

«6D060200 – Информатика» мамандығының PhD докторанты
Мекебаев Нұрбапа Отанұлының
«Сөйлеулерді тану есептерінде машиналық оқытуды қолданып
белгілерді анықтау және өңдеу алгоритмдерін зерттеу және құру»
тақырыбындағы диссертациялық жұмысына

АҢДАТПА

Зерттеу тақырыбының өзектілігі. Қазіргі уақытта информатиканың өзекті мәселелерінің бірі - сөйлеуді тану проблемасы. Компьютерлік жүйелерді пайдалану тиімділігі тікелей осыған байланысты, өйткені сөйлеу - бұл адамның қарым-қатынас жасаудың ең кең таралған және табиғи құбылыс болып саналады және ол ақпаратты енгізу және мобильді жүйелерді басқару процесін едәуір жылдамдатады. Ақпараттық технологиялар қарқынды дамып келеді және ақпарат алмасуда кеңінен қолданылуда. Осыған байланысты сөйлеуді танудың дамуы маңызды рөл атқарады.

Күнделікті өмірде тіл - адамзат жаратылысының табиғи көрінісі. Ғылым мен техниканың дамуы барысында ғалымдар мен инженерлер көптеген жылдар бойы пайдаланушы мен машина арасындағы ауызша байланыс мәселесін зерттеп келе жатқаны бәріне мәлім.

Көптеген компаниялар мен жеке әзірлеушілер сөйлеуді тану технологияларын жасауда біршама жетістіктерге жеткенін мойындау керек, бірақ олардың Қазақстанда әлі де кең қолданыста еместігін мойындау керек. Бұл сөйлеушінің сөйлеу мәнерінің ерекшеліктеріне және дыбыстық кедергілердің болуына байланысты.

Осыған байланысты, мәселені шешудің ең өзекті міндеті - сөйлеуді автоматты тануда сөйлеушіні анықтау.

Қолданыстағы сөйлеуді тану жүйелерін оңтайландыру адамның компьютермен өзара қарым-қатынасының тиімділігін едәуір жеңілдетуге және арттыруға мүмкіндік береді. Сондай-ақ сөйлеуді тану жүйелерін қолдану құқық қорғау қызметінің жұмысында аса үлкен маңызға ие.

Қарастырылған мәселе бойынша зерттеулердің өзектілігі қазіргі жүйелердегі бірқатар факторлармен күрделенеді - тілдердің әр түрлі құрылымы, шуылмен бірге өңделетін сөйлеу сигналдар нәтижелерінің төмендігімен және нәтиженің дикторға тәуелділігімен, жүйелердің жұмыс жылдамдығының жоғары болмау мәселелерімен түсіндіріледі.

Сөйлеуді танудың қазіргі заманғы жүйелері негізінен жасырын Марковтық модельдерді (ЖММ) қолдануға негізделген, олар сөйлеу кезінде бір фонеманың екінші бір фонемамен үйлесу ықтималдығының мүмкін екендігін анықтайды. Осылайша Гаусс құрамасы (ГҚ) арқылы белгілердің ықтимал бөліктерге бөлуін модельдеу көмегімен, бақыланып отырған сигналдың вариативтігін қамтамасыз етеді. Бұл әдісті 1989-жылы Лоуренс Робимер ұсынды және ол ұзақ уақыт сөйлеу сигналын модельдеуге негіз болған. Deep Belief Networks өзінің қарқынды дамуы арқасында ЖММ-не балама болуда және тану үдерісінде жоғарғы дәлдікті қамтамасыз етіп отыр.

Л.Райнердің еңбектері жарияланған уақыттан бері, сөйлеуді автоматты тану жүйелеріндегі сөйлеу сигналын сипаттау үшін мел-кепстралдық коэффициенттер пайдаланылады (MFCC Mel Frequency Cepstral Coefficients), олардың даму негізін Пол Мермельстайн қалаған болатын. Сонымен қатар, соңғы уақыттарда қазіргі қолданыстағы MFCC диктордың сөйлеу ағынының вариациялығына төзімді белгілер балама болуда, ол мейлінше сенімді жүйелерді құруға септігін тигізеді.

Қазіргі уақытта дауыстық биометриялық жүйесін әзірлейтін көптеген шетелдік (Agnitio, Nuance, Voice Security Systems), оның ішінде ресейлік (Speech technologies, Speech Technologies Centre) компаниялар бар, жасалған жүйелердің көпшілігінде сөйлеушіні анықтау қателігінің ықтималдығы 1-3% құрайды, бірақ бұл қосымшалардың бірқатар кемшіліктері бар.

