

Ako algoritmy sociálnych médií získavajú informácie z používateľov a následne ich využívajú vo svoj prospech

David Vach

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

xvachd@stuba.sk

Vedenie: Mirwais Ahmadzai

16. december 2023

Abstrakt

Tento článok sa bude venovať podrobnej analýze získavaniu informácií používateľov sociálnych médií pomocou algoritmov. Práca bude pozostávať z toho ako sa tieto algoritmy snažia získať čo najviac informácií z používateľov, ktoré následne môžu využiť na rôzne marketingové ciele. Jadro tohto článku sa sústreďuje na algoritmy a ich efektivitu v udržiavaní pozornosti používateľov na základe získaných dát a následné zobrazovanie špecifického obsahu za účelom získať čo najviac presnejších dát a následné vymazávanie nepotrebných dát. Ďalšia časť tejto práce ukazuje rôzne spôsoby zneužívania týchto dát vo svoj prospech z dlhodobého hľadiska a ako tieto algoritmy vedú pomocou získaných dát určiť s veľkou pravdepodobnosťou osobnosť používateľa a iné citlivé informácie. Záver práce tvoria dôvody prečo by sme si mali tieto súkromné dáta chrániť a aké spôsoby sa na to využívajú.

Kľúčové slová— sociálne médiá, zber dát, ochrana súkromia, algoritmy, marketing

1 Úvod

V úvode dnešnej digitálnej éry predstavujú sociálne médiá viac než len platformy na zdieľanie osobných príbehov a fotografií. Sú to sofistikované ekosystémy využívajúce algoritmy, ktoré transformujú spôsob, akým komunikujeme, angažujeme sa a dokonca ako sa rozhodujeme. Algoritmy sociálnych médií sú navrhnuté tak, aby boli nenápadnými, no mocnými kustódmi našej online pozornosti. Vychádzajú zo záveru, že každá interakcia na platforme - lajk, zdieľanie, komentár, dokonca aj dĺžka zdržania sa pri príspevku - je cenným údajom, ktorý odhaľuje náš vkus, preferencie a správanie. Tieto údaje sú potom analyzované a transformované na mieru, ktorá nás drží zapojených - a často prílepených - na naše obrazovky dlhšie než sme plánovali. No táto personalizácia má aj svoju cenu. Algoritmy sociálnych médií sú takisto navrhnuté na podporu obchodných modelov ich prevádzkovateľov, čo znamená, že informácie získané z používateľských interakcií sú často využívané na cieľovú reklamu a obsahové odporúčania, ktoré podporujú angažovanosť a zisky. V tomto procese sú používatelia niekedy nevedomky manipulovaní k tomu, aby konzumovali obsah a produkty, ktoré možno nepotrebujú alebo dokonca nechcú. Tento úvod bude skúmať, ako algoritmy sociálnych médií zbierajú a využívajú používateľské informácie a aké sú implikácie tohto procesu na súkromie, spoločenské správanie a dokonca na demokraciu. V tejto sekcii by som sa chcel venovať tomu, čo mňa motivovalo k tejto téme. Myslím si, že rýchlym rozširovaním online sociálnych médií si veľa ľudí vyberá, aby svoje názory na určité témy vyjadrovali práve v týchto sieťach, vrátane komentárov k produktom a službám. V období veľkých dát poskytuje nárast používateľmi vytvorených komentárov k produktom či iným veciam. Preto na sociálnych sieťach vznikajú rôzne príležitosti pre firmy na získanie cenných obchodných informácií pre lepšie rozhodovanie. Keďže komentáre zverejnené na Facebooku alebo Instagrame sa veľmi často odrážajú na individuálne postoje ľudí k špecifickým produktom alebo službám, môžu byť zneužitú na riešenie obchodných problémov rôznych firiem.

