

MOBIL DASTURLARDA NUTQNI MATNGA O'TKAZISH TEXNOLOGIYALARINING TAHLILI

To'ychiboyev Hamidullo Habibullo o'g'li

Namangan viloyati, Namangan shaxar

Namangan Davlat Texnika Universiteti, RTTM xodimi

<https://doi.org/10.5281/zenodo.17406523>

Annotatsiya. Mazkur tezisda mobil dasturlarda nutqni matnga o'tkazish texnologiyalarining ishlash tamoyillari, asosiy bosqichlari va zamonaviy neyron modellar asosidagi yondashuvlar tahlil qilinadi. Nutqni matnga o'tkazish jarayonining akustik modeli, til modeli va xususiyat ajratish bosqichlari ilmiy jihatdan o'rganilgan. Shuningdek, Android platformasida STT texnologiyasini joriy etish usullari va amaliy kod namunasi keltirilgan. Tadqiqot natijalari STT texnologiyalarining mobil tizimlardagi qulaylik darajasini oshirishi, foydalanuvchi bilan muloqotni tabiiylashtirishi hamda o'zbek tili uchun ham keng istiqbolli yo'nalish ekanini ko'rsatadi.

Kalit so'zlar: nutqni matnga o'tkazish, Speech-to-Text, STT, mobil dasturlash, sun'iy intellekt, tabiiy tilni qayta ishlash, neyron tarmoq, akustik model, til modeli, DeepSpeech, Whisper, Android API, ovozli interfeys, real vaqt tanib olish, o'zbek tili.

So'nggi yillarda mobil qurilmalar inson hayotining ajralmas qismiga aylandi. Ularning imkoniyatlari kengaygani sayin foydalanuvchi bilan o'zaro aloqani soddalashtirish va tabiiylashtirish masalasi dolzarb bo'lib bormoqda. Aynan shu maqsadda nutqni matnga o'tkazish (Speech-to-Text, STT) texnologiyasi mobil ilovalarning asosiy komponentlaridan biriga aylandi. Ushbu texnologiya foydalanuvchining ovozli nutqini tahlil qilib, uni raqamli matn shakliga aylantiradi. Bu esa mobil chat-botlar, ovozli yordamchilar, diktovka tizimlari, ta'lim ilovalari va ovoz orqali boshqariladigan servislar uchun asosiy yechim bo'lib xizmat qiladi. Mazkur tezisda nutqni matnga o'tkazish texnologiyasining ishlash prinsiplari, texnik bosqichlari, qo'llanilayotgan sun'iy intellekt modellarining tahlili hamda mobil dasturlashdagi amaliy yechimlari ko'rib chiqiladi.

Nutqni matnga o'tkazish (Speech-to-Text, STT) texnologiyasi dastlab 1950–1960-yillarda fonetik analizatorlar orqali sodda buyruqlarni aniqlashdan boshlangan. Dastlabki tizimlar fonetik signallarni sonli kodlar yordamida tahlil qilgan, ammo nutqning grammatik murakkabligi, urg'u, ohang va kontekstni tushunish imkoniyatiga ega bo'lmagan. Bugungi kunda esa STT tizimlari sun'iy intellektning chuqur o'rganish (Deep Learning) yondashuvlariga asoslanadi. Ayniqsa, Transformer va Self-Attention arxitekturalari nutq signallarini kontekstuallash orqali matnga yuqori aniqlikda aylantirish imkonini berdi. Masalan, Wav2Vec 2.0, Whisper, HuBERT, SpeechBrain kabi modellar hozirgi STT tizimlarining asosini tashkil etadi. Ular tovush signallarini fonemalarga emas, balki vektor fazodagi semantik vakilliklarga aylantiradi. Bu yondashuv tildan mustaqil ishlash imkonini beradi va ko'p tillilikni qo'llab-quvvatlaydi.

Nutqni matnga o'tkazuvchi mobil tizimlar uch pog'onali arxitekturaga ega:

1. Audio capture layer — foydalanuvchi nutqini real vaqt rejimida yozib oluvchi modul. U mobil qurilmaning mikrofonidan kiruvchi signalni PCM (Pulse Code Modulation) formatida qabul qiladi va 16 kHz chastotada namunalanadi.

2. Processing layer — bu qatlamda signalga raqamli ishlov berish amalga oshiriladi. Hamming oynasi, Fourier transformatsiyasi (FFT) va Mel filterbank funksiyalari yordamida nutqning spektral xaritasi olinadi. Shundan so'ng modelga yuborish uchun MFCC yoki log-Mel spektrogramlar shakllantiriladi.

