

**ՏԵՂԵԿԱՏՎԱԿԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ, ԷԼԵԿՏՐՈՆԻԿԱ,
ՌԱԴԻՈՏԵԽՆԻԿԱ**

ԻՆՖՈՐՄԱՏԻԿԱ ԵՎ ՀԱՇՎՈՂԱԿԱՆ ՏԵԽՆԻԿԱ

ՀՏԴ 004.8

Ն.Հ. ԱԲՐՈՅԱՆ

**ՆԵՅՐՈՆԱՅԻՆ ՑԱՆՅԵՐԻ ԿԻՐԱՌՈՒՄԸ ԵՎ ԱՐԴՅՈՒՆՔՆԵՐԻ
ՀԱՄԵՄԱՏՈՒՄԸ ԻՐԱԿԱՆ ԺԱՄԱՆԱԿԻ ՏՎՅԱԼՆԵՐԻ ԴԱՍԱԿԱՐԳՄԱՆ
ԴԵՊՔՈՒՄ**

Վերջին մի քանի տարիների ընթացքում մեքենայական ուսուցման ոլորտում գրանցվել է մեծ առաջընթաց: Մասնավորապես, հայտնի են խոր նեյրոնային, այդ թվում՝ նաև փաթույթային և ռեկուրենտ ցանցերի կիրառումը պատկերների, ձայնի և ազդանշանների մշակման համար: Սակայն գործնականում չկան կատարված հետազոտություններ ֆինանսական գործընթացներում իրական ժամանակի տվյալների դասակարգման մեթոդների և միջոցների ուղղությամբ: Հետազոտվում են խոր նեյրոնային ցանցերը՝ իրական ժամանակի տվյալների դասակարգման համար, ինչպես նաև առաջարկվում են նոր մոդելներ և արդյունավետ մեթոդներ՝ գործարքների արդյունքում ստացված տվյալները վերլուծելու համար:

Առանցքային բաներ. մեքենայական ուսուցում, դասակարգում, իրական ժամանակի տվյալներ, փաթույթային նեյրոնային ցանցեր, ռեկուրենտ նեյրոնային ցանցեր:

Ներածություն: Սկսած 2006թ.՝ խոր մեքենայական ուսուցումը (որը կրճատ կոչվում է խոր ուսուցում) առաջ եկավ, որպես մեքենայական ուսուցման նոր ուղղություն [1]: Վերջին տասնամյակի ընթացքում խոր ուսուցման տեխնոլոգիաներն ու մեթոդները սկզբնավորեցին մեքենայական ուսուցման նոր փուլը: Այդ ամենին մեծապես նպաստեցին նաև համակարգիչների հզորության աճը (այդ թվում՝ նաև գրաֆիկական պրոցեսորների առաջ գալը) և հասանելի տվյալների էքսպոնենտային աճը: Չնայած այդ առաջընթացին, ներկայումս շատ հետազոտություններ չեն կատարվում իրական ժամանակում մեքենայական ուսուցման կիրառման ուղղությամբ, որն ունի մի շարք առանձնահատկություններ [2]: Սակայն, մյուս կողմից՝ իրական ժամանակի համակարգերի կիրառումը հետզհետե սկսում է ընդլայնվել, և հետևաբար՝ ավելի նպատակահարմար է դառնում մեքենայական ուսուցման կիրառումը նման համակարգերում՝ անընդհատ նոր գեներ-

