УДК 004.65

Ә.Ф. Галимҗанов,
Р.Ә. Галимҗанов

ЗУР МӘГЪЛҮМАТ МАССИВЛАРЫН ЭШКӘРТКӘНДӘ КУЛЛАНЫЛУЧЫ НЕЙРОН ЧЕЛТӘРЛӘРЕ ТИГЕЗЛӘМӘЛӘРЕ ҺӘМ АЛАРНЫ ЧИШҮ ЫСУЛЛАРЫ

Для моделирования нейронной сети с Residual архитектурой предложено дробно-интегральное уравнение со специальным дробным оператором. Предложенное уравнение решено прямым методом, получена оценка погрешности приближенного решения к точному.

Ключевые слова: нейронные сети, residual архитектуры, дифференциальные уравнения, приближенные методы

To simulate a neural network with Residual architecture, a fractional-integral equation with a special fractional operator is proposed. The proposed equation is solved by a direct method, an estimate of the error and an approximate solution to the exact one is obtained.

Keywords: neural networks, residual architectures, differential equations, approximate methods.

Residual архитектуралар. Нейрочелтәр архитектуралар арасында residual элемтәләр механизмын файдаланучы архитектуралар популярлашты. Зур мәгълүмат массивларын анализлаганда андый архитектура мисалы булып ResNet тора. Бу архитектура сурәтләрне анализлаганда бик актив файдаланыла, ләкин аны теләсә нинди структуралы зур мәгълүмат массивларында, шул исәптән, мәсәлән, иске татар язуында закончалыклар эзләү һәм кластеризация өчен файдаланырга мөмкин.

ResNet ны кыскача калдык нейрон челтәре дип әйтергә була. Калдык нейрон челтәре ResNet ясалма нейрон челтәре булып тора, аның прототибы – баш мие кабыгындагы пирамидаль күзәнәкләр. Калдык нейрон челтәрләре кайбер катламнар аша үтү өчен тоташулар яисә ярлыкларны калдырып китүләре белән үзенчәлекле. ResNet ның гадәти модельләре ике яисә өч катламны калдырып китүләре белән аерыла, аларга шулай ук нәсызыкчалыклар һәм пакет нормализациясе керә. Алар өчен өстәмә матрица файдаланылырга мөмкин, ул калдырып кителгән элемтәләр микъдарын билгеләү өчен кулланыла. Бу модельләр HighwayNets исеме белән билгеле. Берничә параллель калдырып китүләр булган модельләр DenseNets дип атала. Калдык нейрон челтәрләре белән нисбәтле рәвештә калдык булмаган нейрон челтәрен гади челтәр дип атарга мөмкин.

Калдырып кителүче элемтәләр өстәү сәбәпләре түбәндәгечә. Беренчедән, бу – югалучы градиентлар проблемасын булдырмау яисә деградация проблемасын (төгәллек туенучанлыгын) киметү зарурлыгы. Ә бусы исә күбесе бик якынча булган зур сандагы мәгълүмат өчен бик актуаль. Икенчедән, җитәрлек дәрәҗәдә тирән булган модельгә зур сандагы катламнар өстәү өйрәтүдә зуррак хаталарга китерә. Өйрәтү вакытында микъдарлар күтәрелүче катлам комачауларын өзү һәм элегрәк калдырып кителгән катламны көчәйтү өчен адаптацияләнәләр (ярашалар). Иң гади очракларда исә күрше катламны тоташтыру өчен микъдарлар гына яраша, өстәрәк катлам өчен микъдарлар анык ярашмый. Бу бары бер нәсызыкча катлам калдырып кителгәндә, яисә барлык арадаш катламнар сызыкча булганда оптималь. Башкача калдырып кителгән катлам микъдарларының анык матрицасын өйрәнү, ягъни өстәмә матрицаны файдалану мәгъкуль.

Калдырып китү шулай ук челтәр структурасын гадиләштерә, чөнки өйрәтүнең башлангыч стадияләрендә азрак сандагы катламнар файдаланыла. Шулай итеп, өйрәтү тизләнә, югалучы градиентлар тәэсире кими, чөнки таралу өчен азрак сандагы катламнар кулланыла. Ләкин соңрак челтәр объектлар пространствосын өйрәнү барышында катламнарны тергезә. Өйрәтү процессының ахырына, барлык катламнар җәелгәндә, ул коллекторга якынрак булып кала, шулай итеп, тизрәк өйрәнә. Калдык өлешләре булмаган нейрон челтәре күбрәк функцияләр пространствосын тикшерә, бу исә аны коллекторны калдырып китәргә мәҗбүр итүче комачауларга бирешүчән ясый һәм тергезү өчен өстәмә мәгълүмат таләп итә. ResNet архитектурасы авторлары тарафыннан билгеләп үтелгәнчә, тасвирланган катлам челтәргә элекке катламнардан информацияне җиңел «сакларга» мөмкинлек бирә. Шуңа күрә сыйфатны югалтмыйча челтәр тирәнлеген сизелерлек арттыруга ирешелде.

