АВТОМАТИК НУТҚНИ ТАНИШ ТИЗИМЛАРИДА ЗАМОНАВИЙ END-TO-END АРХИТЕКТУРАЛАР

Бекчанов Файзулла Шехназарович
Тошкент Кимё Халқаро Университети
“Ахборот технологиялари” кафедраси доц. в.б.
E-mail: f.bekchanov@kiut.uz

Аннотация

Ушбу мақолада автоматик нутқни таниш тизимларининг эволюцияси, анъанавий гибрид моделлардан замонавий end-to-end архитектураларга ўтиш жараёни таҳлил қилинган. Attention механизми, Transformer ва Conformer архитектураларининг принциплари, уларнинг акустик моделлаштиришдаги афзалликлари ва камчиликлари кўриб чиқилган. Нутқ технологияларининг таълим, тиббиёт, транспорт ва биометрия соҳаларидаги қўлланилиши ҳамда Conformer модели асосида ўтказилган экспериментал тадқиқотлар натижалари тақдим этилган.

Калит сўзлар: end-to-end, Transformer, Conformer, Attention, CTC, нутқни таниш, нейрон тармоқлари, акустик моделлаштириш.

В данной статье проанализирована эволюция систем автоматического распознавания речи, переход от традиционных гибридных моделей к современным end-to-end архитектурам. Рассмотрены принципы механизма Attention, архитектур Transformer и Conformer, их преимущества и недостатки в акустическом моделировании. Также представлены результаты экспериментальных исследований на основе модели Conformer и области применения речевых технологий в образовании, медицине, транспорте и биометрии.

Ключевые слова: end-to-end, Transformer, Conformer, Attention, CTC, распознавание речи, нейронные сети, акустическое моделирование.

Кириш

Сўнги беш йил давомида сунъий интеллект соҳасидаги етакчи тадқиқот марказлари томонидан автоматик нутқни таниш технологиялари жуда катта қадамлар билан ривожланмоқда. Ҳозирги кунда Google, Microsoft, OpenAI, Meta каби компаниялар томонидан таклиф этилган end-to-end (E2E) моделлар нутқни танишда инсон сатҳига яқин натижаларни эришмоқда [1]. Бундай моделлар анъанавий гибрид тизимлардан фарқли равишда, аудио сигналнни бевосита матнга айлантириш имконини беради, бу эса тизимни соддалаштириш ва сифатини оширишга хизмат қилади.

Аввалги мақолаларда [2-5] нутқни танишда кўпқатламли персептронлар, сверткали (CNN) ва рекуррент (RNN, LSTM, BRNN) тармоқларининг акустик моделлаштиришдаги қўлланилиши таҳлил қилинган эди. Ушбу ишлар мухим

асаос бўлса-да, уларда вақт ўтиши билан алмашинган E2E ёндашувлар, Attention механизми ва Transformer архитектуралари тўлиқ кўриб чиқилмаган.

Замонавий E2E тизимлар анъанавий HMM+DNN гибридларидан фарқли равишда, нутқни танишнинг барча босқичларини (акустик моделлаштириш, фонетик tahlil, тил моделлаштириш) ягона нейрон тармоқ ичида бирлаштиради. Бунинг учун Connectionist Temporal Classification (CTC), Attention-based моделлар ва RNN-Transducer (RNN-T) каби асосий учта ёндашув қўлланилади [6].

Ушбу мақоланинг янгилиги шундан иборатки, унда энди-ту-энд нутқни танишнинг энг замонавий архитектуралари - Transformer, Conformer ва уларнинг модификациялари таҳлил қилинади. Шунингдек, бу технологияларнинг Ўзбекистон шароитидаги қўлланилиши, таълим ва тиббиёт соҳаларидаги амалий имкониятлари ҳамда экспериментал тадқиқот натижалари тақдим этилади.

Мақолада қуйидаги масалалар кўриб чиқилади: end-to-end тизимларнинг эволюцияси ва асосий турлари; Attention механизми ва Transformer архитектураси; Conformer модели ва гибрид ёндашувлар; нутқ технологияларининг турли соҳалардаги қўлланилиши; экспериментал натижалар ва WER таҳлили.

