

ANALYSIS OF AUTOMATIC DOCUMENT CLASSIFICATION ALGORITHM BASED ON USING ARTIFICIAL NEURAL NETWORKS

Khonarbaev David Kalbaevich

Assistant teacher, Nukus branch of the Tashkent University of
Information Technologies named after Muhammad al-Khorazmi
<https://doi.org/10.5281/zenodo.11112272>

ARTICLE INFO

Received: 27th April 2024

Accepted: 03rd May 2024

Online: 04th May 2024

KEYWORDS

Neural network, technologies,
models, algorithm, formulas
and functions.

ABSTRACT

This article provides more information about neural network technologies and how they are used. The authors provide important information on boosting algorithms, using genetic algorithms, minimum spanning trees, and clustering algorithms. The use of neural networks in the classification of electronic document flow attracts more developers of tools. Neural network mathematical models and their working principles are correctly shown. In particular, the formulas and functions of the input and output mechanics of the neural network are described. This information can be useful for learning and researching neural network technologies.

SUN'IY NEYRON TARMOQLARIDAN FOYDALANISHGA ASOSLANGAN AVTOMATIK HUJJAT TASNIFI ALGORITMINI TAHLIL QILISH

Qonarbaev David Xalbaevich

Assistent-o'qituvchi, Muxammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari
universiteti Nukus filiali

<https://doi.org/10.5281/zenodo.11112272>

ARTICLE INFO

Received: 27th April 2024

Accepted: 03rd May 2024

Online: 04th May 2024

KEYWORDS

Neyron tarmog'i,
texnologiyalar, modellar,
algoritm, formulalar va
funktsiyalar.

ABSTRACT

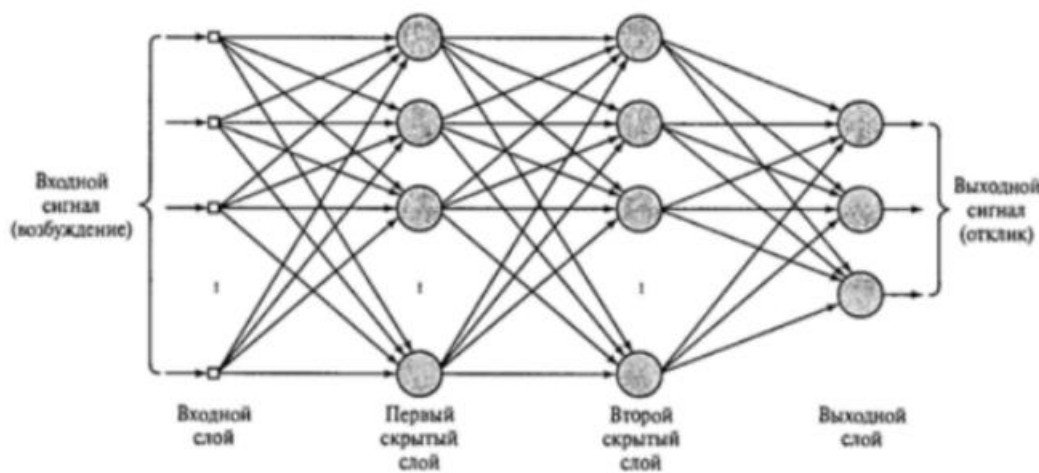
Ushbu maqolada neyron tarmog'i texnologiyalari va ularning foydalanish usullari haqida ko'proq ma'lumot berilgan. Mualliflar algoritmlarni oshirish, genetik algoritmlardan foydalanish, minimal qoplamali daraxtlar va klasterlash algoritmlari haqida muhim ma'lumotlar ko'rsatadilar. Elektron hujjat oqimini tasniflashda neyron tarmoqlaridan foydalanish vositalarini ishlab chiquvchilarni ko'proq jalb qiladi. Neyron tarmog'i matematik modellari va ularning ishlovchi prinsiplari to'g'ri ko'rsatilgan. Ayniqsa, neyron tarmog'ining kiritish va chiqish mexanikasi bo'yicha formulalar va funksiyalar tasvirlangan. Ushbu ma'lumotlar neyron tarmoq texnologiyalarini o'rganish va tadqiq qilish uchun foydali bo'lishi mumkin.

Bugungi kunda AOT sohasidagi ko'pgina tadqiqot yoki amaliy loyihalar neyron tarmoq texnologiyalaridan u yoki bu tarzda foydalanadi. Ayniqsa, tarqalgan so'z vektorlarining modellari, masalan, arifmetik amallarni bajarishingiz mumkin bo'lgan bunday vektorlar bilan so'zning ma'nosini ifodalashga imkon beradi, masalan, qo'shish yoki ayirish. Borisjuk F. V. hujjat to'plashning eng keng tarqalgan usullari haqida umumiy ma'lumot beradi [5].

