

Dokumentácia algoritmu strojového učenia AQ11

Volodymyr Novokhatskyi Roiko Oleksii Nikita Shalashkov
Mykhailo Ruzmetov

10. decembra 2023

Abstrakt

Algoritmus strojového učenia AQ11 predstavuje významný pokrok v oblasti analýzy dát a rozpoznávania vzorov. Vyvinutý na základoch teórie strojového učenia sa AQ11 vyznačuje robustnými schopnosťami spracovania dát a adaptabilitou v rôznych doménach. Tento algoritmus exceluje v klasifikácii zložitých dátových súborov, čo ho robí mimoriadne vhodným pre sektory ako financie, zdravotníctvo a technológie.

Metodológia AQ11 je založená na vylepšenom pravidlovom prístupe, ktorý umožňuje efektívnu tvorbu modelov a presné predpovede. Využíva jedinečnú kombináciu induktívneho uvažovania a štatistickej analýzy, čím sa odlišuje od konvenčných algoritmov. Táto vlastnosť umožňuje AQ11 efektívnejšie spracovať veľké a zložité dátové súbory, poskytujúc nuancované vhľady a presné predpovede.

Napriek svojim silným stránkam čelí AQ11 výzvam, ako je vysoká výpočtová intenzita a obmedzená použiteľnosť na určité typy neštruktúrovaných dát. Tieto obmedzenia však otvárajú cesty pre ďalší výskum a vývoj, najmä v optimalizácii algoritmu pre širšie typy dát a znížení výpočtovej záťaže.

Keďže oblasť strojového učenia neustále evoluje, algoritmus AQ11 vyniká svojou schopnosťou prispôbiť sa a fungovať v rýchlo sa meniacej dátovej krajinke. Jeho neustály vývoj pravdepodobne povedie k vytvoreniu efektívnejších verzií, čím ešte viac zvýši jeho uplatniteľnosť v rôznych priemyselných odvetviach. Algoritmus AQ11 preto nie je len súčasným technologickým úspechom, ale aj priekopníkom pre budúce inovácie v strojovom učení.

Obsah

| | | |
|----------|---|----------|
| 1 | Úvod a prehľad | 3 |
| 1.1 | História a vývoj | 3 |
| 1.2 | Primárne funkcie a aplikácie | 3 |
| 2 | Teoretické pozadie | 4 |
| 2.1 | Indukcia pravidiel a obálka G | 4 |
| 2.2 | Selektor a viactriedna klasifikácia | 4 |
| 2.3 | Absorpčný zákon v AQ11 | 4 |
| 3 | Implementácia a aplikácie | 5 |
| 3.1 | Pokyny pre implementáciu | 5 |
| 3.2 | Reálne aplikácie | 5 |
| 4 | Analýza výkonnosti | 6 |
| 4.1 | Benchmarking a výkonnostné metriky | 6 |
| 4.2 | Výhody a obmedzenia | 6 |
| 5 | Budúci vývoj a záver | 7 |
| 5.1 | Budúce smerovanie | 7 |
| 5.2 | Záverečné poznámky | 7 |

1 Úvod a prehľad

1.1 História a vývoj

Algoritmus AQ11 je špecializovaným rozšírením Michalského algoritmu AQ, ktorý zohráva kľúčovú úlohu v širšom spektre učenia klasifikácie. Tento algoritmus vyniká svojím jedinečným prístupom k riešeniu úloh klasifikácie, najmä v komplexných prostrediach, kde je potrebné kategorizovať dáta do viacerých, často prekrývajúcich sa skupín. Vývoj AQ11 mal osobitný význam pre systémy vyžadujúce presné a prispôsobiteľné metódy klasifikácie, ako je napríklad diagnóza chorôb sóje v poľnohospodárskych kontextoch.