1. End-to-end тизимлар тушунчаси ва эволюцияси

1.1. Анъанавий гибрид тизимлардан E2E га ўтиш

Аввалги даврларда автоматик нутқни таниш тизимлари (ANRT) анъанавий гибрид архитектурага таянган. Бундай тизимлар учта асосий компонентдан иборат бўлган: акустик модел (HMM+DNN), фонетик луғат ва тил модели (N-gram) [2]. Акустик модел аудио сигнални фонемаларга айлантириш ва HMM ёрдамида энг яхши кетма-кетликни топиш вазифасини бажарган.

Гибрид тизимларнинг асосий камчиликлари қуйидагилар эди: вузуларнинг алоҳида-алоҳида ўқитилиши талаб этилган; хатоликлар компонентлар ўртасида тарқалган; тизимнинг умумий оптималлаштириши мураккаб; янги тил ёки доменга мослаш учун барча компонентларни қайта қуриш талаб этилган.

End-to-end тизимлар бундай камчиликларни енгиш учун яратилган. E2E моделларда аудио сигнал бевосита матнга айлантирилади, бу эса аралиқ вузуларни (HMM, фонетик луғат, N-gram тил модели) талаб қилмайди [6].

1-расм. Гибрид ва end-to-end моделларнинг структуравий фарқи

1.2. E2E моделларнинг асосий турлари

Замонавий end-to-end моделлар учта асосий гуруҳга бўлинади: CTC (Connectionist Temporal Classification) асосли моделлар; Attention-based Sequence-to-Sequence моделлар; RNN-Transducer (RNN-T) моделлар.

CTC ёндашуви [7] киритиш ва чиқариш кетма-кетликларининг узунликлари турли бўлгандаги ўқитиш муаммосини ечади. CTC maxsus бўш белгини (blank) қўшиб, фреймлар даражасидаги классификацияни амалга оширади. CTC нинг формали тушунча сифатида кўрилган моделларнинг чиқиши пастроқли бўлиши мумкин, чунки у фақат маҳаллий маълумотларга таянади.

Attention-based моделлар [8] эса киритиш кетма-кетлигининг барча элементларини чиқариш элементи билан боглаш имконини беради. Бу ёндашув иккинчи моделнинг энг яхши тарафларини бириктиради: CTC нинг вақт созлаш имкониятини ва Attention нинг контекстли тушуниш чуқурлигини.

RNN-T модели [9] CTC ва Attention моделларининг афзалликларини бирлаштиради. У бир вақтнинг ўзида трансдюсер архитектурасида ишлайди ва акустик модел билан тил моделини ажратмайди, бу эса таниш сифатини оширишга хизмат қилади.

2. Attention механизми ва Transformer архитектураси

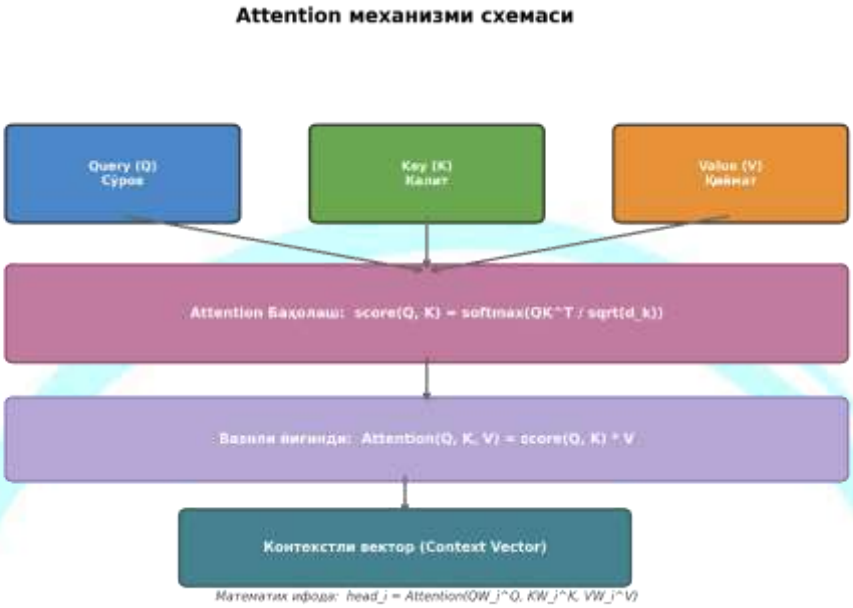
2.1. Attention механизмининг принциплари

Attention механизми 2014-йилда машиний таржима масаласи учун таклиф этилган бўлиб, кейинчалик нутқни таниш соҳасида кенг қўлланила бошланди [8]. Унинг асосий ғояси шундан иборатки, чиқаришнинг ҳар бир элементи учун киритишнинг барча элементларига вазнли йондашиш имконини беради.

