

ISSN 1728-7901

Абай атындағы Қазақ ұлттық педагогикалық университеті
Казахский национальный педагогический университет имени Абая
Abai Kazakh National Pedagogical University

ХАБАРШЫ

«Физика-математика ғылымдары» сериясы
Серия «Физико-математические науки»
Series of Physics & Mathematical Sciences
№1(69)

Алматы, 2020

ABAI UNIVERSITY

BULLETIN

Ser. Physics & Mathematical Sciences

№ 1 (69)

Editor-in-Chief

Dr. Sci. Bektemesov M.A.

Deputy Editor-in-Chief:

Dr. Sci. Ualiyev G.,

Dr. Sci. (Ped.), Bidaibekov Ye.Y.,

Dr. Sci., Corresponding member

of the NAS of RK Kosov V.N.,

Cand.Sci. Bekpatshayev M.Zh.

Responsible editorial secretary:

Cand. Sci. (Ped.) Shekerbekova Sh.

Cand. Sci. (Ped.) Abdulkarimova G.A.

Editorial board:

Dr.Sci. Alimhan K. (Japan),

Phd.d. Cabada A. (Spain),

Phd.d Kovatcheva E. (Bulgaria),

Phd.d. Ruzhansky M. (England),

Dr. Sci. (Ped.), Corresponding member of the

NAS of RK Abylkasymova A.Ye.,

Dr.Sci.(Engineering) Amirgaliyev Ye.,

Dr. Sci. Berdyshev A.S.

Dr.Sci. Grigoriev S.G. (Russia),

Dr.Sci. Grinshkun V.V. (Russia),

Dr. Sci. Mukhambetzhano S.T.,

Dr.Sc. Kabanikhin S.I. (Russia),

Dr. Sci., Academician of the NAS of RK

Kalimoldayev M.N.,

Dr. Sci. Kozhamkulov B.A.,

Dr. Sci. Komarov F.F.,

(Republic of Belarus),

Dr.Sci.(Engineering) Kulbek M.K.,

Dr. Sci. (Ped.) Lapchik MP (Russia),

Dr. Sci. Lisicin V.M. (Russia),

Dr. Sci. (Ped.) Mambetkunov E.M.

(Kyrgyz Republic),

Dr. Sci. (Ped.) Pak N.I. (Russia),

Dr.Sc. Sakhiev S.K.,

Dr. Sci. (Ped.) Sedova Ye.A. (Russia),

Dr. Sci. (Ped.) Sydykov B.D.,

Dr.Sci.(Engineering) Tuleshov A.K.,

Dr.Sci. Ualiyev Z.G.,

Cand.Sci. Khamraev Sh.I.

© Abai University, 2020

Registered in the Ministry of Information of the

Republic of Kazakhstan,

№ 4824 - Ж - 15.03.2004

(Periodicity: 4 issues per year)

Published since 2000

Signed to print 10.03.2020 г.

Format 60x84 1/8. Vol. 57,3 p.

Printing 300 copies. Order 256.

Publishing and Editorial:

050010, 13 Dostyk av.,

Almaty, Kazakhstan

Publisher "Ulagat"

Abai University

Косов В.Н., Мукамеденкызы В., Федоренко О.В., Туцен М. Изоконцентрационные распределения компонентов в тройных газовых смесях при наличии особых режимов диффузионного смешения.....	230
Кушербаева М.Р. Физикалық білімнің қолданбалы бағыты....	236
Минглибаев М.Дж., Байсбаева О.Б. Поступательно-вращательное движения трехосного тела с переменными сжатиями при наличии реактивных сил и моментов	241
Минглибаев М.Дж., Бижанова С.Б. Массасы мен өлшемі айнымалы өстік симметриялы дененің эволюциялық теңдеулерін зерттеу.....	247
Молдабекова М.С., Ж.М.Битибаева Некоторые особенности формирования исследовательских умений студентов в контексте практико-ориентированного подхода.....	253
Опахай С., Кутербеков К.А., Нуркенов С.А. Тіреуіш металл негізіндегі қатты оксидті отын элементтері.....	258
Сайлаубеков Е.К., Морзабаев А.К. Альфа-бөлшектерді беру реакцияларын зерттеу.....	264
Темирбеков Е.С., Тукешова Г.А. Стержневое моделирование конструкций рычажных механизмов с распределенной инерции.....	269

