

АНДАТПА

**6D070400 «Есептеу техникасы мен және бағдарламалық қамтамасыз ету»
мамандығы бойынша «Қазақ сөздерін дыбыс бойынша тануға арналған
машиналық оқу әдістері мен алгоритмдерін және мобильдік қосымшаны
құрастыру» тақырыбындағы кандидаттық диссертациясы**

Ғылыми жұмыстың өзектілігі

Сөйлеуді автоматты түрде тану (ASR) - бұл жасанды интеллект саласындағы динамикалық бағыт. Соңғы жарты ғасырда осы салада айтарлықтай жетістіктерге қол жеткізілді - осы саладағы инвестицияларды ақтайтын және табысты коммерциялық қосымшалар өте көп. Осындай қосымшалардың ішінде, ең алдымен, байланыс орталықтарын немесе IVR жүйесін (интерактивті дауыс мәзірі) - операторды айналып өтіп, ақпаратқа автоматты түрде қол жеткізу жүйелерін енгізуді атап өтуге болады. Қазіргі заманғы байланыс орталықтарында пайдаланушының сұрақтары табиғи тілде тұжырымдалса, оның жауабы да компьютер арқылы пайдаланушының тілінде синтезделеді. Байланыс орталықтарын енгізу біраз операторларды босатып, көптеген әуежайлар мен вокзалдарда қызмет көрсету сапасын жақсартты. Сөйлеуді автоматты түрде тану жүйелері медициналық зерттеулерде, мәселен, оператордың қолы бос емес кезде (рентген) немесе ішкі органдарды тексеруге арналған автономды аппаратпен жұмыс істеген кезде мәліметтерді енгізуді қажет ететін жағдайларда кеңінен қолданылады. Үздік медициналық мекемелерде орта буын қызметкерлерінің медициналық карталарды толтыруы да дауыспен жүзеге асырылады.

Сөйлеуді автоматты түрде тану және синтездеу жүйелерін қолданудың маңызды бағыты - бұл қимыл-тірек аппараты және көру қабілетінің бұзылуынан (көмекші технологиялар) зардап шегетін мүгедектерге көмек.

Қазақстанда сөйлеуді автоматты түрде тану жүйелері іс жүзінде қолданылмайды, бұл әзірлеушілер үшін осы саланы белсенді зерттеуге және игеруге жол ашатынын атап өткен жөн. Қазақ тілінде сөйлеуді тану саласындағы нашар зерттеулердің басты себебі - сөйлеу деректерінің жетіспеуі. Ағылшын, испан және қытай сияқты танымал тілдерде, сапалы ASR жүйелері өте көп деректерді қажет етеді. TIMIT немесе Switchboard сияқты танымал сөйлеу корпустарында телефонмен сөйлесу, ауызекі сөйлеу немесе таза

микрофон сөйлеу сияқты көптеген транскрипцияланған аудио жазбалар бар. Қазақ тілінде веб-дереккөздерде лайықты сөйлеу корпустары жоқтың қасы. Қолжетімді сөйлеу корпустары, әдетте, ақылы немесе қуатты әрі тиімді ASR модельдерін алу үшін жеткіліксіз. ASR жүйесіне лайықты сөйлеу корпусын құру үшін көп уақытты, жақсы құрылымдалған ортаны және берік бақылау жүйесін қажет етеді. Алайда, деректер тапшылығына байланысты мәселені толықтай шешу үшін нейрондық желінің құрылымы мен тәсілін қарастырған жөн.

Ресурстары аз көптеген тілдердегі сөйлеу деректері тіпті жоқ. Сондықтан сөйлеу корпустарын әзірлеу өте күрделі әрі өте көп уақытты қажет етеді. Қазақ тілі танымал болмағандықтан, ресурсы төмен тілдердің біріне жатады.

Адамдар арасындағы коммуникация сөйлеу, бейнелеу тілі, ым, ишарат тілі сияқты әр түрлі формада болуы мүмкін. Осы формалардың ішінде сөйлеу коммуникациясы тиімді әрі танымал болып есептеледі. Бұл адам мен компьютердің өзара әрекеттесуі сөздік қатынас арқылы жүзеге асатындығына әкеледі. Осылайша, бұл сөйлеуді автоматты түрде тану жүйесін құрудың маңыздылығын көрсетеді.

ASR жүйесінің жұмысы сөйлеу деректерінің сапасына тікелей байланысты. Дегенмен, сөйлеу корпустары жалпы тілге немесе пән аймақ тіліне негізделген болуы мүмкін. Пәндік аймаққа байланысты мәліметтер жиынтығын қолдану тану процесінің мүмкіндіктерін арттыруға үлес қосады. Сондай-ақ, пайдаланушының нақты сөйлеуін деректер жиынтығы ретінде пайдалану жалпы ASR өнімділігін арттырады.

Бұл қазақ тілі сияқты, ресурстары төмен тілдерге арналған ASR жүйесін құру кезінде сөйлеу деректерінің барабар жиынтығын табу үшін өте маңызды әрі шешуші фактор. Алайда ресурстар деңгейі төмен тілдер үшін сөйлеу деректерінің қажетті мөлшерінің жоқтығы айқын. Осылайша, сөйлеу мәліметтерін жинауға арналған құралдар сөйлеуді тану жүйелерін құруда маңызды рөл атқарады.

Зерттеудің мақсаты мен міндеттері

Деректер тапшылығын болдырмайтын тиісті ASR жүйесін құру үшін біздің негізгі міндеттеріміз:

- Веб-платформаны қолдана отырып, сөйлеу мәліметтерін жинауға арналған жақсы жобаланған ортаны құру
- Шағын сөйлемдерге арналған автоматты түрде сөйлеу жинау жүйесін құру үшін сөйлеу синтезінің моделін құру
- Қазақ тіліне арналған транскрипциясы бар сөйлеу деректерінің маңызды көлемін жинау
- Жиналған мәліметтерді өңдеу және нейрондық желі жұмысына арналған файлдарды құрылымдау
- Коннекциялық уақытша классификациясының жоғалту функциясы негізінде рекурренттік нейрондық желіні қолданып нейрондық желіні құру
- Трансферлік оқыту тәсілін қолдана отырып, орыс тілімен көптілді техниканы құру

Зерттеу нысаны

Зерттеу жұмысы сөйлеуді тану жүйелеріне арналған автоматты түрде сөйлеу мәліметтерін жинау әдістері мен тәсілдеріне бағытталған.