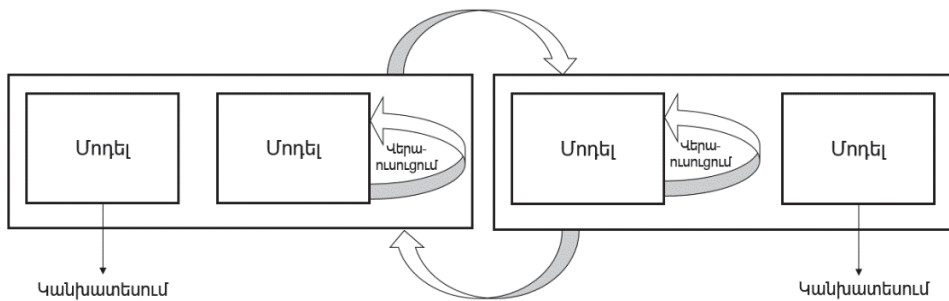
րացվող տվյալներով աշխատելու համար: Ինչպես ցույց է տրվել մեր նախորդ հետազոտություններում, դասական նեյրոնային ցանցերից նոր տեսակի նեյրոնային ցանցերին անցումը բավարար չէ համակարգի ճշտությունը բարձրացնելու համար [3]: Ավելին, ինչպես ասվում է «no free lunch» թեորեմում, ոչ մի ալգորիթմ համապիտանիորեն ավելի արդյունավետ չէ, քան մյուսները: Այդ պատճառով էլ նեյրոնային ցանցի կառուցվածքը պետք է ընտրել կոնկրետ խնդրի համար: Այս աշխատանքում ողջ հետազոտությունը կատարվում է խոր նեյրոնային ցանցերի ուղղությամբ՝ շեշտադրելով փաթայթային և ռեկուրենտ նեյրոնային ցանցերի կիրառումը:

Խնդրի դրվածքը: Աշխատանքի շրջանակներում իրական ժամանակի համակարգ ասելով հասկացվում է այնպիսի համակարգ, որի ճշտությունը կախված է ոչ միայն համակարգի տրամաբանական ճշտությունից, այլ նաև այն ժամանակից, որում այն աշխատում է [2]: Նման համակարգերի օրինակները առավել շատ են ֆինանսական ոլորտում, որտեղ տվյալները քանակապես անընդհատ աճում են և ժամանակի ընթացքում կրում են որակական փոփոխություն: Որպես օրինակ կարելի է դիտարկել խարդախ գործարքների հայտնաբերման խնդիրը: Այդ խնդրի լուծման եղանակներից է հետևյալ մոտեցումը. գառայան բաշխման միջոցով դուրս են բերվում կասկածելի գործարքները և դասակարգվում որպես ռիսկային [4], որից հետո համապատասխան որակավորում ունեցող մասնագետները պետք է կարողանան տարբերել ռիսկային գործարքների խարդախ կամ իրական լինելը: Ստացվում է, որ մասնագետը կատարում է դասակարգման խնդիր՝ իրական ժամանակում: Որպես մեկ ուրիշ օրինակ կարելի է դիտարկել բանկային օնլայն համակարգի միջոցով վարկի հարցման բավարարման կամ մերժման խնդիրը: Երկու խնդիրներն էլ հեշտությամբ մոդելավորվում են որպես մեքենայական ուսուցման դասակարգման խնդիրներ, և այս աշխատանքում հետազոտությունները և փորձերը կկատարվեն այդ խնդիրների օրինակով:

Քանի որ իրական ժամանակի համակարգում տվյալները ժամանակի ընթացքում կարող են կրել որակական փոփոխություն, առաջանում է մոդելի անընդհատ թարմացման անհրաժեշտություն: Որպես այդ խնդրի լուծում կարող է լինել հետևյալ մոտեցումը. ամակարգում պահվում է մեկի փոխարեն երկու մոդել: Սկզբում մոդելներից առաջինը գտնվում է կանխատեսումներ կատարելու, իսկ երկրորդը՝ տվյալների հավաքագրման վիճակում: Որոշ քանակությամբ տվյալների հավաքագրումից հետո երկրորդ մոդելը սկսում է վերաուսուցանվել, որից

հետո անցնում է կանխատեսումներ կատարելուն, վիճակին, իսկ առաջին մոդելը սկսում է հավաքագրել տվյալներ: Որոշ չափով տվյալներ հավաքելուց հետո առաջին մոդելը կատարում է վերաուսուցում, որից հետո անցնում է կանխատեսումներ կատարելու, իսկ երկրորդ մոդելը՝ տվյալների հավաքագրման վիճակին և այդպես շարունակ: Այդ գործընթացը պատկերված է նկ. 1-ում:

Սակայն այս աշխատանքի շրջանակում այս մոտեցումը չի կիրառվում (դա կհետազոտվի հետագա աշխատանքներում): Այսինքն՝ ուսուցումը կկատարվի ոչ իրական ժամանակում՝ նախօրոք հավաքված տվյալներով, իսկ կանխատեսումները կկատարվեն իրական ժամանակում:



