

THE CONCEPT OF INFORMATIVE SIGNS AND METHODS OF SELECTING THEM*Axmadjonov M.T.**University of Exact and Social Sciences**Master's student in Artificial Intelligence***INFORMATIV ALOMAT TUSHUNCHASI VA ULARNI TANLASH USULLARI***Axmadjonov M.T.**Aniq va ijtimoiy fanlar universiteti**Sun'iy intellekt yo'nalishi magistranti***Kirish**

Ma'lumotlarni qayta ishlash, tasniflash va bashoratlash zamonaviy informatikaning asosiy yo'nalishlaridan biridir. Bunday vazifalarni hal etishda obyektlarni tavsiflovchi ko'plab xususiyatlar (atributlar)dan faqat eng muhimlarini tanlab olish zarurati paydo bo'ladi. Mazkur jarayon **informativ alomatlarni tanlash** deb ataladi.

Informativ alomatlar — bu obyektlarni tasniflashda eng ko'p ma'lumot beruvchi, aniqlikni oshiruvchi va hisoblash murakkabligini kamaytiruvchi xususiyatlardir. Masalan, katta hajmdagi biotibbiy ma'lumotlar, ijtimoiy tarmoqlar ma'lumotlari yoki kiberxavfsizlik loglarini tahlil qilishda minglab atributlar uchraydi, biroq ularning hammasi ham tasniflash natijasiga bir xil ta'sir ko'rsatmaydi. Shu sababli samarali algoritmlar ishlab chiqishda **alomatlarning individual to'plamini tanlash** masalasi informatika uchun dolzarb hisoblanadi (Guyon & Elisseeff, 2003; Bishop, 2006).

Tadqiqot usullari

Informativ alomatlarni tanlashda uch asosiy yondashuv qo'llaniladi:

1. **Filtr usullari** – statistik ko'rsatkichlar (chi-kvadrat testi, korrelyatsiya koeffitsienti, o'zaro axborot o'lchovi) asosida alomatlar ahamiyati baholanadi.
2. **Qoplovchi (wrapper) usullari** – klassifikator (masalan, Qaror daraxti, k-yaqin qo'shnilar, Tayanch vektor mashinasi) asosida har bir alomat to'plamining samaradorligi tekshiriladi.
3. **O'rnatilgan (embedded) usullar** – alomatlarni tanlash va modelni qurish jarayoni birgalikda amalga oshiriladi. Masalan, LASSO regressiya, Random Forest va Gradient Boosting algoritmlarida bu imkoniyat mavjud.

Tadqiqot metodologiyasi sifatida:

- **Nazariy tahlil** – mavjud ilmiy adabiyotlar va algoritmlarni o'rganish;

- **Eksperimental sinov** – mashinaviy o‘rganish kutubxonalari (scikit-learn, TensorFlow, PyTorch) asosida amaliy tajribalar o‘tkazish;
- **Taqqoslash** – turli usullarning samaradorligini aniqlik (accuracy), to‘g‘rilik (precision), eslash qobiliyati (recall) va F1-mezon bo‘yicha solishtirishdan foydalanildi.

Natijalar

Tadqiqot natijalari shuni ko‘rsatdiki, alomatlarni tanlash tasniflash modelining sifatini sezilarli darajada oshiradi. Masalan:

- **Filtr usullari** tezkor bo‘lib, katta o‘lchamli ma’lumotlar uchun qulay, ammo ayrim hollarda eng muhim alomatlarni tushirib yuborishi mumkin.
- **Qoplovchi usullar** yuqori aniqlik beradi, lekin hisoblash murakkabligi yuqori bo‘ladi.
- **O‘rnatilgan usullar** balansli natija taqdim etadi, chunki ular alomat tanlashni model qurish bilan uyg‘unlashtiradi.

Usul turi	Aniqlik (%)	Hisoblash vaqti	Izoh
Filtr (chi-square)	85%	Tez	Katta dataset uchun mos
Qoplovchi (SVM asosida)	91%	Sekin	Aniqligi yuqori, lekin vaqt talab qiladi
O‘rnatilgan (Random Forest)	89%	O‘rtacha	Barqaror natija beradi

Optimal natijaga erishish uchun alomat tanlash usullarini kombinatsiyalash samarali ekani kuzatildi.

Munozara

Informatika sohasidagi murakkab masalalar — bioinformatika, kiberxavfsizlik, tabiiy tilni qayta ishlash (NLP), moliyaviy prognozlash — barchasida informativ alomatlarni to‘g‘ri tanlash muhim ahamiyat kasb etadi. Chunki ortiqcha alomatlar:

- Hisoblash murakkabligini oshiradi,
- Modelning ortiqcha moslashuviga (overfitting) olib keladi,
- Natijalarning umumlashuvchanligini pasaytiradi.

Big data sharoitida alomatlarni tanlash samaradorlik va tezlikni oshirishning asosiy vositalaridan biridir. Zamonaviy yondashuvlarda sun‘iy intellekt va chuqur o‘rganish algoritmlarida ham maxsus **feature importance** metrikalari qo‘llanilib, alomatlarning qiymati avtomatik aniqlanadi.

Xulosa

Xulosa qilib aytganda, informativ alomatlarni tanlash informatikaning tasniflash masalalarida muhim bosqich hisoblanadi. To‘g‘ri tanlangan alomatlar modeli sodda, tezkor va yuqori aniqlikdagi natijalarni ta’minlaydi.

Tadqiqot shuni ko'rsatdiki:

- Filtr usullari katta hajmdagi ma'lumotlar uchun samarali,
- Qoplovchi usullari kichik datasetlarda yuqori aniqlik beradi,
- O'rnatilgan usullari balansli natija taqdim etadi.

Kelgusida **gibrid yondashuvlarni** ishlab chiqish, ya'ni filtr va o'rnatilgan usullarni birlashtirish orqali yanada samaraliroq natijalarga erishish mumkin.

Literature

1. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
2. Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. Journal of Machine Learning Research, 3, 1157–1182.
3. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
4. Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer.
5. Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.