Зерттеу әдістері

Зерттеу жұмысы сөйлеуді тану, сөйлеу синтезі, артықшылықтар мен кемшіліктерін анықтайтын табиғи тілді өңдеу өрістеріндегі қазіргі заманғы туындылардың нәтижелерін талдау және түсіндіру арқылы орындалды. Белгіленген мақсаттарына машиналық оқыту алгоритмдері, рекуррентті нейрондық желісінің жуырдағы жетістіктері және sequence-to-sequence моделі арқылы қол жеткізілді.

Ғылыми жаңалық

Мақсаттарды орындау барысында алынған жаңалықтар:

- Жақсы ойластырылған бақылау және мониторинг жүйесі бар сенімді веб-сайт құру арқылы кез-келген тіл үшін сөйлеу мәліметтерін жинауда жаңа тәсілді қолдану

- Сөйлеу мен транскрипцияны нейрондық желіні оқытуды ескере отырып, СТС негізінде құрылымдайтын, жинау процесі аяқталғаннан кейін сөйлеу мәліметтеріне арналған пост-өңдеудің автоматты сатысын құру
- Алдын ала дайындалған орыс тіліндегі модель бойынша алған білімді беру арқылы көптілді тәсілді қолдана отырып, нейрондық желісін оқыту
- сөйлеудің енгізу ретін транскрипцияның шығу ретіне
автоматты түрде туралау

Зерттеу жұмысы бойынша ұсынылатын тұжырымдар

- Сөйлеу синтезін модельдеу негізінде сөйлеу мәліметтерін автоматты түрде жинауға арналған платформа
- Сәйкес мәтін транскрипциясы бар сөйлеу деректерін автоматты түрде алдын-ала өңдеу және құрылымдау.
- Рекурентті нейрондық желіні оқытуға арналған коннекциялық уақытша классификация алгоритм әдісі
- Орысша сөйлеуді тану моделін қолдана отырып, трансферлік оқыту тәсілін құру әдістері мен алгоритмдері
- Ұзақ мерзімді жадыны (LSTM) және екі бағытты LSTM рекурентті нейрондық желі негізіндегі үлгіде салыстырмалы талдау

Жұмыстың апробациясы

Зерттеудің нәтижелері халықаралық конференцияларда баяндалып, талқыланды: «Профессор Р.Г. Бияшевтың 80-жылдық және профессор М.Б. Айдархановтың 70-жылдық мерейтойына арналған «Информатика және қолданбалы математика» атты III Халықаралық ғылыми конференциясы» (Алматы, Қазақстан, 2018); Т.Н. Бияровтың, В. Вальдемардың 70 жылдық және И.Н. Амирғалиевтің 60 жылдық мерейтойына арналған «Информатика және қолданбалы математика» атты IV Халықаралық ғылыми-практикалық конференциясы» (Алматы, Қазақстан, 2019); «Заманауи ғылымның тенденциялары» (Шеффилд, Ұлыбритания, 2019);

Жарияланымдар

Диссертация тақырыбы бойынша 14 жарияланған жұмыс ұсынылды, оның ішінде:

- 1 монография
- Скопус индекстелген халықаралық журналдағы 2 мақала
- Қазақстан Республикасы Білім министрлігінің жоғары аттестаттау комиссиясының талаптарына сәйкес келетін 4 мақала
- Халықаралық конференцияларда жарияланған 4 мақала
- 2 авторлық куәлік / патент
- 1 шетелдік журнал басылымы

1. Амиргалиев Е.Н., Куанышбай Д.Н., “Вербальный робот”, Алматы: ИИВТ, 2020. -143 с, ISBN 978-601-08-0090-8
2. Kuanyshbay D.N., Amirgaliyev Y, Shoiynbek A., Yedilkhan D., “Automatic speech recognition system for kazakh language using connectionist temporal classifier”, Journal of Theoretical and Applied Information Technology, Vol. 98 No.04, PP 703-713, ISSN: 1992-8645(print), ISSN: 1817-3195(online) (SCOPUS: CiteScore:1.2, Quartile: Q3, percentile: 37)
3. Kuanyshbay D.N., Amirgaliyev B., Kutubayeva M., Baimuratov O., “Development of Automatic Speech Recognition for Kazakh Language using Transfer learning”, International Journal of Advanced Trends in Computer Science and Engineering, Vol. 9 No.04, ISSN: 2278-3091(online) (SCOPUS: CiteScore:1.2, percentile: 36, Quartile: Q3)
4. Kuanyshbay D.N., Kozhakhmet K, Shoiynbek A., “Comparison of two speech signal features for deep neural networks to classify emotions”, Вестник КазННТУ. Том 132 №2 2019, с.343-350 ISSN 1680-9211 (Импакт-фактор по Казахстанской базе цитирования за 2017 год = 0,045)
5. Kuanyshbay D.N., Amirgaliyev Y.N., Musabaev T.R., Kenshimov Sh., “Разработка голосовой идентификации диктора с использованием статистик контура основного тона для применения в робото-вербальных системах”, Вестник КазННТУ. Том 132 №2 2019, с.424-433 ISSN 1680-9211 (Импакт-фактор по Казахстанской базе цитирования за 2017 год = 0,045)

