

MOTIVÁCIA Sociálne siete sú veľmi populárne a v prípade mikrobloggerov ide o uverejňovanie obrovského množstva krátkych vyjadrení, ktoré môžu odkazovať aj externý obsah. Aby mohol používateľ z tohto informačného zdroja vyťažiť čo najviac, je potrebné ho vhodne podporiť v navigácii rozmanitým obsahom so zameraním na jeho motiváciu.

PROBLÉM

V súvislosti s jedným dopytom sa tweety obvykle týkajú viacerých tém. Pre používateľa môže byť užitočné, ak by vedel z takýchto súvisiacich tém rýchlejšie vybrať tie, ktoré ho zaujímajú.

Rovnako môžu ušetriť čas odporúčania obsahu, ktorý používateľa zaujíma v danom momente. Pri hľadaní inšpirácie, ktoré miesta navštíviť môžu byť zaujímavé fotografie od používateľov Twitteru. Inokedy môže používateľ hľadať články o aktuálnych udalostiach alebo zistiť, aký názor prevláda na výrobok, ktorý si chce kúpiť.

YEAH! We are playing in Bratislava tomorrow with my great heroes My Bloody Valentine!!! This is huge!! (-...
fb.me/1cflk3Kbq

Today's pics from flooded #Bratislava by Sme daily. Sad Janka Krala (park on Danube's S bank) is underwater
sme.sk/c/6826485/pozr... #Slovakia



event

"Bratislava"?

Just reported a jam on D1, Bratislava, bit.ly/11unmiu using @waze - Social GPS.



@mikegogulski with his freshly married Eva and a wedding pony in #bratislava pic.twitter.com/eKYVTvmzbj

METÓDA

Používateľ po zadaní dopytu číta relevantné tweety a môže ich hodnotiť na stupnici z hviezdíček.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer **a risus erat**. Aliquam ac rhoncus libero.



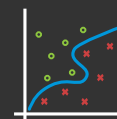
Poskytovanie explicitnej spätnej väzby (čiže hodnotenia tweetov) používateľa zaťažuje.

Pre neohodnotenú tweetovú sa preto ich zaujímavosť pre používateľa odhadne na základe **modelu interpretujúceho implicitnú spätnú väzbu**.

Implicitná spätná väzba je sledovaná ako atribúty:

- čas pozornosti pre text tweetu,
- kliknutie na URL,
- čas pozornosti pre URL.

Pre vystihnúť aktuálnej motivácie používateľa a jeho cieľov pri vyhľadávaní sa iteratívne vytvára model jeho preferencií pre aktuálne sedenie.



$$y(x) = \text{sign} \left(\sum_{i=1}^N \alpha_i y_i e^{-\gamma \|x_i - x\|^2} + b \right)$$

Prečítané tweety tvoria trénovaciu množinu pre druhý **SVM** model. Tento krát sa sleduje **príslušnosť k témam, metadáta tweetov a emocionálne zafarbenie**. Tweety, ktoré používateľ ešte neprečítal sú klasifikované ako zaujímavé alebo nezaujímavé pre daného používateľa v danom sedení.

OVERENIE

Pri experimentálnom overovaní si používatelia zvolili vlastný dopyt a čítali tweety k nim. Vybraté tweety mohli ohodnotiť podľa zaujímavosti.

Po prečítaní 40 tweetov bolo vytvorených niekoľko alternatívnych klasifikátorov pre **model interpretujúci implicitnú spätnú väzbu** a **model preferencií používateľa**.

Následne používateľ musel ohodnotiť až do 60 neprečítaných tweetov. Predikcie modelov boli porovnané s poskytnutým hodnotením používateľa.

VYHODNOTENIE Okrem presnosti (**accuracy**) bola sledovaná tiež **Kappa** štatistika, ktorá berie do úvahy aj zhodu danú náhodou. Podobne neskôr bola použitá aj miera **balanced accuracy** kompenzujúca nárast úspešnosti spôsobený len nerovnomerným rozdelením trénovacej a testovacej množiny.

	presnosť	balanced accuracy	Kappa (s náhodnými prediktormi)	Kappa (podľa rozdelenia)
model interpretujúci implicitnú spätnú väzbu	68,02%	65,13%	22,17%	29,91%
najlepšie alternatívy modelu preferencií	59,65%	62,61%	13,05%	22,23%
	61,14%	61,53%	15,75%	21,43%
	60,99%	61,32%	18,76%	21,09%

Už po prečítaní 40 tweetov je možné správne predikovať zaujímavosť časti nečítaného obsahu.

Rovnako je možné využiť implicitnú spätnú väzbu a **používateľ tak nie je nútený hodnotiť každý tweet**.