

Main methods of identifying a speaker through speech

N. Mirzaev¹^a, J.K. Urinboev¹^b, M.A. Nugmanova¹^c

¹Digital technologies and artificial intelligence research institute, Tashkent, Uzbekistan

Abstract: This paper provides an overview of methods used for voice recognition. Special attention is given to the robust structure of recognition systems. Additionally, a brief description of the most widely used feature extraction methods (such as MFCC and LPCC) and classification techniques (vector quantization method, Gaussian mixture model, support vector method) is presented. Methods for evaluating recognition systems and approaches for presenting the results of such evaluations are discussed.

Keywords: speech recognition, biometric identification, speech signal processing, mel-frequency cepstral coefficients (MFCC), linear prediction coefficients, feature extraction, K-means algorithm, discrete Fourier transform, Gaussian mixture model

Diktor shaxsini nutqi orqali tanib olishning asosiy usullari

Mirzayev N.¹^a, O'rinboyev J.K.¹^b, Nugmanova M.A.¹^c

¹Raqamli texnologiyalar va sun'iy intellekt rivojlantirish ilmiy-tadqiqot instituti, Toshkent, O'zbekiston

Annotatsiya: Ushbu ishda ovoz orqali tanib olish masalasini yechish uchun qo'llaniladigan usullar sharhi keltirilgan. Tanib olish tizimlarining barqaror tuzilishiga alohida e'tibor qaratilgan. Shuningdek, belgilarni ajratib olishning eng keng tarqalgan usullari (masalan, MFCC va LPCC)ning qisqacha tavsifi hamda tasniflash usullari (vektorli kvantlash usuli, Gauss qorishmalari modeli, tayanch vektorlar usuli) haqida ma'lumot berilgan. Tanib olish tizimlarini baholash usullari va bunday baholash natijalarini taqdim etish yo'llari muhokama qilingan.

Kalit so'zlar: nutq orqali tanib olish, biometrik identifikatsiya, nutq signaliga ishlov berish, mel-chastotali kepsstral koeffitsiyentlar (MFCC), chiziqli bashoratlash koeffitsiyentlari, belgilarni ajratib olish, K-meanslar algoritmi, diskret Fure almashtirish, Gauss qorishmalar modeli

1. Kirish

Diktorni tanib olish masalasi nutq texnologiyalari sohasining dolzarb masalasidir. Diktorni tanib olishning nutqni qayta ishlashning boshqa sohalari bilan aloqasi quyidagi munosabatlar bilan aks ettirilishi mumkin [6].

Nutqni qayta ishlashni quyidagi masalalarga ajratish mumkin

- analiz/sintez,
- tanib olish,
- kodlash.

Tanib olish quyidagi kichik vazifalarni qamrab oladi

- nutqni tanib olish,
- diktorni tanib olish,
- tilni identifikatsiyalash.

Diktorni tanib olish diktorni identifikatsiyalash va verifikatsiyalash jarayonlarini birlashtiradi.

Diktorni identifikatsiyalash - bu ovoz namunasini bazada saqlangan namunalarni bilan taqqoslash orqali shaxsni tanib olish jarayoni. Identifikatsiya jarayonining natijasi nomzodlar ro'yxati hisoblanadi. Amalga oshiruvchi tizim belgilangan miqdordagi ro'yxatni berishi yoki foydalanuvchini belgilangan chegara asosida nomzodlar ro'yxatiga kiritish to'g'risida qaror qabul qilishi mumkin.

Agar identifikatsiya jarayonida tizimda ro'yxatdan o'tmagan foydalanuvchi ishtirok etishi mumkinligi ko'zda tutilgan bo'lsa, ochiq to'plamda identifikatsiya haqida so'z yuritiladi. Ideal holda, bunday foydalanuvchi uchun tizim bo'sh ro'yxat taqdim etishi lozim. Identifikatsiya jarayonidan o'tayotgan barcha foydalanuvchilar tizimda ro'yxatdan o'tgan bo'lsa, yopiq to'plamda identifikatsiya haqida gap boradi.

Diktorni verifikatsiyalash - taqdim etilgan namunani bazada saqlanayotgan namuna bilan taqqoslash orqali so'ralgan shaxsni tanib olish jarayoni. Tekshirish natijasi ijobiy yoki salbiy qaror bo'ladi. Ba'zan "ovoz orqali tanib olish" atamasi ham qo'llaniladi (speaker detection [8]). Tanib olish masalasida biroz boshqacha atamalar va ustuvorliklar qo'llanilsa-da, aslida, tekshirish va tanib olish bir xil vazifadir.

Ushbu tasnifdan tashqari, tanib olish tizimlarining o'zi foydalanuvchi aytishi kerak bo'lgan matn tizimga ma'lum yoki noma'lumligiga va tizim bu ma'lumotlardan foydalanishiga qarab matnga bog'liq va matnga bog'liq bo'lmagan turlarga bo'linadi. Matnga bog'liq tanib olishda belgilangan iboralar, shuningdek tizim tomonidan yaratilgan va foydalanuvchiga taklif etilgan iboralar ham ishlatilishi

^a <https://orcid.org/0009-0006-1561-9393>

^b <https://orcid.org/0009-0000-2305-8887>

^c <https://orcid.org/0009-0006-1956-9856>



mumkin. Matnga bog'liq bo'lmagan tizimlar esa ixtiyoriy nutqni qayta ishlashga mo'ljallangan.

Ovoz orqali tanib olish bilan bog'liq boshqa masalalar ham mavjud. Shular qatoriga quyidagi vazifani kiritish mumkin. Faraz qilaylik, signalda ikki yoki undan ortiq shaxsning suhbatlari yozilgan. Signalning bir qismi qo'lda yoki o'qituvchisiz o'rganish algoritmlari yordamida, kim qaysi paytda gapirayotganini ko'rsatish uchun belgilanadi. Signalning qolgan qismi esa avtomatik tarzda belgilanishi kerak. Bu masalada belgilarni tanib olish dan tashqari, ularning chegaralarini ham belgilash talab etiladi. Bunday vazifa ingliz tilidagi adabiyotlarda "speaker diarization" ("so'zlovchilarni protokollashtirish" [1]) deb nomlanadi.

