daha informativ əlamətlər sinfinin tətbiq olunması ilə tanıma keyfiyyətinin yüksəldilməsi, hesablamaların həcmnin azaldılması ilə zamana və resurslara qənaətin təmin olunmasıdır. Təklif olunan struktur əlamətlərin müxtəlif sistemlərdə sınaqdan keçirilməsinin nəticələri və digər əlamət sinifləri ilə müqayisəsi sonda daha geniş şərh olunmuşdur.

Azərbaycan dilinin morfoloji tərkibinə uyğun tanıma sisteminin formalaşdırılması məqsədilə lüğət yanaşmasından istifadə olunmuşdur. Dilimizin sözlərindən ibarət lüğət tərtib olunur, tanıma zamanı qərar qəbuletmənin etibarlılığını artırmaq üçün bu lüğətdəki sözlərə müraciət olunur. Belə ki, hansısa bir sözün tanınması zamanı aşkar təyin edilməsi mümkün olmayan simvol varsa, bu sözə ən yaxın söz lüğətdən tapılır və təsnif edilməsi qeyri-aşkar olan simvol lüğət əsasında təyin olunur. Beləliklə, simvolların klassifikasiyası zamanı müəyyən çəki ilə Azərbaycan dilində mövcud sözlərdən ibarət lüğət sistemindən də istifadə olunur.

Məsələnin qoyuluşu

DVÜ-da optimal ayırıcı hipermüstəvi müxtəlif sinifləri ən böyük fərqlə ayıran müstəvi kimi təyin olunur. Optimal hipermüstəviləri aşağıdakı şərti minimallaşdırma məsələsinin həllindən təyin etmək olar:

$$\begin{aligned} \text{Min} : & \frac{1}{2} w^T w, \\ & y_i (w^T x_i + b) \geq 1, i = 1, \dots, l \end{aligned} \quad (1)$$

Xətti ayrıla bilməyən hallar üçün yeni verilənlər çoxluğunun $x_i, i = 1, \dots, l$ daxil edilməsi ilə modifikasiya olunmuş minimallaşdırma məsələsindən istifadə olunur:

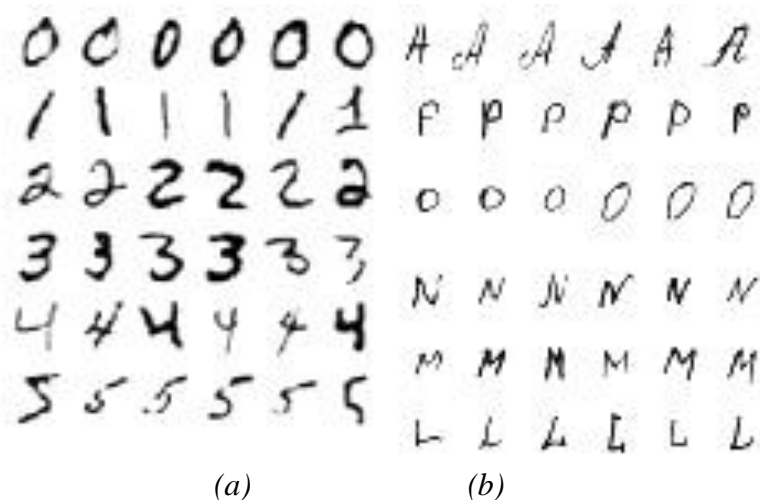
$$\begin{aligned} \text{Min} : & \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i, \\ & y_i (w^T x_i + b) \geq 1 - \xi_i, i = 1, \dots, l \end{aligned} \quad (2)$$

burada C tənzimləmə parametridir və $\xi_i = \max(0; 1 - y_i f(x_i)), i = 1, \dots, l$ şəklində təyin olunur.

DVÜ klassifikatorunun əsas üstünlüyü onun parametrsiz olmasıdır. Üsulun məqsədi ayırıcı hipermüstəvinin təyini məsələsinin optimallaşdırılması olduğuna görə giriş verilənlərinin statistik paylanmasına qarşı həssaslıq nümayiş etdirmir. Bu baxımdan obyektlərin əlamətlərinin seçilməsi və DVÜ klassifikatorunun birgə istifadəsi məqsədilə Bootstrap Resampling (BR) üsulunun tətbiqi daha əlverişlidir [10]. Məqalədə Azərbaycan dilində çap əlyazma simvollarının tanınması üçün BR-DVÜ üsulunun BISSP (Backward Input Space Selection Procedure) prosedurundan istifadə olunmuşdur.