6. Куанышбай Д.Н., Амиргалиев Е, Едилхан Д., Баймуратов О., “Об одном улучшенном подходе автоматического распознавания казахской речи с использованием трансферного обучения”, Вестник КазНУ. Том 141 №5 2019, с.183-189 ISSN 1680-9211 (Импакт-фактор по Казахстанской базе цитирования за 2017 год = 0,045)
7. Kuanyshbay D.N., Kozhakhmet K, Shoiynbek A., “Various Languages impact on the problem of emotion recognition in speech”, Вестник КазНУ. Том 141 №5 2019, с.182-185 ISSN 1680-9211 (Импакт-фактор по Казахстанской базе цитирования за 2017 год = 0,045)
8. Kuanyshbay D.N., Amirgaliyev Y.N., Kozhakhmet K, Shoiynbek A., “Comparison of optimization algorithms of connectionist temporal classifier for speech recognition system”, "Informatyka, Automatyka, Pomiar w Gospodarcei-OchronieŚrodowiska" - IAPGOS, Vol. 9 No.3, PP 54-57, ISSN: 2083-0157(print), ISSN: 2391-6761 (online)
9. Куанышбай Д.Н., Кожамет К.Т., Шойынбек А., “Создание корпуса эмоций на казахском и русском языке для голосового распознавания эмоций”, MATERIALS OF XV INTERNATIONAL RESEARCH AND PRACTICE CONFERENCE «TRENDS OF MODERN SCIENCE - 2019» Vol. 14, PP 9-11, Science and Educational LTD, Sheffield, UK ,2019 ISBN 978-966-8736-05-6
10. Kuanyshbay D.N., Kozhakhmet K, Shoiynbek A., “Comparison of classification algorithms svm vs logistic regression for detecting crime”, МАТЕРИАЛЫ III Международной научной конференции «Информатика и прикладная математика», посвященная 80-летию юбилею профессора Бияшева Р.Г. и 70-летию профессора Айдарханова М.Б. 2018 года, Алматы, Казахстан, Часть 2 , с 37-41, ISBN 978-601-332-165-3
11. Kuanyshbay D.N., Amirgaliyev Y, Shoiynbek A., “Speech recognition preprocessing, background removal”, МАТЕРИАЛЫ III Международной научной конференции «Информатика и прикладная математика», посвященная 80-летию юбилею профессора Бияшева Р.Г. и 70-летию профессора Айдарханова М.Б. 2018 года, Алматы, Казахстан, Часть 2 , с 7-18, ISBN 978-601-332-165-3
12. Куанышбай Д.Н., Амиргалиев Е.Н., Шойынбек А., “Построение языковой модели для казахского языка на основе рекуррентных нейронных сетей”, МАТЕРИАЛЫ IV международной научно -практической конференции "Информатика и прикладная математика", посвященной 70-летию

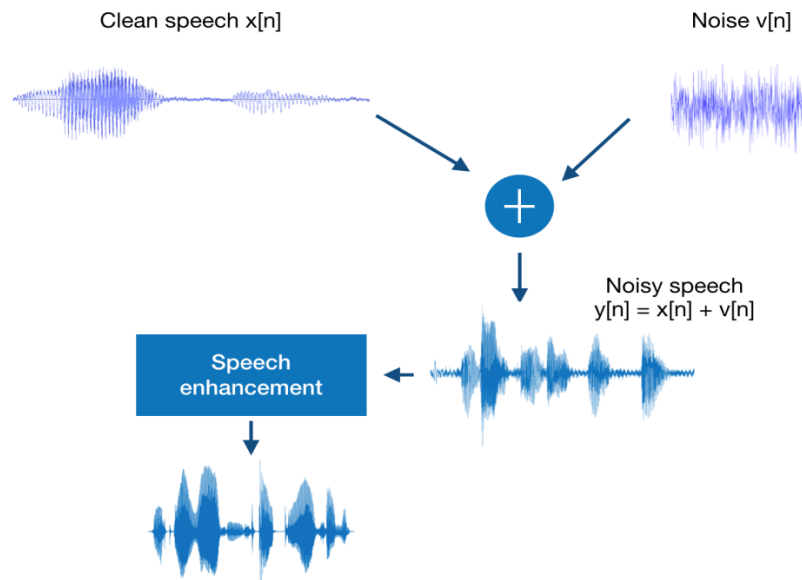
юбилею профессоров Биярова Т.Н., Вальдемара Вуйцика и 60-летию профессора Амиргалиева Е.Н. 2019, Алматы, Казахстан Часть 1 с. 401 – 406, ISBN 978-601-332-384-8

13. Куанышбай Д.Н., Баймуратов О., “Алгоритм синтеза речи казахского языка”, Свидетельство о внесении сведений в государственный реестр прав на объекты охраняемые авторским правом. Вид объекта авторского права: программа для ЭВМ. Запись в реестре №15029 от 10 февраля 2021 года
14. Куанышбай Д.Н., Амиргалиев Е.Н, Козбакова А.Х., “Автоматизированная система формирования корпуса речевых данных”, Свидетельство о внесении сведений в государственный реестр прав на объекты охраняемые авторским правом. Вид объекта авторского права: программа для ЭВМ. Запись в реестре №15236 от 17 февраля 2021 года

Зерттеу жұмысының құрылымы

Аталмыш диссертация кіріспеден, бес тараудан, қорытындыдан, пайдаланылған әдебиеттер тізімінен тұрады. Дипломдық жұмыстың көлемі 120 беттен, 30 суреттен, 124 әдебиет тізімінен және 10 қосымшадан тұрады.

Бірінші тарау сөйлеу сигналы өз-өзін өңдеу және сөйлеу сигналын өңдеу техникасының шолуынан тұрады. Онда шудың кең жолақты шу, кедергі сөйлеу және мезгілді шу сияқты түрлері туралы жалпы ақпарат беріледі. Сонымен қатар, тарауда сөйлеуді жақсарту жүйесі көрсетілген (1-сурет) және сөйлеуді күшейту техникасының әр түріне талдау жасалған.



1-сурет. Сөйлеуді жетілдірудің негізгі жүйесі

Шубасу және сөйлеуді жақсарту әдістерінің әртүрлі түрлері, мәселен, сызықтық болжамдық кодтау, сигналдың ішкі кеңістік әдісі, DFT негізіндегі әдістер және т.б. ұсынылды.