Նկ. 1. Փոփոխվող տվյալները մշակելու համակարգի հնարավոր տարբերակը

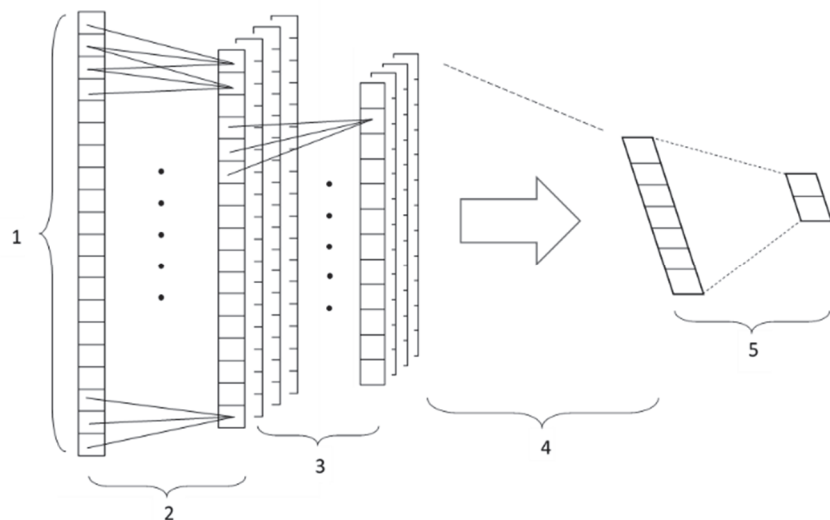
Կատարված հետազոտությունները և առաջարկվող մեթոդները: Ուսումնասիրելով գրականությունը՝ պարզ է դառնում, որ արդեն իսկ կան կատարված աշխատանքներ՝ խոր նեյրոնային ցանցերն իրական ժամանակի տվյալների դասակարգման խնդիրների լուծման ուղղությամբ [5...7]: Մասնավորապես՝ հետազոտությունները կատարված են ֆինանսական խարդախ գործարքների հայտնաբերման խնդրի լուծման համար: Աշխատանքներից մեկում կատարվում է նաև դասական և փաթայթային նեյրոնային ցանցերի արդյունքների համեմատում [5]:

Այդ աշխատանքում նշվում է, որ փաթայթային նեյրոնային ցանցերի՝ բանկային գործարքների տվյալների վրա կիրառման արդյունքում ստացվել է առավել մեծ ճշտություն՝ հայտնի այլ ալգորիթմերի համեմատ: Այդ ալգորիթմների շարքում է նաև դասական նեյրոնային ցանցը: Փաթայթային նեյրոնային ցանցերը հիմնականում կիրառվում են երկչափ տվյալներով աշխատանքի համար [8]: Այդ պատճառով էլ վերը նշված աշխատանքում մուտքային տվյալները վերածնափոխվում են մեկչափանիից երկչափանիի, որից հետո միայն տրվում նեյրոնային ցանցի մուտքին: Դա, բնականաբար, պահանջում է հավելյալ գոր-

ծողություններ և դանդաղեցնում է ողջ համակարգի աշխատանքը: Բացի այդ, ինտուիտիվ չէ կիրառել երկչափանի նեյրոնային ցանց՝ միաչափ տվյալների մշակման համար: Հետևաբար՝ առաջարկվում է ֆինանսական տվյալների հետ աշխատանքի կատարման համար կիրառել մեկչափանի նեյրոնային ցանց, որի ընդհանուր կառուցվածքը պատկերված է նկ.2-ում:

Պատկերում բերված համարակալումներին համապատասխանում են՝

1. ուսուցման տվյալներից մեկ օրինակ,
2. փաթուփավորման գործողություն,
3. մաքսիմում կամ միջին միավորման գործողություն,
4. այն փաստի ցուցադրումը, որ կարող են լինել մեկից ավելի փաթուփավորման և միավորման գործողություններ,
5. լրիվ կապակցված ցանց:



