

Something Something DARK SIDE

Sonja Grdan, Veljko Srdarević

Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva
Unska 3, 10000 Zagreb, Hrvatska
{sonja.grdan, veljko.srdarevic}@fer.hr

Sažetak

sažetak

1. Uvod

Mentalni poremećaji zanimljivo su područje proučavanja u kognitivnoj znanosti budući da nude mogućnost uvida u drugačije funkcioniranje mozga. Shizofrenija se ističe sveobuhvatnošću mentalnih funkcija koje zahvaća, od čega ističemo promjene u mišljenju koje se očituju kao brzo skakanje misli s ideje na ideju, bez jasne povezanosti, "ljepljivost" misli za jednu ideju i njeno teško mijenjanje, izmišljanje riječi i slično. Bipolarni poremećaj primarno je poremećaj raspoloženja, no također može utjecati na kognitivne funkcije kao što su pažnja i verbalno pamćenje, a dobivena je i povezanost sa kreativnošću. Navedene karakteristike oboljelih od ovih bolesti mogu se očitovati i u njihovom tekstualnom izričaju, što u usporedbi sa tekstovima mentalno zdravih osoba otvara mogućnost automatskog kategoriziranja takvih tekstova.

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 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.

2. Dosadašnja istraživanja

Strous et al [1] radili su automatsku klasifikaciju tekstova na temu *Važna osoba u mom životu*, koje su napisali 36 shizofrenih sudionika nasuprot 36 ne-shizofrenih 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

Diederich et al [2] 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 pretpikan, 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 ?????? (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

2.1. Uzorak

Uzorke čine tekstovi poznatih pisaca engleskog govornog područja i tekstovi prikupljeni sa blogova. Psihički poremećaji kod pisaca (ili odsudnost istih) identificirani su detaljnom analizom njihovih dostupnih biografija, te su u uzorak uzeti oni tekstovi koji se bave upravo autorovim osvrto 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 prema dostupnosti.

Blogovi autora koji imaju neki psihički poremećaj pronađeni su preko web-stranica koje se bave tim poremećajima¹ i na njima se autori deklariraju kao oboljeli od određenog poremećaja (bipolarni, shizofrenija). Blogovi psihički zdravih 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 stila, a ne sadržaja).

Dužina teksta varirala je od ???? do ????, a u prosjeku iznosila ??????. Sve značajke određivane su uzimajući u obzir duljinu teksta.

Veličina uzorka: ????

3. Značajke teksta

4. Izvedba ili slično

dijagram

4.1. Pretprocesiranje: obrada teksta

Obrada teksta je izvršena pomoću skripte pisane u programskom jeziku perl. Obrada teksta se sastoji od izvlačenja određenog broja mjera koje se poslije koriste za daljnu obradu teksta. Mjere 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,

¹npr. www.psychologytoday.com,
www.schizophrenia.com

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.

4.2. Obrada podataka

5. Rezultati

6. Buduća istraživanja

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 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. 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.

7. Pozdravi i glazbene želje

fale qouteove, qouth the Raven, "Nevermore", jel imamo EAPa?

8. Zaključak

Rad je super.