Калдык челтәрләр баш миен яхшырак модельли, чөнки кабык катламы нейроннары кереш мәгълүматны, арадаш катламнарны калдырып китеп, турыдан-туры беренче катламнан ала.

Тасвирланган элемтә механизмының математик моделен карыйк.

Теорема (Универсаль аппроксимация теоремасы, K. Hornik, 1991) Теләсә нинди өзлексез функция өчен для f(x) функциясен бирелгән төгәллек белән якынайтучы сызыкча чыгышлы a(x) нейрон челтәре бар.

Стандарт катлам керешенә челтәрнең баштагы t катламы эше нәтиҗәсен кабул итә дип фараз итик – бу күрсәтелешне h_t дип тамгалыйк. Ул вакытта челтәр чыгышын болай билгеләп була [Kaiming He..., 2015]:

$$h_{t+1} = f(h_t, \theta_t),$$

биредә $f(h_t, \theta_t)$ дип ниндидер сызыкча үзгәртү тамгаланган, аны практикада тулы бәйләнешле конволюцион катламнар, аларның комбинациясе рәвешендә күрсәтергә мөмкин.

$f(h_t, \theta_t)$ үлчәме h_t үлчәменә тәңгәл дип фаразлыйк. Ул вакытта residual элемтәле катлам дип мондый үзгәртү атала:

$$h_{t+1} = f(h_t, \theta_t) + h_t, \qquad (1)$$

Вектор очракта кушу элементлап башкарыла.

h_t күрсәтелешен t дан функция кебек карыйк, ягъни $h_t = h(t)$. Ул вакытта, катлаулы булмаган үзгәртүләр башкарып, табабыз:

$$h(t+1) - h(t) = f(h(t), \theta_t).$$

$$\frac{h(t+1) - h(t)}{(t+1) - t} = f(h(t), \theta_t).$$

Табылган аңлатманы үзгәртик. $h(t)$ функциясен [a, b] кисемтәсендә өзлексез аргументтан өзлексез функция кебек карый һәм чыгарылмага күчик, ягъни мондый рәвешле языйк

$$\frac{h(t+\Delta t) - h(t)}{\Delta t} = f(h(t), \theta_t),$$

Катламнар санын чиксезлеккә омтылдырып, табабыз:

$$\frac{dh(t)}{dt} = f(h(t), \theta_t) \qquad (2)$$

Модельләүнең сыгылмалырак ысулын карыйк. Бөтен тәртипле чыгарылмаларны гомимуләштерү буларак вакланма тәртипле чыгарылмалар кулланыла. Алар практикада, мәсәлән, нейтроннар таралуны модельләгәндә, биологиядә, оптикада киң кулланылыш таптылар. Ягъни, (2) тигезләмәсе урынына мондый модель тигезләмә тәкъдим ителә [Galfmyanov, Vorontsova, Gorskaya]

$$D^{(\alpha)}h(t) = f(h(t), \theta_t), \;\; h, f \in L_2[0,2\pi] \qquad (3)$$

Белгәнебезчә, сызыкча үзгәртү юлы белән һәр кисемтәне $[0,2\pi]$ кисемтәсенә китерергә һәм периодик дәвам итәргә мөмкин, ягъни (2) дә күрсәтелгән шартлар гомумилекне бозмый.

Биредә $D^{(\alpha)} = \frac{D_+^{(\alpha)} + D_-^{(\alpha)}}{2\cos\left(\frac{\alpha\pi}{2}\right)}$ симметрик операторы $D_\pm^{(\alpha)}$ вакланма дифференциаллау операторлары ярдәмендә билгеләнә. Алар исә [a,b] кисемтәсендә билгеләнгән $\varphi(t)$ функциясе өчен мондый формулалар ярдәмендә билгеләнәләр [3]:

$$\left(D_{a+}^{(\alpha)}\right)\varphi(t) = \frac{1}{\Gamma(1-\alpha)}\frac{d}{dt}\int_a^t \frac{\varphi(x)dx}{(t-x)^\alpha}, -\infty < t < \infty,$$

$$\left(D_{b-}^{(\alpha)}\right)\varphi(t) = \frac{1}{\Gamma(1-\alpha)}\frac{d}{dt}\int_t^b \frac{\varphi(x)dx}{(x-t)^\alpha}, -\infty < t < \infty,$$

болар Риман-Лиувилль сул һәм уң вакланма чыгарылмалары, $0 < \alpha < 1$, Γ(x) – Эйлер гамма-функциясе. Табигый ки, вакланма чыгарылмаларны теләсә нинди $\alpha \geq 1$ өчен билгеләп була һәм барлык нәтиҗәләр дөрес булыр иде, ләкин биредә без $0 < \alpha < 1$ очрагы белән чикләнәбез.