**ИНФОРМАТИКА. ИНФОРМАТИКАНЫ ОҚЫТУ
ӘДІСТЕМЕСІ. БІЛІМ БЕРУДІ АҚПАРАТТАНДЫРУ
ИНФОРМАТИКА, МЕТОДИКА ПРЕПОДАВАНИЯ
ИНФОРМАТИКИ. ИНФОРМАТИЗАЦИЯ
ОБРАЗОВАНИЯ**

Абуярова А.А. Earthquake time prediction with multi-agent systems	275
Артыкбаева Е.В., Бактыбаев Ж.Ш., Тусубаева Ж.М., Арыстанова А.Ж. Готовность преподавателей к внедрению дистанционных образовательных технологий в высшем образовании.....	280
Астамбаева Ж.Қ., Жұмабаева Ә.Е. Болашақ бастауыш сынып мұғалімдерінің алгоритмдік сауаттылығын дамыту жолдары.....	285
Берікқызы Р., Рахимжанова Л.Б., Исабаева Д.Н. Жаратылыстану - математикалық бағыттағы жаңартылған мазмұны бойынша информатиканы оқыту әдістемесі.....	291
Гусманова Ф.Р., Абдулкаримова Г.А. Блоктық шифрлаудың дамуына шолу.....	295
Денгельбаева Н.Б., Исенгалиева А.Г., Атантаева А. Цифрлық жаһандану жағдайында жоғары оқу орындары кітапханаларының даму үрдістері.....	302
Джолдасбаев С.К., Куламбаев Б.О. Применение алгоритмов балансировки нагрузки для повышения качества предоставления услуг.....	308
Досжанов Б.А. Блокчейн технологиясы мен биткойн криптовалютасы: жұмыс ұстанымы және ерекшеліктері.....	314
Zhanibek Zh.A. Balakayeva G.T. Web application for processing a large amount of data in the field of business.....	319
Zhapsarbek N.B. Modeling of large volumes of data with the use of NoSQL.....	323
Жолдас Н.А., Дарибаев Б.С. Ауылшаруашылық нысандарының (жылыжайлардың) өнімділігін арттыру үшін ИОТ және Big Data технологияларын қолдану	327
Заурбеков Н.С., Бодык А.М. Методы применения программного обеспечения Maple и Mathcad в решении математических задач.....	333
Заурбеков Н.С., Шерхан Г.А. О проблемах и методике обучения учащихся старших классов основам алгоритмизации и программирования	339

Зейнулласва И.Д., Керімбаев Н.Н., Бейсов Н.К., Азыбаев М. Дәріс беру барысында студенттермен виртуалды кері байланыс орнату.....	345
Ильясова Р.А., Даулеткулова А.У., Тохтахунова Д.Я. Системы компьютерно-ориентированных задач в курсе дифференциальных уравнений.....	351
Иманбаев К.С., Джанузаков С.Д., Кожамкулова Ж.Ж., Джанузаков А.С. Задача построения оптимальной структуры информационной системы иерархической структуры.....	355
Камалова Г.Б., Шайбасов К. Python как эффективное средство разработки цифровых ресурсов для численного решения систем линейных алгебраических уравнений.....	361
Қадырбек Н.Қ., Мансурова М.Е., Қырғызбаева М.Е. Қазақ тіліндегі құжаттар үндестігін талдауда LSTM желілерін қолдану	366
Маликова Ф.Ө., Төлеушова А.Т., Рыскелді Р.С. Қолтаңбаны визуализациялау әдістемесі.....	370
Маликова Ф.Ө., Жанат Н.Ж., Сағинаева А.К., Рыскелді Р.С. Бет әлпетті тану ерекшеліктері	374
Nurmukhanov T.A., Daribayev B.S. Recognition of the text by means of Deep Learning	378
Неверова Е.Г. Исследование динамики спроса на кредитование физических лиц с помощью инструментов языка R.....	383
Нугманова С.А., Ерболат М. Мектеп оқушыларын оқытуда микроконтроллерлерді қолдану.....	387
Нуруллаев Н.М., Турғунбоев Д.А., Жолдасов Е.Н. Кедір-бұдырлы қатты денелерді қармауға арналған манипуляторларды жетілдіру мүмкіншіліктерін бағалау.....	392
Оразбеков Ж.Н., Мошкалов А.Қ., Сабраев Қ.Ж. Корпоративтік портал ортасында өндіріс деректерін өңдеу мен алмасу процессінде кезекті басқару алгоритмін оңтайландыру	395
Оспан Ә.Ғ., Мансурова М.Е., Какимжанов Е.Х. Разработка гибридной модели для эффективного распределения водных ресурсов на основе модели прогнозирования.....	399
Салғараева Г.И., Асан Г.Е. Педагогикалық зерттеулерде цифрлық білім беру технологияларын қолдану.....	404
Сапанов Н.А., Бектемесов А.Т. Қалалық агломерацияның логистикалық инфрақұрлымын басқару негіздері.....	409
Сарсимбаева С.М. Vuforia платформасында кеңейтілген шындық қосымшаларын құру және оқу процесінде қолдану...	414
Сарсимбаева С.М., Бекеева С.И., Аханова М.Б. Исследование вопросов разработки системы «умный дом» на платформе Arduino.....	417
Сыдыхов Б.Д., Касиегова А.Б., Діқамбай Н.Б. Болашақ мұғалімнің сандық білім беру ресурстарын қолдануының теориялық-әдіснамалық мәселелері.....	421
Сыдыхов Б.Д., Қойшыман Г., Батырхан З.Ә. Оқушыларға робототехника негіздерін оқытудың әдістемелік ерекшеліктері....	426
Toleugazy R.T., Balakayeva G.T. Application of the regression analysis method for modelling the processing of large amounts of data	431
Тұльбасова Б.Қ., Салықова А.Н. Цифрлық білім ресурстарын орта мектепте қолдану ерекшеліктері.....	436
Турганбаева А.Р., Болысбекова Ф.Қ. 3D Studio Max редакторының көмегімен компьютерлік модельдеу.....	441
Турганбаева А.Р., Рахымжанова А.А., Черикбаева А.С. Информатика пәні бойынша жаңартылған бағдарламамен оқытумен бағалаудың жолдары.....	445
Шекербекова Ш.Т., Исабаева Д.Н., Тілеубергенов М.А. Мектеп оқушыларын компьютерлік ойындарын құруға оқыту әдістемесі.....	450