U algoritmni "oshirish oynasini surish" usuli asosida amalga oshirishni taklif qilmoqdalar. Bu algoritm dinamik gugurt darajasini sozlash va shunga o'xshash variantlarni topadi, va u grammatik xatolar va o'zgartirilgan so'z tartibi bilan ishlash mumkin. Qog'oz genetik algoritm asosida loyqa klassifikatordan foydalanishni tahlil qiladi. Shuningdek, mualliflar minimal qoplamali daraxtning algoritmilaridan birini ko'rib chiqadilar. Bu algoritmda grafik yaratilgandan so'ng yeng yuqori og'irlikdagi qirralar olib tashlanadi.

Klasterlash algoritmлари bu algoritmni amalga oshirishning katta usullarini tahlil qilib, bitta klasterlash usuliga yega bo'lish mumkin yemas, degan xulosaga keldilar. Elektron hujjat oqimini tasniflashda sun'iy neyron tarmoqlaridan foydalanish har yili 174 kataloglashtirish vositalarini ishlab chiquvchilarni ko'proq jalb qiladi. Tilshunoslarning neyron tarmoqlariga bo'lgan e'tiborining konsentratsiyasi bir qator sabablar bilan bog'liq, masalan, neyron tarmoqlari matnli hujjatlar bilan ishlash tezligini sezilarli darajada oshiradi, ular bilan ishlashning murakkabligini kamaytiradi.

O'z-o'zini o'qitish neyron tarmog'i-bu mashina o'qitishning turli jihatlarida qo'llaniladigan matematik modeldir. Neyron tarmog'ining soddalashtirilgan tasviri 1-rasmda ko'rsatilgan [2].



1 – rasm. Ko'p katlamli neyron tarmoq

Rasmda ko'rsatilganlarga asoslanib, neyron tarmog'i ma'lum arxitekturaning grafigidan boshqa narsa yemasligi aniq. Tepaliklar va tugunlar neyronlar deb ataladi. Ularning soni hal qilinayotgan muammoga bog'liq. Ta'kidlash joizki, neyron tarmog'ining muhim ko'rsatkichlari neyronning kirish signaliga javobini ma'lum minimal holatga ushlab turuvchi va unga yetganda axborotni signal va undan keyingi shaklda uzatish imkonini beruvchi chekka og'irligi (sinaps) va neyronning aktivlanish funksiyasidir. Og'irlik bir neyronning bir-biridan oquvchi axborotlarning "ustunligini" belgilaydi. Tarmoqdagi so'zning matematik modeli ham vektor bo'lgan satri matritsa bilan ifodalanadi. Buning sababi neyron tarmog'i faqat 0 dan 1 gacha



bo'lgan sonlar qatori bilan ishlaydi. Neyron tarmog'ining ishi quyidagi formulalar bilan tasvirlanadi:

$$H_{1input} = (I_1 W_1) + (I_2 W_2), \quad (1)$$

$$H_{2input} = f_{activation}(H_{1input}), \quad (2)$$

I – kirish neyron soni, H – yashirin qatlam W- vazin

Birinchi formulada kiritish axboroti deb, ularning og'irliklari bo'yicha ko'paytirilgan kiritish ma'lumotlari yig'indisiga aytiladi. Chiqish ma'lumotlari neyron faollashuvining o'tkazib yuborilgan funksiyasidir. Tahlil namunasi quyidagi elektron hujjatlardan iborat: iqtisodiy sharh, mashinasozlik bo'yicha analitik maqola, algebrada teoremaning isboti va falsafiy dalillar work. Matn tasnifi ustida ishlashning butun jarayonini bir necha bosqichlarga ajratadi: - ma'lumotlar bilan ishlash - oldindan tayorlash; - tarmoq sifatini loyihalash, o'qitish va baholash - foydalanish va diagnostics. At ma'lumotlar bilan ishlash bosqichi, kerakli hajmdagi ma'lumotlarning terminologik bazasi yig'iladi [1].

Neyron tarmoqlari bo'yicha Klassifikatorni ishlab chiqishda yeng muhim qism hisoblanadi. Namunalar tegishli sinflar bilan korrelyatsiyalanadi va yeng kam talab qilinadigan namunalar soni yaratiladi, chunki ortiqcha hollarda tizim ishdan to'xtaydi va yetishmovchilikda tizim ishonchli ishlamaydi. Hosil bo'lgan terminologik to'plam bir necha to'plamlarga bo'linadi. Ushbu to'plamlar mashinani o'qitishning aniqligi talablariga qarab ikki yoki uch bo'lishi mumkin: ta'lim, sinov va tasdiqlash.