1.2 Primárne funkcie a aplikácie

Primárnou funkciou AQ11 je sofistikovaný prístup k klasifikácii. Funguje tak, že inštancie každej kategórie považuje za pozitívne príklady a inštancie ostatných kategórií, spolu s generalizáciami urobenými pri učení predchádzajúcich kategórií, za negatívne príklady. Tento spôsob zabezpečuje odlišnosť každej kategórie, čím sa predchádza prekrývaniu a zaisťuje presnejšiu klasifikáciu.

2 Teoretické pozadie

2.1 Indukcia pravidiel a obálka G

Centrálным konceptom v prevádzke algoritmu AQ11 je 'obálka G', ktorá sa používa na nájdenie najšpecifickejších ($G(e1/e5)$) a najvšeobecnejších ($G(E1/E2)$) generalizácií pre pozitívne príklady oproti protipríkladom. Táto obálka zohráva kľúčovú úlohu pri definovaní hraníc, v rámci ktorých algoritmus AQ11 funguje, a zaisťuje, že generované pravidlá nie sú príliš široké ani príliš úzke.

- $G(e1/e5)$: Predstavuje najšpecifickejšiu generalizáciu, kde jeden pozitívny príklad ($e1$) je kontrastovaný s jedným protipríkladom ($e5$).
- $G(e1/E2)$: Toto rozširuje kontrastovaním pozitívneho príkladu so súborom všetkých protipríkladov ($E2$), čím poskytuje širšiu generalizáciu.
- $G(E1/E2)$: Ponúka najvšeobecnejšiu formu kontrastovaním súboru všetkých pozitívnych príkladov ($E1$) so súborom všetkých protipríkladov, čím obaluje výstup algoritmu.

2.2 Selektor a viactriedna klasifikácia

Selektory v AQ11 sú vyjadrenia vo forme $A_i \# R_i$, kde A_i je atribút a R_i je disjunkcia hodnôt atribútov, s $\#$ symbolizujúcim rovnosť alebo nerovnosť. Táto štruktúra umožňuje AQ11 riešiť problémy viactriednej klasifikácie tak, že inštancie jednej triedy považuje za pozitívne príklady a všetky ostatné za negatívne, iteruje cez každú triedu postupne.

2.3 Absorpčný zákon v AQ11

Absorpčný zákon je logické zjednodušenie používané v AQ11 na zlepšenie súboru pravidiel. Pomáha eliminovať redundancie a protirečenia v generovaných pravidlách, zaisťuje, že pravidlá sú čo najstručnejšie a najpresnejšie. Tento proces je nevyhnutný pre udržanie integrity súboru pravidiel, najmä keď algoritmus škáluje na spracovanie väčších a zložitejších dátových súborov.

3 Implementácia a aplikácie

3.1 Pokyny pre implementáciu

Implementácia algoritmu AQ11 v Python je rozdelená do niekoľkých modulov, z ktorých každý zodpovedá za rôznu časť procesu strojového učenia. Hlavný modul `main.py` koordinuje načítanie dát, ich predspracovanie, generovanie pravidiel a vyhodnotenie výkonnosti.

```
import sys
from data_preprocessing import load_data, preprocess_data
from aq11_algorithm import generate_rules
from inference import evaluate_performance

sys.path.append('src')

def main():
    # Main code to run the AQ11 algorithm
    eval_samples = 100
    data = load_data('data/student_stress.csv').iloc[eval_samples:]
    preprocessed_data = preprocess_data(data)
    print("Training AQ11 algorithm...")
    rules = generate_rules(preprocessed_data['positive'].iloc[:, :-1],
                          preprocessed_data['negative'].iloc[:, :-1])
    with open('data/rules.txt', 'w') as f:
        f.write(rules)
    print("Generated Rules!")
    print("Evaluating performance...")
    metrics = evaluate_performance('data/rules.txt', 'data/student_stress.csv',
                                  'stress_level', eval_samples)

    print(metrics)

if __name__ == "__main__":
    main()
```

3.2 Reálne aplikácie

Algoritmus AQ11 môže byť aplikovaný v rôznych oblastiach strojového učenia, napríklad pri klasifikácii úrovne stresu študentov, ako je uvedené v ukázkovom datasete `student_stress.csv`. Predspracovanie dát a generovanie pravidiel umožňujú algoritmu naučiť sa rozlišovať medzi rôznymi stavmi študentov na základe poskytnutých dát.