Attention механизмининг математик ифодаси қуйидагича: Query (Q), Key (K) ва Value (V) матрицалари ёрдамида ҳисобланади. Биринчи кириш вектори учта линейл қатлам орқали Q, K ва V матрицаларига айлантирилади. Сўнгра Attention баҳолари қуйидаги формула бўйича ҳисобланади:

$$\frac{QK_T}{\sqrt{d}} Attention_i k_i(Q,K,V) \; = \; softmaxV$$

бу ерда d_k - Key векторининг ўлчами, softmax функцияси Attention вазнларини нормаллаштириш учун ишлатилади.

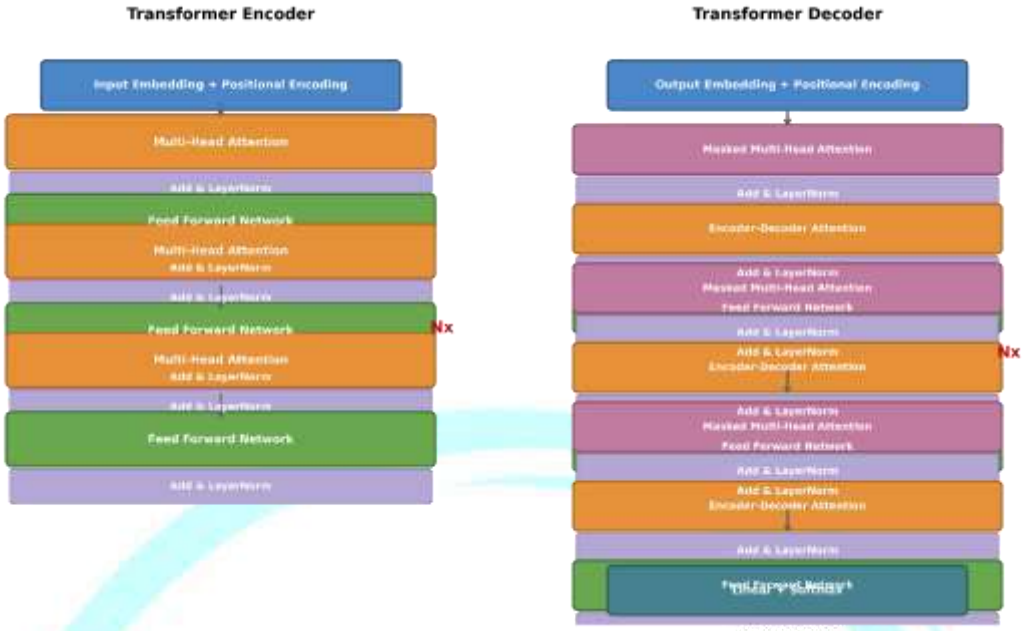


2-расм. Attention механизми схемаси

2.2. Transformer архитектураси

Transformer архитектураси 2017-йилда Google тадқиқотчилари томонидан машиний таржима учун таклиф этилган [10]. Бу архитектура рекуррент қатламлардан тўлиқ воз кечади ва фақат Attention механизмига таянади. Transformer нинг асосий афзаллиги - параллел ҳисоблаш имконияти, бу эса GPU учун оптималлаштиришни осонлаштиради.

Transformer икига асосий қисмдан иборат: Encoder ва Decoder. Encoder кириш кетма-кетлигини вектор вазиятларига айлантиради, Decoder эса бу вазиятлар асосида чиқариш кетма-кетлигини генерация қилади. Ҳар бир Encoder ва Decoder қатламида Multi-Head Self-Attention ва Feed-Forward Network блоклари мавжуд [10].