(3) тигезләмәсен якынча чишү методын төзегәндә без $D^{(\alpha)}$ ның симметрик һәм уңай билгеләнгән оператор булуын файдаланырбыз [Galfmyanov, Vorontsova, Gorskaya, 2015]. Моны исәпкә алып, скаляр тапкырчыгыш һәм норма кертәбез:

$$[h,v] = \left(D^{(\alpha)}h, v\right), [h] = \left(D^{(\alpha)}h, h\right)^{1/2}.$$

$D = L_2[0,2\pi]$ пространствосын кертелгән норма буенча тутырып энергетик пространство табабыз һәм аны H_D дип тамгалыйбыз.

θ_t ны параметр кебек карап һәм башлангыз (3) тигезләмәсен ирекле $v \in H_D$ функциясенә тапкырлап, тигезләмә табабыз:

$$[h,v] = (f,v). \tag{4}$$

Табылган (4) тигезләмәсен Бубнов-Галеркин методы белән чишәбез. Бу метод буенча H_D энергетик пространствосында $\varphi_j, j = 1..N$ базис функцияләре системасы сайлана. Якынча чишелеш базис функцияләренең сызыкча комбинациясе рәвешендә эзләнә:

$$h_N = \sum_{j=1}^{N} c_j\varphi_j. \tag{5}$$

Ә билгесез коэффициентлар $c_j, j = 1..N$ түбәндәге сызыкча алгебраик тигезләмәләр системасыннан табыла:

$$[h_N, \varphi_k] = (f, \varphi_k),\ k = 1..N \tag{6}$$

Якынча чишелеш рәвешен һәм кертелгән һәм гадәти скаляр тапкырчыгышларның сызыкча булуын исәпкә алып, мондый сызыкча алгебраик тигезләмәләр системасын табабыз:

$$\sum_{j=1}^{N} c_j[\varphi_j, \varphi_k] = (f, \varphi_k),\ k = 1 \dots N \tag{7}$$

Тәкъдим ителгән метод җыела һәм аның өчен түбәндәге теорема үтәлә:

Теорема. (3) тигезләмәсе теләсә нинди уң як өчен бердәнбер чишелешкә ия һәм $L(u,v) = [u,v]$ формасы H_D – билгеләнгән һәм H_D – чикләнгән булсын, ягъни түбәндәге шартлар үтәлсен:

$L(u,u) \geq \gamma_0^2[u]^2, L(u,v) \leq \gamma_1^2[u][v], \gamma_0, \gamma_1 \equiv const$, моннан тыш последовательность подпространств $H_N - \varphi_j, j = 1..N$ функцияләренең сызыкча тышчасы булган аспространстволар эзлеклелеге H_D да чикләмә тыгыз булсын.

Ул вакытта теләсә нинди чикле N өчен (7) системасы бердәй чишелүчән һәм якынча чишелешләр h_N төгәл чишелеш h ка $N \to \infty$ булганда кертелгән энергетик пространство метрикасы буенча омтыла һәм

$$[h - h_N] \leq c\varepsilon(h,N),$$

биредә $\varepsilon(h,N)$ – N га бәйле функция (якынча чишелешнең төгәлсезлек бәясе), һәм ул мондый тигезсезлекне канәгатьләндерә:

$$\min_{c_j}\left\|D^{(\alpha)}\left(h-\sum_{j=1}^{N} c_j\varphi_j\right)\right\| \leq \varepsilon(h,N) \to 0.$$

Бу вакытта, билгеле, f функциясенең θ параметрларын оптимальләштерү мәсьәләсе дә туа. Аны [Kaiming He, 2015; Vaswani.., 2017; Chen.., 2018] чыганакларында китерелгән нәтиҗәләргә таянып башкарырга мөмкин.

Әдәбият

Kaiming He and Xiangyu Zhang and Shaoqing Ren and Jian Sun. (2015) Deep Residual Learning for Image Recognition. CoRR

Galfmyanov A.F., Vorontsova V.L., Gorskaya T.Y.. Approximate methods for the equations with fractional differential operator//Global Journal of Pure and Applied Mathematics. 2015. Vol. 11, Is. 6. P. 5133–5144.

Самко С.Г., Килбас А.А., Маричев О.Н. Интегралы и производные дробного порядка и некоторые их приложения. Минск: Наука и техника, 1987. 688 с.

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. (2017). Attention Is All You Need. NIPS.

Chen, Tian Qi and Rubanova, Yulia and Bettencourt, Jesse and Duvenaud, David K. (2018). Neural Ordinary Differential Equations. Advances in Neural Information Processing Systems 31, 6571–6583 NIPS2018

Галимҗанов Әнис Фоат улы,
ТР ФА Гамәли семиотика фәнни-тикшеренү институты фәнни хезмәткәре

Галимҗанов Ринат Әнис улы,
Казан дәүләт энергетика универститеты аспиранты