3. Model inference layer — bu bosqichda neyron model (masalan, Whisper yoki DeepSpeech) audio xususiyatlarini matnga aylantiradi. Natija probability decoding orqali CTC (Connectionist Temporal Classification) algoritmi yordamida yakuniy matn shakliga keltiriladi.

Bu qatlamlarning mobil qurilmada ishlashi uchun yengillashtirilgan modellar, masalan, TensorFlow Lite, ONNX Mobile yoki PyTorch Mobile versiyalari qo'llaniladi. Ular RAM va CPU resurslarini tejaydi, bu esa real vaqtli ishlashni ta'minlaydi.

O'zbek tili fonetik jihatdan agglutinatив bo'lgani sababli, so'zlarning ko'plab shakllari mavjud. Bu esa STT modelining morfologik umumlashtirish imkoniyatini murakkablashtiradi.

Shuning uchun o'zbek tili uchun quyidagi yondashuvlar samarali hisoblanadi:

1. Subword tokenization (masalan, Byte-Pair Encoding — BPE) orqali yangi yoki qo'shma so'zlarni avtomatik tahlil qilish;

2. Custom language model yordamida o'zbekcha matnlar korpusidan til qoidalarini o'rganish;

3. Data augmentation — nutq tezligini o'zgartirish, shovqin qo'shish va ovoz diapazonini kengaytirish orqali ma'lumotlar bazasini boyitish.

Bugungi kunda o'zbek tili uchun ochiq manbali loyihalar paydo bo'lmoqda. Masalan, "Vosk Uzbek Model" va "Whisper uz-small" modellarining mavjudligi STT texnologiyalarini lokal loyihalarda joriy etish imkonini bermoqda.

Mobil dasturlarda STT texnologiyasi foydalanuvchi tajribasini sezilarli darajada yengillashtiradi. Quyidagi amaliy yechimlar bunga misol bo'la oladi:

1. Diktovka ilovalari: foydalanuvchi nutqini yozuvga aylantiruvchi oflayn rejimli dasturlar (masalan, Notta.ai, Gboard Voice Input).

2. Ovozli boshqaruv: IoT asosidagi ilovalarda foydalanuvchi ovoz orqali buyruqlar beradi.

3. Ta'lim platformalari: talaffuzni tahlil qilish, o'quvchi nutqini avtomatik yozuvga o'tkazish (masalan, Duolingo Speaking Practice).

4. Nogironligi bor foydalanuvchilar uchun ilovalar: STT texnologiyasi eshitish yoki qo'l harakatida cheklovga ega foydalanuvchilarga imkoniyat yaratadi.

Mobil dasturlashda STT integratsiyasining samarali yo'li — hibrid arxitekturalardan foydalanishdir: foydalanuvchi nutqi avval lokal model yordamida qayta ishlanadi, so'ng Internet mavjud bo'lsa, aniqlikni oshirish uchun serverdagi AI modeliga yuboriladi.

Nutqni matnga o'tkazish texnologiyasi mobil dasturlarda foydalanuvchi va tizim o'rtasidagi aloqa samaradorligini oshiruvchi muhim innovatsion yechimdir. Ushbu texnologiyaning rivojlanishi sun'iy intellekt, tabiiy tilni qayta ishlash va ovozli interfeyslarning kelajakdagi istiqboli bilan chambarchas bog'liq. O'zbek tilida STT tizimlarini takomillashtirish uchun maxsus ovozli korpuslar yaratish, mahalliy akustik modellarni o'qitish va ochiq manbali platformalar bilan integratsiya qilish zarur. Bu yo'nalishda olib boriladigan izlanishlar raqamli transformatsiya jarayonida muhim ahamiyat kasb etadi.

Foydalanilgan adabiyotlar:

1. Google Cloud. Speech-to-Text API Documentation. 2024. [elektron resurs]
2. OpenAI. Whisper Model Overview. 2024. [elektron resurs]
3. Jurafsky, D. & Martin, J. H. Speech and Language Processing. Pearson, 2023.
4. Mozilla. DeepSpeech: Open Source Speech Recognition Engine. 2022. [elektron resurs]
5. Baevski, A. et al. Wav2Vec 2.0: A Framework for Self-Supervised Speech Recognition. Facebook AI, 2021.
6. O‘zbekiston Respublikasi Raqamli texnologiyalar vazirligi. Sun’iy intellektni rivojlantirish strategiyasi 2021–2030. Toshkent, 2021.