2. Tadqiqot metodikasi

2.1. Diktorni tanib olish tizimlarining tuzilishi

Tanib olish tizimlarining ishlashi ikkita asosiy bosqichdan iborat: tizimda foydalanuvchilarni ro'yxatdan o'tkazish va tanib olish jarayonining o'zi (identifikatsiyalash yoki verifikatsiyalashga urinish). Foydalanuvchilar avvaldan tizimda o'z ovozlari yozib qoldirib ro'yxatdan o'tadilar. Har bir gapiruvchining ovoz namunasi tanib olishda foydalanish mumkin bo'lgan belgilarni ajratib olish maqsadida qayta ishlanadi. Ajratib olingan belgilar asosida foydalanuvchilarning modellari yaratiladi (ba'zi hollarda "namuna" atamasi ko'proq mos keladi). Model berilgan belgilar bo'yicha o'xshashlik darajasini baholash yoki tezda qaror qabul qilish imkonini beruvchi ma'lum bir tuzilmadir.

Verifikatsiya holatida foydalanuvchi identifikator va ovoz namunasini taqdim etib, tizimga kirishga harakat qiladi. Taqdim etilgan namunadan olingan belgilar bazada saqlangan tegishli model bilan, shuningdek, ba'zi foydalanuvchilarning belgilangan to'plamini yoki ushbu ovozga eng yaqin bo'lgan etalon model bilan taqqoslanadi. Natija belgilangan chegara bilan solishtiriladi va ruxsat berish haqida ijobiy yoki salbiy qaror qabul qilinadi.

Identifikatsiya jarayonida taqdim etilgan namunadan o'ziga xos belgilar ajratib olinadi va keyin ular tizimda ro'yxatdan o'tgan yoki oldindan tanlangan barcha foydalanuvchilarning modellari bilan taqqoslanadi. Shunday qilib, tanib olish tizimining umumiy sxemasi quyidagi asosiy bosqichlar yoki darajalar orqali amalga oshiriladi:

- Signallarga ishlov berish bosqichi. Bu bosqichda signal tanib olish vazifasi uchun muhim bo'lgan belgilarni ajratib olish maqsadida qayta ishlanadi. Nutq signali belgilar vektorlari ketma-ketligi yordamida ifodalanadi.

- Modellar bosqichi. Foydalanuvchini ro'yxatdan o'tkazish bosqichi signallarga ishlov berish bosqichidan olingan belgilar vektorlari ketma-ketligidan foydalanib model yaratadi. Modellashtirish belgilar vektorlarini oddiy nusxalashdan tortib ehtimollik modellari yoki boshqa tuzilmalarni qurishgacha bo'lishi mumkin. Shundan so'ng, berilgan belgilar va saqlangan model o'rtasidagi o'xshashlik darajasini hisoblash imkoniyati paydo bo'ladi.

- Qaror qabul qilish bosqichi. Qaror qabul qilish funksiyalari odatda alohida bosqichga ajratiladi, garchi u oddiy funksiyalarni bajaradi, agar yakuniy yechimlar modellar darajasida hisoblab chiqilsa umuman bo'lmayligi ham mumkin. Qaror qabul qilish uchun modellar darajasida hisoblangan o'xshashlik darajalari zarur bo'lsa, belgilangan chegaralardan foydalaniladi.

2.2. Namuna olish va uni dastlabki qayta ishlash

Ovoz orqali tanib olishda nutq signalining yozuvi ishlov beriladigan namuna hisoblanadi. Impuls-kodli modulyatsiya bilan kodlashda analog signal amplituda qiymatlarining lahzali o'lovchilari (hisoblar) ketma-ketligi bilan ifodalanadi. Nutq signalini yozish va qayta ishlash uchun odatda 8 yoki 16 kHz diskretlash chastotasi qo'llaniladi, yuqoriroq diskretlash chastotasi esa ko'proq hisoblash xarajatlari talab etadi. Hisoblarni ifodalash uchun 8, 12 yoki 16 bit ishlatiladi, boshqa qiymatlar ham qo'llanilishi mumkin.

Nutq signalini yozib olish va uzatish bilan bog'liq bir qator omillar tanib olish sifatiga ta'sir ko'rsatadi. Bular orasida quyidagilarni ajratib ko'rsatish mumkin:

- kanalning mos kelmasligi,
- xonaning yomon akustikasi,
- mikrofondan turli masofada bo'lish va boshqalar.

Misol uchun, telefon kanali orqali uzatilayotgan ovozni tanib olishda, odatda, ro'yxatga olish va identifikatsiya qilish uchun bir xil mikrofon va uzatish kanalidan foydalanishni kafolatlash mumkin emas. Bundan tashqari, tashqi xalaqitlarning ta'sirini ham hisobga olish lozim. Masalan, xonaga kirishni nazorat qilish uchun ovoz orqali tasdiqlash tizimida yuqori sifatli yozuvdan foydalanish imkoniyati mavjud. Bunday holda kanal mikrofon, uning kabeli va analog-raqamli o'zgartiruvchi qurilmalardan tashkil topadi. Signalga dastlabki ishlov berish nutq bo'lmagan qismlarni olib tashlash, shuningdek, chastota filtri yordamida qayta ishlashdan iborat bo'lishi mumkin.

