

Something Something DARK SIDE a.k.a. Klasificiranje mentalnog zdravlja autora analizom teksta

Anonimni autori

Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva
Unska 3, 10000 Zagreb, Hrvatska

Sažetak

Proučavanje psihičkih poremećaja nudi bogate informacije o funkcioniranju ljudskog mozga, a takve podatke moguće je analizirati i široko primjenjivim metodama strojnog učenja. U ovom radu proučavani su pisani tekstovi, kao reprezentanti kognitivnog funkcioniranja, autora različitog mentalnog statusa (sudionici oboljeni od shizofrenije, bipolarnog poremećaja, depresije i kontrolna skupina bez tih poremećaja). Koristeći stroj s potpornim vektorima, stablo odluke, neuronsku mrežu i Bayesov klasifikator izgrađeni su klasifikatori tekstova autora s različitim psihičkim statusom. REZ. ZAKLJUČAK

1. Uvod

Čovjekova želja za spoznavanjem samog sebe oduvijek je bila veliki motivator istraživanja ljudskih sposobnosti. Vrijedne podatke u takvom pothvatu pružaju kako proučavanja normalnih, prototipičnih pojava, tako i onih koja po nekim svojim karakteristikama odskakuju po nekom od kriterija normalnosti. Mentalni poremećaji zanimljivo su područje proučavanja u kognitivnoj znanosti budući da nude mogućnost uvida u rad mozga izvan uobičajenih okvira. Oni mogu zahvaćati sve aspekte ljudskog funkcioniranja: ponašanje, emotivne doživljaje, tjelesne simptome i kogniciju. Pojedina područja različito su zahvaćena ovisno o vrsti i težini poremećaja. U ovom radu posvećujemo se posljedicama na kogniciju koje neki poremećaji mogu imati, pa su u nastavku navedeni neki mogući simptomi shizofrenije, bipolarnog poremećaja i depresije.

Shizofrenija se ističe sveobuhvatnošću mentalnih funkcija koje zahvaća, od čega ističemo promjene u mišljenju koje se očituju kao disocirano mišljenje: slijed misli bez smislene povezanosti, brzo skakanje misli s ideje na ideju, "ljepljivost" misli za jednu ideju i njeno teško mijenjanje i slično (?). Bipolarni poremećaj primarno je poremećaj raspoloženja, no također može utjecati na kognitivne funkcije kao što su pažnja i verbalno pamćenje, rezultirajući u maničnoj fazi ubrzanim mislima i bijegom ideja: brze promjene teme misli, no koje u podlozi imaju logičan slijed. Javljaju se i osjećaj velike vlastite važnosti i posebnosti. Također, dobivena je i povezanost sa kreativnošću (?). Kod depresije kognitivni simptomi su prvenstveno loša koncentracija i pesimistične, beznadne misli (?). Zbog jasne povezanosti procesa misli i jezika, smatramo da bi se navedene karakteristike oboljelih od ovih bolesti mogle očitovati i u njihovom tekstualnom izričaju.

Simptomi bolesti vidljivi u mišljenju očituju se i na formalnoj, i na sadržajnoj razini. Primjerice, sadržaj misli shizofrenih bolesnika može biti posve besmislen, no gramatička struktura očuvana. Također, misaoni tok depresivnih osoba često je zatvoren u krug tužnih, negativnih misli. Međutim, zbog svoje opsežnosti analiza semantičkog aspekta tekstova nadilazi okvire ovog rada, stoga se koncentriramo na formalne karakteristike teksta. Iznimku čini

uzimanje u obzir vlastite važnosti kao simptoma, koje smo pomoću jednostavne operacionalizacije uzimali u obzir. U usporedbi sa tekstovima mentalno zdravih osoba tekstovi oboljelih mogli bi se razlikovati po nekim karakteristikama izvedenim iz simptoma, što otvara mogućnost automatskog kategoriziranja takvih tekstova.

Automatski klasifikator je računalni sustav koji na temelju značajki danih uzoraka određuje njihovu pripadnost određenim kategorijama. Izrada klasifikatora uključuje odabir i izvlačenje značajki iz uzorka, određivanje seta podataka za treniranje i testiranje, odabir algoritma klasifikacije, treniranje samog klasifikatora na podacima za treniranje, te potom testiranje njegove uspješnosti na podacima za testiranje.

