

NUTQNI TANISH TIZIMINI CHUQUR NEYRON TARMOQLARI YORDAMIDA YARATISH BOSQICHLARI

Daminova Barno Esanovna

Qarshi davlat universiteti
Algoritmlar va dasturlash texnologiyalari kafedrası dotsenti,
Qarshi shahri, O'zbekiston

Omonov Javohir Ma'murjon o'g'li

Qarshi davlat texnika universiteti
Raqamli texnologiyalar va sun'iy intellekt kafedrası kompyuter injenering
(AT-servis) yo'nalishi 4-kurs talabasi,
Qarshi shahri, O'zbekiston

Norqo'chqorov Yusufbek Yashin o'g'li

Qarshi davlat texnika universiteti
Raqamli texnologiyalar va sun'iy intellekt kafedrası kompyuter injenering
(AT-servis) yo'nalishi 4-kurs talabasi,
Qarshi shahri, O'zbekiston

Annotatsiya. Ushbu maqolada nutqni tanish tizimlarini chuqur neyron tarmoqlari asosida yaratish bosqichlari tizimli ravishda tahlil etilgan. Ma'lumotlarni tayyorlash, xususiyatlarni ajratib olish, model arxitekturasini tanlash, o'rgatish, baholash va optimallashtirish jarayonlari yoritilgan. Zamonaviy texnologiyalar va innovatsion yondashuvlarning amaliy samaradorligi haqida fikr yuritilgan.

Kalit so'zlar: nutqni tanish, chuqur neyron tarmoqlar, MFCC, CTC, wav2vec 2.0, til modeli, dekodlash.

STAGES OF CREATING A SPEECH RECOGNITION SYSTEM USING DEEP NEURON NETWORKS

Daminova Barno Esanovna

Karshi State University

Associate Professor of the Department of Algorithms and Programming

Technologies,

Karshi, Uzbekistan

Omonov Javokhir Mamurjon ugli

Karshi State Technical University

Department of Digital Technologies and Artificial Intelligence Computer

Engineering

4th year student of the (IT-service) direction,

Karshi, Uzbekistan

Norkuchkarov Yusufbek Yashin ugli

Karshi State Technical University

Department of Digital Technologies and Artificial Intelligence Computer

Engineering

4th year student of the (IT-service) direction,

Karshi, Uzbekistan

Abstract. This article systematically analyzes the stages of developing speech recognition systems based on deep neural networks. It discusses data preparation, feature extraction, model architecture selection, training, evaluation, and optimization processes. The practical effectiveness of modern technologies and innovative approaches is explored.

Keywords: speech recognition, deep neural networks, MFCC, CTC, wav2vec 2.0, language model, decoding.

Аннотация. В данной статье систематически анализируются этапы создания систем распознавания речи на основе глубоких нейронных сетей. Рассматриваются этапы подготовки данных, выделения признаков, выбора архитектуры модели, обучения, оценки и оптимизации.

Ключевые слова: распознавание речи, глубокие нейронные сети, MFCC, CTC, wav2vec 2.0, языковая модель, декодирование.

Nutqni avtomatik tarzda tanish texnologiyalari sun'iy intellekt va chuqur o'rganish yutuqlari natijasida sezilarli darajada rivojlanmoqda. Ayniqsa, chuqur neyron tarmoqlarga asoslangan tizimlar inson nutqini aniqlik bilan tahlil qilish va matn shaklida ifodalash imkonini beradi. Ushbu maqolada nutqni tanish tizimini yaratishning asosiy bosqichlari ko'rib chiqiladi.

Ma'lumotlar bazasini tayyorlash

Nutqni tanish tizimi uchun sifatli ma'lumotlar bazasi hal qiluvchi ahamiyatga ega.

Ma'lumot turi	Tavsifi	Manba
Nutq yozuvlari	Audio fayllar (.wav, .flac)	LibriSpeech, CommonVoice
Transkriptlar	Nutqdagi so'zlarning matn ko'rinishi	Professional transkriptorlar
Anotatsiyalar	Lahja, shovqin darajasi va boshqa metama'lumotlar	Maxsus belgilar

Nutq korpusining hajmi va xilma-xilligi modelning umumlashuv qobiliyatiga bevosita ta'sir qiladi. Misol uchun, LibriSpeech korpusi 1000 soatdan ortiq inglizcha nutq namunalarini o'z ichiga oladi.

2. Akustik xususiyatlarni ajratish

Nutq signallaridan foydali ma'lumotlarni ajratib olish uchun akustik xususiyatlar hisoblanadi.