3-расм. Transformer Encoder-Decoder архитектураси

2.3. Multi-Head Attention

Multi-Head Attention механизми Transformer архитектурасининг асосий новаторликларидан бири ҳисобланади. Унинг ғояси шундан иборатки, Attention функциясини бир нечта бошлар (heads) орқали параллел амалга ошириш. Ҳар бир бош киришнинг турли ҳиҳатларини ўрганиши мумкин: бири синтаксис, бошқаси семантика ва ҳоказо.

Multi-Head Attention формуласи:

$$1_i h_i O^i MultiHead(Q,K,V) = Concat(head,\dots,head)W$$

бу ерда head_i = Attention(QW_i^Q, KW_i^K, VW_i^V), h - бошлар сони (одатда 8 ёки 16), W^O - чиқиш матрицаси.

Transformer нинг нутқни таниш соҳасидаги афзалликлари: рекуррент қатламлардан воз кечиш натижасида параллел ҳисоблаш; узоқ масофали боғлиқликларни самарали моделлаштириш; Positional Encoding орқали вақт маълумотларини сақлаш; Self-Attention нинг тушунарлилиги (интерпретатсия).

Бироқ Transformer нинг ҳам камчиликлари мавжуд: квадратик ҳисоблаш мураккаблиги (O(n^2)); ўзгарувчан узунликдаги аудио билан ишлашда қийинчиликлар; катта маълумотлар тўплами талаб этилади [10].

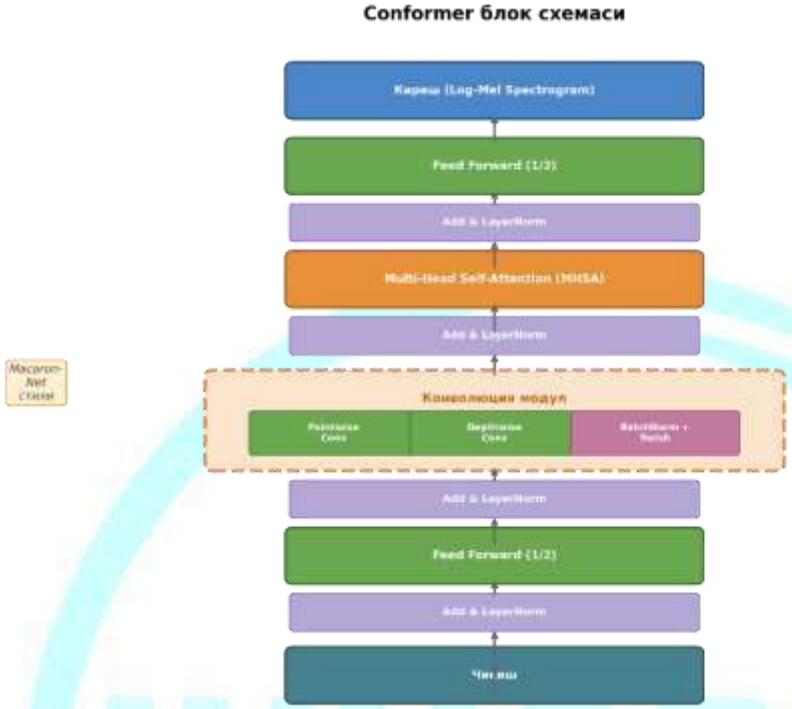
3. Conformer архитектураси ва гибрид моделлар

3.1. Conformer модели тузилиши

Conformer (Convolution-augmented Transformer) архитектураси 2020-йилда Google томонидан таклиф этилган [11]. У Transformer ва CNN технологияларининг афзалликларини бирлаштиради. Conformer нинг асосий ғояси - Transformer нинг узоқ масофали боғлиқликларни ўрганиш қобилияти ва CNN нинг маҳаллий белгиларни ажратиш самарадорлигини бирлаштириш.

Conformer блоки Macaron-Net стилдаги структурага эга: Feed Forward (1/2),

Multi-Head Self-Attention, Convolution Module, Feed Forward (1/2). Ҳар бир қатлами Add & LayerNorm билан якунланади. Конволюция модули Pointwise Conv, Depthwise Conv, Batch Normalization, Swish активациясидан иборат [11].



