

MA'LUMOTLARNI INTELLEKTUAL TAHLIL QILISHDA MASHINAVIY O'RGANISH USULLARI

Dexqonov Alisher

(Farg'ona Davlat Texnika Universiteti)

Email: alisherbekdehqonov43@gmail.com

Aminjonov Nurulloh

(Farg'ona Davlat Texnika Universiteti)

Email: aminjonovnurulloh25@gmail.com

<https://doi.org/10.5281/zenodo.17959518>

ANNOTATSIYA

Zamonaviy axborot texnologiyalari rivojlanishi bilan ma'lumotlar hajmi eksponensial sur'atda o'sib bormoqda. Ushbu tadqiqot ishida ma'lumotlar intellektual tahlilida mashinaviy o'rganish usullarining qo'llanilishi va samaradorligi keng qamrovli o'rganildi. Qaror daraxtlari, gradient boosting, neyron tarmoqlar (CNN, LSTM) va transformer asosidagi modellar tahlil qilindi. Eksperimental ishlar davomida Kaggle platformasidan olingan "Customer Churn" va "Stock Price Prediction" ma'lumotlar to'plamlari ustida turli modellarning bashorat va tasniflash samaradorligi sinovdan o'tkazildi. Natijalar shuni ko'rsatdiki, transformer modellari tasniflashda 92% aniqlikka, bashoratda esa 0.033 RMSE ko'rsatkichiga erishdi.

Kalit so'zlar: ma'lumotlar intellektual tahlili, mashinaviy o'rganish, chuqur o'rganish, bashorat, tasniflash, CNN, LSTM, transformer

1. KIRISH

XXI asrda raqamli texnologiyalarning jadal rivojlanishi natijasida har kuni ulkan hajmdagi ma'lumotlar yaratilmoqda. Ma'lumotlar intellektual tahlili zamonaviy fan va texnologiyaning eng dolzarb yo'nalishlaridan biri hisoblanadi. Bu soha katta hajmdagi ma'lumotlardan yashirin naqshlar, qonuniyatlar va foydali bilimlarni ajratib olishga qaratilgan.

Mashinaviy o'rganish usullari ma'lumotlar intellektual tahlilining asosiy vositalari bo'lib, murakkab bog'lanishlarni aniqlash, kelajakni bashorat qilish, ob'ektlarni tasniflash va qaror qabul qilishni avtomatlashtirishda muhim rol o'ynaydi. An'anaviy statistik metodlar asosan chiziqli bog'lanishlarni modellashtiradi va katta hajmdagi murakkab ma'lumotlarda samarali natija bermaydi. Zamonaviy chuqur o'rganish va ansambl usullari natijalarning aniqligi, ishonchliligi va barqarorligini sezilarli darajada oshiradi.

Tadqiqotning maqsadi: Ma'lumotlar intellektual tahlilida mashinaviy o'rganish metodlarini chuqur tahlil qilish, ularning samaradorligini qiyosiy baholash va turli amaliy sohalarda qo'llash imkoniyatlarini aniqlash.

Asosiy vazifalar:

- Mashinaviy o'rganishning asosiy metodlarini nazariy jihatdan o'rganish
- Real ma'lumotlar to'plamlarida eksperimental tadqiqotlar o'tkazish
- Turli modellarning bashorat va tasniflash samaradorligini qiyoslash

2. MASHINAVIY O'RGANISH USULLARI

2.1. Qaror daraxtlari va ansambl modellari

Qaror daraxtlari ma'lumotlarni ketma-ket bo'limlash orqali tasniflash yoki regressiya vazifalarini bajaradi. Har bir tugun ma'lumotlarni ikki yoki undan ko'p bo'lakka ajratadigan test shartini o'z ichiga oladi.

Asosiy afzalliklari:

- Interpretatsiya oson va vizual tushunarli
- Ham sonli, ham kategorik ma'lumotlar bilan ishlaydi
- Chiziqli bo'lmagan bog'lanishlarni yaxshi modellashtiradi
- Ma'lumotlarni oldindan masshtablash talab qilmaydi

Cheklovlari:

- Overfitting xavfi yuqori, ayniqsa chuqur daraxtlarda
- Kichik o'zgarishlar butun daraxt strukturasini o'zgartirishi mumkin
- Murakkab bog'lanishlarni ifodalash uchun juda katta daraxtlar kerak

Random Forest: Ko'plab qaror daraxtlarini yaratadi va ularning bashoratlarini ovoz berish yoki o'rtacha qiymat orqali birlashtiradi. Har bir daraxt ma'lumotlarning tasodifiy namunasi va tasodifiy xususiyatlar to'plami asosida o'qitiladi. Bu usul overfitting xavfini sezilarli kamaytiradi va katta hajmdagi ma'lumotlarda yuqori aniqlik ko'rsatadi.