Xususiyat turi	Izoh
MFCC	Mel chastota kepsstral koeffitsientlari, inson quloq sezgirligiga moslab olingan
Spectrogram	Chastota va vaqt o'zgarishini tasvirlovchi rasm
Log-Mel Spectrogram	Logarifmik o'lchovdagi energiyani tasvirlaydi

MFCC hisoblash formulasi:

Bu yerda $S(f, t)$ - signalning spektral quvvati, N - umumiy koeffitsientlar soni.

3. Model arxitekturasini tanlash

Model tanlash tizim samaradorligiga sezilarli ta'sir qiladi.

Model turi	Afzalliklari	Kamchiliklari
RNN/LSTM	Uzoq ketma-ketliklarni o'zida saqlay oladi	Hisoblash xarajati yuqori
CNN	Lokal xususiyatlarni tez aniqlaydi	Sekvensial kontekstni to'liq ushlamaydi
Transformer/wav2vec 2.0	Parallel hisoblash va yuqori aniqlik	Katta ma'lumot va resurs talab qiladi

4. Modelni o'qitish va izlanishlar

Eksperimentlar uchun quyidagi shartlar o'rnatildi:

- O'quv to'plam: 960 soat LibriSpeech
- Optimallashtiruvchi: Adam, boshlang'ich o'rganish tezligi 0.0001
- Batch size: 32

Tajriba natijalari:

Model	WER (%)
LSTM-CTC	10.5
CNN-CTC	8.9

wav2vec 2.0 fine-tuned	5.2

O'qitish jarayoni grafigi:

Tajribada wav2vec 2.0 modeli ancha tez konvergenstsiyaga erishdi.

5. Dekodlash va transkripsiya

Dekodlash strategiyalari transkripsiya sifati uchun muhimdir.

Dekodlash turi	Xato darajasi (WER)
Greedy decoding	6.0%
Beam Search + 5-gram LM	5.2%

Greedy decoding formulasi:

$$\hat{y} = \arg \max_y P(y|x)$$

Bu yerda x - kiruvchi signal, y- transkribatsiya.

6. Baholash va optimallashtirish

Model sifatini baholash uchun quyidagi metrikalar ishlatiladi:

• **Word Error Rate (WER):**
$$WER = \frac{S + D + I}{N}$$

Bu yerda:

- S- almashtirishlar soni,
- D- o'chirishlar soni,
- I- qo'shishlar soni,
- N- asl so'zlar soni.

• **Character Error Rate (CER)** xuddi shunday hisoblanadi, faqat belgilar asosida.

Optimallashtirish uchun:

- Model siqish (quantization)
- Tarmoqlarning nozarur vaznlarini olib tashlash (pruning)
- Distillation orqali yengil modellar yaratish



1-rasm. Nutq tanib olishni amalga oshirish bosqichlari

Dastlabki ishlov berish – signalni qabul qilish, analogdan raqamli ko‘rinishga o‘zgartirish, filtrlashdan iborat. Xususiyatlarni ajratish – jimlik sohalarini olib tashlash, oynalardan o‘tqazish, spektr qiymatlarini va MFCC(ing. Mel Frequency Cepstral Coefficient - Mel chastotaviy kepsstral koeffitsienti) parametrlarini hisoblashni tashkil etadi. o‘qitish va tanib olish – bunda har bir nutq so‘zi olingan parametralari bo‘yicha mashinali o‘qitish va sun‘iy intellekt algoritmlari orqali o‘qitiladi va tanib olinadi. Nutqni avtomatik tanib olishning eng muvaffaqiyatlilari ajratilgan so‘zlarni tanib olish uchun suxandonga sozlangan tizimlar bo‘lib chiqdi. Bunday tizimlar ikki rejimdan birida ishlaydi. o‘qitish rejimida foydalanuvchi tizimga mikrofondan tanib olish uchun har bir so‘zni aytib, ovozi tanib olishga o‘rgatadi. Tizim har bir so‘zni na‘munasini(etalonini) yaratadi va xotirada saqlaydi. Tanib olish rejimida har bir aytilgan so‘z yana raqamlashtiriladi va uning na‘muna etaloni xotiradan olingan na‘muna etalonlar bilan taqqoslanadi. Agar mos keladigan bo‘lsa, unda so‘z tanib olinadi va tizim bu haqda foydalanuvchini xabardor qiladi yoki biron bir harakatni amalga oshiradi. Bunday tizimlarning ishlashi suxandonning har bir so‘z oldidan yetarlicha pauza qilishiga, shovqinli fonga va so‘zlarning aniq va to‘g‘ri talaffuz qilinishiga bog‘liq. Tanib olish paytida nutq signallariga raqamli ishlov berishning eng muhim amallari, xususiyatlarni ajratish, og‘zaki nutqlardan alohida namunalar olinib etalonlar yaratish va namunalardan kerakligini tanlab olish, etalonlar xotirada saqlanadigan yozuvlar bilan taqqoslash bajariladi (2-rasm).