2.3. Belgilarni ajratib olish usullari

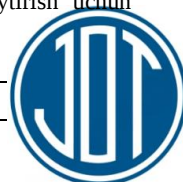
Mavjud ilovalarda signalni qayta ishlash nutq signalidagi ovozni tanib olish masalasi uchun muhim bo'lgan ma'lumotni, ya'ni inson ovozinin o'ziga xos belgilarini ajratib olishni maqsad qiladi. Ajratilgan belgilar namunani shakllantirish yoki mavjud namunalar bilan taqqoslash uchun qo'llaniladi. Oldindan qaysi belgilar tanib olish uchun ko'proq mos kelishini baholab bo'lmaydi. Mos keluvchi belgilarni tanib olish jarayoni mumkin bo'lgan belgi variantlarini ko'rib chiqish va ularni keyinchalik tajriba orqali baholashdan iborat.

Belgilarning ikki turi ajratiladi: quyi darajadagi (nutq apparatining anatomik tuzilishi bilan bog'liq) va yuqori darajadagi (orttirilgan, talaffuz uslubi bilan bog'liq).

Nutq signalini qayta ishlash jarayoniga nisbatan shakllangan yondashuv qisqa muddatli tahlildan foydalanishdan iborat. Bunda signal belgilangan o'lchamdagi vaqtinchalik oynalarga bo'linadi va bu oynalarda signal parametrlari o'zgarimas deb taxmin qilinadi. Nutq signali uchun oyna o'lchami odatda 10-30 millisekund oralig'ida tanlanadi. Signalni aniqroq ifodalash uchun oynalar orasida oyna uzunligining yarmiga teng qoplama hosil qilinadi. So'ngra har bir oynaga spektral tahlil, chiziqli bashorat usuli yoki boshqa belgilarni ajratib olish algoritmlari qo'llaniladi.

Mel-chastotali keprstral koeffitsiyentlar.

Ushbu belgilarni ajratib olish usuli ham diktorni, ham nutqni tanib olish tizimlarida eng keng tarqalgan usullardan biridir. Algoritmning kirishiga joriy iteratsiyada o'rganilayotgan signal qismining $x_0, \dots, x_{N-1}$ hisoblash ketma-ketligi beriladi. Ushbu ketma-ketlikka avval vazn funksiyasi, so'ngra diskret Fure almashtirilishi qo'llaniladi. Vazn funksiyasi tanlanmaning cheklanganligidan kelib chiqadigan Fure tahlili buzilishlarini kamaytirish uchun



ishlatiladi. Amalda vazn funksiyasi sifatida ko'pincha Xemming oynasi qo'llaniladi, u quyidagi ko'rinishga ega:

$$\omega_n = 0,54 - 0,46 \cdot \cos\left(2\pi \frac{n}{N-1}\right), n = 0, \dots, N-1,$$

bu yerda N - sanoqlarda ifodalangan oyna uzunligi.

U holda muallaq signalning diskret Fure almashtirishini quyidagi ko'rinishda yozish mumkin:

$$X_k = \sum_{n=0}^{N-1} x_n \omega_n e^{-\frac{2\pi i}{N} kn}, k = 0, \dots, N-1.$$

k indekslarining qiymatlari chastotalarga mos keladi

$$f_k = \frac{F_s}{N} k, k = 0, \dots, N/2,$$

bu yerda F_s - signalni diskretlash chastotasi.

Chastota sohasida olingan signal tasviri uchburchakli filtrlar to'plami yordamida diapazonlarga ajratiladi. Filtrlarning chegaralari mel shkalasida hisoblanadi. Ushbu shkala inson qulog'ining turli chastotalardagi tovushlarni idrok etish qobiliyatini o'rganish natijasida yaratilgan. Mel-chastota sohasiga o'tkazish quyidagi formula orqali amalga oshiriladi:

$$B(f) = 1127 \cdot \ln\left(1 + \frac{f}{700}\right).$$

Teskari almashtirish quyidagicha ifodalanadi:

$$B^{-1}(b) = 700(e^{b/1127} - 1).$$

N_{FB} - filtrlar soni (odatda 24 ta filtrdan foydalaniladi), (f_{low}, f_{high}) - o'rganilayotgan chastotalar diapazoni bo'lsin. U holda ushbu diapazon mel shkalasiga o'tkaziladi, bir tekis taqsimlangan bir-birini qoplaydigan diapazonlar N_{FB} ga bo'linadi va chiziqli chastotalar sohasida tegishli chegaralar hisoblanadi. Olingan filtrlarning og'irlik koeffitsiyentlarini $H_{m,k}$ orqali belgilaymiz. Filtrlar Fure almashtirish koeffitsiyentlari modullarining kvadratlarga qo'llaniladi. Olingan qiymatlar logarifmlanadi

$$e_m = \ln\left(\sum_{k=0}^N |X_k|^2 H_{m,k}\right), m = 0, \dots, N_{FB} - 1.$$

MFCC koeffitsiyentlarni hisoblashning yakuniy bosqichi diskret kosinus almashtirish hisoblanadi

$$c_i = \sum_{m=0}^{N_{FB}-1} e_m \cos\left(\frac{\pi i(m+0,5)}{N_{FB}}\right), i = 1, \dots, N_{MFCC}.$$

c_0 koeffitsiyent sifatida ishlatilmaydi, chunki u signal energiyasini ifodalaydi. Amaliyotda N_{MFCC} koeffitsiyentlari soni 12 ga yaqin qilib tanlanadi.

Chiziqli bashoratlash asosidagi kepsral koeffitsiyentlar.

Chiziqli bashoratning mohiyati shundan iboratki, oldingi o'lchov natijalarining ma'lum miqdordagi chiziqli kombinatsiyasi yordamida joriy o'lchovni taxminiy ifodalash mumkin

$$x_n \approx \sum_{k=1}^p a_k x_{n-k}.$$

Chiziqli kombinatsiyaning $a_1, \dots, a_p$ og'irlik koeffitsiyentlari chiziqli bashoratlash koeffitsiyentlari deb ataladi. Chiziqli bashoratlash koeffitsiyentlarini aniqlash Darbining rekursiv algoritmi yordamida amalga oshiriladi [2].