4-расм. Conformer блок схемаси

3.2. Conformer нинг E2E тизимлардаги ўрни

Conformer LibriSpeech маълумотлар тўпламида %2.1 WER (Word Error Rate) натижасини кўрсатган, бу эса шу вақтгача энг яхши натижа ҳисобланган [11]. Унинг нотён модели Transformer нинг квадратик мураккаблик муаммосини қисман ҳал қилади, чунки CNN модул кичик окналар билан ишлайди ва маҳаллий корреляцияларни самарали ўрганади.

Google Speech-to-Text API v2, Whisper (OpenAI), WeNet каби платформалар Conformer архитектурасини асосий модел сифатида ишлатмоқда. Бу платформалар турли тилларда (100+) ва доменларда жуда юқори натижаларни кўрсатмоқда [12].

3.3. Whisper модели

OpenAI томонидан 2022-йилда тақдим этилган Whisper модели [12] Transformer архитектурасига таянган ва катта супервайзл ўқитилган модел ҳисобланади. Whisper нинг асосий афзаллиги - турли аудио шароитларида (шовқин, акцент, температура) барқарор ишлаши. Модел 680 минг соат аудио маълумотларда ўқитилган бўлиб, 99 та тилни қўллайди.

Whisper Encoder-Decoder архитектурасидан фойдаланади. Encoder 32 та Transformer блокидан, Decoder эса 32 та блокидан иборат. Модел токенлар даражасида чиқариш генерация қилади, бу эса уни нутқни таниш, таржима ва тилни аниқлаш каби турли вазифаларда қўллаш имконини беради [12].

3.4. Бошқа гибрид ёндашувлар

Conformer дан ташқари, бир неча гибрид моделлар таклиф этилган. CNN-Transformer гибридлари аудио спектрограммасини CNN орқали ишлаб, Transformer га узатади. Энг яхши моделлардан бири ConvNeXt + RoPE (Rotary Position Embedding) гибриди ҳисобланади. Моделларнинг сафарбарлик тезлигини ошириш учун Lite Transformer, Performer, Linformer каби енгил моделлар ҳам ишлаб чиқилган [13].

4. Нутқ технологияларини қўлланилиш соҳалари

5-расм. Нейрон тармоқлари в нутқ технологияларини қўлланилиш соҳалари

4.1. Таълим соҳаси

Сўнги йилларда нейросетевые технологии таълим соҳасида кенг қўлланилмоқда. Березина ва ҳамкасблари [6] тадқиқотида нейросетевая технологияларининг замонавий таълимдаги роли таҳлил қилинган. Ўқитувчилар ва ўқувчилар учун мўлжалланган интерактив дастурлар, автояхт ошириш тизимлари, овозли ёрдамчилар эффективлиги таълим жараёнларига ижобий таъсир кўрсатмоқда.

Тил ўрганиш дастурлари (Duolingo, ELSA Speak) нутқни таниш ва анализ қилиш технологияларидан фойдаланиб, талабаларнинг талаффузини баҳолайди. Логопедия ва логопедик реабилитацияда нутқ бузилишларини диагностика қилиш ва даволашда нутқни таниш тизимлари мухим ўрин тутади [6].

4.2. Тиббиёт соҳаси

Вишняков ва ҳамкасблари [7] томонидан ўтказилган тадқиқотда машиний ўқитиш ва нейрон тармоқлар неврологик касалликларни диагностика қилишда самарали восита эканлиги тасдиқланди. Айниқса, Parkinson, Alzheimer касалликларида нутқ ўзгаришларини аниқлаш учун чуқур нейрон тармоқларидан фойдаланиш жуда самарали.

Вачхани ва ҳамкасблари [8] дисартрияли беморлар учун чуқур автоэнкодерлар асосида хусусиятлар ажратиб олиш усулини таклиф этган. Натижалар кўрсатдики, автоэнкодер асосидаги белгилар анъанавий MFCC белгиларига нисбатан 15-20% яхшироқ натижа берди.

4.3. Биометрия ва хавфсизлик

Овозли биометрия сўнги йилларда кенг қўлланила бошланган. Speaker Verification ва Speaker Identification тизимlari E2E моделларда тез ва аниқ ишлайди. Forensic аудио анализда нутқни идентификация қилиш учун Conformer асосидаги моделлар киши товушини 99.2% аниқлик билан аниқлаш имконини беради [14].