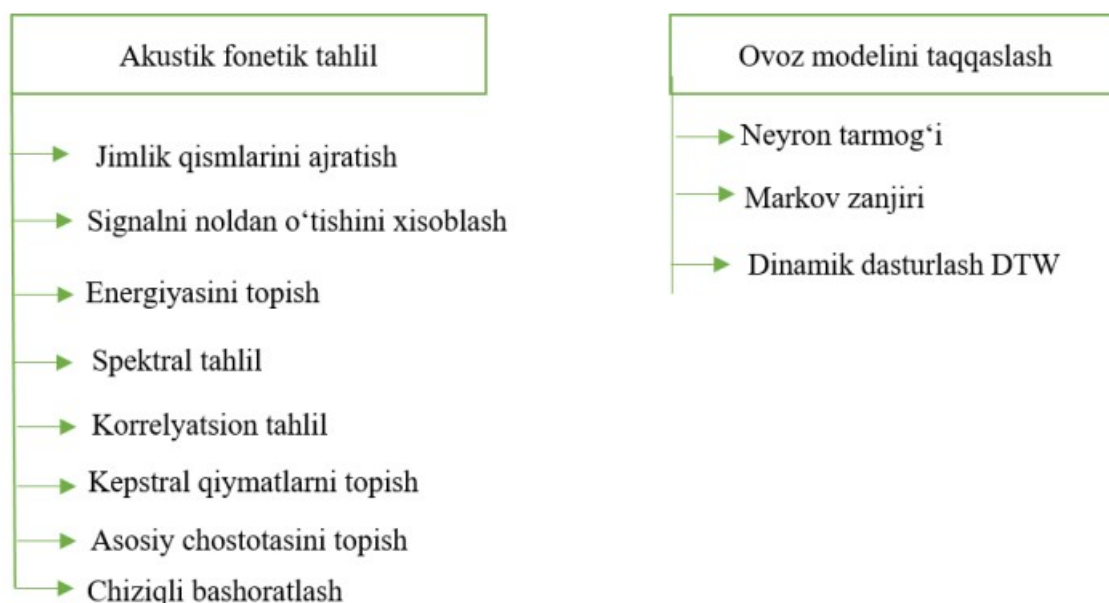


2-rasm. Nutqni tanib olish tizimining umumiy blok-sxemasi

Nutq signallariga ishlov berishning dastlabki va nisbatan eng sodda darajasi bu akustik va fonetik tahlil bo'lib, asosan gap bo'lagini ichidagi alohida ajratilgan so'zlarni tanib olishga bag'ishlangan. Nutq buyruq so'zlarini tanib olishda akustik ishlov berish bosqichini amalga oshirish yetarli hisoblanadi. Keyingi nisbatan murakkab darajasi esa bu leksik, sintaktik va semantik tahlil hisoblanib, qo'shma murakkab gaplarni tushunishga qaratilgan

Bajarilgan ilmiy ishlanmalar va eksperimental tadqiqotlarni tahlil qilib qaraydigan bo'lsak, nutqni tahlil qilish va tanib olish usullari va algoritmlari orasida alohida ajratilgan nutqlarga ishlov berishning quyidagi protseduralarini ajratib ko'rsatishimiz mumkin(3-rasm). Akustik tahlil muhitda tarqalgan akustik axborotni aniqlash, tahlil qilish va tasniflashga qaratilgan usullarni anglatadi. Akustik tahlil qilish avtomatik kuzatuv, aqlli videokonferensiya, inson-kompyuter aloqasi va eshitish apparatlari texnologiyalari kabi bir necha xil ilovalarga ega bo'lgan usullarni o'z ichiga olishi mumkin. Akustik hodisalarni aniqlash va tasniflash usullari bu tovushlarning vaqt chegaralarini aniqlash va ularni sinf bo'yicha turkumlashga qaratilgan. Akustik tahlilning asosiy muammolaridan biri signallarning nutq faolligi aniqlash, nutqli va nutqsiz sinflarga ajratishdir. Nutq signallarida energiya chegaralari va chastotaning noldan o'tish qiymatlarini aniqlash orqali nutqning aktivligini aniqlash eng sodda usullaridir. Bundan tashqari nutq signali aktivligini aniqlashning chuqur o'rganishiga asoslangan usullari keng tarqalgan. Nutq signallariga dastlabki ishlov berishning bir nechta asosiy algoritmlari mavjud. Ushbu algoritmlarning barchasi, shu jumladan eng murakkablari ham bir xil arifmetik amallarni takrorlaydi. Signallarga dastlabki ishlov berish amallari yuqorida keltirilgan ma'lumotlarni yig'ish(o'qib olish), segmentlash, filtrlash, o'zgartirishlar(almashtirishlar) orqali xususiyatlarin ajratish, tanib olish bosqichlarida amalga oshiriladi. Nutq signallariga ishlov berish barcha asosiy amallari uchun faqat oddiy arifmetik amallar - ko'paytirish, qo'shish, ayirish va o'zgartirish amallarini bajarish talab qilinadi. Svyortka signallarga