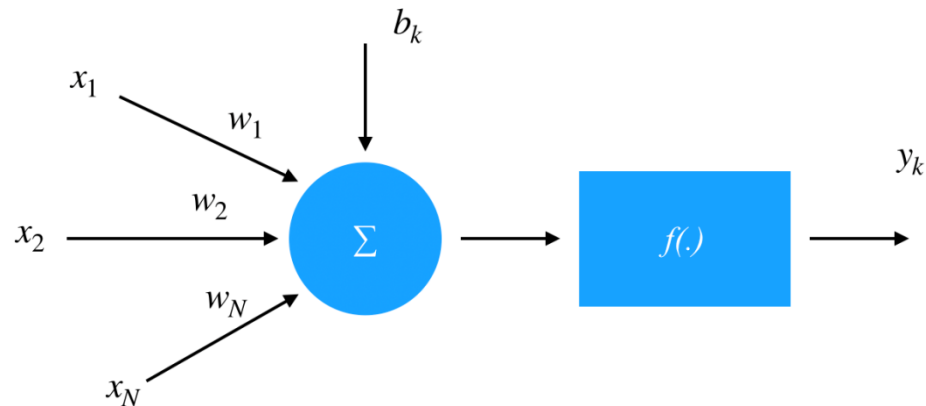
Екінші тарауда сөйлеуді автоматты түрде танудың қолданыстағы жүйелерін талдау қарастырылған. Біріншіден, ықтималдық әдіснама бойынша салынған Марковтың жасырын моделіне негізделген акустикалық модел қарастырылған. Тану үдерісінде көзге көрінбейтін сөздер бұрынғы шаблондармен салыстыру және ең жақынын сөзін табу арқылы танылады. Бірақ бұл шаблондар акустикалық өзгерістерге сәйкес келмейді. Сондықтан акустикалық модельдер акустика бойынша ықтималдықтың үлестірілуімен құрылады. Ең қарапайым тәсілі - ұсыну параметрлерін табатын Гаусс үлестірмесін қолдану.

Сөйлеуді тану тапсырмасы бастапқыда статистикалық жіктеу міндеті ретінде қарастырылды. W - үлкен сөздік қорынан сөздерінің тізбегі және X - сөйлеу ерекшеліктері ретінде анықталған енгізу деректері. Теңдеу келесідей:

$$P(W|X) = \frac{P(X|W)P(W)}{P(X)} \quad (8)$$

Екіншіден, тарау ANN негізіндегі акустикалық модельдерге шолу жасайды. Сөйлеуді тануда ANN-нің басқа модельдеу әдістеріне қарағанда басты артықшылықтарының бірі ол сызықтық емес динамикалық жүйелерді жақындата алады. Сөйлеу - бұл сызықтық емес жүйе арқылы жасалатын сызықтық емес сигнал.

ANN is basically an interconnection of computational elements (neurons) and this nonlinear system is distributed through the network. The basic nonlinear neuron model is shown in Figure 2.



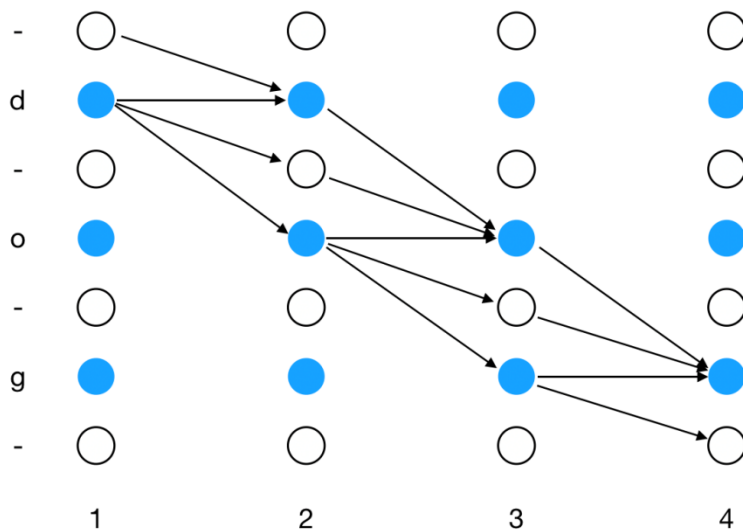
2-сурет. Сызықтық емес нейрондық модель

Бастапқы кезеңде, сөздерді тануға арналған ANN қарапайым цифрларды, біршама фонемалар мен көп қабатты перцептронды тану үшін кеңінен қолданылған.

Үшіншіден, тарау терең нейрондық желіге негізделген акустикалық модельдерге ұласады.

Терең сенім нейрондық желісі (DBNN) көптеген машиналық оқыту есептерін шешуде, сондай-ақ HMM / ANN негізінде сөйлеуді тану жүйесінде акустикалық модельдеуде қуатты әрі тиімді. Алғашқыда DBNN акустикалық модельдеу тапсырмалары үшін ұсынылды, өйткені оның модельдеу қабілеті қарапайым GMM-ге қарағанда әлдеқайда жоғары болды. Сонымен қатар, DBNN атқарымды табу үшін бақылаусыз оқыту және атқарымды оңтайландыру мен баптау үшін бақыланатын оқытуды біріктіретін өте тиімді дайындық қабілетін көрсетті. DBNN тиімді болуының тағы бір себебі - төменгі деңгейдегі объектісінің сипаттамалары төменгі деңгейлерде есептелінеді, ал енгізу деректерінің сызықтық емес құрылымы жоғары деңгейде ескеріледі. Бұл қажетті сипаттамаларды көрсету үшін бірнеше қабатты қолданатын адамның сөйлеу тануына ұқсас болуы мүмкін.

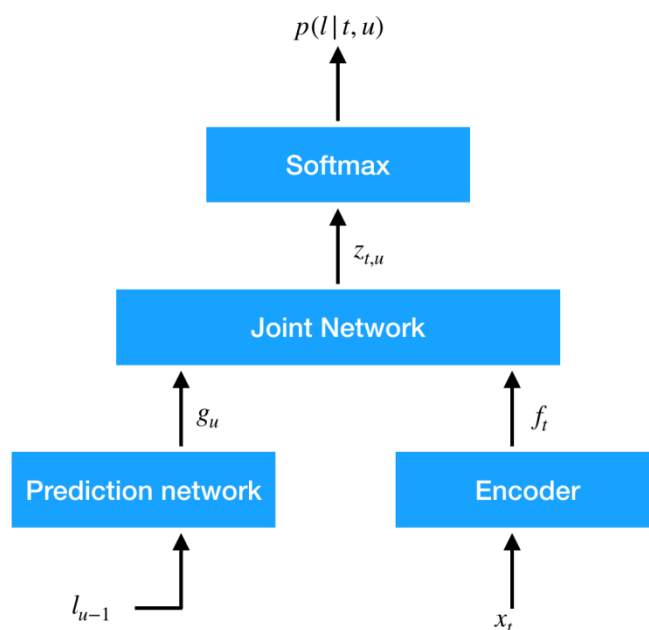
Үшінші тарауда, коннекциялық уақытша классификация және RNN түрлендіргіш негізіндегі ұштық моделдері сияқты нейрондық желіні оқытудың жаңа тәсілдері қарастырылады. CTC негізінде сөйлеуді тану моделі бойынша тренингтер талданып, түсіндірілді. CTC жолдың ықтималдығын есептей отырып, екі ретті сәйкестендіру үшін бос символды енгізеді. Сәйкесінше жолды біріктіру алгоритмі орындалады (3-сурет).