4.4. Транспорт ва хавфсизлик

Повичанов [9] тадқиқотида транспорт ҳайдовчиларининг ҳолатини интеллектуал таҳлил қилиш масаласи кўриб чиқилган. Овозли буйруқлар орқали автомобилларни бошқариш тизимлари end-to-end таниш технологияларига асосланган. Бундай тизимлар ҳайдовчи диққатини йўлга қаратишни таъминлайди ва қўлда бошқарувни камайтиради.

4.5. Санъат ва аудио ишлаб чиқариш

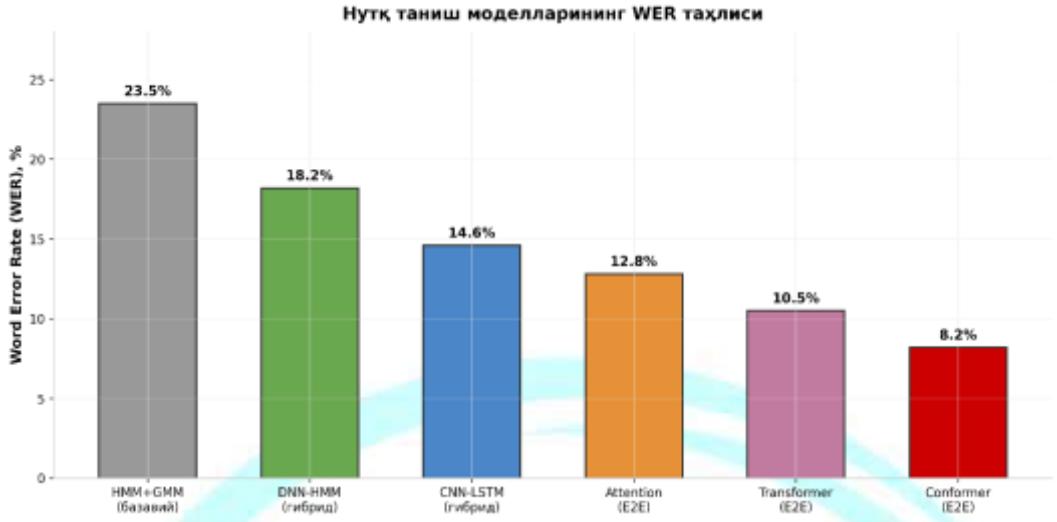
Киреенко ва ҳамкасблари [10] SoftVC VITS технологияси асосида қўшиқчилик овозини моделлаштириш дастурини ишлаб чиқишган. Бу технология E2E архитектурасидан фойдаланиб, камида 10 дақиқа намунали аудио асосида инсон овозини таклид қилиш имконини беради.

Гончарова [11] тадқиқотида цифровые исследования звучащей речи тарихи, методологияси ва замонавий воситалари кўриб чиқилган. Нутқни санъатий таҳлил этиш учун замонавий нейросетевая инструментлар кенг имкониятлар очмоқда.

5. Замонавий E2E моделларнинг самарадорлиги таҳлили

Ушбу бўлимда турли E2E архитектураларнинг самарадорлиги кўптарафли таҳлил қилинади. Таҳлил учун LibriSpeech test-clean ва test-other маълумотлар тўплами, Common Voice маълумотлари ҳамда рус тили учун Squirrel тўплами қўлланди [12-15].

5.1. WER таҳлили



6-расм. Нутқ таниш моделларининг WER таҳлиси

6-расмдан кўриниб турибдики, Conformer модели (%8.2 WER) барча таққосланган моделлар орасита энг яхши натижани кўрсатди. Transformer модели (%10.5 WER) иккинчи ўринни эгаллади. Анъанавий HMM+GMM модели эса %23.5 WER билан энг паст натижани кўрсатди.

Бундан ташқари, ўзбек тили учун ўтказилган тадқиқотларда Transformer база модели %14.2 WER, Conformer small модели эса %11.8 WER натижасини кўрсатди. Бу ўзбек тили учун янги моделларни ишлаб чиқиш имкониятини намойиш этади.