Olingan chiziqli bashoratlash koeffitsiyentlari asosida kepsral koeffitsiyentlar hisoblanadi. Bunday kepsral koeffitsiyentlar soni chiziqli bashoratlash koeffitsiyentlari sonidan ko'proq bo'lishi mumkin.

$$c_n = \begin{cases} a_n + \sum_{k=1}^{n-1} \frac{k}{n} c_k a_{n-k}, & 1 \leq n \leq p; \\ \sum_{k=n-p}^{n-1} \frac{k}{n} c_k a_{n-k}, & n > p. \end{cases}$$

Diskretlash chastotasi 8000 Hz bo'lgan signal uchun taxminan 12 ta chiziqli bashoratlash koeffitsiyentlaridan foydalaniladi, ulardan esa taxminan 18 ta kepsral koeffitsiyentlar hosil qilinadi.

2.4. Belgilarga ishlov berish

Yuqorida tavsiflangan belgilarni ajratib olish usullari kichik qismda belgilarni tanib olish uchun mo'ljallangan. Nutq dinamikasi haqidagi ma'lumotni saqlash maqsadida belgilar vektorlarini ularning birinchi va ba'zan ikkinchi hosilalari bilan birlashtirish yondashuvi qo'llaniladi. Bunday hosilalar Δ - va $\Delta\Delta$ -koeffitsiyentlar (delta- va delta-delta-koeffitsiyentlar) deb nomlanadi.

Belgilarga keyingi ishlov berish bosqichida o'rganilayotgan obyektning barcha belgilar vektorlari to'plamidan foydalanadigan normallashtirish usullari ham qo'llaniladi. Kanalning ta'sirini kamaytirish maqsadida qo'llaniladigan eng keng tarqalgan normallashtirish usuli kepsal o'rta qiymatni ayirish (Cepstral Mean Subtraction; CMS) usulidir. Ushbu usul sessiyalar orasidagi o'zgarishlarni kompensatsiya qilish uchun mo'ljallangan bo'lib, doimiy sharoitlarga qo'llanilganda, aksincha, samaradorlikni pasaytiradi.

2.5. Tasniflash usullari

Ovozni tanib olish ko'plab biometrik tizimlardan shunisi bilan farq qiladiki, bu holda tanib olish obyekti barmoq izlari, yuz yoki ko'z qorachig'ini tanib olishdagi kabi statik tasvir emas, balki jarayon hisoblanadi. Shu sababli, ovoz namunasi ko'pincha yagona belgilar vektori sifatida emas, balki nutq signalining kichik qismlarini tavsiflovchi belgilar vektorlarining ketma-ketligi ko'rinishida taqdim etiladi. Signalga ishlov berish bosqichidan so'ng olingan vektorlar ketma-ketligi diktir ovozi namunasi yoki modelini yaratish yoxud mavjud namunalar bilan taqqoslash uchun ishlatiladi. Tekshirish va tanib olish vazifalari uchun taqdim etilgan namunaning bir yoki bir nechta namunalar bilan o'xshashlik darajasini hisoblash usuli belgilanishi mumkin. O'xshashlik darajasi ma'lum bir mezon yoki ehtimollikni baholash asosida hisoblanishi mumkin.

Tanib olish masalasi uchun modellarni tasniflashning bir necha usullari mavjud. Ilmiy adabiyotlarda modellar ko'pincha generativ yoki diskriminativ deb ataladi. Generativ deb nomlanadigan modellarning mohiyati o'rganish uchun olingan ma'lumotlarni modellashtrish, masalan, ehtimollik zichligi funksiyasini baholash orqali amalga oshiriladi. Bunga Gauss qorishmalar modeli misol bo'la oladi. Diskriminativ modellar esa sinflar o'rtasidagi chegarani qurishga asoslanadi, masalan, tayanch vektorlar usulida qo'llanilganidek.

Masofani hisoblash.

Masofani hisoblash usulini aniqlash namunaviy modellar uchun asos hisoblanadi. Bunday modellarda aniqlanadigan obyekt saqlanadigan obyektlardan birining noaniq nusxasi sifatida ko'rib chiqiladi. Vektorlar orasidagi masofani hisoblashning eng keng tarqalgan usullaridan biri quyidagilardir:



• L1-me'yor (shahar kvartallari masofasi, Manhattan masofasi)

$$d(x, y) = \sum_{i=1}^D |x_i - y_i|;$$

• Yevklid masofasi

$$d(x, y) = \sum_{i=1}^D (x_i - y_i)^2;$$

• Mahalonobis masofasi

$$d(x, y) = (x - y)^T \times W^{-1} \times (x - y),$$

bu yerda W – kovariatsiya matritsasi hisoblanadi. Agar W birlik matritsaga teng bo'lsa, masofa Yevklid masofasi bilan bir xil bo'ladi.

Yaqin qo'shnilar usuli.

Ushbu usulda diktir namunasi sifatida o'rgatuvchi ketma-ketlikning barcha vektorlari to'plamidan foydalaniladi. Namunaning bunday andoza bilan taqqoslanishi quyidagicha amalga oshiriladi. Sinov ketma-ketligining har bir vektori eng kichik masofani aniqlash uchun andozaning har bir vektori bilan solishtiriladi. Olingan masofalar yakuniy bahoni shakllantirish maqsadida o'rtacha qiymatga keltiriladi.

$$d_k = \sum_{i=1}^L \min_{x_j \in Sp_k} d(x_i, x_j).$$

Hisoblash murakkabligini kamaytirish maqsadida turli xil usullar qo'llaniladi. Bular jumlasiga namunani qisqartirish yoki qidiruvni tezlashtirish uchun ma'lumotlarni saqlash usullari kiradi. Masalan, kd-daraxt yoki boshqa usullar shular qatoridandir [5].