Gradient Boosting: Ketma-ket ravishda kuchsiz o'rganuvchilarni qo'shib boradi, har bir yangi model oldingi modellarning xatolarini to'g'rilashga harakat qiladi. XGBoost, LightGBM va CatBoost bu algoritmnining optimallashtirilgan versiyalari bo'lib, tezkor ishlash va yuqori aniqlik bilan ajralib turadi.

2.2. Neyron tarmoqlar

Neyron tarmoqlar inson miyasining neyronlari tuzilishidan ilhomlangan model bo'lib, bir nechta qatlamlardan iborat.

Konvolyutsion neyron tarmoqlar (CNN): Asosan tasvirlarni tahlil qilishda qo'llaniladi, lekin strukturalangan ma'lumotlar bilan ham samarali ishlaydi. Konvolyutsiya qatlamlari lokal naqshlarni avtomatik aniqlaydi va pooling mexanizmi parametrlar sonini kamaytiradi.

Afzalliklari:

- Lokal naqshlar va xususiyatlarni avtomatik aniqlaydi
- Transfer learning imkoniyati
- Tasvirlar va ketma-ketliklar bilan yuqori samaradorlik

Cheklovlari:

- Katta ma'lumotlar to'plami talab qiladi
- O'qitish resurslari talab qiladi
- Qarorlar jarayonini tushunish qiyin

LSTM (Uzoq qisqa muddatli xotira): Ketma-ketliklar bilan ishlashga mo'ljallangan maxsus neyron tarmoq turi. Vaqt bo'yicha uzoq bog'lanishlarni yodda saqlaydi va gradient yo'qolishi muammosini hal qiladi.

Afzalliklari:

- Uzoq muddatli bog'lanishlarni saqlaydi
- Vaqt seriyali ma'lumotlarni bashoratda yuqori aniqlik
- Ikki tomonlama LSTM orqali yanada yuqori natijalar

Cheklovlari:

- O'qitish sekin va resurs talab qiladi
- Hiperparametrlarni sozlash murakkab
- Juda uzun ketma-ketliklarda samaradorlik pasayishi mumkin

2.3. Transformer arxitekturalari

Transformerlar 2017-yilda taqdim etilgan va tabiiy tilni qayta ishlash sohasida inqilob yaratgan. Asosiy g'oyasi self-attention mexanizmi bo'lib, barcha pozitsiyalar orasidagi global bog'lanishlarni parallel hisoblash imkonini beradi.

Self-attention mexanizmi: Har bir element uchun boshqa barcha elementlar bilan bog'lanish darajasini hisoblaydi, bu global kontekstni hisobga olish imkonini beradi.

Afzalliklari:

- Ketma-ketlikning istalgan elementlari orasidagi bog'lanishni to'g'ridan-to'g'ri hisoblaydi
- Parallel hisoblash imkoniyati - tezkor o'qitish
- Uzun ketma-ketliklar bilan samarali ishlaydi
- Murakkab va ko'p qatlamli bog'lanishlarni yaxshi modellashtiradi

Cheklovlari:

- Ko'p parametrlar va xotira talab qiladi
- Katta ma'lumotlar to'plami zarur
- Kvadratik murakkablik
- Kichik ma'lumotlarda overfitting xavfi

3. EKSPERIMENTAL TADQIQOTLAR VA NATIJALAR

3.1. Ma'lumotlar to'plamlari

Customer Churn Dataset: Telekommunikatsiya kompaniyasi mijozlarining ma'lumotlarini o'z ichiga oladi. 7043 ta yozuv va 21 ta xususiyat mavjud. Ma'lumotlar tayyorlash jarayonida yetishmayotgan qiymatlar to'ldirildi, kategorik o'zgaruvchilar raqamlashtirildi va sonli xususiyatlar standartlashtirildi. Ma'lumotlar 80:20 nisbatida o'qitish va test to'plamlariga bo'lingan.

Stock Price Prediction Dataset: Kompaniyalar aksiyalar narxining tarixiy ma'lumotlarini o'z ichiga oladi. Texnik indikatorlar qo'shildi (MA, EMA, RSI, MACD), ma'lumotlar normalizatsiya qilindi va 60 kunlik sliding window yondashuvi qo'llandi.

3.2. Tasniflash natijlari: Customer Churn

Mijozlarni yo'qotishni bashorat qilish uchun to'rt xil model sinovdan o'tkazildi:

Qaror daraxti:

- Aniqlik: 81%
- Precision: 78%, Recall: 75%, F1-score: 76.5%

Qaror daraxti nisbatan sodda model sifatida yaxshi asosiy natija ko'rsatdi. Overfittingni kamaytirish uchun max_depth=10 va min_samples_split=50 parametrlari qo'llandi.