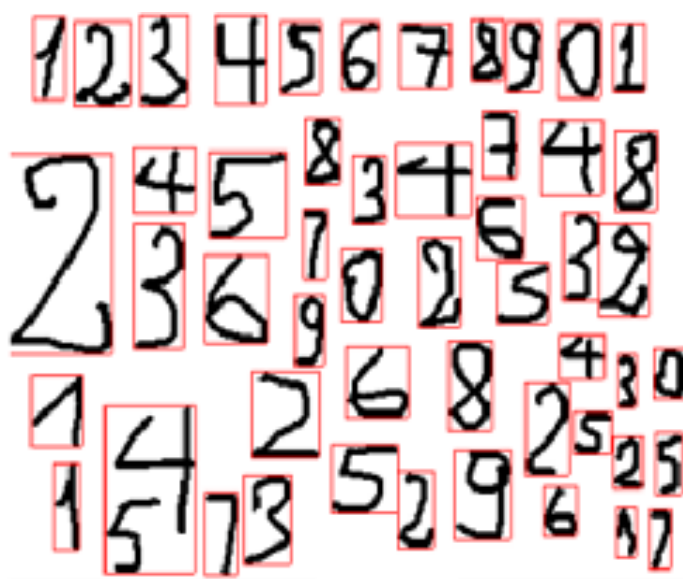
Simvolların emalı

Təcrübələr zamanı verilənlər bazası olaraq müxtəlif şəxslər tərəfindən yazılmış Azərbaycan dilində çap əlyazma simvolları nümunələrindən istifadə olunmuşdur (32 hərfin və 10 rəqəmin hər biri üçün 100 nümunə, cəmi 4200 simvol) (şəkil 1).



Şəkil 1. Verilənlər bazasının simvollarına nümunələr a) rəqəmlər b) hərflər

Əlamətləri hesablamadan öncə mühüm məqamlardan biri ilkin sənədin emal olunması, başqa sözlə, ümumi sənəddən tanınacaq obyektin seçilməsi, obyektin küylərdən təmizlənməsi, skeletləşdirilməsi və eyni ölçüyə gətirilməsidir (şəkil 2). Seçilən simvollar eyni ölçüyə gətirildikdən sonra skeletləşdirmə məqsədilə Zonqa-Sunya (Zhang-Suen) alqoritmindən istifadə olunmuşdur [11].



Şəkil 2. Obyektlərin seçilməsi

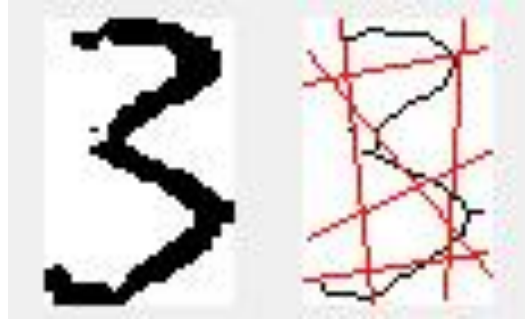
Əlamətlərin çıxarılması

Tanınma sistemlərinin qurulmasında əsas addımlardan biri əlamətlərin formalaşdırılması, daha dəqiq desək, onların keyfiyyətlilik dərəcəsinin qiymətləndirilməsidir. Əlamətlər elə seçilməlidir ki, onlar obyekt haqqında tam təsəvvür yarada bilsin və bir-birinə yaxın olmasın. Verilmiş obyekt üçün daha informativ əlamətlərdən ibarət vektorların müəyyən edilməsi üçün AdDel alqoritmindən istifadə olunmuşdur [12]. Nəticə olaraq Azərbaycan dilində çap əlyazma simvollarının tanınması üçün aşağıdakı əlamətlər informativ olaraq seçilmişdir.