Dijagnosticiranje mentalnih bolesti zahtjevan je problem, pa bi izrada sustava koji procjenjuje pripadnost nekog teksta kategorijama bolesti i zdravlja mogla pružiti korisne dodatne informacije za dijagnostičare. Cilj rada stoga je izraditi klasifikator tekstova dobivenih od sudionika različitog mentalnog stanja (psihički zdravi, shizofreni sudionici, depresivni sudionici i sudionici s bipolarnim poremećajem). Svrha rada je provjeriti jesu li karakteristike navedenih mentalnih poremećaja dovoljno izražene na tekstovima psihički bolesnih sudionika da bi se oni razlikovali od tekstova psihički zdravih sudionika, te, ukoliko jesu, doprinijeti postupku dijagnostike.

U idućem odjeljku dan je pregled dosadašnjih istraživanja ove teme. U trećem odjeljku opisan je metodološki pristup klasifikaciji tekstova s obzirom na mentalni status autora korišten u ovom radu. U četvrtom odjeljku prikazani su alati i algoritmi korišteni za klasifikaciju te rezultati vrednovanja. Potom je u petom odjeljku iznesen zaključak, uz prijedloge za buduća istraživanja.

2. Srodni radovi

Problemom klasifikacije mentalnog statusa autora teksta bavilo se nekoliko radova koji su u uzorku imali tekstove pacijenata psihijatrijskih ustanova. (?) radili su automatsku klasifikaciju tekstova na temu *Važna osoba u mom životu*, koje su napisali 36 shizofrenih sudionika nasuprot 36 neshizofrenih sudionika, pomoću stroja s potpornim vektorima (engl. *support vector machine, SVM*) i Bayesove

regresije. Korištene se značajke najčešćih riječi, tri-grama slova (zbog specifičnosti hebrejskog jezika u kojima su gramatički značajni) i ponavljanje riječi, te je dobivena uspješnost klasifikacije tekstova 83,3%.

?) kao skup za učenje koristili su tekstove 31 shizofrenog i 16 maničnih pacijenata psihijatrijske ustanove i 9 kontrolnih sudionika, koji su bilo verbalizirali tekst koji je kasnije pretipkan, bilo sami napisali tekst. Tekstovi su bili pisani na jednu od dvije zadane teme (*Piknik u parku* i *Lopov*) uz zadatak da se u tekst ukomponira unaprijed određena lista riječi. Tekstovi su pretvoreni u hrpu riječi (engl. *bag of words*) oblik i korištena je frekvencija riječi, a učenje je izvedeno pomoću SVM i stabla odluke. Uspješnost klasifikacije shizofrene nasuprot kontrolne skupine bila je 77%, dok u klasifikaciji koja je uključivala manične sudionike uspješnost nije bila bolja od slučajne.

Rješavanje ovog problema donekle je slično rješavanju problema određivanja i/ili razlikovanja autora teksta, budući da na neki način pretpostavljamo postojanje određenih stilskih sličnosti u tekstovima autora koji pate od bipolarnog poremećaja, shizofrenije i depresije. ?) navode tri glavna pristupa problemu raspoznavanja autora i njima pripadne metode. Prvi pristup je usmjeren na uspoređivanje tekstova po mjerama stila, što se odnosi na značajke poput duljine riječi i rečenica i bogatstvo vokabulara, koje ovise o žanru teksta i duljini teksta. Drugi pristup koristi sintakсна obilježja kao značajke, analizirajući sintakсну strukturu u pozadini teksta, što je najčešće računalno zahtjevno. Naposljetku, treći pristup kao značajke koristi osobine korištenih riječi u tekstu. Tako se mogu koristiti funkcijske riječi nezavisne od teme teksta, ili pristup hrpe riječi koji je dobar ukoliko postoji povezanost između autora i tema o kojima piše.

?) uspješno su rješavali problem automatskog prepoznavanja autora teksta, pri čemu koriste velik broj mogućih značajki koje određuju tekst, a koje nisu ovisne o temi. Najbolje rezultate producirala je sljedeća kombinacija značajki: broj interpunkcijskih znakova, broj funkcijskih riječi, broj samoglasnika i duljine riječi.

Govoreći o duljini tekstova potrebnoj da bi raznolikost rječnika bila stabilno određena i predstavljala potencijalno dobru značajku, ?) navode da je minimalna potrebna duljina 1000 riječi.