raqamli ishlov berishda eng ko‘p ishlatiladigan amallardan biri bo‘lib, raqamli filtrlash asosini tashkil etadi va ko‘paytirib jamlash amalidan tashkil topadi.



3-rasm. Tahlil qilish usullari.

Filtrlashning maqsadi bu foydali signalda shovqinni yo‘qotish yoki kamaytirishdan iborat. Zamonaviy nutq signallariga ishlov berishda va ayniqsa ovozli boshqaruv tizimlarida real sharoitlarda murakkab filtrlash amallarini talab etadi. Bunday murakkab filtrlash amallarini adaptiv filtrlar, ya’ni ma’lum bir muhitga moslashuvchi filtrlar orqali amalga oshiriladi. Nutq signallarini tanib olish va uni qurilmaga integratsiya qilish yangi chuqur neyron tahlilga asoslangan yondashuvlarni talab etadi. Bu orqali qisqa lug‘atga asoslangan nutq buyruqlarini tanib olishda yuqori natijalarga erishish mumkin

Xulosa

Nutqni tanish tizimlarini yaratishda chuqur neyron tarmoqlarga asoslangan yondashuvlar yuqori natijalar bermoqda. Tajriba natijalari shuni ko'rsatadiki, wav2vec 2.0 modeli an'anaviy metodlardan sezilarli ustunlikka ega. Kelgusida o'z-o'zini nazorat qilishga asoslangan o'qitish va multimodal tizimlarni rivojlantirish nutqni tanish texnologiyalarining yangi bosqichini belgilashi mumkin.

Foydalanilgan adabiyotlar

1. Graves, A., Mohamed, A., Hinton, G. (2013). Speech recognition with deep recurrent neural networks. ICASSP.
2. Baevski, A., Zhou, Y., Mohamed, A., Auli, M. (2020). wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. NeurIPS.
3. Hannun, A. et al. (2014). Deep Speech: Scaling up end-to-end speech recognition.
4. Vaswani, A. et al. (2017). Attention is All You Need. NIPS.
5. O‘rinboy Naziraliyev (2024). NUTQNI TANIB OLISHNI AMALGA OSHIRISH BOSQICHLARI.
6. Daminova, B. E., Bozorova, I. J., Shodmonqulova, F. O., & Shodmonqulova, Z. O. (2024). USE OF ONLINE ELECTRONIC DICTIONARIES IN ENGLISH LANGUAGE LESSONS. *Экономика и социум*, (5-1 (120)), 193-196.
7. Daminova, B. E., Bozorova, I. J., & Jumayeva, N. X. (2024). FORMATION OF TEXT DATA PROCESSING SKILLS. *Экономика и социум*, (4-2 (119)), 110-119.
8. Daminova, B. E., Bozorova, I. J., Tursunova, S. N., & Sunnatova, H. R. (2024). USING DIGITAL TECHNOLOGIES IN FOREIGN LANGUAGE LESSONS. *Экономика и социум*, (5-1 (120)), 226-229.
9. Daminova, B. E., Bozorova, I. J., Turaeva, G. N., Husanova, G. Z., & Urolova, S. U. (2024). USING COMPUTER PRESENTATIONS IN TEACHING FOREIGN LANGUAGES. *Экономика и социум*, (5-1 (120)), 211-215.
10. Daminova, B. E., Bozorova, I. J., Tog‘Ayeva, R. T., Tajibayeva, A. K., & Zokirova, L. O. (2024). USE OF ELECTRONIC EDUCATIONAL RESOURCES IN THE PROCESS OF TEACHING A FOREIGN LANGUAGE. *Экономика и социум*, (5-1 (120)), 230-232.