A. Simvolun yerləşdiyi düzbucaqlıya çəkilmiş düz xətlərin sayı

Emal olunduqdan sonra simvolların yerləşdiyi düzbucaqlının sərhədindən üfüqi, şaquli və

müxtəlif bucaq altında maili xətlər çəkilir, bu xətlərin simvolla kəsişdiyi nöqtələrin sayı təyin olunur (şəkil 3). Bu xətlər eksperimental yolla təyin olunmuşdur, yəni başlanğıc və son nöqtələrinin koordinatları simvolun yerləşdiyi düzbucaqlının sərhədində təsadüfi şəkildə seçilən çoxlu sayda xətlər çəkilmiş və eksperimentlər aparılmışdır. Nəticədə simvol siniflərini daha yüksək dəqiqliklə klasifikasiya edən 6 düz xətt saxlanılmış və onların simvolla kəsişmə nöqtələrinin sayı əlamət vektoruna daxil edilmişdir.



Şəkil 3. Kəsişmə nöqtələrinin təyin olunması

Əlamətlər vektorunun optimallaşdırılması nəticəsində müxtəlif nöqtələrdən keçən düz xətlərin içərisindən aşağıdakı nöqtələrdən keçən düz xətlərin simvollarla kəsişmə nöqtələrinin sayının Azərbaycan dilində çap əlyazma simvollarının klasifikasiyası üçün daha əlverişli olması aşkarlanmışdır (cədvəl 1 – burada $(x_0; y_0)$ – başlanğıc nöqtənin, $(x_1; y_1)$ – son nöqtənin koordinatlarıdır).

N	1	2	3	4	5	6
$(x_0; y_0)$	(0; 20)	(0; 40)	(21; 0)	(0; 50)	(0; 50)	(0; 50)
$(x_1; y_1)$	(42; 20)	(42; 40)	(21; 63)	(42; 40)	(42; 25)	(42; 10)

Cədvəl 1. Düz xətlərin keçdiyi nöqtələrin koordinatları

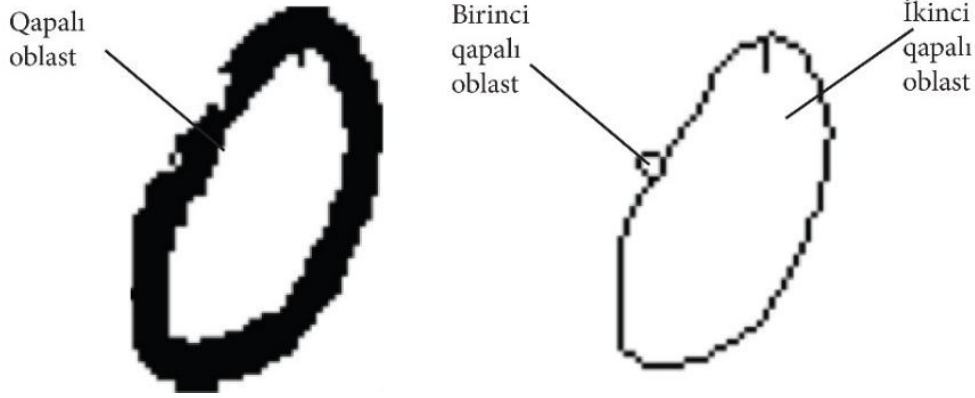
B. Simvolların qapalı oblastı

Simvolların klasifikasiyası zamanı qapalı oblastların təyin olunması çox vacibdir. Burada qapalı oblastın təyin olunması üçün aşağıdakı alqoritmdən istifadə olunmuşdur. Simvolun ilk rəngsiz pikseli təyin olunur və qonşu rəngsiz simvolla birlikdə rənglənir, sonra digər rəngsiz piksel təyin olunur və qonşu piksellərlə birlikdə başqa rənglə rənglənir. Nəticədə əgər iki rəng alınarsa, simvolun bir qapalı oblastı var (digər rəng simvolun fonunu təyin edir), üç rəng verilmiş simvolda iki qapalı oblastın olmasını göstərir, və s. (şəkil 4).



Şəkil 4. Qapalı oblastın təyin olunması

Qeyd edək ki, skeletləşdirmədən sonra səhv informasiyanın əldə olunması mümkündür, məsələn “o” hərfi və ya “0” rəqəmlərində bəzən iki qapalı oblast alınır, lakin əslində bir qapalı oblast olmalıdır (şəkil 5). Bu tip problemlərin aradan qaldırılması məqsədi ilə oblastların diaqonalı təyin olunur. Əlamət olaraq ən böyük oblastın diaqonalı götürülür.