МРНТИ 20.23.17; 20.23.21; 20.23.25
УДК 004.912; 004.62

Н.Қ. Қадырбек¹, М.Е. Мансурова¹, М.Е. Қырғызбаева¹

әл-Фараби атындағы Қазақ Ұлттық университеті, Алматы қ., Қазақстан

ҚАЗАҚ ТІЛІНДЕГІ ҚҰЖАТТАР ҮНДЕСТІГІН ТАЛДАУДА LSTM ЖЕЛІЛЕРІН ҚОЛДАНУ

Аңдатпа

Әлеуметтік медиа-ресурстардағы ақпараттарға деген сенімнің артуына байланысты үндестікті талдау саласына деген қызығушылық күн өткен сайын артуда. Өйткені үндестікті талдау миллиондаған әлеуметтік желі қолданушыларының пікірлеріне мониторинг жүргізудегі басты технологиялардың бірі болып табылады.

Мақалада қазақ тіліндегі мәтіндер үндестігін талдауда LSTM желілерін қолдану қарастырылған. Нейрондық желіні оқыту үшін ұялы телефондар пайдаланушыларының жалпы саны 1000 пікірі қолданылды. Зерттеу жұмысы екі түрлі жолмен жүргізілді: бірінші жағдайда талданатын пікірлер алдын-ала өңдеуден (preprocessing) өткізілді, екінші жағдайда алдын-ала өңдеу жүргізілген жоқ. Модель алдын-ала өңдеуден өткізілген жағдайдағы сапаны бағалау өлшемінің орташа мәні 80%-ке жетті. Бұл көрсеткіш алдын-ала өңдеу жүргізілмеген мәліметпен оқытылған моделмен салыстырылғанда 11%-ға жоғары. Зерттеу нәтижелері мәтіндерді алдын-ала өңдеуден өткізу модельдің сапасын жақсартады деген қортынды жасауға мүмкіндік берді.

Түйін сөздер: үндестікті талдау, табиғи тілдерді өңдеу, терең оқыту, нейрондық желілер, LSTM архитектурасы.

Аннотация

Н.К. Қадырбек¹, М.Е. Мансурова¹, М.Е. Қырғызбаева¹

¹Казахский национальный университет имени аль-Фараби, г.Алматы, Казахстан

ИСПОЛЬЗОВАНИЕ СЕТЕЙ LSTM В АНАЛИЗЕ ТОНАЛЬНОСТИ ДОКУМЕНТОВ НА КАЗАХСКОМ ЯЗЫКЕ

В связи с растущим доверием к информации в социальных медиа-ресурсах растет и интерес к области анализа тональности. Потому что анализ тональности является одной из основных технологий для мониторинга мнений миллионов пользователей социальных сетей.