Сөйлеуді тану есептерінің бірі сөйлеушіні анықтауда әлемдік тәжірибеде біршама жетістіктерге қол жеткізген Карнеги-Меллон университеті (АҚШ), Иллинойс университеті (АҚШ), Орегон университеті (АҚШ), шығыс Финляндия университетінде Томи Кинунен бастаған зерттеушілер тобы және т.б. шетелдік университетер біршама жетістіктерге жеткен. Украинаның Донецк қаласындағы «Информатика және жасанды интеллект мемлекеттік университетінің» ғалымы В.Ю.Шелепова, Ресей ғылым академиясының Санкт-Петербург мемлекеттік университетінің ғалымы А.А.Карпов және де Аграновский А.В., Леднов Д.А., Балакирев Н.Е., Малков М.А., ал отандық ғалымдардан Л.Н.Гумилев атындағы Еуразия ұлттық университетінің ғалымы А.Шәріпбай, Әл-Фараби Қазақ ұлттық университетінің ғалымы У.А. Түкеев, Назарбаев университетінің ғалымы Ж.А.Есенбаев, ҚР БҒМ ҒК Ақпараттық және есептеуіш технологиялар институтының ғалымдары Әмірғалиев Е.Н., Мусабаев Р.Р., Мамырбаев О.Ж. айналысқан.

Машиналық оқыту жасанды интеллект саласының бір бағыты ретінде сөйлеулерді тану үдерісінің негізгі бір бөлшегі болып қалыптасты. Ақпараттық-коммуникациялық технологияларды жылдам басқаруда сөйлеу интерфейстерін құру, әртүрлі сөйлеулерді тану есептері бүгінгі күнде толығымен шешілмеген, өзекті мәселелердің бірі болып табылады.

Елімізде жасанды интеллект жүйесі кешенді түрде дамуда. Осы үдерісте аса өзектілікті әліде шешуін күтетін мәселелер аз емес. Солардың бірі – сөйлеулерді автоматты тануда, сөйлеушіні анықтау әлі жете зерттелмегені тақырыптың өзектіліктігін анықтайды.

Диссертациялық жұмыстың мақсаты: Сөйлеуді тануда машиналық оқытуды қолдана отырып, сөйлеуді танудың оңтайлы алгоритмдері мен модельдерін құру және зерттеу.

Зерттеудің міндеттері. Белгіленген мақсатқа қол жеткізу үшін төмендегі міндеттерді шешу қажет:

- Сөйлеуді тану саласындағы сөйлеу белгілерін және сөйлеушіні анықтауға арналған әдістер мен заманауи жүйелерге сараптау жүргізу;

- Сөйлеуді тану үдерісінде сөйлеушінің дыбыстық сөйлеу белгілерін және сөйлеушінің мәліметтерін анықтауға арналған акустикалық корпусын құру;

- Сөйлеушіні танудағы заманауи классификациялық моделдер мен алгоритмдерді салыстыру;

- Нейрондық желілер негізіндегі гендерлік ерекшелігі мен сөйлеушіні анықтаудың моделі мен алгоритмін құру;

Зерттеу нысаны. Сөйлеуді тану және анықтау жүйесі.

Зерттеу пәні: Сөйлеудің акустикалық деректері және сөйлеуді автоматты түрде тануда сөйлеушіні анықтау әдістері мен алгоритмдері болып табылады.

Зерттеу әдісі: Ақпараттар теориясы, сигналдар теориясы, нейрондық желілер теориясы және технологиялары, сөйлеулерді тану әдістері мен технологиялары, бейне тану теориясы мен технологиялары, бағдарламалық камтама жобалау және құру технологиялары.

Жұмыстың ғылыми жаңалығы:

- Сөйлеуді тану үдерісінде сөйлеушінің дыбыстық сөйлеу белгілерін және сөйлеушінің мәліметтерін анықтауға арналған акустикалық корпусы құрылды;

- Машиналық оқыту саласындағы классификациялық алгоритмдер көмегімен және осы алгоритмдердің дәлдігін арттыра отырып сөйлеуші анықталды;

- Нейрондық желілер негізіндегі гендерлік ерекшелігі мен сөйлеушіні анықтаудың моделі мен алгоритмі құрылды;

- Зерттеу барысында алынған модель мен алгоритм көмегімен сөйлеу белгілерін және сөйлеушіні анықтауға арналған бағдарламалық қосымша құрылды.