K-yaqin qo'shnilar usuli ham o'rgatuvchi vektorlar ketma-ketligini saqlashni talab etadi, biroq o'xshashlik darajasini hisoblash biroz boshqacha usulda amalga oshiriladi. Har bir y_i sinov vektori uchun saqlangan namuna vektorlarigacha bo'lgan masofalarni hisoblagandan so'ng, k ta eng yaqin vektorni aniqlash mumkin. Faraz qilaylik, k_{ij} - topilgan k ta eng yaqin vektorlar orasida j sinfga (bizning holatimizda j diktorga) tegishli vektorlar soni bo'lsin. Har bir sinfga o'rgatish vektorlarining taxminan bir xil sonini nazarda tutib, i vektorning j sinfga tegishlilik ehtimolini baholash quyidagicha amalga oshirilishi mumkin

$$\hat{P}(C_j|y_i) = \frac{k_{ij}}{k}.$$

Shunda vektorlar ketma-ketligini quyidagi qoida asosida tasniflash mumkin bo'ladi:

$$C = \arg \max_{1 \leq j \leq N} \prod_{i=1}^L \hat{P}(C_j|y_i).$$

Bunday usul ovoz berish sxemasi deb ataladi, chunki ketma-ketlik eng ko'p "ovoz" to'plagan sinfga kiritiladi [10].

Vektorli kvantlash.

Vektorli kvantlash usulida, eng yaqin qo'shni usulidan farqli o'laroq, o'rgatuvchi vektorlar to'plami to'liq saqlanmaydi, balki kod vektorlari to'plamiga (odatda belgilangan o'lchamli) aylantiriladi. Bunday to'plamni yaratishning keng tarqalgan usuli, shuningdek kod kitobi deb ham ataladigan, K-meanslar algoritmi hisoblanadi.

K-meanslar algoritmi dastlabki to'plamni K ta klasterga ajratadi, bunda K oldindan berilgan son hisoblanadi. Buning uchun avval o'rtacha qiymatlar dastlabki to'plamdagi ayrim vektorlar bilan initsializatsiya qilinadi. So'ngra algoritim har

bir iteratsiyasida vektorlar o'zlariga eng yaqin klasterlarga taqsimlanadi (buning uchun vektor va o'rtacha qiymatlarining joriy holati orasidagi masofa hisoblanadi) va har bir klasterdagi o'rtacha qiymat qayta hisoblab chiqiladi. Keyingi iteratsiyada klasterlar holati o'zgarmaganda yoki belgilangan maksimal iteratsiyalar soniga yetganda algoritim yakunlanadi. Olingan o'rtacha qiymatlar namuna tuzish uchun ishlatiladigan kod vektorlari hisoblanadi. Vektorlarning kirish ketma-ketligi va kod kitoblari orasidagi masofani hisoblash eng yaqin qo'shnilar usuli kabi amalga oshiriladi.

Gauss qorishmalar modeli.

Gauss qorishmalar modeli diktorni tanib olish sohasida keng qo'llaniladi. Bu model Gauss taqsimotlarining vaznli yig'indisidan iborat.

$$p(x|\lambda) = \sum_{i=1}^M \omega_i p_i(x),$$

bu yerda λ - diktir modeli, M - model komponentlarining soni, ω_i - komponentlarning og'irligi

$$\sum_{i=1}^M \omega_i = 1.$$

Har bir komponentning ehtimollik zichligi funksiyasi quyidagi formula orqali ifodalanadi:

$$p_i = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp \left(-\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right),$$

bu yerda D - belgilar fazosining o'lchami, μ_i - matematik kutish vektori, Σ esa kovariatsiya matritsasi. Ushbu modelni amalga oshiruvchi tizimlarda ko'pincha kovariatsiyaning diagonal matritsasiidan foydalaniladi. Shuningdek, so'zlovchi modelining barcha komponentlari uchun yagona kovariatsiya matritsasi yoki barcha modellar uchun bitta matritsani qo'llash ham mumkin.

Shunday qilib, diktir modelini yaratish uchun o'rtacha vektorlarni, kovariatsiya matritsalarini va komponentlar og'irliklarini aniqlash zarur. Bu masala EM-algoritmi yordamida hal etiladi. Kirish sifatida $X = \{x_1, \dots, x_T\}$ vektorlarining o'rgatuvchi ketma-ketligi beriladi. Model parametrlari boshlang'ich qiymatlar bilan initsializatsiya qilinadi va keyin algoritimning har bir iteratsiyasida parametrlarni qayta baholash amalga oshiriladi.

Boshlang'ich parametrlarni aniqlash uchun odatda K-means [3] kabi klasterlash algoritmi qo'llaniladi. O'rgatuvchi vektorlar to'plamini M ta klasterga ajratib olish orqali model parametrlarini quyidagicha boshlang'ich holatga keltirish mumkin. μ_i ning dastlabki qiymatlari klaster markazlari bilan mos keladi, kovariatsiya matritsalarini esa shu klasterga tegishli vektorlar asosida hisoblanadi. Komponentlarning og'irliklari esa mazkur klasterga tegishli vektorlarning umumiy o'rgatuvchi vektorlar soniga nisbati bilan aniqlanadi.

Parametrlarni qayta baholash quyidagi formulalar bo'yicha amalga oshiriladi:

• aposterior ehtimolliklarni hisoblash (Estimation-step)

$$p(i|x_t, \lambda) = \frac{\omega_i p_i(x_t)}{\sum_{k=1}^M \omega_k p_k(x_t)};$$

• modelning yangi parametrlarini hisoblash (Maximization-step)

$$\omega_i = \frac{1}{T} \sum_{t=1}^T p(i|x_t, \lambda);$$



$$\mu_i = \frac{\sum_{t=1}^T p(i|x_t, \lambda)}{\sum_{t=1}^T p(i|x_t, \lambda) x_t};$$

$$\Sigma_i = \frac{\sum_{t=1}^T p(i|x_t, \lambda) (x_t - \mu_i)(x_t - \mu_i)^T}{\sum_{t=1}^T p(i|x_t, \lambda)};$$

Ushbu qadamlar parametrlar o'zaro moslashguncha takrorlanadi.