В статье рассматривается использование сетей LSTM при анализе тональности текстов на казахском языке. Для обучения нейронной сети было использовано 1000 отзывов пользователей мобильных телефонов. Эксперименты были проведены двумя способами: в первом случае была проведена предварительная обработка (preprocessing) анализируемых отзывов, во втором случае предварительная обработка не проводилась. Среднее значение метрики для оценки качества модели с предварительной обработкой достигло значения 80%. Этот показатель на 11% выше, чем для модели, обученной на данных без предварительной обработки. Результаты исследования позволили заключить, что предварительная обработка текстов способствует повышению качества модели.

Ключевые слова: анализ тональности, обработка естественного языка, глубокое обучение, нейронные сети, архитектура LSTM.

Abstract

USING OF LSTM NETWORKS IN SENTIMENT ANALYSIS OF DOCUMENTS IN KAZAKH LANGUAGE

Kadyrbek N.K.¹, Mansurova M.E.¹, Kyrgyzbayeva M.E.¹

Al-Farabi Kazakh National University, Almaty, Kazakhstan

Due to the growing trust in information in social media resources, interest in the field of sentiment analysis is growing. Because sentiment analysis is one of the main technologies for monitoring the opinions of millions of users of social networks.

The article discusses the use of LSTM networks in the analysis of the tonality of texts in the Kazakh language. For training the neural network, 1000 user reviews of mobile phones were used. The experiments were carried out in two ways: in the first case, preprocessing of the analyzed reviews was carried out, in the second case, the preprocessing was not carried out. The average value of the metric for assessing the quality of the pre-processed model reached 80%. This indicator is 11% higher than for a model trained on data without preprocessing. The results of the study allowed us to conclude that the preprocessing of the texts improves the quality of the model.

Keywords: sentiment analysis, natural language processing, deep learning, neural networks, LSTM architecture.

Кіріспе

Қазіргі таңда ақпараттық технологиялардың қарқынды дамуына байланысты әлеуметтік желілердегі пікірлер нарықты бағалау, қандайда бір нақты өнім, қызмет, шоу-бизнес, спорт, тіпті, саяси ұстанымдардың танымалдылығы мен дәрежесін анықтауда жиі қолданылады. Мұндай пікірлер позитивті, негативті немесе бейтарап болуы мүмкін. Осындай пікірлердің қай топқа жататындығын анықтау компьютерлік лингвистиканың бір саласы – үндестікті талдау (sentiment analysis) арқылы жүргізіледі. Үндестікті талдау – табиғи тілдерді өңдеу (NLP) әдістерінің, статистика, машиналық оқыту көмегімен пікірлердің үндестілігін анықтау. Сонымен қатар үндестікті талдау пікірлердегі спамдарды анықтау, пікірлердің пайдалылығын талдау, салыстырымдарды іздеуде қолданылады. Әдетте, әлеуметтік желілерден алынған пікірлер грамматикалық ережелерге сәйкес емес, түрлі белгілер, қысқартылған сөздер және т.б. болуы мүмкін. Сондықтан мұндай жағдайда деректерді алдын-ала өңдеуден өткізу жақсы нәтижелерге қол жеткізуге мүмкіндік береді [1, 2]. Дегенмен біз бұл зерттеу жұмысында алынған пікірлер алдын-ала өңдеуден өткізілген және алдын-ала өңдеуден өткізілмеген екі жағдайды да қарастырып, нәтижелерін салыстыратын боламыз. Соңғы жылдары нейрондық желілер машиналық оқытудың қуатты модельдері ретінде қайта кең жанданып келеді, бейнені тану және табиғи тілді өңдеу сияқты салаларда үздік нәтижелер көрсетуде [3].

"Bag of words", байес әдісі сияқты дәстүрлі моделдерді пайдаланатын классификаторлармен бірге үндестік талдауы есептерінде өте дәл болжамдарды алу үшін ұтымды пайдаланылды [4]. Терең оқыту (deep learning) технологияларының пайда болуымен және оларды табиғи тілді өңдеуде қолдануымен осы әдістердің дәлдігін екі негізгі бағытта жақсарту мүмкіндігі туды: деректерді алдын ала өңдеу және кластеризатор мен классификаторларды оқытуда оқытушымен және оқытушысыз нейрондық желіні пайдалану.

