

NOMINAL (KATEGORIYALI) BELGILARNI SAMARALI SARALASH MUAMMOLARI VA ZAMONAVIY ALGORITMIK YONDASHUVLAR

Rajapboyeva Hulkaraxon Sotiboldiyevna

(Urganch Ranch texnologiyalari universiteti magistranti)

+998974540391, hulkararajapboyeva@gmail.com

<https://doi.org/10.5281/zenodo.17950959>

Annotatsiya: Ushbu tezisda tibbiy ma'lumotlarda keng uchraydigan nominal belgilarni saralash masalalari va zamonaviy algoritmik yondashuvlar tadqiq qilinadi. Raqamli transformatsiya jarayonida tibbiy sohada to'planayotgan katta hajmdagi yuqori o'lchamli ma'lumotlarni qayta ishlash, keraksiz belgilarni ajratib tashlash va diagnostik modelarning aniqligi, tezligi va izohlanishini oshirish masalalari yoritiladi. Nominal belgilarni kodlash (masalan, one-hot encoding) va ularni tanlash (feature selection) uchun turli usullar – statistik (o'zaro axborot), modelga asoslangan (Lasso, RFE) va izohlash vositalari (SHAP, LIME) tahlil qilinadi. Yurak kasalliklarini tashxislash misolida nominal belgilarni saralashning amaliy bosqichlari ko'rsatilgan. Xulosa qilib, nominal belgilarni to'g'ri tanlash va izohlash tibbiy sun'iy intellekt tizimlarining samaradorligi, ishonchliligi va klinik qarorlar sifatini oshirishda muhim ahamiyatga ega ekanligi ta'kidlanadi.

Kalit so'zlar: belgilarni saralash, nominal ma'lumotlar, mashinali o'qitish, tibbiy tahlil, algoritmlar, baholash metrikalari. nominal belgilar, tibbiy ma'lumotlar, kategoriyali ma'lumotlar, one-hot encoding, rekursiv belgilarni chiqarib tashlash, Lasso regressiyasi, SHAP, LIME, izohlanadigan sun'iy intellekt, tibbiy diagnostika.

Kirish qismi

Belgilarni saralash masalasi mashinali o'qitish va ma'lumotlar intellektual tahlilida muhim rol o'ynaydi, ayniqsa tibbiy ma'lumotlar muhitida. Ushbu tezisda Rutbalash masalasining turlari, algoritmlari va nominal tibbiy ma'lumotlardagi qo'llashdagi qiyinchiliklar tahlil qilinadi[1]. Tadqiqotning dolzarbligi tibbiy tashxis aniqligini oshirish va klinik qaror qabul qilishni yaxshilashdan iborat[2]. Maqsad - belgilarni saralash masalalarini yechish uchun samarali algoritmlarni taklif qilish va ularning tibbiy natijalarga ta'sirini baholashdan iborat.

Nominal tibbiy ma'lumotlar muhitida belgilarni saralashning zamonaviy algoritmik yondashuvlarini tahlil qilish, ularning samaradorligini taqqoslash va klinik qarorlar qabul qilish jarayonida izohlanishi (interpretability) yuqori modellar yaratish imkoniyatlarini aniqlashdan iborat.

Tibbiyotda raqamli ma'lumotlarning ko'payishi, shuningdek sun'iy intellekt va mashina o'rganish modellarining qo'llanilishi tobora kengayib bormoqda. Biroq, ko'p hollarda ma'lumotlar nominal shaklda (kategoriyali) bo'lib, ularni to'g'ri qayta ishlash, saralash va izohlash klinik jihatdan ishonchli tizimlar yaratish uchun muhim shartdir. Samarali belgilarni saralash model aniqligini oshiradi, ortiqcha hisoblash xarajatlarini kamaytiradi, overfitting xavfini pasaytiradi va shifokorlar tomonidan tushuniladigan (izohlanishi mumkin) natijalar taqdim etadi.

Reyting vazifalari nuqtaviy, juftlik va ro'yxat bo'yicha turlarga bo'linadi. Nuqtaviy usullar alohida elementlarning ahamiyatini baholaydi, juftlik usullari ikkita elementni taqqoslaydi, ro'yxat bo'yicha esa guruhni baholaydi. Statistik o'rganishda reyting masalalari tasniflash (diskret sinflar) va regressiya (doimiy natijalar) orqali yechiladi. Axborot qidirishda dolzarblik darajasi qidiruv natijalarini optimallashtiradi, masalan, grafik o'rnatish orqali.

Jadval 1: Reyting vazifalari turlari

Turi	Tavsif	Javob o'zgaruvchisi	Maqsad
Yorliqlar reytingi	Sinf yorliqlarini bashorat qiladi	Kategorik	Xususiyatlarni teglar bilan moslashtirish
Ob'ektlar reytingi	Tartibni bashorat qiladi	Juftlik afzalliklari	Ob'ektlarni taqqoslash
Namuna reytingi	Ballarni tayinlaydi	Uzluksiz/diskret	Ball asosida tartiblash
Ikki tomonlama	Ikkilik misollarni tartiblaydi	Ikkilik	Ehtimollik tartibi
Ko'p qismli	Bir nechta sinf bilan tartiblaydi	Toifali	Sinf ehtimoli bo'yicha tartiblash
Doimiy	Uzluksiz misollarni tartiblaydi	Uzluksiz	Natijalar asosida tartiblash

2-bo'lim: Vazifalarni tartiblash algoritmlari

An'anaviy algoritmlar (Qarorlar daraxtlari, SVM) tadbqiq qilinadigan va ishonchli, ammo murakkab ma'lumotlarda cheklangan. Zamonaviy yondashuvlar (neyron tarmoqlar, chuqur o'rganish) murakkab naqshlarni aniqlaydi, masalan, GPT-4 kabi LLM tavsiya tizimlarida. RankRLS va RankSVM kabi algoritmlar hisoblash samaradorligini oshiradi.