U našem radu, za razliku radova (?) i (?), korišteni su tekstovi koje su napisali profesionalni pisci i autori blogova, čime pokušavamo utvrditi koliko su pokazatelji navedenih psihičkih bolesti u tekstu prožimajući u životu ljudi s tim poremećajima - budući da se radi o ljudima koji su u vrijeme pisanja očito adekvatno funkcionirali da bi mogli objaviti knjigu, odnosno održavati blog. Koristili smo i veći broj značajki i pokušali utvrditi koje najviše doprinose uspješnosti klasifikacije (JESMO LI?). Korištene značajke bazirane su na onima navedenima u (?), koje su odabrane na temelju moguće povezanosti sa simptomima psihičkih poremećaja.

3. Klasifikacija mentalnog statusa autora koristeći tekstove

Cilj rada pokušali smo ostvariti koristeći značajke forme teksta izvedene na temelju simptoma pojedinih psihičkih

poremećaja i značajki spomenutih u srodnim radovima. Tekstove smo potom obradili generirajući numeričke vrijednosti za svaku značajku, te tako dobivene podatke o značajkama analizirali različitim algoritmima, da bismo mogli usporediti rezultate. Svaki od ovih koraka detaljnije je opisan u nastavku.

3.1. Odabir značajki

Korištene su sljedeće značajke: broj različitih riječi, broj zamjenica *ja*, *meni*, *mene* itd. gdje je to imalo smisla, broj interpunkcijskih znakova, broj funkcijskih riječi i prosječna dužina riječi u tekstu. Prve dvije značajke mogu se dovesti u vezu sa simptomima, dok su preostale tri korištene zbog svoje uspješnosti u srodnim radovima (?) dok je povezanost sa psihičkim bolestima samo spekulativna. Primjerice, moguće je da dezorganiziranost misli kod shizofrenije rezultira različitom količinom sadržaja u tekstovima, što se može očitovati na broju funkcijskih riječi. Također, ekspresivnost i ubrzanost kod bipolarnih bolesnika možda je vidljiva i u korištenju većeg broja interpunkcija. S druge strane, sklonost kontemplaciji kod depresivnih autora možda rezultira duljim riječima i rečenicama (no analiza duljine rečenica po zahtjevnosti nadilazi okvire ovog rada). Međutim, kako za ove pretpostavke nemamo potvrdu u radovima drugih autora, korištenje ovih značajki posve je eksploratorno.

Korištenje broja različitih riječi ima svoju podlogu u povezanosti kreativnosti i bipolarnog poremećaja (?). Budući da se kreativnost često operacionalizira pomoću raznolikosti produkcije, pretpostavljamo da će tekstovi autora s bipolarnim poremećajem imati veći broj različitih riječi u tekstu nego tekstovi drugih kategorija. Također, sklonost ka *ljepljenju* na temu može se (iz različitih razloga) javiti kod oboljelih od shizofrenije ili depresivnog poremećaja, čime mogu biti ograničeni na manji skup riječi vezan uz tu temu, pa u tim skupinama vjerojatno možemo očekivati manji broj različitih riječi.

Motivacija za korištenje broja zamjenica nalazi se u simptomima grandioznosti prisutnih kod boljelih od bipolarnog poremećaja. Također, u radu (?) razlikovanje između tekstova shizofrenih i ne-shizofrenih pacijenata rađeno je prema frekvenciji pojedinih riječi, pri čemu je zamjenica *ja* bila najfrekventnija u shizofrenoj skupini. Dakako, ova značajka korištena je samo na tekstovima pisanima u prvom licu (svim blogovima i dijelu knjiga).

3.2. Predobrada podataka

Obrada teksta je izvršena pomoću skripte pisane u programskom jeziku perl. Obrada teksta se sastoji od izvlačenja određenog broja značajki koje se poslije koriste za daljnu obradu teksta. značajke koje se izvlače su:

1. Frekvencija pojavljivanja funkcijskih riječi u tekstu u odnosu na broj svih riječi u tekstu,
2. Frekvencija pojavljivanja interpunkcijskih znakova u odnosu na dužinu teksta,
3. Frekvencija pojavljivanja zamjenica za prvo lice jednine u odnosu na broj svih riječi u tekstu,
4. Postotak različitih riječi u tekstu,

5. Prosječna dužina riječi u tekstu.

Potrebno je napomenuti da nisu korišteni svi postojeći interpunkcijski znakovi engleskog jezika, već nešto manja grupa koja se sastoji od: točke (.), upitnika (?), uskličnika (!), zareza (,), dvotočja (:), točke-zareza (;), apostrofa ('). Može se primjetiti da nismo koristili znakove navoda, kao ni crticu zbog različite forme navođenja teksta među autorima.