4 Analýza výkonnosti

4.1 Benchmarking a výkonnostné metriky

Výkonnosť algoritmu AQ11 je vyhodnotená na základe metrík ako presnosť, precíznosť, odvolanie a F1 skóre. Tieto metriky sú vypočítané z matice zámen, ktorá porovnáva predpovedané štítky s pravdivými štítkami.

```
# Výstup metrík z konzoly  
{'accuracy': 0.85, 'precision': 0.84, 'recall': 0.88, 'f1_score': 0.86}
```

4.2 Výhody a obmedzenia

Výhodou algoritmu AQ11 je jeho schopnosť prispôbiť sa zmenám v dátach a generovať pravidlá, ktoré reflektujú aktuálne vzory v dátach. Je vhodný pre aplikácie, kde sú dáta dynamické a menia sa v čase. Na druhej strane, obmedzenie algoritmu spočíva v jeho výpočtovej náročnosti a potrebe manuálneho predspracovania dát.

5 Budúci vývoj a záver

Oblasť strojového učenia je neustále v pohybe s pravidelne sa objavujúcimi novými algoritmami a metodológiami. Algoritmus AQ11 so svojim robustným indukovaním pravidiel a adaptabilitou na koncept drift stojí ako významný príspevok do tohto dynamického poľa. Budúcnosť však sľubuje ďalších vylepšení a aplikácií.

5.1 Budúce smerovanie

Pohľad do budúcnosti vývoja AQ11 môže smerovať do niekoľkých sľubných oblastí. Jedným z oblastí zamerania by mohla byť škálovateľnosť a efektívnosť algoritmu, optimalizácia pre väčšie dátové sady, ktoré sú bežné v aplikáciách big data. Vylepšenia by sa tiež mohli uskutočniť v oblastiach spracovania v reálnom čase a streamovania dát, čo by AQ11 umožnilo ponúkať okamžitejšie vhlády.

Existuje tiež potenciál v zdokonaľovaní schopnosti algoritmu zvládať neštruktúrované dáta, čo by rozšírilo jeho aplikovateľnosť na nové oblasti, ako je spracovanie prirodzeného jazyka a rozpoznávanie obrazu. Okrem toho by integrácia AQ11 s inými metódami strojového učenia, ako sú neurónové siete alebo hlboké učenie, mohla viesť k hybridným modelom, ktoré využívajú silné stránky pravidlového učenia s rozpoznávacími schopnosťami týchto zložitejších modelov.

5.2 Záverečné poznámky

Algoritmus AQ11 v sebe ukrýva princípy adaptability a presnosti v učení klasifikácie. Jeho vývoj odráža neustále hľadanie algoritmov, ktoré dokážu nielen spracovávať dáta, ale tiež sa učiť a vyvíjať spolu s dátami. Keďže algoritmus bude naďalej prepracovávaný a prispôbovaný, jeho príspevky do oblasti strojového učenia sa očakávajú trvalé a vplyvné.

Literatúra

- [1] M. A. Maloof, *Metódy AQ pre koncept drift*, v J. Koronacki, Z. W. Ras, S. T. Wierchon, & J. Kacprzyk (Eds.), *Pokroky v strojovom učení I: Venované pamiatke profesora Ryszarda S. Michalského*, Studies in Computational Intelligence, vol. 262, str. 23–48, Springer, Berlín, Heidelberg, 2010.
- [2] R. S. Michalski & Y. Chilausky, *Učenie sa povedaním a učenie sa z príkladov: Experimentálne porovnanie dvoch metód získavania vedomostí v kontexte vývoja expertného systému pre diagnostiku sójových chorôb*, International Journal of Policy Analysis and Information Systems, vol. 4, č. 2, str. 125–161, 1980.
- [3] *Pravidlá klasifikácie a strojové učenie*, KKUI TU Košice, Prezentácia prednášok, 2023.