Նկ. 2. Մեկչափանի փաթուփային նեյրոնային ցանցի ընդհանուր կառուցվածքը

Նեյրոնային ցանցի մուտքին տրվում է ուսուցման տվյալներից մեկ օրինակ: Յուրաքանչյուր օրինակի համար կատարվում է փաթուփավորման գործողությունը՝ մի քանի ֆիլտրերի կիրառմամբ: Դրանից հետո կատարվում է մաքսիմում կամ միջին միավորման գործողությունը: Փաթուփավորման և միավորման բոլոր շերտերին հաջորդում է լրիվ կապակցված ցանցը [6]: Փորձերի արդյունքում պարզ է դառնում, որ իրական ժամանակի տվյալների դասակարգման խնդրի դեպքում մեկչափանի նեյրոնային ցանցերի կիրառումը բարձրացնում է ոչ միայն ցանցի արագագործությունը, այլ նաև ճշտությունը:

Ֆինանսական խարդախ գործարքների հայտնաբերման խնդրի շրջանակներում գրականությունում արդեն իսկ կան հետազոտություններ նաև ռեկուրենտ նեյրոնային ցանցերի կիրառման վերաբերյալ [6]:

Նեյրոնային ցանցը համարվում է ռեկուրենտ, եթե նրանում կա առնվազն մեկ հետադարձ կապ [9]: Գրականությունից հայտնի աշխատանքում կիրառվում են փականով (gated) LSTM (long short-term memory) տիպի ռեկուրենտ նեյրոնային ցանցեր՝ վերացող կամ պայթող գրադիենտի (vanishing or exploding gradient) խնդրից խուսափելու համար [6]: Աշխատանքում ցույց է տրվում, որ ռեկուրենտ նեյրոնային ցանցերն ապահովում են ավելի բարձր ճշտություն, քան հայտնի SVM (support vector machines) ալգորիթմը: Սակայն նշված աշխատանքում չկա ոչ մի համեմատություն այլ տեսակի նեյրոնային ցանցերի հետ: Ավելին, համեմատությունը SVM ալգորիթմի հետ կատարվում է միայն ROC (receiver operating characteristic) կորերի միջոցով, և չկա թվային ոչ մի արդյունք, որը կարելի է համեմատել այլ ալգորիթմերի հետ: Մեր աշխատանքի շրջանակներում կատարվում է նման փորձ՝ թվային արդյունքներ ստանալու և դրանք այլ ալգորիթմերի հետ համեմատելու համար:

Կատարված փորձերը և ստացված արդյունքները: Նեյրոնային ցանցերի իրականացման և փորձարկման համար օգտագործվել է Keras գրադարանը՝ Tensorflow-ի հիման վրա [10]: Որպես փորձարկային ուսուցման տվյալներ օգտագործվել են «German Credit Card» տվյալները [11]: Քանի որ օգտագործվող տվյալների քանակը մեծ չէ, նեյրոնային ցանցի հիպերպարամետրերի ընտրության համար կիրառվում է k-fold cross validation ալգորիթմը, որից ընտրվում է k=10 [8]: Որպես կորստի ֆունկցիա (loss function) ընտրվում է երկուական խաչաձև էնտրոպիան (binary crossentropy) [8]: Որպես ակտիվացման ֆունկցիա ընտրվում է ուղղված գծային միավոր ֆունկցիան (rectified linear unit) [8]: Քանի որ լուծում ենք երկուական դասակարգման խնդիր, որպես ելքային շերտի ակտիվացման ֆունկցիա ընտրվում է սիգմոիդ ֆունկցիան [8]: Որպեսզի մոդելը լինի լավ ընդհանրացվող, կիրառվում է dropout մեթոդը [8]: Բոլոր նշված արժեքներն ընտրվել են՝ ելնելով առավել հայտնի հետազոտական աշխատանքներից [1, 4, 8, 10]: Մեր կողմից ստացված արդյունքներն արդեն իսկ հայտնի արդյունքների հետ համեմատելու համար օգտագործվում է F_1 պարամետրը [4]՝

