

Evaluacija postupka otkrivanja izuzetaka za klasificirane primjere

Evaluaciju implementiranog postupka smo izvršili na dva načina. U prvom smo testirali sposobnost detekcije krive klase primjera i to tako da smo određenom broju primjera namjerno promijenili pridruženu klasu. U drugom smo testirali sposobnost detekcije slučajno generiranih primjera tako da smo generirali određeni broj novih primjera slučajnim odabirom i vrijednosti atributa i vrijednosti klase koristeći vrijednosti koje se pojavljuju u stvarnim primjerima.

Testiranje je izvršeno na pet javno dostupnih domena na OpenML platformi (www.openml.org). Odabrani su skupovi podataka sa velikim brojem atributa (do preko 22.000 atributa) i različitim brojem primjera (96-3468). Prva domena (D1) je prepoznavanje rukom pisanih znamenki a preostalih četiri (D2-D5) su skupovi podataka ekspresije gena. Sažetak karakteristika domena s obzirom na broj atributa, broj primjera i broj klasa je u Tablici 1.

Tablica 1 Karakteristike domena korištenih za testiranje.

| ID | Skup podataka | Broj atributa | Broj primjera | Broj klasa |
|----|-----------------------|---------------|---------------|------------|
| D1 | GINA PRIOR2 | 784 | 3468 | 10 |
| D2 | LYMPHOMA 11CLASSES | 4026 | 96 | 11 |
| D3 | AP COLON PROSTATE | 10936 | 355 | 2 |
| D4 | OVA BREAST | 10936 | 1545 | 2 |
| D5 | BURKITTLYMPHOMA | 22283 | 220 | 3 |

U prvom testiranju smo u svakom eksperimentu po jednom primjeru namjerno promijenili klasu te smo testirali da li će se taj primjer pojaviti u skupu izdvojenih 10 najznačajnijih izuzetaka i koji će biti njegov rang u tom popisu izuzetaka. Skup od deset izuzetaka je odabran zato jer se očekuje da ljudi trebaju analizirati rezultat selekcije izuzetaka te se može očekivati da će neki ekspert za domenu biti spreman analizirati do najviše deset potencijalnih izuzetaka. Za svaku domenu napravljeno je po 10 odvojenih eksperimenata i sumarni rezultat je prikazan u Tablici 2.

Tablica 2 Uspješnost detekcije izuzetaka u vrijednosti klase.

| ID | Broj primjera | Postotak uspješno detektiranih izuzetaka | Prosječni rang detektiranih izuzetaka |
|----|---------------|--|---------------------------------------|
| D1 | 3468 | 50% | 4 |
| D2 | 96 | 100% | 4 |
| D3 | 355 | 100% | 2 |
| D4 | 1545 | 20% | 5 |
| D5 | 220 | 90% | 2 |

U Tablici 3 je sumarni rezultat za drugo testiranje. U tom testiranju smo u svakom eksperimentu namjerno ubacili po jedan slučajno generirani primjer. Kao i u prvom testiranju mjerili smo pojavu generiranog izuzetka u listi od 10 detektiranih izuzetaka.

Tablica 3 Uspješnost detekcije slučajno generiranih izuzetaka.

| ID | Broj primjera | Postotak uspješno detektiranih izuzetaka | Prosječni rang detektiranih izuzetaka |
|----|---------------|--|---------------------------------------|
| D1 | 3468 | 0% | - |
| D2 | 96 | 40% | 7 |
| D3 | 355 | 20% | 3 |
| D4 | 1545 | 0% | - |
| D5 | 220 | 40% | 6 |

Usporedbom tablica jasno se vidi da je otkrivanje izuzetaka puno uspješnije kada je primjer izuzetan u pogledu klasifikacije. U dvije od pet domena u svim eksperimentima je izuzetak uvijek uspješno otkriven. Analizom Tablice 2 može se uočiti da je uspješnost detekcije lošija za domene sa velikim brojem primjera. Razlog je što u tim domenama postoji i originalno veliki broj izuzetaka pa se umjetno ubačeni izuzetak ponekad ne može ubaciti u listu 10 najznačajnijih izuzetaka. Lošiji rezultat za neke domene zbog toga nije posljedica metodologije već posljedica načina izvođenja testiranja. Da su prethodno eliminirani svi postojeći izuzeci u skupovima podataka pa da su zatim ubačeni samo umjetno generirani izuzeci, rezultat bi sigurno bio dobar i za domene sa velikim brojem primjera.

Skroman rezultat za umjetno generirane primjere je očekivan jer kod tih primjera postoji velika vjerojatnost da je umjetno generirana klasa u skladu sa umjetno generiranim vrijednostima atributa. I u ovom slučaju je najmanja mogućnost uspješne detekcije u domenama sa velikim brojem primjera (domene D1 i D5).

Zaključak testiranja je da je realizirani postupak indukcije velikog skupa pravila sa ugrađenom detekcijom izuzetaka vrlo pogodan za detekciju primjera koji su izuzetni u pogledu klasifikacije ali da je postupak vrlo nepouzdan za detekciju

izuzetaka u vrijednosti atributa čak i u slučaju kada su svi atributi slučajno generirani. Pri tom treba imati na umu da je posljednji slučaj vrlo težak zbog činjenice da su vrijednosti slučajno odabrane ali od vrijednosti koje se stvarno pojavljuju u originalnim primjerima. Zbog toga vrijednosti atributa nisu izuzetne po svom iznosu već po kombinaciji vrijednosti u raznim atributima.