3 - сурет. «dog» белгісіне арналған жолдар мысалы

Келесі бөлімде RNN түрлендіргіш негізіндегі ұштық моделдері қарастырылады . RNN түрлендіргіш негізіндегі ұштық моделдері құрылымы бойынша CTC негізіндегі модельмен көптеген ұқсастықтарға ие. Мысалы, олардың екеуінде шығын функциясын бір; Екеуі де енгізу және шығару реттілігі арасындағы қолмен сегментациялау есептерін шешеді; олар «қара таңба» элементін пайдаланады; Олардың екеуі де барлық жолдардың ықтималдығын бағалайды және белгілердің ретін алу үшін жолдарды біріктіреді. Бірақ, жолды құру және жолдың ықтималдығын бағалау процестері өте әртүрлі. Сондай-ақ, тарауда бұл алгоритмнің CTC негізіндегі модельмен салыстырғанда артықшылықтары мен кемшіліктері анықталған.

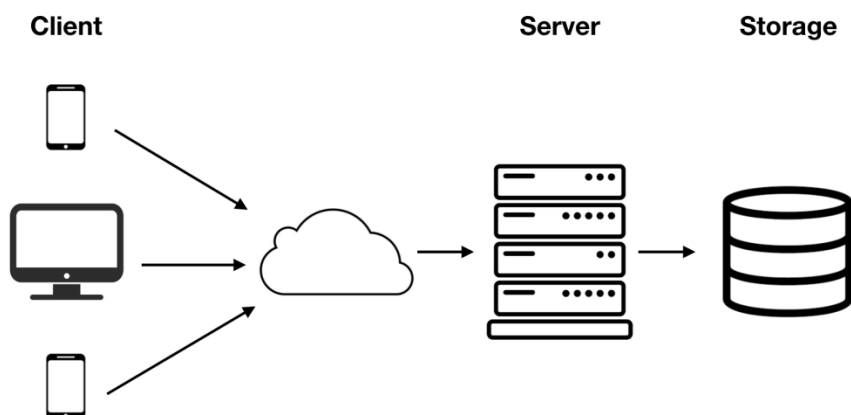
RNN түрлендіргіш негізіндегі модельдерінің үш маңызды компоненттері бар: транскрипцияға арналған желі ($F(x)$); болжам желісі ($P(y, g)$); бірлескен желі ($J(f, g)$) . RNN түрлендіргіш моделінің құрылысы 4-суретте көрсетілген.



4 – сурет. RNN түрлендіргіш құрылысы

Төртінші тарауда веб-қосымшаға негізделген сөздерді жинаудың автоматты жүйесін құру туралы ақпарат келтірілген (5-сурет).

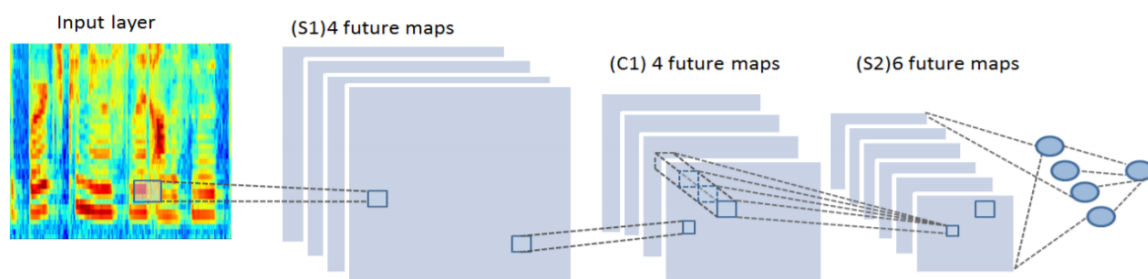
Бұл қосымша қазіргі заманғы нарықтағы кез-келген құрылғыға немесе компьютерге жарамды ең жаңа технологиялармен жасалған. Қосымшаның артқы жағы php7.2 тіліндегі yii2-негіздеменің көмегімен жасалды. Бұл жобаның архитектурасын әзірлеу Model-View-Controller (MVC) үлгісіне негізделген.



5 - сурет. Жүйе архитектурасының көрінісі

HTTP сервері ретінде біз NGINX қозғалтқышын пайдаландық. Қосымшаның дерекқорларын басқару үшін біз тілді MySQL негізінде қолдандық. Басқару панелі phpMyadmin және heidiSql арқылы өңделді. Кодты редакциялау үшін біз phpstorm технологиясын қолдандық. Қосымшалардағы әрбір веб-парақ ажыратымдылықты ағымдағы құрылғы негізінде өзгерту мүмкіндігіне ие.

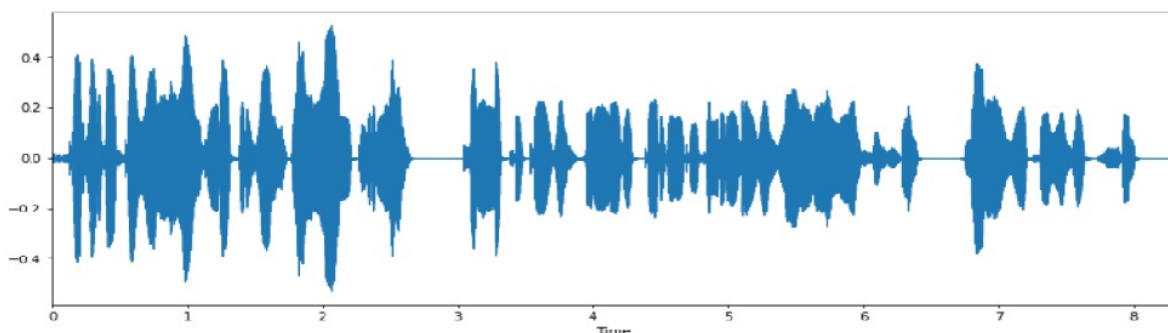
Веб қосымшасы қолданушы жүктеген кітаптан алынған шағын сөйлемдерді автоматты түрде айтатын сөйлеу синтезінің моделінен тұрады. Бұл сөйлеу синтезінің моделі конволюциялық нейрондық желіні (CNN) қолдана отырып, қазақ әдебиеті кітаптарында оқытылады (6-сурет).