3.3. Metode klasifikacije

Provedba postupaka strojnog učenja na obrađenim tekstovima rađena je u programskom alatu *Weka*¹, koji podržava velik broj algoritama analize podataka i prediktivnog modeliranja. S obzirom na jednostavnost uporabe različitih algoritama nadziranog učenja u *Weka*-i korišteno ih je nekoliko te su uspoređeni rezultati.

Naivni Bayesov klasifikator pridjeljuje primjeru onu klasu za koju je vjerojatnost najveća na temelju njegovog vektora svojstava, uz pretpostavku da su svojstva nezavisna i temelji se na Bayesovom teoremu (?). Iako je jako jednostavan klasifikator, koji počiva na često pogrešnoj pretpostavci o nezavisnosti svojstava, u upotrebi redovito daje dobre rezultate (?).

Naivni Bayesov klasifikator s procjenom jezgrene funkcije se temelji na klasičnom Bayesovom klasifikatoru uz razliku da ne pretpostavlja normalnu raspodjelu neprekidnih varijabli, već radi procjenu jezgrene funkcije, tj. distribucije iz koje proizlaze kontinuirane varijable (?).

Stroj s potpornim vektorima (engl. *Support Vector Machine, SVM*) je nadzirana metoda strojnog učenja koja se temelji na konstrukciji hiperravnine ili hiperravnina koje odjeljuju primjere. Hiperravnine se pronalaze traženjem *potpornih vektora*, tj. onih primjera koji pripadaju različitim klasama, a najmanje su međusobne udaljenosti (?).

Umjetne neuronske mreže (engl. *artificial neural networks*) su mreže međusobno povezanih *neurona*, nezavisnih isprogramiranih djelova koji obavljaju neku jednostavnu funkciju, da bi kao cjelina obavljali složeniju zadaću. Osnovni model umjetne neuronske mreže je perceptron (?).

4. Rezultati

Podaci za analizu napravljeni su iz uzorka tekstova poznatih pisaca i tekstova prikupljenih sa blogova, oboje iz engleskog govornog područja. Psihički poremećaji kod pisaca (bipolarni poremećaj i depresija) identificirani su detaljnom analizom njihovih dostupnih biografija, te su u uzorak uzeti oni tekstovi koji se bave upravo autorovim osvrtnom na psihičku bolest (ukoliko takvi postoje) ili su pisani u vremenskom razdoblju najbližem onom u kojem je poremećaj identificiran. Tekstovi pisaca koji ne pate od depresije, bipolarnog poremećaja ili shizofrenije odabrani su također analizom biografija, uz preferenciju prema opširnim biografijama te onima u kojima se spominje neko drugo zdravstveno stanje autora, nevezano uz psihičku bolest.

Blogovi autora koji imaju neki psihički poremećaj pronađeni su preko web-stranica koje se bave tim

poremećajima² i na njima se autori deklariraju kao oboljeli od određenog poremećaja (bipolarni, shizofrenija). Blogovi psihički zdravih autora odabrani su s lista popularnih osobnih blogova, a koji sadrže odjeljak u kojem se autor opisuje i u kojem navodi podatke o sebi, no ne navodi neku psihičku bolest, te također na kojima pretraživanjem nisu pronađene ključne riječi (npr. *mental, depression, illness* itd.) u kontekstu psihičke bolesti autora. Svi blogovi bave se osobnim doživljajima autora i pisani su u prvom licu. U slučaju da autor bloga piše samo kratke zabilješke, spojeno je nekoliko zabilješki u jedan tekst, budući da tekstove pokušavamo klasificirati prema značajkama forme, a ne sadržaja (iako ova aspekta nisu potpuno nepovezana).