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}};$$

Փորձերի արդյունքում ստացվեցին F_1 -ի հետևյալ արժեքները՝ մինչև 0,8՝ փաթույթային նեյրոնային ցանցերի դեպքում (միայն մեկ փաթույթավորման և միավորման գործողությունների կիրառմամբ): Ստացված արդյունքն ավելի լավն է, քան [5]-ում ստացվածը: Իսկ ռեկուրենտ նեյրոնային ցանցերի կիրառմամբ F_1 -ի արժեքը ստացվեց մինչ 0,92՝ ընդամենը 2 թաքնված շերտի կիրառմամբ: Ստացվում է, որ ռեկուրենտ նեյրոնային ցանցերի կիրառման դեպքում ստացված արդյունքն ավելի լավն է, քան փաթույթային նեյրոնային ցանցերի կիրառման դեպքում ստացված արդյունքները, և էլ ավելի լավն է, քան արդեն իսկ հայտնի արդյունքը [5]-ում, որն էլ, իր հերթին, ավելի լավն է, քան մյուս հայտնի ալգորիթմների արդյունքները: Այսպիսով, կարելի է եզրակացնել, որ ռեկուրենտ նեյրոնային ցանցերի միջոցով ստացվում է առավել մեծ ճշտությունը: Ավելին, համոզվելու համար, որ LSTM-ի միջոցով ստացված մոդելը լավ ընդհանրացվող է, այն փորձարկվել է նաև մեկ այլ տվյալների համեմատությամբ [12], որտեղ F_1 -ի արժեքը ստացվել է մինչև 0,91:

Եզրակացություն: Կատարվել են հետազոտություններ՝ տարբեր տեսակի խոր նեյրոնային ցանցերն իրական ժամանակի տվյալների դասակարգման խնդրի լուծման համար: Առաջարկվել է նոր մեթոդ՝ մեկչափանի փաթույթային նեյրոնային ցանցերն իրական ժամանակի տվյալների դասակարգման խնդրի լուծման համար, որն ունի ավելի բարձր ճշտություն, քան գրականությունում հայտնի տարբերակը: Տարբեր տեսակի նեյրոնային ցանցերի արդյունքները համեմատելու համար կատարվել է գրականությունում հայտնի փորձի կրկնությունը՝ թվային արժեքներ ստանալու համար, որից հետո ստացված արդյունքների հիման վրա կարելի է եզրակացնել, որ իրական ժամանակի տվյալների դասակարգման խնդրի լուծման համար առավել լավ արդյունք կարելի է ստանալ ռեկուրենտ նեյրոնային ցանցերի կիրառմամբ, երբ հնարավոր է ստանալ մինչև 0.92 F_1 արժեք:

ԳՐԱԿԱՆՈՒԹՅԱՆ ՑԱՆԿ

1. **Deng L., Yu D.** Deep Learning: Methods and Applications // Foundations and Trends in Signal Processing.- 2014.- Vol. 7, Nos. 3–4.- P. 197–230.
2. **Abroyan N., Hakobyan. R.** A Review of the Usage of Machine Learning in Real-time Systems // Proceedings of NPUA, Information technologies, Electronics, Radio engineering.- 2016.- № 1.- P. 46–54.
3. **Abroyan N.** Classification of Real-time Data Using Deep Learning // Proceedings of the 13th International Conference of Science and Technology, New Information Technologies and Systems.- Penza, Russia, 2016.- P. 109-112.

4. **Andrew N.** CS 229 Machine Learning Course Materials.- Stanford University [<http://cs229.stanford.edu/materials.html>].- 2016.
5. **Fu K., Cheng D., Tu Y., Zhang L.** Credit Card Fraud Detection Using Convolutional Neural Networks // Proceedings of 23rd International Conference, ICONIP.- Part III.- Kyoto, Japan, 2016.- P. 483-490.
6. **Abroyan N., Hakobyan R.** Convolutional Neural Networks for Real-time Data Classification // Proceedings of the 11th International Conference on Computer Science and Information Technologies (CSIT 2017).- Yerevan, Armenia, 2017.- P. 267-269.
7. **Wiese B., Omlin C.** Credit Card Transactions, Fraud Detection, and Machine Learning: Modelling Time with LSTM Recurrent Neural Networks // Innovations in Neural Information Paradigms and Applications.- 2009.- P. 235-272.
8. **Goodfellow I., Bengio Y., Courville A.** Deep Learning.- Cambridge, Massachusetts, The MIT Press, 2016.
9. **Haykin S.** Neural Networks and Learning Machines.- 3rd ed.- McMaster University, Hamilton, Ontario, Canada, 2009.
10. **Chollet F.** Keras: Deep Learning library for Theano and TensorFlow // Github, [<https://github.com/fchollet/keras>].- 2015.
11. **Lichman M.** UCI Machine Learning Repository // [<http://archive.ics.uci.edu/ml>], Irvine, CA: University of California, School of Information and Computer Science.- 2013.
12. **Pozzolo A., Caelen O., Johnson R., Bontempi G.** Calibrating Probability with Undersampling for Unbalanced Classification // Symposium on Computational Intelligence and Data Mining (CIDM), IEEE.- 2015.- P. 159-166.