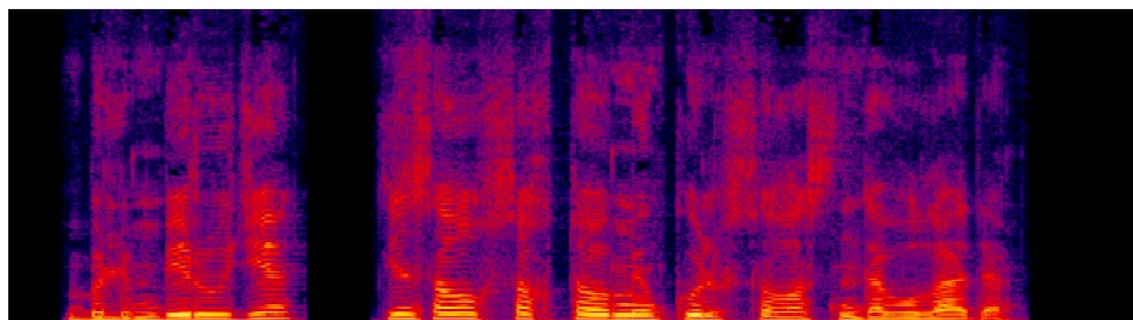
6 - сурет. CNN көп өлшемді желісінің архитектурасы

Сөйлеу синтезі жүйелерін құру кезіндегі маңызды міндеттердің бірі - сөйлеу сигналдарының мәліметтер базасын семантикалық және фонетикалық маңызды сөйлеу бірліктеріне бөлу және таңбалау, ал біздің жағдайда, бұл фонемалар. Алынған сегменттер мәліметтер базасында сақталады және интеллектуалды жүйеде акустикалық модельдерді машинада оқыту үшін мәтіннен сөйлеуге синтез жүйесінде спикер дауысының кейінгі генерациясымен қолданылады. Оқытылған жүйелермен жұмыс істеудің нақты әдістерінің бірі - белгілі бір тілге жүйелік параметрлерді орнату және алфавит таңдау. Фонемалар қазақ сөйлеу синтезаторының үлгісінде алфавиттің кіру таңбасы ретінде қолданылатындықтан, графиканың ережелері бойынша мәтіндерді фонемаларға қазақ тілінің фонетикалық ерекшеліктерін ескере отырып транскрипциялау мәселесін шешу қажет болды. Мәселе фонетикалық транскрипция модулін құру арқылы шешілді, содан кейін дайындалған мәтінді оқыту үлгісі транскрипцияланды.

Осылайша, 3500 сөйлемнен тұратын оқу-тәжірибелік базасы құрылды, олардың әрқайсысы wav форматындағы аудио файлға сәйкес келеді, оның іріктеу жиілігі 22050 Гц. Осыдан кейін жүйелік модуль іске қосылды, ол аудио файлдардың спектрограммаларын қабылдауға жауап береді, оның негізінде желілерді терең оқыту жүзеге асырылады (7-сурет).



білім беруде, денсаулық сақтау мен көлік саласында болады.



7 - сурет. Оқу базасынан ұсыныс үлгісі

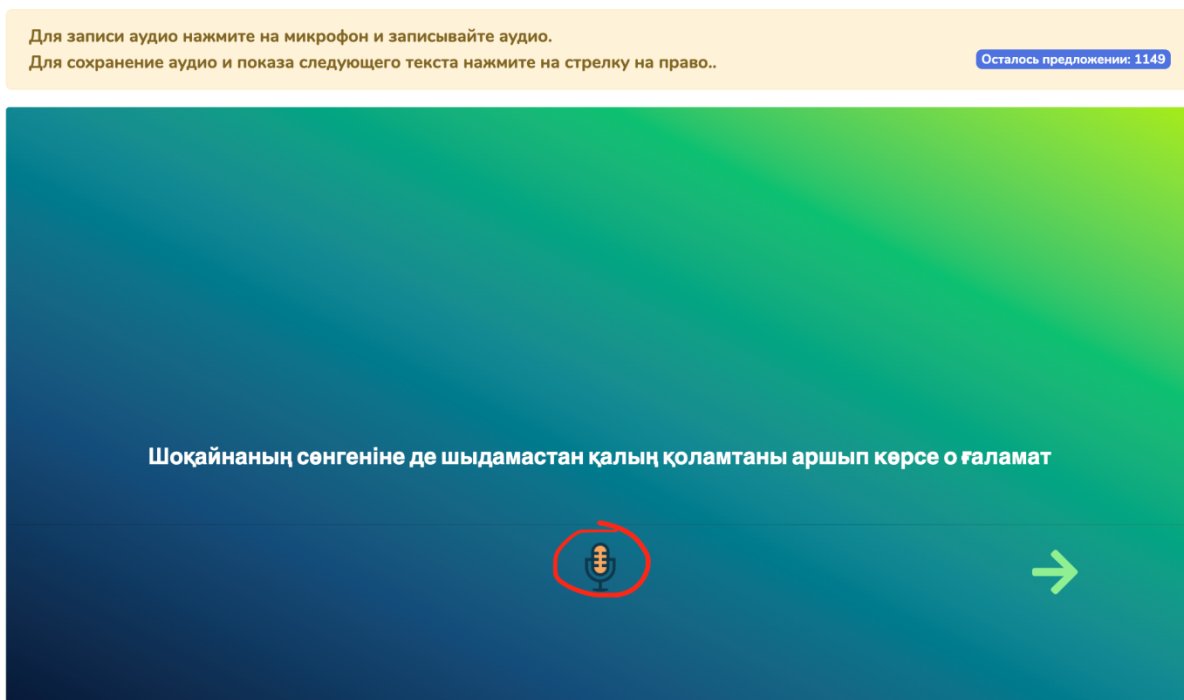
Келесі бөлімде веб-қосымшаның барлық функциялары мен мүмкіндіктерін сипаттайтын шолу жасалады.