Şəkil 5. Skeletləşmədən əvvəl və sonra qapalı oblastlar

C. Qapalı oblastın yerləşdiyi hissə

Verilmiş simvolun qapalı oblastının yerləşdiyi hissə ağırlıq mərkəzinə əsasən təyin olunur. Sonra oblastın ağırlıq mərkəzindən simvolun yerləşdiyi düzbucaqlının aşağı sərhədinə qədər olan məsafə hesablanır, bu da müəyyən sinif simvolları fərqləndirmək üçün çox faydalıdır. Məsələn “p” hərfinin ən böyük qapalı oblastı yuxarı hissədir, “ə” hərfinin qapalı oblastı isə aşağı hissədə yerləşir.

Beləliklə, məqalədə təklif olunan “soft” əlamətlər vektoru 8 elementdən ibarətdir:

$F = \{f_1, \dots, f_6 - \text{kəsişmə nöqtələrinin sayı}, f_7 - \text{ən böyük qapalı oblastın diaqonalı}, f_8 - \text{qapalı oblastın ağırlıq mərkəzindən olan məsafə}\}$

Eksperimentlərin nəticələri

Məqalədə təklif olunan əlamətlər sinfinin üstünlüyünün əsaslandırılması üçün müxtəlif tədqiqatçılar tərəfindən Azərbaycan dilində çap əlyazma simvollarının tanınmasına tətbiq edilmiş aşağıdakı əlamətlər istifadə edilmiş, nəticələr müqayisəli analiz olunmuşdur.

I sinif əlamətlər. Azərbaycan dilində çap əlyazma simvollarının tanınması üçün Ayda-zadə və Mustafayev tərəfindən təklif olunmuş üsulda simvollar 28×28 ölçüsünə normallaşdırılaraq, boz rəngin çalarları şəklində verilənlər bazasında saxlanılır. Əlamət olaraq simvolun yerləşdiyi düzbucaqlıdakı bütün piksellərin rəng (RGB) qiyməti götürülür. Beləliklə, hər bir nümunə üçün 0-dan 255-ə qədər qiymət alan $784 (28 \times 28 = 784)$ əlamətlər vektoru hesablanır [2].

II sinif əlamətlərin hesablanması zamanı simvol 16×24 piksel ölçülü təsvirə normallaşdırılır. Normallaşdırılmış təsvir 4 istiqamət - 1 şaquli, 1 üfüqi, 2 diaqonal üzrə 22 hissəyə ayrılır (4 şaquli, 6 üfüqi və diaqonalların hər biri üçün). Hər bir hissə üzrə içərisində ən azı bir qara və bir ağ piksel olan (tam qara və tam ağ olmaz) 4×4 piksel ölçülü kvadrat hissəciklər götürülür. Bu kvadrat hissəciklərin piksellərə görə bütün mümkün variantlarının sayı 14-dür, bu variantlarda ağ və qara piksellərin yerləşmə formasından asılı olaraq, istiqamət təyin edilir. Beləliklə, bu 14 variant üzrə 4 istiqamətə şaquli, üfüqi, soldan-sağa diaqonal və sağdan-sola diaqonal istiqamətlər məlum olur. Alqoritmin sonu olaraq, 22 hissə üzrə sonda qeyd etdiyimiz 4 istiqamətin sayı hesablanır və beləliklə təsvirin kontur üzrə istiqamətini təyin edən, $22 \times 4 = 88$ elementdən ibarət kontur istiqamət vektoru əldə edilir. Bu üsulla 88 ədəd əlamət alınır [3].