Зерттеу нысандары мен әдістері

Зерттеу жұмысының барысында нейрондық желіні оқыту үшін ұялы телефондар пайдаланушыларының жалпы саны 1000 пікірі қолданылды. Жиынтықта әрбір пікір «клас:пікір» құрылымында сақталған, мұндағы 0-позитивті және 1-негативті пікірлер (1-сурет).

Бұл мәліметтер алдын-ала өңдеудің келесі қадамдарынан өтеді:

- 1) артық таңбаларды алып тастау: тек әріптерді қалдыру
- 2) Сегментация – әрбір пікірді сөйлемдерге, ал сөйлемдер токендерге ажыратылады.
- 3) Лемматизация – токендерді бастапқы қалпына келтіру процесі (нормализация).

Class	Data
0	0 камерасы әлсіз мегапикселін жаңарту қажет зам...
1	0 не деген сұмдық жады аз нәліктен коробкасында...
2	0 телефон не деген ауыр темірден жасаған ба жең...

Сурет 1. Пікірлер жіктелімі

Мысалы: «телефондардың» токени үшін лемма «телефон». Бұл жерде анализатор инструмент ретінде біздің осыған дейін жоба барысында жасалған инструмент қолданылды [5].

Жұмыс барысында лемматизацияны қолдана отырып және қолданбай тәжірибе жасаймыз.

Әдеттегі нейрондық желілердің рекуррентті желілерден негізгі айырмашылығы рекуррентті желінің уақытпен байланысты аспектісі болып табылады. Рекуррентті торларда әрбір сөз кіріс кезектілігі белгілі бір уақыт қадамымен байланысты болады. Іс жүзінде уақыт қадамдарының саны тізбектің максималды ұзындығына тең болады (2-сурет).

Байланыс нашар ұстайды ... тез бұзылады

x_0	x_1	x_2	x_{18}	x_{19}
t_0	t_1	t_2	t_{18}	t_{19}

Сурет 2. Уақыт қадамдарының тізбек ұзындығына сәйкестік мысалы

Әрбір h_t уақыт қадамымен жасырын күй векторы (hidden state vector) деп аталатын жаңа компонент байланысты. Өзінен жоғарғы деңгейден бұл вектор алдыңғы уақыт қадамдарында байқалған барлық ақпаратты инкапсуляциялауға және жинақтауға ұмтылады.

Демек, x_t нақты сөзге қатысты барлық ақпаратты қамтитын вектор, h_t – бұл алдыңғы уақыт қадамдарынан ақпаратты жинақтайтын вектор.

Жасырын күй – бұл ағымдағы сөз векторының, сондай-ақ алдыңғы уақыт қадамындағы жасырын күй векторының функциясы. Сигма екі мүшенің қосындысы активация функциясы арқылы (әдетте сигмоид немесе тангенс) орналастырылатынын көрсетеді.

$$h_t = \sigma(W^H h_{t-1} + W^X x_t)$$

W мүшелері – салмақ матрицалары. Кіріс векторын W^X салмақ матрицасына, ал алдыңғы уақыт қадамындағы жасырын күй векторына W^H рекурренттік салмақ матрицасы көбейтіледі. W^H – бұл барлық уақыт қадамдарында өзгеріссіз қалатын матрица, ал W^X өлшеу матрицасы әрбір кіріс сигналы үшін өзгеше болады.

Осы салмақтық матрицалары жасырын күй векторының не ағымдағы, не алдыңғы жасырын күйіне әсер етеді.

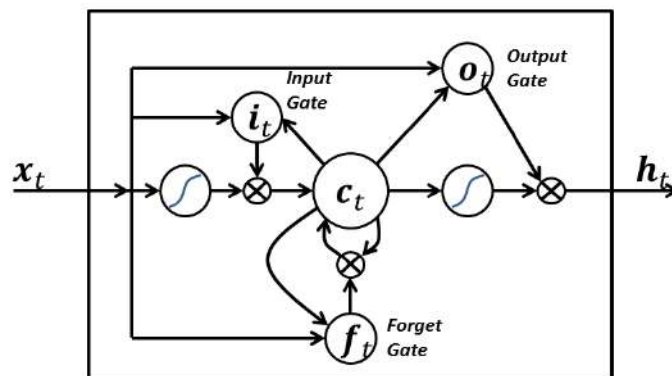
Long Short Term Memory Units - бұл рекуррентті нейрондық желілер ішінде орналастыруға болатын модульдер. Жоғары деңгейде олар h жасырын күй векторының мәтіндегі ұзақ мерзімді тәуелділік туралы ақпаратты инкапсуляциялауға қабілетті болуын бақылайды [6].