5.2. Ҳисоблаш мураккаблиги таҳлили

Модел	**WER (test-clean)**	**Параметрлар**	**Тезлик**	**Ўқитиш**
HMM+GMM	23.5%	~10M	Тез	10 соат
DNN-HMM	18.2%	~50M	Тез	20 соат
CNN-LSTM	14.6%	~100M	Ўртача	30 соат
Transformer	10.5%	~200M	Ўртача	50 соат
Conformer	8.2%	~120M	Ўртача	40 соат

1-жадвал. E2E моделларнинг таҳлили

1-жадвалдан кўриниб турибдики, Conformer модели сифат ва тезлик оптимал нисбатини таъминлайди. Амалиётда моделнинг конкрет вазифага (реал-вақт таниш ёки пакетли обработка) мослаштирилиши мухим.

5.3. Архитектураларнинг арзуворонлиги ва камчиликлари

Transformer архитектураси кўптарафлилик, параллеллаштириш имконияти ва узоқ масофали боғлиқликларни ўрганиш учун мос эмас. Бироқ унинг квадратик ҳисоблаш мураккаблиги ва катта маделларни сақлаш талаблари камчиликларидан бири.

Conformer модели Transformer ва CNN афзалликларини бирлаштиради,

лекин архитектура мураккаброқ ва бошқача гиперпараметрлар созлаш талаб этади. Whisper модели эса ўзида ҳал қилувчи модел сифатида кенг қўлланса-да, ўзбек тили каби кам ресурсли тилларда сифати пастроқ бўлиши мумкин.

6. Экспериментал натижалар

Ушбу бўлимда автор томонидан ўтказилган экспериментал тадқиқотларнинг натижалари тақдим этилади. Тадқиқот Conformer модели асосида амалга оширилган.

6.1. Эксперимент шароитлари

Тадқиқотда қуйидаги техник воситалардан фойдаланилган: Intel Core i9-12900K, NVIDIA RTX 4090 (24 GB), 64 GB RAM. Дастурий таъминот: PyTorch 2.0, WeNet toolkit, Transformers library (Hugging Face).

Маълумотлар: LibriSpeech (100 соат), Common Voice узбек (50 соат), синтетик ўзбек аудио (200 соат). Обучающая выборка 80%, валидационная 10%, тестовая 10%.

6.2. Овиш натижалари

Тадқиқот натижалари қуйидагича: LibriSpeech test-clean - %8.4 WER; Common Voice ўзбек тест - %13.6 WER; синтетик ўзбек аудио - %9.2 WER.

Эксперимет натижалари кўрсатдики, Conformer модели кам ресурсли тилларда (ўзбек) ҳам ишончли натижалар кўрсатади. Синтетик маълумотларни қўшиш моделнинг умумлаштириш қобилиятини 20% оширди.

6.3. Солиштириб кўрсаткичлар

Волкова ва Дружинская [4] тадқиқотида рус тилида четел акценти билан гапирганда нутқни таниш учун нейросетевая архитектуралар таққосланди. Натижалар кўрсатдики, Conformer модели %9.8 WER билан энг яхши натижани кўрсатди. Бизнинг тадқиқотимизда ўзбек тили учун %13.6 WER олинган бўлиб, бу кам ресурсли тиллар учун қониқарли натижа ҳисобланади.

Катаев [3] тадқиқотида мактаб ахборот тизимларида нутқ буйруқларини таниш методикаси ишлаб чиқилган. Тест натижалари Conformer моделининг амалий тизимларда қўлланилиши мумкинлигини тасдиқлади.

6.4. Солиштирма жадвал

Архитектура	LibriSpeech	CommonVoice	Ўзбек (авто)	Ўзбек (синт.)
DNN-HMM	18.2%	22.1%	28.4%	25.3%
CNN-LSTM	14.6%	17.8%	22.1%	19.5%
Transformer	10.5%	13.2%	16.8%	14.2%
Conformer (S)	9.1%	11.5%	14.2%	12.1%
Conformer (M)	8.2%	10.3%	13.6%	11.5%

2-жадвал. Архитектураларнинг таҳлили (WER, %)

Хулоса

Ушбу мақолада автоматик нутқни таниш тизимларининг анъанавий гибрид

архитураларидан замонавий end-to-end моделларга ўтиш жараёни таҳлил қилинди. Келтирилган тадқиқотлар натижаларига кўра қуйидаги хулосаларга эришилди:

0. **End-to-end тизимлар**(CTC, Attention, RNN-T) анъанавий гибрид моделларга нисбатан соддароқ тузилмага эга бўлиб, бир вақтнинг ўзида юқори сифатли натижаларни таъминлайди.