Жұмыстың теориялық және практикалық маңызы. Зерттеу жұмысының теориялық маңызы сөйлеулерді тануда гендерлік ерекшеліктерін анықтауға арналған нейрондық желі модельдері мен алгоритмдерін жетілдіру болып табылады. Сонымен қатар қазіргі кезде пайдаланылып келе жатқан әдістердің жетілдірілген және ерекшелігі бар сөйлеу белгілерін анықтайтын жаңа әдістер әзірлеумен және эксперименттік зерттеумен сипатталады.

Диссертациялық зерттеудің практикалық маңызы сөйлеу сигналдарының белгілерін анықтау кезінде әзірленген, жетілдірілген нейрондық желі моделдерін қолдану арқылы құрылған акустикалық корпус сөйлеуді тану саласында зерттеу жұмыстарын жүргізуге мүмкіндік береді.

Қорғауға шығарылған негізгі тұжырым. Сөйлеулерді тану және сөйлеу белгілерін анықтауға арналған 36 сағаттан астам сөйлеулерден тұратын қазақ тілінің акустикалық корпусы жасалған. Деректер базасында тіркелген сөйлеушілердің дыбыс ерекшеліктерін автоматты түрде сүзгілеуден өткізіп, сөйлеулер белгілерін анықтау.

Сөйлеу белгілерін анықтауда нейрондық желілердің жалпы құрылымы мен жетілдірілген архитектуралары қарастырылып, гендерлік ерекшеліктерін анықтауда қолданатын алгоритмдер мен модельдер құрылды.

Зерттеушінің жеке үлесі. Ізденушінің жеке өзі диссертациялық жұмыстың қойылған міндеттерін шешті. Гендерлік ерекшелігі мен сөйлеушіні анықтаудың алгоритмі мен моделін әзірледі. Әзірленген алгоритмдердің тиімділігін эксперименттік бағалау жүргізілді. Гендерлік ерекшелігі мен сөйлеушіні анықтау моделі құрылды.

Диссертация тақырыбының ғылыми-зерттеу жұмыстарының жоспарларымен байланысы. Диссертация бойынша зерттеу жұмыстары Білім және ғылым министрлігі Ғылым комитеті Ақпараттық және есептеуіш технологиялар институтында «Терең нейрондық желілерді пайдаланатын мультитілдік автоматты сөйлеуді тану технологиясын құру» гранттық қаржыландыру жобасы аясында орындалды. (2018-2020 жж., мемлекеттік тіркеу нөмірі: 0118РК00139).

Жұмыс көлемі мен құрылымы. Диссертациялық жұмыс кіріспе, 3 тарау, қорытынды және пайдаланылған әдебиеттерден тұрады. Диссертацияның толық көлемі: 108 бет жазба мәтіні, соның ішінде 35 сурет, 10 кесте. 110 пайдаланылған әдебиеттер тізімінен және 2 қосымшадан тұрады.

Кіріспе бөлімінде жұмыстың өзектілігі анықталды, зерттелетін тақырыпқа байланысты мәселелер көрсетілді. Жұмыстың идеясы, зерттеудің мақсаты мен міндеттері, жұмыстың ғылыми жаңалығы және практикалық құндылығы, зерттеу әдістері берілген.

Бірінші бөлімде сөйлеу сигналын алдын ала өңдеудің жолдары мен белгілерін анықтаудың ерекшеліктері, сөйлеу сигналдарының сипаттамасы, сөйлеуді тану және белгілерін анықтауға арналған әдіс-тәсілдер мен моделдерге талдау жасалды.

Екінші бөлімде машиналық оқыту алгоритмдері мен модельдерін сөйлеуді тану есептерінде қолдану және машиналық оқытудағы нейрондық желілер қарастырылды. Сөйлеуді тану және белгілерін анықтауға арналған акустикалық корпус құрылды. Сөйлеушіні анықтауға арналаған классификациялық алгоритмдерге талдау жасалып SVC жоғары нәтиже көрсетті.

Үшінші бөлімде сөйлеулерді тану есептерінде машиналық оқытуды қолданып белгілерді анықтау және өңдеу алгоритмдері мен модельдері құрылды. Сөйлеу сигналын алдын ала өңдеуде MFCC-ті қолданып гендерлік ерекшелігі анықталды. Гендерлік ерекшелігі мен сөйлеушінің дыбыс ерекшеліктерін тануға арналған MLP және CNN нейрондық желі архитектураларына салыстырма жасалып CNN жақсы нәтиже көрсеткені анықталды.