Random Forest:

- Aniqlik: 88%
- Precision: 85%, Recall: 83%, F1-score: 84%

Random Forest 100 ta daraxt yordamida sezilarli yaxshilanish ko'rsatdi va barqarorroq natijalar berdi.

LSTM:

- Aniqlik: 90%
- Precision: 88%, Recall: 86%, F1-score: 87%

LSTM modeli ikki qatlamli arxitektura (128 va 64 neyron) bilan qurildi. Dropout qatlamlari overfittingni oldini oldi.

Transformer:

- Aniqlik: 92%

- Precision: 90%, Recall: 89%, F1-score: 89.5%

Transformer arxitekturasi 4 ta attention head va 2 ta encoder qatlami bilan eng yuqori aniqlikni ko'rsatdi. Self-attention mexanizmi barcha xususiyatlar orasidagi murakkab bog'lanishlarni samarali aniqladi.

3.3. Bashorat natijalari: Stock Price Prediction

Aksiyalar narxini bashorat qilishda uchta model qo'llandi:

Gradient Boosting (XGBoost):

- RMSE: 0.045, MAE: 0.034, R^2 : 0.87

XGBoost modeli 100 ta daraxt bilan tez o'qitildi va yaxshi bashorat natijalari berdi.

LSTM:

- RMSE: 0.038, MAE: 0.029, R^2 : 0.91

LSTM modeli 3 qatlamli arxitektura bilan vaqt bo'yicha ketma-ketlikni yaxshi saqladi va trend o'zgarishlarini aniq bashorat qildi.

Transformer:

- RMSE: 0.033, MAE: 0.025, R^2 : 0.93

Transformer modeli 6 ta attention head bilan eng yaxshi natijani ko'rsatdi. Uzoq muddatli bog'lanishlarni va murakkab naqshlarni eng aniq aniqladi.

Natijalar tahlili: Vaqt seriyali ma'lumotlarda Transformer va LSTM modellar eng samarali ishladi. Transformer global bog'lanishlarni parallel hisoblash tufayli biroz ustun keldi. Barcha modellar 60 kunlik oynadan foydalanib keyingi kun narxini bashorat qildi.

4. XULOSA VA TAVSIYALAR

Tadqiqot natijalari shuni ko'rsatdiki, ma'lumotlar intellektual tahlilida mashinaviy o'rganish metodlari an'anaviy statistik yondashuvlardan sezilarli darajada samaraliroqdir. Transformer va LSTM modellari murakkab naqshlarni aniqlash va bashorat qilishda yuqori natija beradi, lekin katta hajmdagi ma'lumotlar va hisoblash resurslari talab qiladi. Random Forest va Gradient Boosting esa muvozanatli yechim sifatida ko'plab amaliy vazifalarda samarali qo'llanilishi mumkin.

Amaliy qo'llanish sohalari:

- Moliya: aksiyalar narxi va bozor tendentsiyalarini bashorat qilish
- Tijorat: mijozlar xatti-harakatini tahlil qilish va churn bashorati
- Tibbiyot: kasalliklarni diagnostika qilish va davolash natijalarini bashorat qilish
- Sanoat: jihozlarning nosozliklarini oldindan aniqlash va texnik xizmat rejalashtirish
- Marketing: reklama kampaniyalari samaradorligini baholash

Kelajakdagi yo'nalishlar:

- Modellarini optimallashtirish va yengil variantlarini ishlab chiqish
- Real vaqt tizimlarida qo'llash imkoniyatini oshirish
- Explainable AI texnologiyalari bilan integratsiya
- Federatsiyalashgan o'rganish yordamida maxfiylikni saqlash
- AutoML vositalari yordamida avtomatlashtirish

Mashinaviy o'rganish usullari doimiy rivojlanib bormoqda va ularning kelajakda yanada keng qo'llanilishi kutilmoqda. Tadqiqotchilar va amaliyotchilar uchun muhim vazifa - bu texnologiyalarni real muammolarni hal qilishda to'g'ri qo'llash va samaradorligini oshirishdir.

Adabiyotlar, References, Литературы:

1. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016.
2. Breiman L. Random Forests. Machine Learning, 2001.
3. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. KDD, 2016.
4. Vaswani A., et al. Attention Is All You Need. NeurIPS, 2017.
5. Hochreiter S., Schmidhuber J. Long Short-Term Memory. Neural Computation, 1997.
6. He K., et al. Deep Residual Learning for Image Recognition. CVPR, 2016.
7. LeCun Y., Bengio Y., Hinton G. Deep learning. Nature, 2015.