Jadval 2: Algoritmlarni taqqoslash

Algoritm	Tavsif	Ishlash	Ilova
Qarorlar daraxtlari	Ierarxik qaror qabul qilish	Talqin qilinadigan	Tasniflash, resurs taqsimlash
SVM	Optimal giperplanlar yaratadi	Yuqori o'lchamli ma'lumotlarda yaxshi	Murakkab tasniflash
Neyron tarmoqlar	Chuqur o'rganish arxitekturasi	Murakkab naqshlarni aniqlaydi	Multimodal ma'lumotlar
QuickLexSort	Leksikografik saralash	Barqaror, tez	Ma'lumotlar bazalari, Bayes tarmoqlari

Reyting uchun baholash ko'rsatkichlari

Baholash metrikalari aniqlik, eslab qolish, F1 ball, MAP, NDCG va MRR ni o'z ichiga oladi. Ushbu ko'rsatkichlar reyting sifatini, adolatni va foydalanuvchi tajribasini baholaydi. XGBoost kabi algoritmlar yuqori natijalarni beradi.

Jadval 3: Baholash metrikalari

Metrik	Ta'rif	Formula
Aniqlik	Tegishli qismi	$TP / (TP + FP)$
Eslab qolish	Tegishli misollar ulushi	$TP / (TP + FN)$
F1 ball	Garmonik o'rtacha	$2 * (aniqlik * eslab\ qolish) / (aniqlik + eslab\ qolish)$

Metrik	Ta'rif	Formula
MAP	O'rtacha aniqlik ballari	$(1/Q) * \sum AP(q)$
NDCG	Reyting sifatini o'lchaydi	$DCG@k / IDCG@k$
MRR	Birinchi tegishli elementning o'rtachasi	$(1/Q) * \sum (1/darajali_i)$

4-bo'lim: Nominal tibbiy ma'lumotlarda belgilarni saralash qiyinchiliklari

Nominal ma'lumotlardagi nomuvofiqliklar (matbaa xatolari, noaniq toifalar) saralashni murakkablashtiradi. Normallashtirish (min-maks, z-skor) roli muhim. Samarali usullar: RFE, ANOVA, Random Forest. Qo'lda saralash xatolarga moyil, avtomatlashtirilgan (AI asosida) esa tez va aniq.

Jadval 4: Qo'lda va avtomatlashtirilgan saralash taqqoslashi

Metrik	Qo'lda	Avtomatlashtirilgan
O'rtacha vaqt (daqqa)	6:81	3:22
Jami xatolar	163	46

Tibbiy natijalarga ta'siri va amaliy tadqiqotlar

Noto'g'ri saralash diagnostikani buzadi, bemor natijalarini yomonlashtiradi. Amaliy tadqiqotlar (yurak kasalliklari, CKD) XGBoostning 98.5% aniqligini ko'rsatadi. AI integratsiyasi klinik qarorlarni yaxshilaydi.

Xulosa va takliflar

Tezida belgilarni saralash masalalari va algoritmlari tahlil qilinib, tibbiy ma'lumotlarda AI ning roli ta'kidlanadi. Takliflar: normallashtirishni kuchaytirish, avtomatlashtirilgan tizimlarni joriy etish, yangi algoritmlarni ishlab chiqish. Kelajakdagi tadqiqotlar - multimodal ma'lumotlar va real vaqt rejimida saralash.

Adabiyotlar, References, Литературы:

- Guyon, I., & Elisseeff, A. (2003). "An Introduction to Variable and Feature Selection". Journal of Machine Learning Research, 3, 1157-1182.
- Saeyns, Y., Inza, I., & Larrañaga, P. (2007). "A review of feature selection techniques in bioinformatics". Bioinformatics, 23(19), 2507-2517.
- Liu, H., & Motoda, H. (Eds.). (2007). "Computational Methods of Feature Selection". Chapman & Hall/CRC.
- Bishop, C. M. (2006). "Pattern Recognition and Machine Learning". Springer.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). "The Elements of Statistical Learning: Data Mining, Inference, and Prediction". Springer (2nd ed.).
- Géron, A. (2019). "Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow". O'Reilly Media (2nd ed.).
- Molnar, C. (2022). "Interpretable Machine Learning: A Guide for Making Black Box Models Explainable" <https://christophm.github.io/interpretable-ml-book/>
- Rudin, C. (2019). "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead". Nature Machine Intelligence, 1(5), 206-215.
- Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day

readmission".Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

10. Jiang, F., et al. (2017). "Artificial intelligence in healthcare: past, present and future".Stroke and Vascular Neurology, 2(4).

11. Rajkomar, A., Dean, J., & Kohane, I. (2019). "Machine Learning in Medicine". The New England Journal of Medicine, 380 (14), 1347-1358.

12. Journall: Journal of Biomedical Informatics, Artificial Intelligence in Medicine, IEEE Journal of Biomedical and Health Informatics, Nature Machine Intelligence.

13. Konferensiyalar:Conference on Health, Inference, and Learning (CHIL), Machine Learning for Healthcare (MLHC), AMIA Annual Symposium.