Жоғарыда келтірілген RNN туралы тұжырым салыстырмалы тұрғыда қарапайым. Мұндай тәсіл бірнеше уақыт қадамдарына бөлінген ақпаратты біріктіре алмайды.

LSTM бірліктерін(units) техникалық тұрғыдан қарастырғанда, бірліктер x_t сөзінің ағымдағы векторын қабылдап, h_t жасырын күй векторын шығарады.

Осы бірліктерде h_t тұжырымдамасы типтік RNN қарағанда біршама күрделі болады.

Есептеу 4 компонентке бөлінеді: кіріс элементі (input gate), ұмыту элементі (forget gate), шығыс элементі (output gate) және жана жады контейнері (3-сурет).



Сурет 3. LSTM бірліктері

Әрбір элемент x_t және h_{t-1} (суретте көрсетілмеген) кіріс деректер ретінде қабылдайды және аралық күйлерді алу үшін кейбір есептеулерді орындайды.

Әрбір аралық күй әртүрлі желіге келіп түседі және ақыр соңында ақпарат h_t қалыптастыру үшін агрегацияланады.

Мұнда әрбір элемент өзіндік рөл атқарады: кіріс элементі әрбір кіріске қаншалық көңіл бөлу керектігін анықтайды, ұмыту элементі біз алып тастайтын ақпаратты анықтайды, ал шығыс элементі аралық күй негізінде соңғы h_t анықтайды.

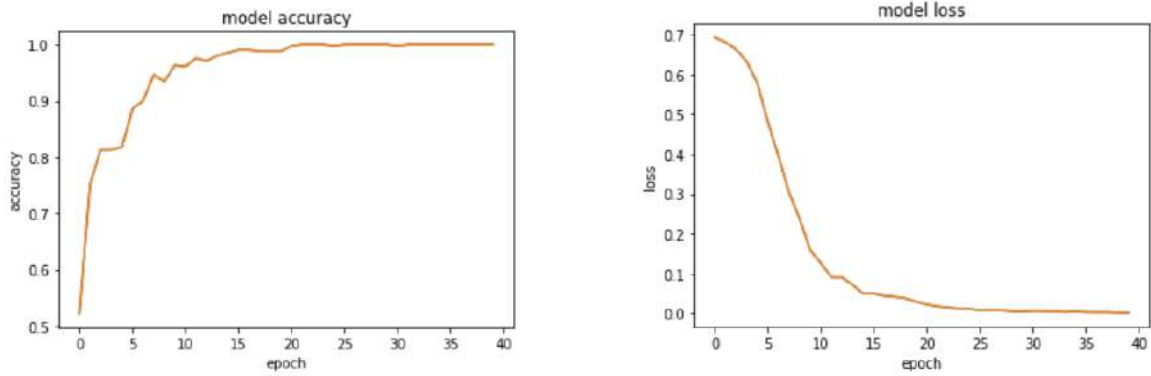
Нәтижелер мен оларды талқылау

LSTM арқылы оқытылған моделімізде *batch size* шамасы, яғни оқытуға алынатын пікірлер мөлшері – 100, ал эпоха (epoch) саны – 40.

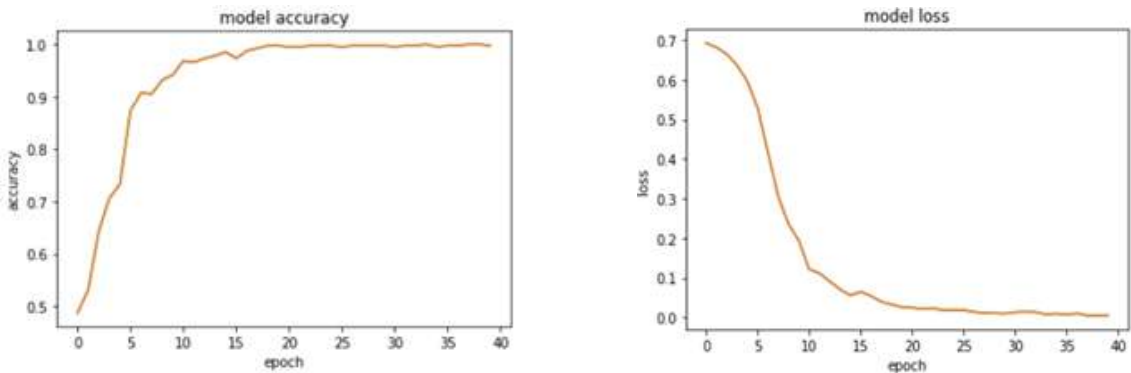
Активация функциясы ретінде *softmax* функциясы пайдаланылды. Себебі желі категориялық кроссэнтропияны (categorical crossentropy) пайдаланады және softmax біз үшін оңтайлы шешім болып табылады.