Жүйе бүкіл мәліметтерді жинау процесін аяқтайтын екі маңызды рөлден тұрады. Бұл админ (тек 1-2 қолданушы) және саны өте көп клиенттер. Админ кез-келген форматтағы кез-келген кітап санын қосуға, жоюға және түзеуге құқылы. Админ белгілі бір кітаптың толықтық пайызын көре алады және клиенттердің барлық жазбаларын тыңдай алады. Ол кез-келген кітаптың жазбаларын, сондай-ақ кез-келген сөйлеушілерді тізімдей алады. Ол қажет емес немесе бұзылған жазбаларды өшіріп, қате жіберген сөйлеушілерге электрондық хат жібере алады. Барлық жазу процесі аяқталғаннан кейін админ барлық жазбаларды жүктей алады. Жүктелген zip-файл барлық қажетті транскрипциямен және аудиофайлдармен құрылымдалған.

Name	^	Date Modified	Size	Kind
 0_Farukh Iskalinov.txt		4/24/20	70 bytes	Plain Text Document
 0_Farukh Iskalinov.wav		4/24/20	352 KB	Waveform audio
 1_Farukh Iskalinov.txt		4/25/20	13...ytes	Plain Text Document
 1_Farukh Iskalinov.wav		4/25/20	639 KB	Waveform audio
 2_Baimolda Aray.txt		4/13/20	11...ytes	Plain Text Document
 2_Baimolda Aray.wav		4/13/20	524 KB	Waveform audio
 3_Farukh Iskalinov.txt		4/25/20	12...ytes	Plain Text Document
 3_Farukh Iskalinov.wav		4/25/20	606 KB	Waveform audio
 4_Farukh Iskalinov.txt		4/25/20	44 bytes	Plain Text Document
 4_Farukh Iskalinov.wav		4/25/20	336 KB	Waveform audio

8 - сурет. Файлдардың құрылымы

Клиент тек веб-сайтқа аты-жөнін, жынысы мен жасын енгізу арқылы тіркеліп, одан әрі жазба процедурасы үшін админ кез-келген кітапты таңдай алады. Клиенттің жақсарту мақсатында өзінің жазбаларын тыңдаудан басқа кез-келген кітапты қосуға, жоюға немесе түзеуге құқығы жоқ. Жазу процесінде клиент өзінің жазбасын бірнеше рет тыңдағаннан кейін сөйлемді қайта жаза алады.



9 - сурет. Жазба беті

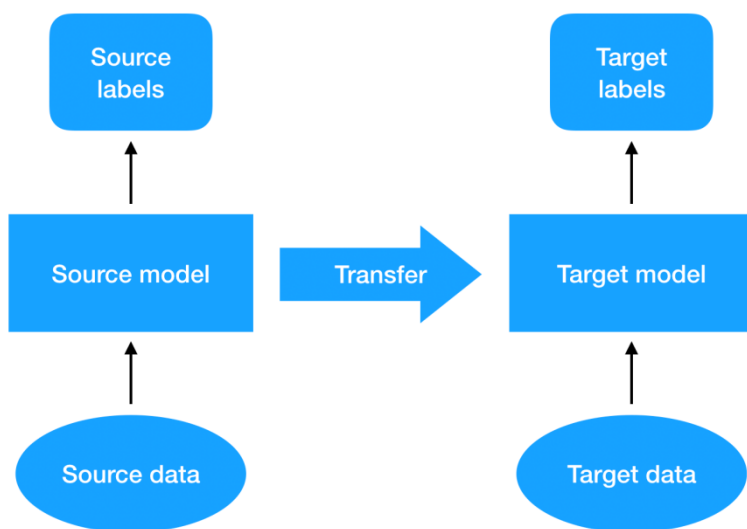
Келесі бөлімдерде сөйлеушілерді жазу (9-сурет), сөйлемдерді қайта жазу, бүлінуді қолмен тексеру, жазбалар тізімі (10-сурет), сақталған файлдардың құрылымы (8-сурет) және т.б. сияқты сөйлеушілердің функциялары түсіндіріледі.

Зерттеушілерге мұндай технологияның пайдасы өте зор, себебі ол көптеген халықты қамтамасыз етіп, қысқа мерзім аралығында деректер жинауды ұсынады. Болеe того, он обеспечивает хорошее качество аудиозаписи речи, благодаря возможности контролировать и контролировать весь процесс записи. Қазір барлық нұсқаулар мен ережелер қазақ тілінде ұсынылған, бірақ оларды ресурстары төмен кез-келген тілге бейімдеуге болады. Осылайша, сөздерді ыңғайлы түрде жинауға көмектесу арқылы бұл құрал танымал тілдерді кеңінен насихаттауға үлкен үлес қосады.

Предложения книги	Аудиозапись	Дата создание	Скачать все
Сілекейімді жия алмай қойдым № 1		01-05-2020 00:10:14	Скачать Удалить
- дедім газеттің ішкі бетін ашып жатып № 2		01-05-2020 00:10:08	Скачать Удалить
Он екі биені күніне бес рет сауғанда нешеу болады отыз бола ма № 3		01-05-2020 00:09:52	Скачать Удалить

10 - сурет. Жазбалардың тізімі

Бесінші тарау сөйлеуді автоматты түрде тану (ASR) жүйесін алдыңғы тарауда жиналған сөйлеу мәліметтерін қолдана отырып жүзеге асырады. Жалпы мәліметтер 100 сағаттық баяндамалардан тұрады. Нейрондық желіні құрудың негізгі идеясы - трансферлік оқыту тәсілі (11-сурет).



11 - сурет. Трансферлік оқытудың жоғарғы деңгейі

Алдын-ала дайындалған орысша сөйлеуді тану моделін қолдана отырып, барлық салмақ біздің жеке нейрондық желімізге берілді. Орысша сөйлеуді тану моделі 100 сағаттық сөйлеу мәліметтерін қамтитын VoxForge деректер жиынтығында оқытылды.