Radi odsutnosti dovoljnog broja poznatih pisaca koji boluju od shizofrenije ta skupina nije prikupljena, kao ni skupina tekstova autora blogova koji boluju od depresije, zbog ponekad olakog korištenja riječi depresija u svakodnevnom govoru i time uzrokovane pretpostavljene niske pouzdanosti procjene stvarne depresivnosti kod autora. Tako su se podaci za treniranje i testiranje sastojali od šest kategorija: poznati pisci s bipolarnim poremećajem, poznati pisci koji boluju od depresije, poznati pisci bez identificiranog psihičkog poremećaja, autori blogova s dijagnosticiranom shizofrenijom, autori blogova oboljeli od bipolarnog poremećaja i autori blogova bez identificiranog psihičkog poremećaja. Slijedeći kriterij o 3 do 5 puta većem broju primjera u pojedinoj kategoriji od broja značajki (?) u svakoj kategoriji skupa podataka za treniranje prikupljeno je 20 primjera, osim u kategoriji blogova autora sa shizofrenijom, gdje ih zbog loše dostupnosti ima 14. U skupu podataka za testiranje svaka je kategorija imala 15 primjera, osim, ponovno, kategorije blogova autora sa shizofrenijom, u kojoj ih ima 12.

Veličina tekstova varirala je od 3 kB do 24 kB, te su snimani u ANSI standardu. Pretežno su imali iznad 5 kB, kada god je ta količina teksta bila dostupna, što odgovara graničnoj vrijednosti od 1000 riječi potrebnih za ispravnost nekih značajki (?). Sve značajke određivane su uzimajući u obzir duljinu teksta.

U ovom radu je korištena 10-prolazna kros validacija. Testiranje dobivenih modela je izvršeno na odvojenim skupovima za testiranje.

SVM je u Weki korišten pomoću biblioteke WLSVM³ (?).

Može se primjetiti da rezultati Neuspjeh izrade automatskog klasifikatora mentalnog statusa autora tekstova u ovom radu u konačnici ima optimističnu notu: s obzirom da su tekstovi uzeti od autora koji imaju dijagnozu, no u vrijeme pisanja tekstova najvjerojatnije nisu bili u akutnoj fazi bolesti, moguće je da utjecaj bolesti van akutne faze nije dovoljno jak da bi dovoljno ugrozio kognitivno funkcioniranje bolesnika da bi se ono očitovalo na njihovom pisanju.

Dakako, postoji mnogo prostora za poboljšanja metodološkog pristupa. U budućim istraživanjima bilo bi dobro koristiti tekstove za koje pouzdano znamo da su ih autori pisali u akutnoj fazi svoje bolesti, kada su kognitivne funkcije najviše pogođene, čime bi se povećala izglednost

¹<http://www.cs.waikato.ac.nz/ml/weka/>

²npr. www.psychologytoday.com, www.schizophrenia.com

³www.cs.iastate.edu/~yasser/wlsvm

Tablica 1: Postotci točno klasificiranih primjera na različitim metodama na skupu blogova

| Metoda | Točno klasificiranih |
|---|------------------------|
| Naivni Bayesov klasifikator | 50% |
| Naivni Bayesov klasifikator s procjenom jezgrene funkcije | 50% |
| SVM | Tekst u trećem retku |
| Nesto | Tekst u četvrtom retku |

Tablica 2: Postotci točno klasificiranih primjera na različitim metodama na skupu pisaca

| Metoda | Točno klasificiranih |
|---|----------------------|
| Naivni Bayesov klasifikator | 52.17% |
| Naivni Bayesov klasifikator s procjenom jezgrene funkcije | 56.52% |
| SVM | rezultat |
| Nesto | rezultat |

dobivanja uspješnijeg razlikovanja među tekstovima. Također, bilo bi dobro koristiti tekstove koji su pisani na istu temu (na tragu korištenja osobnih blogova), kako bi se smanjila mogućnost da semantički kontekst utječe na rezultate, ili pak da se lakše koriste neke potencijalno korisne semantičke značajke. Osim toga moglo bi se kontrolirati vještinu pisanja autora tekstova koristeći autore koji nisu izrazito vješti u tome kao što su profesionalni pisci, te koristiti nezavisne procjene stručnjaka za određivanje kvalitete teksta. Moguće je da bi i veći uzorak tekstova donio bolje rezultate, no problematično je njegovo prikupljanje s obzirom na malu populaciju oboljelih. Također, mogle bi se koristiti još neke značajke, a čija je implementacija bila preopširna za okvire ovog rada: frekvencije izraza (engl. *term frequency*, *TF*), inverzne frekvencije dokumenta (engl. *inverse document frequency*) i njihovih međuodnosa sa frekvencijom funkcijskih riječi.

5. Zaključak

Rad je super.