Tayanch vektorlar usuli.

Tayanch vektorlar usuli binar tasniflagich bo'lib, ajratuvchi funksiyani quyidagi ko'rinishda quradi:

$$f(x) = w \cdot x + b.$$

$(x_1, y_1), \dots, (x_N, y_N)$ o'qitish ketma-ketligi berilgan bo'lsin, bu yerda x_i - belgilar fazosining nuqtalari, y_i esa 1 yoki -1 qiymatlarni qabul qiluvchi sinflardan biriga tegishlilikni bildiruvchi belgilar. Avval ma'lumotlarning chiziqli ajralish holatini ko'rib chiqaylik. Bunday cheklov quyidagi ko'rinishda ifodalanishi mumkin:

$$\begin{cases} w \cdot x_i + b \geq +1, & y_i = +1; \\ w \cdot x_i + b \leq -1, & y_i = -1; \end{cases} \quad (1)$$

yoki bitta tengsizlik orqali ifodalansa

$$y_i(w \cdot x_i + b) - 1 \geq 0, i = 1, \dots, N.$$

Mumkin bo'lgan ajratuvchi gipertekisliklar orasidan sinflar o'rtasida eng katta oraliqni hosil qiluvchi gipertekislik izlanadi. Bunda ajratuvchi gipertekislikdan har bir sinfning eng yaqin nuqtalarigacha bo'lgan masofa maksimal bo'lishi lozim. Oraliq kattaligini $2/||w||$ sifatida hisoblash mumkin. Shunday qilib, maksimal oraliqli ajratuvchi gipertekislikni topish masalasini (1) ifoda asosida $||w||^2$ ni minimallashtirish masalasiga keltirish maqsadga muvofiq. Ushbu masala (shuningdek, keyinchalik ta'riflanadigan chiziqli ajralmaslik va yadro funksiyasi yordamida nochiqizlilikni kiritish holatlari uchun umumlashtirishlar ham) kvadratik dasturlash usullari bilan yechilishi mumkin [11].

Chiziqli ajralmaslik holati uchun masalani umumlashtirish maqsadida, cheklovlar quyidagi ko'rinishda qayta yoziladi:

$$\begin{cases} w \cdot x_i + b \geq +1 - \xi_i, & y_i = +1; \\ w \cdot x_i + b \leq -1 + \xi_i, & y_i = -1; \\ \xi_i \geq 0 \end{cases}$$

barcha $i = 1, \dots, N$ uchun. Maqsad funksiyasi quyidagi ko'rinishda bo'ladi:

$$\frac{1}{2} w \cdot w + C \sum_{i=1}^N \xi_i \rightarrow \min,$$

bu yerda C - xatolarning paydo bo'lishi uchun jarima darajasini belgilaydigan musbat doimiy.

Chiziqli ajralmaydigan to'plamlarni aniqlashning yana bir usuli yadro funksiyasini kiritishdir. Bu g'oyaning mohiyati shundaki, boshlang'ich fazoni yuqori o'lchamli fazoga aks ettirish orqali, to'plamlar ajraladigan bo'lishi mumkin. Bunda, o'qitish va tanib olish algoritmlarida belgilar alohida emas, balki skalyar ko'paytmalar shaklida ishlatilganligi sababli, bunday o'zgartirishni aniq ko'rinishda yaratishning hojati yo'q. Yangi fazoda skalyar ko'paytmani belgilaydigan yadro funksiyasini berish kifoya qiladi.

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j).$$

Keng tarqalganlar orasida quyidagi yadrolarni keltirish mumkin:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) - \text{Gauss radial bazis}$$

funksiyalarining yadrosi,

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^n - \text{polinomial yadro.}$$

Usul parametrlari (masalan, C va yadro parametrlari) odatda ma'lum qiymatlar to'plamini tanlash va krossvalidatsiya usuli bilan baholash orqali aniqlanadi. Bu jarayon qiymatlarni ketma-ket sinab ko'rish va natijalarni tahlil qilish yo'li bilan amalga oshiriladi.

Yuqorida tavsiflangan usullar binar tasniflash masalasini hal qiladi. Ushbu usullarni ko'p sinfli aniqlash masalasiga qo'llash uchun "bittasi qolganlarga qarshi" yoki "bittasi bittasiga qarshi" kabi strategiyalardan foydalaniladi. Aytaylik, o'qitish uchun q ta sinf berilgan. "Bittasi qolganlarga qarshi" strategiyasini qo'llaganda q ta tasniflagich yaratiladi, ularning har biri ushbu sinfni boshqalaridan ajratib olishga o'rgatiladi. Aniqlash jarayonida obyekt $f(x)$ funksiyasining eng katta qiymatini bergan sinfga kiritiladi. "Bittasi bittasiga qarshi" ("har biri har biriga qarshi") strategiyasi esa ikki sinfni ajratadigan $q(q-1)/2$ ta tasniflagichdan foydalanadi.

Muayyan sinfni qolgan barcha sinflar bilan taqqoslash natijalari jamlanadi va so'ngra boshqa sinflarning shunga o'xshash natijalari bilan solishtiradi. Bunda $f(x)$ funksiyani $P(C|x)$ ehtimollikka o'zgartirish uchun turli usullardan ham foydalaniladi.

Belgilar vektorlari ketma-ketligini aniqlash uchun har bir kadmi tasniflash natijalarini birlashtiruvchi qoidalar qo'llanilishi mumkin. Shuningdek, vektorlar ketma-ketligidan generativ modellarni o'rgatish uchun foydalaniladigan, keyinchalik tayanch vektorlar usuli yordamida tasniflanadigan yondashuvlar ham qo'llaniladi.