Тәжірибе екі түрлі әдіс арқылы жүзеге асырылды. Алдын-ала өңдеусіз (4-сурет) және алдын-ала өңдеумен (5-сурет).

Суреттерде нейрондық желіні оқыту барысы келтірілген. Бірінші жағдайда оқыту дәлдігі, ал екіншісінде оқыту қателігі бейнеленген.



Сурет 4. Нормализациялаумен оқыту барысы



Сурет 5. Нормализациялаусыз оқыту барысы

Кесте 1. Нәтижелерді бағалау

	Алдын-ала өңдеусіз			Алдын-ала өңдеумен		
	precision	recall	f1-score	precision	recall	f1-score
негативті	0.68	0.77	0.72	0.81	0.80	0.80
позитивті	0.71	0.62	0.66	0.78	0.80	0.79
дәлдігі	0.69			0.80		

Қорытынды

Үндестік талдауы есептеуіш лингвистиканың іргелі есебі болып табылады. Қазақ тілі аз ресурсты тілдер қатарына жатқандықтан, бұл бағыттағы зерттеу жұмыстары үлкен еңбекті қажет етеді. Қарастырылған жұмыста жоба аясындағы жасалынған морфологиялық анализатор көмегімен өңдеуден өткізілген мәтінге үндестік талдауы жасалынды. Мұнда LSTM архитектурасы арқылы құрастырылған моделдің дәлдігі 80% -ке жетті. Бұл көрсеткіш нормализациядан өтпеген мәліметпен оқытылған моделмен салыстырылғанда 11%-ға жоғары. Бұл бір жағы оқыту үшін қолданылған мәліметтердің көп болмауымен түсіндіруге болады. Өз кезегінде нормализация арқылы моделдің жинақы болуына және көптеген мәліметтерді жалпылау мүмкіндігіне қол жеткізіледі.

Бұл жұмыс ҚР БҒМ О.0856 BR05236340 «Қазақстан Республикасының цифрлы экономикасын қалыптастыру шеңберінде «логистикалық-агломерациялық» жүйесінің талдау және шешім қабылдау жоғары өнімді зияткерлік технологияларын құру» және AP05132933 «Шешім қабылдау сапасын жақсарту үшін деректердің гетерогенді көздерінен білімді алу жүйесін құру» жобалары шеңберінде жасалды.

Пайдаланылған әдебиеттер тізімі:

1 Hemalatha, G. P. Saradhi Varma, A.Govardhan Preprocessing the Informal Text for efficient Sentiment Analysis // International Journal of Emerging Trends & Technology in Computer Science (IJETTCS). Volume 1, Issue 2 July-August 2012. – P. 58–61.

- 2 Muhammad Javed, Shahid Kamal Normalization of Unstructured and Informal Text in Sentiment Analysis // International Journal of Advanced Computer Science and Applications // (IJACSA), Vol. 9, No. 10, 2018. – P. 78–85.
- 3 Харламов А.А., Ле Мань Ха «Нейросетевые подходы к классификации текстов на основе морфологического анализа» // ТРУДЫ МФТИ. 2017. Том 9, № 2. – С. 143-150.
- 4 Narayanan V., Arora I., Bhatia A. Fast and accurate sentiment classification using an enhanced Naive Bayes model // International Conference on Intelligent Data Engineering and Automated Learning. 2013. Oct 20. – P. 194–201.
- 5 Мансурова М.Е., Койбагаров К.Ч., Баракшин В.Б., Солтангельдинова М., Бердибеков С. Применение морфологического анализатора казахского языка для извлечения фактов из фактографических систем // Материалы Международной научной конференции «Информатика и прикладная математика», посвященной 25-летию независимости Республики Казахстан и 25-летию Института информационных и вычислительных технологий. Алматы, 21-24 сентября 2016 года. – Часть I. – С. 156-166.
- 6 Jenq-Haur Wang, Ting-Wei Liu, Xiong Luo, Long Wang An LSTM Approach to Short Text Sentiment Classification with Word Embeddings // The 2018 Conference on Computational Linguistics and Speech Processing ROCLING 2018, - P. 214-223.