Birlamchi qayta ishlash bosqichi ma'lumotlar bazasini neyron tarmog'iga qulay shaklda taqdim yetishni nazarda tutadi. Bunday ishlov berishning turli usullari mavjud: chiziqli o'zgarishdan nochiziqli me'yorlash va ko'p o'lchovli tahlilgacha. Axborotni kodlash quyidagi usullarda amalga oshirilishi mumkin sinf komponentalari bilan chiqish vektorini shakllantirish, bu usul yeng soddaga, lekin har doim ham optimal yemas; - chiqish vektori binar shaklda Klaster raqami deb yoziladi, klasterlar orasidagi masofa axborot uzatishning ishonchligini oshirish uchun bolg'alash kodi bilan oshirilishi kerak; - k sinflari bilan bog'liq muammoni $k(k-1)/2$ subproblems bilan ikki sinf (2 tomonidan 2 kodlash) har biriga bo'lish [4].

Uchinchi kodlash opsiyasi ko'p vazifalarda yaxshi natija beradi. Manba vektori chiqish vektorining barcha mumkin bo'lgan tarkibiy qismlarini o'z ichiga olishi uchun har birida ikki komponentdan iborat guruhlariga bo'linadi.

Guruhlar soni quyidagi formula bilan aniqlanadi:

$$A_k = \frac{k(k-1)}{2}$$

k- sinflar soni 4 sinfli muammo uchun sinflarni hisoblash natijasida 6 ta subtaska hosil bo'ladi, ularning har biri o'z chiqish qismlariga yega. Neyron tarmog'ining chiqishi bo'yicha birlik chiqish komponentlaridan birining mavjudligini ko'rsatadi. Bu holda siz sinf raqamiga o'tishingiz mumkin: qaysi kombinatsiyalar bitta (yoki juda yaqin) chiqish qiymatini (ya'ni, qaysi subtaskalar faollashtirilgan) olganini aniqlaymiz va sinf raqami yeng ko'p faollashtirilgan subtaskalarni kiritgan bo'ladi deb taxmin qilamiz. Neyron tarmog'ini shakllantirish ikki yondashuvdan biri bilan mumkin: - destruktiv; - constructive. In birinchi holda, tarmoq minimal olinadi va asta-sekin kengayadi, ikkinchi holda, tarmoq ortiqcha hajmga yega, unda



kichik tugunlar asta-sekin olib tashlanadi. Bunday holda, vaznlar soni misollar sonidan katta bo'lishi kerak, aks holda tarmoq tasniflash qobiliyatini yo'qotadi [4].

Neyron tarmoqlari o'z afzalliklariga yega: avtomatik tasniflash usullarida qo'llaniladigan tarmoqlardan real vaqt ma'noda chiqarish uchun ham, matn tonalligini tahlil qilish uchun ham foydalanish mumkin. Neyron tarmoq texnologiyalaridan foydalanishga asoslangan Google, Microsoft va Yandex qidiruv tarmoqlarining mashina tarjimasini. Kerakli joyda har qanday vazifada neyron tarmoq texnologiyalari matnlarni tasniflash uchun foydalanish mumkin, bu spam filtrlash yoki yo'qligini, firibgarlik aniqlash yoki kredit skoring, matn kayfiyatini aniqlash yoki depressiya hatto muallifning moyillik aniklash mumkin.

Xulosa. Neyron tarmoqlari bo'yicha klassifikatorni ishlab chiqishda yeng muhim qism namunalar bilan korrelyatsiyalanadi va yeng kam talab qilinadigan namunalar sonining yaratilishidir. Ortiqcha hollarda tizim ishdan to'xtatish va yetishmovchilikda tizim ishonchli ishlamaydi. Ma'lumotlar bazasini neyron tarmog'iga qulay shaklda taqdim yetish uchun mashinani o'qitishning aniqligi talablariga moslashtirish mumkin. Neyron tarmog'ining chiqishi bo'yicha birlik chiqish komponentlaridan birining mavjudligini ko'rsatadi.

References:

1. Abduqodirov, A., & Umurzakova, D. (2018). SUN'IY NEYRON TARMOQLARI YORDAMIDA NUTQNI TANISH MASALALARINI TADQIQ ETISH. *Scientific journal of the Fergana State University*, (4), 1
2. VAZIRLIGI, O'. R. I. I., KARIMOV, I., & TURGUNOV, N. (2016). AXBOROT XAVFSIZLIGI ASOSLARI. Jololov, A. J., Samijonov, A. N., Samijonov, B. N., Yo'ldosheva, A. E., & Ibodillayev, A. X. (2022). Sensitive system on the basis of neuran networks. *Проблемы вычислительной и прикладной математики*, (S2), 46-54
3. Turgunbayev, R. (2021). Mashina o'rganishi va uni akademik maqolalardan metama'lumotlarni avtomatik ekstraksiya qilishda ahamiyati. *Science and Education*, 2(11), 302-316.
4. Turayeva, G. (2021). Neyrotarmoqlarga asoslangan mashina ta'limi. *ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu.uz)*, 2(2).
5. Xalilov, D. (2022). SUN'IY INTELLEKT VA RADIAL NEYRON TARMOQLARNING MATEMATIK ASOSLARI. *Science and innovation*, 1(A6), 664-671.