Н.О. АБРОЯН

ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ И СРАВНЕНИЕ РЕЗУЛЬТАТОВ ПРИ КЛАССИФИКАЦИИ ДАННЫХ РЕАЛЬНОГО ВРЕМЕНИ

В течение последних нескольких лет были достигнуты значительные успехи в области машинного обучения. В частности, известны случаи применения глубоких нейронных сетей, включая сверточные и рекуррентные сети, для обработки изображений, речи и сигналов. Однако практически отсутствуют исследования методов и средств транзакций данных реального времени в процессах финансовых операций. В рамках этой работы исследованы глубокие нейронные сети для классификации данных реального времени, новые модели и эффективные методы анализа результатов транзакции.

Ключевые слова: машинное обучение, классификация, данные реального времени, сверточные нейронные сети, рекуррентные нейронные сети.

N.H. ABROYAN

THE USAGE OF NEURAL NETWORKS AND COMPARISON OF RESULTS AT CLASSIFYING THE REAL-TIME DATA

In recent years, notable success has been recorded in the sphere of machine learning. Particularly, different usages of deep neural networks, including convolutional and recurrent neural networks in image, speech and signal processing are known. However, there is practically no research carried out to find methods for using deep neural networks in elaboration of real-time data of financial transactions. In the frames of this work, the deep neural networks for classifying the real-time data, and also finding new models and methods for analyzing the transaction results are investigated.

Keywords: machine learning, classification, real-time data, convolutional neural networks, recurrent neural networks.

ՀՏԴ 004.22:004.421

Բ.Գ. ԱԹԱՅԱՆ

ԱՄՊԱՅԻՆ ՊԱՀՈՒՍՏԱՎՈՐՄԱՆ ՀԱՄԱԿԱՐԳՈՒՄ ՏՎՅԱԼՆԵՐԻ ՊԱՀՈՒՍՏԱՎՈՐՄԱՆ ՕՊՏԻՄԱԼԱՑՄԱՆ ՄԵԹՈԴ

Ներկայացվում է ամպային պահուստավորման համակարգում տվյալների պահուստավորման օպտիմիզացման մեթոդ, որը հիմնված է գաղտնագրված տվյալների դեդուպլիկացիայի վրա: Համակարգն ապահովում է բազմակի պահուստավորվող տվյալների դեդուպլիկացիա՝ կոնվերգենտ գաղտնագրության շնորհիվ, իսկ եզակի տվյալների անվտանգության ապահովման համար օգտագործվում է գաղտնագրության երկրորդ մակարդակը: Համակարգում բոլոր տվյալները խմբավորվում են՝ ըստ տվյալի հայտնիության աստիճանի, որը որոշվում է տվյալը տարբեր օգտագործողների կողմից պահուստավորելու քանակով:

Առանցքային բառեր. ամպային պահուստավորում, կոնվերգենտ գաղտնագրում, դեդուպլիկացիա, բազմաշերտ գաղտնագրում, վեբ կայքերի պահուստավորում:

Ներածություն: Տվյալների քանակը տարեցտարի աճում է գերաբազ կերպով, ինչի հետևանքով ստեղծվում են նոր լուծումներ մեծածավալ տվյալների պահպանման համար: Ներկայումս լայն ճանաչում են ստացել ամպային պահուստավորման միջոցները, որոնք թույլ են տալիս օգտագործողներին պահպանել իրենց տվյալները առցանց: Ամպային պահուստավորման միջոցներն ունեն մի շարք առավելություններ՝ ի տարբերություն տվյալների պահուստավորման ավանդական միջոցների: Սակայն ինչպես այլ ոլորտներում, այստեղ նույնպես