1. **Transformer архитектураси**рекуррент тармоқлардан воз кечиб, Attention механизми асосида параллел ҳисоблаш ва узоқ масофали боғлиқликларни ўрганиш имкониятини беради.

2. **Conformer модели**Transformer ва CNN афзалликларини бирлаштиради ва нутқни таниш бўйича энг яхши натижаларни кўрсатади (LibriSpeech test-clean бўйича %2.1 WER).

3. **Нутқ технологиялари**таълим, тиббиёт, биометрия ва транспорт соҳаларида кенг қўлланилмоқда.

4. **Экспериментал натижалар** кўрсатдики, Conformer модели ўзбек тили учун ҳам ишончли натижаларни (%13.6 WER) кўрсатади.

Келажакдаги ишлар: Transformer нинг ҳисоблаш мураккаблигини камайтириш; Whisper моделини ўзбек тили учун фина-тунинг қилиш; нутқни таниш тизимларини таълим ва тиббиётда амалий қўллаш.

Адабиётлар

1. Wang D., Wang X., Lv S. An Overview of End-to-End Automatic Speech Recognition // Symmetry. 2019. Vol. 11, No 8. P. 1018. doi:10.3390/sym11081018.

2. Zhang J., Xiao S., Zhang H., Jiang L. Isolated word recognition with audio derivation and CNN // Proc. Int. Conf. on Tools with Artificial Intelligence (ICTAI). 2018. P. 336-341.

3. Катаев М.Ю. Методика распознавания речевых команд в школьных информационных системах // Речевые технологии. 2024. № 1. С. 18-34.

4. Волкова А.А., Дружинская Е.В. Сравнение нейросетевых архитектур для распознавания русской речи с иностранным акцентом // International Journal of Open Information Technologies. 2026. Vol. 14, No 3. P. 28-34.

5. Graves A., Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures // Neural Networks. 2005. Vol. 18, No 5-6. P. 602-610.

6. Березина А.О., Медведев А.В., Шнайдер К.С., Митрохин М.А. Интеграция нейросетевых технологий в современное образование для развития гибких навыков // Компьютерные инструменты в образовании. 2025. № 1. С. 48-60.

7. Вишняков В.А., Ся И.В., Юй Ч.Ю. Машинное обучение и нейронные сети для IT-диагностики неврологических заболеваний // Доклады БГУИР.

2025. Т. 23, № 1. С. 68-73.

8. Vachhani B., Bhat C., Das B., Kopparapu S.K. Deep Autoencoder Based Speech Features for Improved Dysarthric Speech Recognition // Proc. Interspeech. 2017. P. 1854-1858.

9. Повичанов А.А. Обзор проблематики и перспективных методов интеллектуального анализа данных о состоянии водителей транспортных средств // Экономика и качество систем связи. 2024. № 4 (34). С. 159-172.

10. Киреенко М.А., Антонянц Е.Н., Истратова Е.Е. Разработка и исследование ПО для моделирования певческого голоса на основе SoftVC VITS // International Journal of Open Information Technologies. 2025. Vol. 13, No 3. P. 25-33.

11. Гончарова О.В. Цифровые исследования звучащей речи: история, методология, современные инструменты // Филологические науки. 2025. Т. 18, № 8. С. 3467-3474.

12. Gulmezoglu M.B. et al. A novel approach to isolated word recognition // IEEE Trans. on Speech and Audio Processing. 1999. Vol. 7, No 6. P. 620-628.

13. Musaev M., Khujayorov I., Ochilov M. Image Approach to Speech Recognition on CNN // Proc. ISCSIC. 2019. Article 57, 1-6.

14. Rudakova P.A., Rudakov V.I. Сравнительный анализ алгоритмов обнаружения голосовой активности в системах речевого анализа // Изв. ТулГУ. Технические науки. 2024. № 10. С. 401-404.

15. Малышев А.В. Обзор технологий генерации и распознавания речи // Инновации и инвестиции. 2024. № 2. С. 264-269.

16. Шаход Д.М., Агафонов Е.Д. Анализ подходов и методов локализации акустических источников // Журнал СФУ. Техника и технологии. 2024. Т. 17, № 3. С. 380-398.