Біз өз тәжірибемізде RNN-нің екі түрін өткіздік: ұзақ мерзімді жад (LSTM) және екі бағытты LSTM.

Алдын-ала дайындалған орысша сөйлеуді тану моделін қолдана отырып, барлық салмақ біздің жеке нейрондық желімізге берілді. Ресейлік сөйлеуді тану моделі 100 сағаттық сөйлеу мәліметтерін қамтитын VoxForge деректер жиынтығында оқытылды.

Біз өз тәжірибемізге RNN-нің екі түрін өткіздік: ұзақ мерзімді жад (LSTM) және екі бағытты LSTM. BiLSTM соңғы зерттеулерде үміті зор нәтижелер көрсеткендіктен және контекстті жақсы түсінетіндіктен, біз оны әдеттегі LSTM-мен салыстыруды жөн көрдік. Жоғарыда айтып өткеніміздей, біз Jupiter Notebook-ты осы эксперименттің ортасы ретінде Python бағдарламалау тілінде қолдандық. Атап айтқанда, біз ASR моделін құру және оқыту үшін Python негізіндегі Tensorflow кітапханасын пайдаландық. Нейрондық желіні құру үшін Keras Tensorflow кітапханасының кеңейтімі пайдаланылды. Сөйлеу жиынтығы GitHub репозиторийінен клондалған. Репозиторий үш папкадан тұрады: пойыз жиынтығы, валидация жиынтығы және тест жиынтығы. Оқыту бірнеше (GPU) Tesla K80 графикалық процессорларында өткізілді. Осы процессорлары бар сервер үш айға пайдалануға жалға алынды.

Біздің экспериментте қолданылған параметрлер тізімі:

- Бөлек LSTM және BiLSTM екі қабаты
- әр қабатта 128 нейрондық бірлік
- 500 оқу кезеңі
- әр LSTM қабатынан кейін 50% ықтималдықпен тастау қабаты
- Топтаманың мөлшері - 4
- Momentum Optimizer импульс мәні 0,9 құрайды

- Оқу деңгейі 0,0005

- жоғалту функциясы - CTC

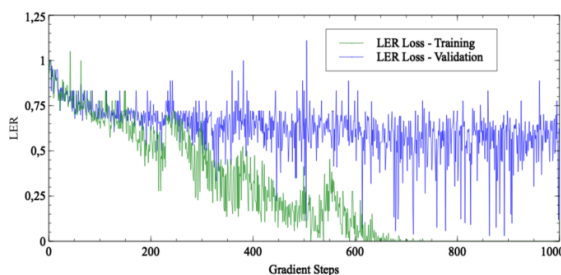
- Көрсеткіш - жапсырманың қателік коэффициенті (LER)

Біз 4 түрлі сценарийді қарастырдық: 1) орыс тілінің моделін қолданбаған LSTM нейрондық желісі; 2) орыс тілінің моделін қолданған LSTM жүйке желісі; 3) орыс тілінің моделін қолданбаған екі жақты LSTM; 4) орыс тілді моделін қолданған екі бағытты LSTM.

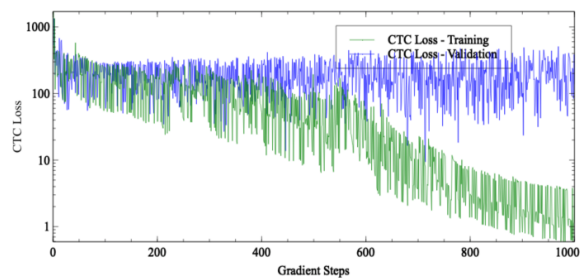
Олар әр қабатта бірдей қабатты және бірдей нейронды қолданды. Оқыту процесінің нәтижелерін 2-кестеде бағалауға болады. Әрбір рекурентті нейрондық жүйенің нәтижесі өте жақын, бірақ біз трансферлік оқудың архитектурасы бәрін жетілдіріп жатқанын көреміз.

Сыртқы моделі бар LSTM қабатты нейрондық желісі оқыту шығындарын 8% -ға дейін төмендетті, ал қателіктер коэффициенті 4% -ға өсті. Екі бағытты LSTM өте перспективалы нәтижелер көрсетті, оқыту құнын 24% -ға дейін арттырды және жапсырманың қателік деңгейі 32% -ға дейін азайды (12, 13-сурет).

Тәжірибе көрсеткендей, орысша ASR жүйесінің сыртқы моделін пайдалану арқылы білімді қазақ тілі жүйесіне ауыстыру өнімділікті айтарлықтай жақсартады.



12 - сурет. LER иллюстрациясы



13 - сурет. CTC шығынының иллюстрациясы

Қорытынды

Сөйлеуді тану саласын зерттеу барысында барлық мақсаттар мен міндеттерге қол жеткізілді:

- Сөйлеу туралы деректерді автоматты түрде жинауға арналған жақсы ойластырылған орта және бақылау жүйесі бар платформа құрылды. Ол конволюциялық нейрондық желімен оқытылған сөйлеу синтезінің моделін қамтитын веб-қосымшаға негізделген.
- Сөйлеу мәліметтерін жинау құралын пайдалану арқылы 50 сағаттан астам деректер жиналды. Мәліметтер жинау барысында сөйлеушілердің жалпы санынан 83% ерлер, 17% әйелдер.
- Орысша сөйлеуді танудың алдын-ала жасалған моделін қолдана отырып, трансферлік оқыту әдісі қазақша ASR жүйесінде қолданылды. Ұзақ мерзімді қысқа мерзімді жадыға негізделген нейрондық желіні құру және барлық салмақтарды орысша сөйлеуді тану моделінен өту арқылы сөйлеуді танудың көптілді моделіне қол жеткізілді.

Сөйлеуді жинау ортасын қолдана отырып, 65 тіл иесі жұмылдырылды және 1,5 айдан аз уақыт ішінде 50 сағаттан астам таза сөйлеу мәліметтері жиналды.

Қазақша сөйлеуді тану жүйесі LSTM және BiLSTM қабаттарына негізделген нейрондық желіні қолдану арқылы оқытылды. Трансферлік оқытуды қолдана отырып көптілділік тәсіл (алдын-ала жасалған орыс моделі) белгілердің қателік коэффициенті тұрғысынан қазақша ASR моделін 24% арттырды.