Қорытындыда диссертацияның негізгі нәтижелері мен тұжырымдары баяндалып, оның болашақ жұмыстармен байланысы көрсетілген.

Сенімділік дәрежесі мен апробациялау нәтижелері. Зерттеудің сенімділігі мен нәтижелерінің негізділігі міндеттердің қойылудың негізделінген жауапкершілігімен, критерийлердің және берілген саладағы зерттеулердің жай-күйінің сарапталуымен, жүргізілген эксперименттер санының көптігімен, сондай-ақ олардың практикаға табысты енгізілуімен

камтамасыз етіліп дәйектеледі. Диссертация нәтижелері мына төмендегі ғылыми-әдістемелік конференцияларда баяндалып талқыланды:

1. «Төртінші өнеркәсіптік революция жағдайындағы дамудың жаңа мүмкіндіктері» атты ҚР Президенті Н. Назарбаевтың Жолдауын іске асыру шеңберінде «Көліктегі инновациялық технологиялар: білім, ғылым, тәжірибе» атты ХЛІІ Халықаралық ғылыми-практикалық конференциясында (Алматы, 18 сәуір, 2018);

2. Профессор Р.Г. Бияшевтың 80 жылдығына және профессор М.Б. Айдархановтың 70 жылдығына арналған «Информатика және қолданбалы математика» атты ІІІ Халықаралық ғылыми-практикалық конференциясында (Алматы, 26-29 қыркүйек 2018);

3. Мамырбаев О.Ж., Н.О.Мекебаев., М.Тұрдалыұлы. Идентификация диктора используя mfcc // 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. – Алматы, 2019. – Т. 2. – С. 384-392.

4. 3rd International Conference Applied Mathematics, Computational Science and Systems Engineering (Рим, Италия, 2018);

5. 11th Asian Conference on Intelligent Information and Database Systems (Yogyakarta, Индонезия, 6-12 сәуір 2019 ж.).

Диссертация тақырыбы бойынша 20 мақала жарияланды және 2 авторлық куәлік алынды:

1. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Тұрдалыұлы. Сөйлеулерді тану үрдісінде MFCC алгоритмін қолдану // ҚазҰТЗУ хабаршысы. – 2018. -№ 2 (126). – Б. 389-392.

2. О.Ж.Мамырбаев, Н.О.Мекебаев, М.Тұрдалыұлы. Генетикалық алгоритм көмегімен сөйлеуді автоматты танудағы гендерлік сәйкестендіру // Алматы энергетика және байланыс университетінің хабаршысы. – 2018. – специальный выпуск. – Б. 120-129.

3. О.Ж.Мамырбаев, М.Тұрдалыұлы, Н.О.Мекебаев. Система распознавания слитной казахской речи на основе глубоких нейронных сетей // Вестник Алматинского университета энергетике и связи. – 2018. – специальный выпуск. – С. 130-135.

4. О.Ж.Мамырбаев, М.Тұрдалыұлы, Н.О.Мекебаев. Кіріккен қазақ сөйлеуін тану жүйесі // ҚБТУ хабаршысы. – 2018. - № 3 (46). – Б. 129-133.

5. О.Ж.Мамырбаев, М.Тұрдалыұлы, Н.О.Мекебаев, Ахметов И. MFCC негізіндегі дикторды анықтау жүйесі // ҚазҰТЗУ хабаршысы. – 2019. – № 2 (132). – Б. 155-160.

6. Мамырбаев О.Ж., М.Тұрдалыұлы, Т.Тұрдалықызы, Шаяхметова А.С. Мекебаев Н.О. Автоматическое распознавание казахской речи с использованием DNN // ҚБТУ хабаршысы. – 2019. – №2 (49). – Б. 134-142.

7. Мамырбаев О.Ж., Кыдырбаева А.С., Ахмедиярова А.Т., М.Тұрдалыұлы. Н.О.Мекебаев. Систематический обзор и анализ особенностей идентификации по голосу // ҚБТУ хабаршысы. – 2019. – №2 (49). – Б. 120-133.

8. Қалимолдаев М.Н., Мамырбаев О.Ж., Н.О. Мекебаев., Тұрдалыұлы М. Машиналық оқытуды қолдануда дауыстың гендерлік жіктелінуі // ҚазҰТЗУ хабаршысы – 2019. – № 6 (136). – Б.229-233.