2.6. Tizimni baholash

Tanib olish tizimlarining aniq ishlashiga bir qator omillar ta'sir ko'rsatadi. Avvalo, ovozning o'zidagi o'zgaruvchanlikni ta'kidlash lozim. Hissiy holat, charchoq, yosh o'zgarishlari, shamollash va boshqa ko'plab omillar ovozga ta'sir qiladi. Ikkinchidan, tanib olish tizimlari uchun atrof-muhit ta'siri hamda yozib olish sharoitlarining o'zgarishi muammo hisoblanadi.

Ekspirimental baholash uchun qo'llaniladigan ma'lumotlar bazalari (korpuslar) doim ham yuqorida sanab o'tilgan vaziyatlarni to'liq modellashtira olmaydi. Shu sababli, natija asosan baza qanchalik reprezentativligi va tajriba qanday tashkil etilganligiga bog'liq bo'ladi. Tadqiqotchilar o'tkazilgan tajribalar haqida batafsil ma'lumot berganliklari sababli, tajribaning haqiqiy qo'llanish sharoitlariga mosligini tushunish va natijalarni taqqoslash imkoniyatini beradi. Bunday ma'lumot, avvalo, yozib olish seanslari sonini va ular orasidagi tanaffuslar davomiyligini ko'rsatishi lozim. Qolaversa, yozib olish shartlarining tavsifi (mikrofon turi, uzatish kanali, xonaning shovqin darajasi va boshqalar) hamda ro'yxatga olish va sinov sessiyalari uchun sharoitlar turlichami (nomuvofiq sharoitlar) yoki yo'qligini aks ettirishi kerak. Shuningdek, natija har bir sinovda modellarni yaratish uchun foydalaniladigan materialning davomiyligiga va bazadagi foydalanuvchilar soniga ham bog'liq bo'ladi.

Identifikatsiya tizimlarini baholash uchun aksariyat hollarda foydalanuvchilarning yopiq to'plami bilan cheklanadi, ya'ni identifikatsiya qilishga urinayotgan barcha foydalanuvchilar tizimda ro'yxatdan o'tgan. Natija ro'yxatdan o'tgan foydalanuvchilar soniga va qaytariladigan ro'yxatning hajmiga (ko'pincha faqat bitta identifikator ishlatiladi) yoki ro'yxatga kirish chegarasiga bog'liq. Identifikatsiyalash ehtimoli (haqiqiy-ijobiy identifikatsiyalash) to'g'ri identifikatorni o'z ichiga olgan



nomzodlar ro'yxati qaytarilishiga olib kelgan identifikatsiyalash urinishlarining ulushi sifatida baholanadi.

Verifikatsiya tizimlarida ikki turdagi xatoliklar yuzaga keladi. Soxta ruxsat xatoligi ikki har xil foydalanuvchi namunalarini taqqoslashda ijobiy qaror (o'xshashlik) qabul qilishdan iborat. Namunalar bir foydalanuvchiga tegishli bo'lsa-da, ularni farqlash to'g'risida qaror qabul qilish noto'g'ri rad etish xatosi deb ataladi. Har ikkala turdagi xatolar qaror qabul qilish chegarasiga bog'liq bo'ladi.

Verifikatsiya sinovlarining natijalari qaror qabul qilish chegarasining parametrik belgilangan egri chizig'ini aks ettiruvchi ishlash xususiyati grafigi orqali ko'rsatilishi mumkin. Bunday grafikda absissalar o'qi bo'yicha soxta-ijobiy qarorlar ehtimolliklarining baholari (noto'g'ri qabul qilish ehtimolliklari), ordinatalar o'qi bo'yicha esa haqiqiy ijobiy qarorlar ehtimolliklarining baholari joylashtiriladi. Diktorni tekshirish sohasida boshqacha usul ommalashgan bo'lib, unda absissalar o'qi bo'yilab noto'g'ri ruxsat berish ehtimolligini baholash, ordinatalar o'qi bo'yilab esa noto'g'ri rad etish ehtimolligini baholash joylashtiriladi. Bunda o'qlar uchun ko'proq ko'rgazmalilik maqsadida normal og'ish shkalasi [9] (ba'zan logarifmik yoki boshqa shkala) qo'llaniladi. Bunday usulda [7] taklif qilingan va diktorni tanib olish (speaker detection) masalasida natijalarni taqdim etish uchun ishlatilgan, shuning uchun u tanib olish xatolari kompromis egri chizig'i (Detection Error Tradeoff; DET) nomini olgan.

Baholash natijalarini yagona parametr ko'rinishida ifodalash uchun quyidagi usullar qo'llaniladi. Ulardan biri soxta yo'l qo'yish (CFA) va soxta yo'l qo'ymaslik (CFR) qiymatlarini belgilash hamda aniqlash qiymati funksiyasini (Detection Cost Function; DCF) hisoblashdan iborat [8].

$$DCF = C_{FR}P_{tar}R_{FR} + C_{FA}P_{imp}R_{FA},$$

bu yerda P_{tar} va P_{imp} - haqiqiy shaxs va "soxta shaxs" urinishlarining aprior ehtimolliklari, R_{FR} va R_{FA} - mos ravishda noto'g'ri rad etish va noto'g'ri qabul qilish xatoliklari ehtimolliklarining olingan baholaridir. Qaror qabul qilish chegarasi qiymat funksiyasining qiymatini minimallashtirish uchun optimallashtiriladi. Keng tarqalgan o'lchov esa teng xatolik ehtimolligi darajasi (Equal Error Rate; EER) bo'lib, u shunday chegarada xatolik ehtimolligining kattaligini ifodalaydi, bunda noto'g'ri qabul qilish va noto'g'ri rad etish xatoliklari ehtimolligi bir-biriga teng yoki qiymati jihatidan eng yaqin bo'ladi.