МРНТИ 20.53.19
УДК 004.93

Ф.Ө. Маликова^{1,2}, А.Т. Төлеушова², Р.С. Рыскелді¹

¹Алматы Технологиялық Университеті, Алматы қ., Қазақстан

²әл-Фараби атындағы Қазақ Ұлттық Университеті, Алматы қ., Қазақстан

ҚОЛТАҢБАНЫ ВИЗУАЛИЗАЦИЯЛАУ ӘДІСТЕМЕСІ

Аңдатпа

Қазіргі заманғы ақпараттық қоғамда адам мен машина интерфейсін жетілдіруге көп көңіл бөлінеді, ол деректер мен білімнің қарапайым, жылдам және қолжетімді жолдармен тиімді өңдеуін қамтамасыз етуі тиіс. Оны ұйымдастыру тәсілдерінің бірі - қолжазба енгізу (мәтінді енгізу, суреттер, сызбалар және т.б.). Оны пайдалану әдеттегідей жылдам, ыңғайлы түрде арнайы оқытуды қажет етпейді. Сонымен қатар, адам-машина интерфейсінің ажырамас бөлігі математикалық және бағдарламалық қамтамасыз ету болып табылады, бұл бастапқы төменгі деңгейдегі деректерден енгізілген ақпаратты тікелей сипаттайтын деректерге көшуге мүмкіндік береді. Қазіргі уақытта қолжазба мәтінін жасау процесін үлгілеудің қазіргі заманғы тәсілдері қарастырылады. Қолтаңбаны зерттеу кезінде модельді пайдалану мысалы келтіріледі. Ұсынылған визуализация техникасын қазіргі заманғы үш өлшемді мониторларда толықтай қолдануға болады. Қазіргі уақытта қолтаңба мәтінін компьютерлік талдау жұмыстары белсенді жүргізілуде. Бұл ретте қолжазба мәтінінен оның көмегімен берілетін ақпарат (қолжазба мәтінін тану), сондай-ақ жазушының жеке басы және оның жағдайы туралы ақпарат (жеке басын жазу және қол қою бойынша сәйкестендіру, психологиялық және медициналық диагностика, графикалық зерттеу) алынады.

Түйін сөздер: қолтаңбаны тану, сәйкестендіру, медициналық диагностика, визуализация әдісі, перспективалық проекция, ортогональді проекция.

Аннотация

Ф.Ө. Маликова^{1,2}, А.Т. Төлеушова², Р.С. Рыскелді¹

Алматынський Технологический Университет¹, г. Алматы, Казахстан

Казахский национальный университет имени аль-Фараби², г. Алматы, Казахстан

МЕТОДИКА ВИЗУАЛИЗАЦИИ ПОДПИСИ

В современном информационном обществе большое внимание уделяется совершенствованию человеческого и машинного интерфейса, которое должно обеспечивать эффективную обработку данных и знаний простыми, быстрыми и доступными способами. Одним из способов его организации является введение рукописи (ввод текста, рисунки, чертежи и т. д.). Его использование, как правило, не требует специального обучения в быстром, удобном виде. Кроме того, неотъемлемой частью человеческого-машинного интерфейса является математическое и программное обеспечение, что позволяет перейти от исходных низких данных к данным, непосредственно характеризующим внесенную информацию. В настоящее время рассматриваются современные подходы к моделированию процесса создания рукописного текста. При исследовании подписи приводится пример использования модели. Предлагаемую технику визуализации можно полностью использовать на современных трехмерных мониторах. В настоящее время активно ведутся работы по компьютерному анализу рукописного текста. При этом из рукописного текста извлекается как передаваемая с его помощью информация (распознавание рукописного текста), так и информация о личности пишущего и его состоянии (идентификация личности по почерку и подписи, психологическая и медицинская диагностика, графологическое исследование).

Ключевые слова: распознавание подписи, идентификация, медицинская диагностика, метод визуализации, перспективная проекция, ортогональная проекция.