9. Bagher BabaAli, Waldemar Wojcik, Orken Mamyrbayev, Mussa Turdalyuly, Nurbapa Mekebayev. Speech Recognizer-Based Non-Uniform Spectral Compression for Robust MFCC Feature Extraction // Przegląd Elektrotechniczny. ISSN: 0033-2097 – 2018. - № 6 (94). – P. 90-93. (Scopus) (Clarivate Analytics)

10. Orken Mamyrbayev, Mussa Turdalyuly, Nurbapa Mekebayev. Automatic Recognition of Kazakh Speech Using Deep Neural Networks // Lecture Notes in Computer Science. 11432 LNAI, с. 465-474. (Scopus)

11. Mamyrbayev O, Toleu A, Tolegen G, Mekebayev N. Neural Architectures for Gender Detection and Speaker Identification // Journal Cogent Engineering. ISSN: 2331-1916. – 2020. Volume 7, - Issue 1. (Scopus)

12. Orken Mamyrbayev, Turdalyuly, Nurzhamal Oshanova, Tolga Ihsan Medeni, Aigerim Yessentay. Voice Identification Using Classification Algorithms // We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists. , June 25, 2019. London.

13. Automatic Recognition of Kazakh Speech Using Deep Neural Networks. 11th Asian Conference on Intelligent Information and Database Systems, ACIIDS 2019 (Scopus).

14. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Тұрдалыұлы. Қазақ сөйлеуін тануда іргелі және қолданбалы зерттеуге арналған фонетикалық мәтін // «Төртінші өнеркәсіптік революция жағдайындағы дамудың жаңа мүмкіндіктері» атты ҚР Президенті Н. Назарбаевтың Жолдауын іске асыру шеңберінде «Көліктегі инновациялық технологиялар: білім, ғылым, тәжірибе» атты ХЛІІ Халықаралық ғылыми-практикалық конференцияның материалдары. – Алматы, 2018. – Т. 2. – Б. 81-87.

15. О.Ж. Мамырбаев, М. Тұрдалыұлы, Н.О. Мекебаев. Қазақ тілі сөйлеуінің акустикалық және тілдік модельдерін құру // Материалы XIV Международной Азиатской школы-семинара «Проблемы оптимизации сложных систем». – Алматы, 2018. – Т. 2. – Б. 344-347.

16. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Турдалыұлы. Алгоритмы и архитектуры систем распознавания речи // Материалы III Международной научной конференции «Информатика и прикладная математика» посвященная 80-летнему юбилею профессора Бияшева Р.Г. и 70-летию профессора Айдарханова М.Б. – Алматы, 2018. – Т. 2. – С. 108-121.

17. Мамырбаев О.Ж., Н.О.Мекебаев., М.Тұрдалыұлы. Идентификация диктора используя mffc // IV международной научно-практической конференции "Информатика и прикладная математика", посвященной 70-летнему юбилею профессоров Биярова Т.Н., Вальдемара Вуйчика и 60-летию профессора Амиргалиева Е.Н. – Алматы, 2019. – Т. 2. – С. 384-392.

18. Orken Mamyrbayev, Mussa Turdalyuly, Nurbapa Mekebayev, Kuralay Mukhsina, Alimukhan Keylan, Bagher BabaAli, Gulnaz Nabieva, Aigerim Duisenbayeva and Bekturgan Akhmetov. Continuous Speech Recognition of Kazakh Language // AMCSE 2018 - International Conference on Applied

Mathematics, Computational Science and Systems Engineering. - Rome, Italy, 2018, v24 - 2019

19. О.Ж. Мамырбаев, А.С. Кыдырбекова, Н.О. Мекебаев, М. Тұрдалыұлы. Обзор методов идентификации и аутентификации пользователей по голосу // Материалы научной конференции ИИВТ КН МОН РК «Инновационные IT и Smart-технологии», посвященной 70-летию профессора Утепбергенова И.Т., Алматы, 2019. – Б. 315-321.

20. О.Ж. Мамырбаев, М. Турдалыұлы, Н.О. Мекебаев, К. Алимхан, Г.С. Набиева, Б.Ж. Мамырбаев. Фонетически представительный текст для создания систем автоматического распознавания казахской речи // Наука и Мир. – 2018. – № 6 (58). – Т. 2 – С. 49-52.

2 авторлық куәлік алынды:

1. ЭЕМ-ге арналған бағдарлама «System of automatic creation vocabulary for ASR» авторлық куәлік 2019 жылғы 22 қаңтар № 1425.

2. ЭЕМ-ге арналған бағдарлама «Мульти язычное распознавание речи MultiSpeech» авторлық куәлік 2020 жылғы 09 қаңтар, № 7844.