3. Xulosa

Diktorni tanib olish tizimlarini rivojlantirish bir necha yo'nalishda amalga oshirilmoqda. Belgilarni ajratib olishning eng keng tarqalgan usullari kepsstral koefitsiyentlarni hisoblash usullari hisoblanadi: mel-chastotali va chiziqli bashoratlashga asoslangan. Shuningdek, asosiy ton va formant chastotalarining statistikasi ham qo'llaniladi [4]. Signal qayta ishlash darajasidagi rivojlanish asosan nutq signalini tashqi shovqinlar va signal uzatish kanali tomonidan yuzaga keladigan buzilishlarga chidamli bo'lgan mustahkam tasvirini yaratishga qaratilgan yangi usullarni izlash yo'nalishida sodir bo'lmoqda. Yuqori darajadagi belgilardan foydalanish yo'nalishi ham rivojlanmoqda. Diktorni modellarini yaratish usullari oddiy belgilar vektorlarini o'rtaqalashtirishdan murakkab generativ va diskriminativ modellargacha takomillashdi.

Hozirgi paytda modellarni yaratishning quyidagi usullari ustunlik qilmoqda:

- matnga bog'liq tizimlar uchun - vaqtning dinamik o'zgarishi (Dynamic Time Warping; DTW) va yashirin Markov modellari (Hidden Markov Model; HMM);

- matnga bog'liq bo'lmagan tizimlar uchun - vektorli kvantlash (Vector Quantization; VQ), Gauss qorishmalar modeli (Gaussian Mixture Model; GMM) va tayanch vektorlar usuli (Support Vector Machine; SVM).

Qarorlar qabul qilish yagona tasniflagichdan foydalangan holda ham, qaror qabul qilish qoidalarini birlashtirish orqali ham amalga oshiriladi. Eksperimental baholash bo'yicha kichik guruhlarda (5-10 kishi) o'tkazilgan laboratoriya sinovlaridan to tizimning haqiqiy qo'llanish sharoitlarini aks ettiruvchi katta hajmli korpuslarni yaratishgacha bo'lgan yo'l bosib o'tildi [12].

Foydalangan adabiyotlar / References

[1] Samatov, R., & Xalilova, G. (2024). Searching for a free Будков В.Ю., Прищеп М.В., Ронжин А.Л., Марков К. Многоканальная система анализа речевой активности участников совещания // Третий междисциплинарный семинар «Анализ разговорной русской речи» АР3. 2009. СПб. 2009. С. 57-62.

[2] Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов / Пер. с англ. М. : Радио и связь, 1981. 496 с.

[3] Садыхов Р.Х., Ракуш В.В. Модели гауссовых смесей для верификации диктора по произвольной речи // Доклады БГУИР. Минск, 2003. №

[4] С. 95–103. 4. Сорокин В.Н., Цыплихин А.И. Верификация диктора по спектрально-временным параметрам речевого сигнала // Информационные процессы. 2010. Т. 10, № 2. С. 87-104.

[5] Arya S., Mount D.M., Netanyahu N.S., Silverman R., Wu A. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions // Journal of the ACM. 1998. V. 45, N 6. P. 891–923

[6] Campbell J.P., Speaker Recognition: A Tutorial // Proceedings of the IEEE. 1997. V. 85, N 9. P. 1437-1462.

[7] Martin A., Doddington G., Kamm T., Ordowski M., Przybocki M. The det curve in assessment of detection task performance // Proc. of Eurospeech. 1997. V. 4. P. 1895-1898.

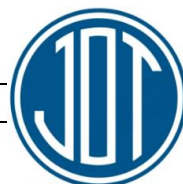
[8] Martin A., Przybocki M. The NIST 1999 Speaker Recognition Evaluation - An Overview // Digital Signal Processing. 2000. V. 10.

[9] Navratil J., Klusacek D. On linear DETs // Internat. Conf. on Acoustics, Speech, and Signal Processing (ICASSP-07). 2007.

[10] Paredes R., Vidal E., Casacuberta F. Local features for speaker recognition // SPR 2004. International Workshop on Statistical Pattern Recognition. LNCS 3138 of Lecture Notes in Computer Science. 2004. P. 1087-1095.

[11] Platt J.C. Fast Training of Support Vector Machines using Sequential Minimal Optimization // Advances in Kernel Methods / Ed. by B. Scholkopf, C.C. Burges, A.J. Smola. MIT Press, 1999. P. 185–208.

[12] Reynolds D.A. An Overview of Automatic Speaker Recognition Technology // The International



Conference on Acoustics, Speech, and Signal Processing
ICASSP 02. 2002. P. 4072–4075.

Mualliflar to'g'risida ma'lumot/ Information about the authors

Mirzayev Raqamli texnologiyalar va sun'iy
Nomaz / intellektni rivojlantirish ilmiy-tadqiqot
Mirzaev instituti katta ilmiy xodimi, t.f.d
Nomaz (professor)
E-mail: nomazmirzaevairi@gmail.com
Tel.: +998909889571
<https://orcid.org/0009-0006-1561-9393>

O'rinboyev Raqamli texnologiyalar va sun'iy
Johongir intellektni rivojlantirish ilmiy-tadqiqot
Kalbay o'g'li / instituti tayanch doktoranti
Urinboev E-mail: jahongir8010@gmail.com
Johongir Tel.: +998911621994
<https://orcid.org/0009-0000-2305-8887>

Nugmanova Raqamli texnologiyalar va sun'iy
Mavluda Avaz intellektni rivojlantirish ilmiy-tadqiqot
qizi instituti tayanch doktoranti,
E-mail: Miss.abduazimova@gmail.com
Tel.: +998909191060
<https://orcid.org/0009-0006-1956-9856>