Üçüncü sinif əlamətlər. Bu əlamətlər Peripheral Directional Contributivity (PDC) alqoritmi ilə təyin olunan əlamətlər çoxluğundan ibarətdir [4]. Simvolun hər bir nöqtəsi özlüyündə 8 və ya 4 ölçülü vektordur. Vektorun hər bir komponenti simvolun və kvadratın sərhədinə qədər olan məsafəni özündə saxlayır. Simvolların əlamətlərini təyin edərəkən 8 istiqamət götürülür, onları birləşdirən düz xəttlərin kəsişmə nöqtəsi təyin olunur. Bu nöqtəni birinci dərəcəli xətti nöqtə adlandırırlar (XN). Bu şəkildə davam edərək ikinci dərəcəli xətti nöqtə təyin olunur. Eyni addımla daha yüksək səviyyəli xətti nöqtələr tapılır. PDC-alqoritmının ölçüsü məsələnin şərtindən asılıdır. Azərbaycan dilində çap əlyazma simvollarının əlamətləri üçün PDC üsulunun parametrləri aşağıdakı şəkildə təyin olunmuşdur: DC-nin ölçüsü 8, baxılan istiqamətlərin sayı 4, dərinlik 1 və seqmentlərin sayı 8. Bu parametrlər çərçivəsində PDC əlamətlərinin ümumi sayı $8 \times 4 \times 1 \times 8 = 256$ olur.

Verilmiş öyrədici nümunələr “soft” əlamətlər və yuxarıda təsvir olunan hər bir 3 əlamət sinfi əsasında DVÜ-in bootstrap nüvəsi üsulu və süni neyron şəbəkələr vasitəsilə tanıma həyata keçirilmişdir. Tanınmanın nəticələri cədvəl 2-də göstərilmişdir:

Cədvəl 2

DVÜ və süni neyron şəbəkələr üsulu ilə tanınmanın nəticələri

	“soft” əlamətlər	I sinif	II sinif	III sinif
Əlamətlərin sayı	8	784	88	256
SVM Bootstrap	93,80 %	90,05 %	88,06 %	79,03 %
Süni neyron şəbəkələr	86,22 %	80,31 %	86,10 %	82,47%

Cədvəl 2-dən görünür ki, əlamətlərin xarakteri dəyişdikcə tanınmanın nəticəsi də dəyişir, ən yaxşı nəticə müəllifin təklif etdiyi “soft” əlamətlər və DVÜ üsulunun tətbiqi nəticəsində əldə edilmişdir.

Ekspərimənt zamanı test bazasındakı simvolların 60 %-i öyrətmə üçün, 20 %-i öyrənmənin qiymətləndirilməsi üçün, 20 %-i isə sistemin sazlanması üçün istifadə olunmuşdur. Ən yaxşı nəticə DVÜ Bootstrap üsulu ilə tanıma zamanı əldə olunmuşdur, lakin bu zaman müəyyən hərflərin tanınmasında problemlər meydana çıxmışdır. Cədvəl 3-də yanlış tanınan simvollar və onlara alternativ göstərilən simvollar təsvir olunmuşdur.

Cədvəl 3

Düzgün tanınmayan simvollar

Yoxlanılan simvol	d	ə	k	z	q
Yanlış təklif edilən simvol	o, a	z, b	r, y	x, f	e, o

Yanlışlığın araşdırılması məqsədilə səhv tanınan simvollar ayrılıqda yenidən öyrədilmiş və yoxlanma təkrar olunmuşdur. Nəticədə yalnız z və x hərflərinin fərqləndirilməsi sistemdə müəyyən problemlər yaratsa da, digər simvolların sistem tərəfindən yüksək səviyyədə tanınması müşahidə olunmuşdur.

Nəticə

Müxtəlif əlamətlər üzərində aparılmış çoxsaylı təcrübələrin nəticəsi göstərir ki, DVÜ kompleksi yalnız çap əlyazma simvollarının tanınmasında deyil, həm də digər tanıma sistemlərinin qurulmasında uğurla istifadə edilə bilər. Məqalədə aparılmış tədqiqatın nəticəsini aşağıdakı müddəalarla dəyərləndirmək olar:

1) Tanıma sistemi tərtib olunarkən, bir deyil, bir neçə əlamət qrupundan istifadə etməklə, “qarşılıqlı yoxlama” üsulu vasitəsilə verilmiş məsələ üçün hansı sinif əlamətin daha faydalı olduğunu təyin etmək olar;

2) Həmçinin əlamət siniflərində müxtəlif dəyişikliklərə yol verməklə, yeni əlamət nüvələrini dəyişdirməklə fərqli nəticələr əldə etmək olar. Sinifləri təmsil edən əlamətlərin sayı öyrənməyə sərf olunan zamana təsir etsə də, tanımanın dəqiqliyinə təsir göstərmir, belə ki, əlamətlər üçün əsas tələb onların informativ və hər sinif üçün maksimal dərəcədə unikal olmasıdır.

Ədəbiyyat

1. Cristianini N., Shawe-Taylor J. An Introduction to support vector machines and other kernelbased learning methods. Cambridge University Press, 2000.
2. Aida-zade K.R., Hasanov J.Z., Cursive Handwritten Azerbaijani Latin Text Segmentation Based on Word baseline. / INISTA, Trabzon, Turkey, 2009, pp.63–66.
3. Aida-zade K.R., Hasanov J.Z., Word base line detection in handwritten text recognition systems. // International journal of computer systems science and engineering, 2009, no.4, pp. 49–53.
4. Aida-zade K.R., Mustafayev E.E., On a hierarchical handwritten forms recognition system on the basis of the neural network. / Proceed. Inter. Conf. TAINN, Canakkale, 2003.
5. Moubtahij H.E., Halli A., Satori K., Review of feature extraction techniques for offline handwriting arabic text recognition // International journal of advances in engineering & technology, 2014, pp.50–58.
6. Hadidi G., Delavari H., Persian handwritten words detection based on features, extraction and fuzzy algorithm. //Electrical and electronics engineering: an international journal (elelij) 2015, vol.4, no.2, pp.93–104.
7. Hussain E., Hannan A., Kashyap K., A zoning based feature extraction method for recognition of handwritten assamese characters. // International journal of computer science and technology 2015, vol. 6, no.2, pp.226–228.
8. Lawgali A., Bouridane A., Angelova M., Ghassemlooy Z., Handwritten arabic character recognition: which feature extraction method?. // International journal of advanced science and technology, 2011, vol.34, pp.1–8.
9. Platt J. Fast training of support vector machines using sequential minimal optimization. Advances in Kernel Methods — Support Vector Learning, MIT Press, 1999, pp.185–208.
10. Atienza F.A. Bootstrap feature selection in Support Vector Machines for ventricular fibrillation detection / ESANN'2006, Belgium, 2006, pp.233–238.
11. Chen W., Sui L., Xu Z., Lang Y. Improved Zhang-Suen thinning algorithm in binary line drawing applications // ICSAI, 2012, doi: 10.1109/ICSAI.2012.6223430
12. Ismayilov E., Ismayilova N., Fuzzy Features Extraction for Hand-printed character/digit recognition system. / INISTA 2014, Italy, pp.249–253.

УДК 004.93

Исмайлов Эльвиз А.

Азербайджанский Государственный Университет Нефти и Промышленности, Баку, Азербайджан

elviz.ismayilov@gmail.com

Применение нового класса признаков и метода опорных векторов в системе распознавания рукопечатных символов азербайджанского языка

Несмотря на наличие широко распространенных латинских символов в нашем алфавите, специальные символы и морфологическая структура языка требуют индивидуального подхода. В статье используется класс «софт» признаков (выявленный с использованием характеристик, близких к человеческому мышлению) и метод опорных векторов для распознавания рукопечатных символов азербайджанского языка. Для классификации символов была использована процедура bootstrap resampling метода опорных векторов. Результаты сравнены с результатами использования других классов признаков и методов.

Ключевые слова: «софт» признаки, метод опорных векторов, рукопечатные символы, система распознавания.

Elviz A. Ismayilov

Azerbaijan State Oil and Industry University, Baku, Azerbaijan

elviz.ismayilov@gmail.com

Application of SVM and new features class to recognition of Azerbaijani hand-printed symbols

Although there are widely spread latin symbols in our alphabet, intending special symbols and morphological content of our language requires individual approach for character recognition. In this paper “soft” (close to human mind, constructed on base of characteristics which we use in alphabet learning) features and SVM for recognition of Azerbaijani hand-printed characters are used. For character classification bootstrap resampling procedure of support vector machines is used. Results are compared with results of other feature classes and methods.

Keywords: “soft” features, SVM, hand-printed symbols, recognition system.