Môžeme si všimnúť, že podrobná analýza sentimentu používateľov na online sociálnych médiách môže poskytovať dôležité informácie (dátá) pre obchodné rozhodovanie. Preto v navrhovanom rámci analýzy sentimentu delíme jednotlivcov (aktérov) sociálnych sietí do troch skupín na základe ich odlišných orientácií názorov na produkt, a to: (a) skupina P obsahuje aktérov, ktorí majú k produktu pozitívny názor (b) skupina N zahŕňa aktérov, ktorí produkt nemajú radi a prejavujú negatívnu orientáciu názorov (c) skupina U obsahuje aktérov, ktorí majú k produktu neutrálny (ambivalentný) postoj. [4]

2 Súvisiaca práca

[4] [1]

3 Ako tieto algoritmy fungujú

Algoritmy sociálnych médií sú navrhnuté tak, aby efektívne zbierali a analyzovali údaje generované užívateľmi pri ich interakciách online. Ich schopnosť poskytovať personalizovaný obsah je založená na dômyselnom sledovaní a vyhodnocovaní činností používateľov. Keď používateľ lajkuje príspevok, komentuje video alebo strávi určitý čas čítaním článku, tieto akcie sú zaznamenané a stávajú sa

časťou veľkého údajového modelu, ktorý tento algoritmus používa na pochopenie a predvídanie preferencií používateľa.

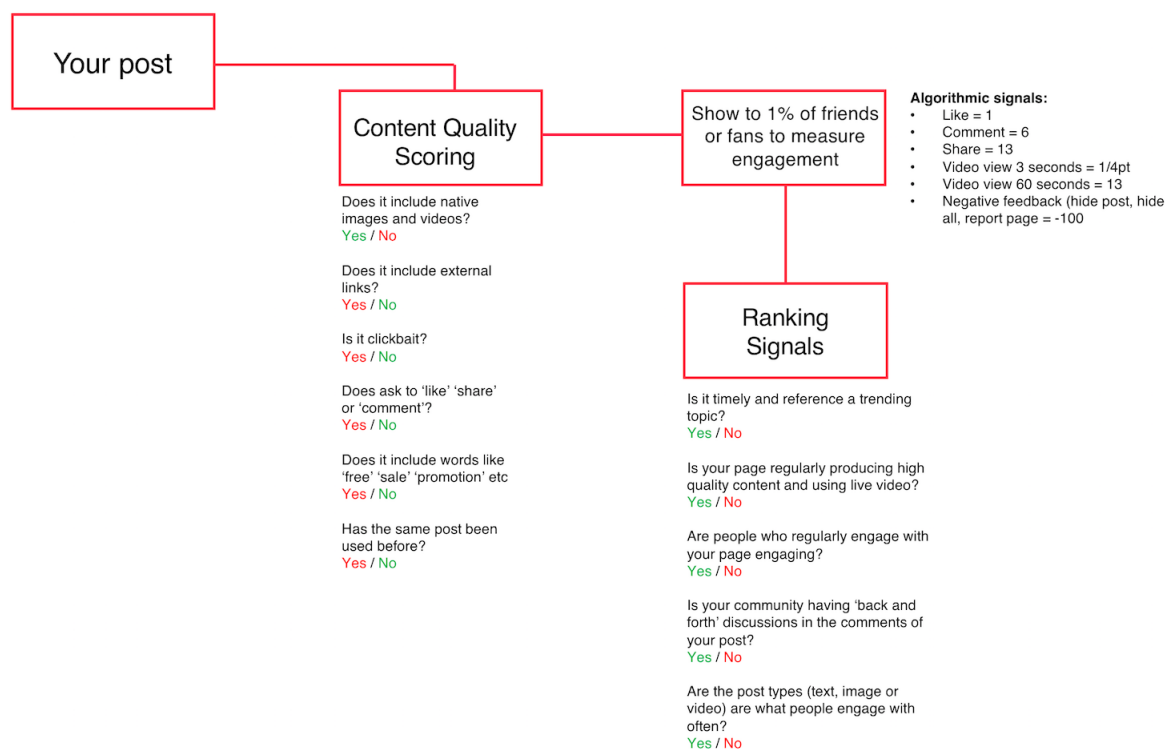
Na základe textu, obrázkov a videí, ktoré používatelia zdieľajú alebo komentujú, algoritmy využívajú pokročilé metódy ako spracovanie prirodzeného jazyka a rozpoznávanie obrazu na to, aby lepšie porozumeli obsahu, ktorý používatelia preferujú. Algoritmy zároveň skúmajú vzťahy a interakcie medzi používateľmi, aby zistili, aké sociálne siete sú pre nich najdôležitejšie a ako sa tie môžu prepojiť s ich obsahovými preferenciami.

Na základe získaných dát tieto algoritmy vytvoria obsah, ktorý sa používateľom zobrazuje vo feedoch, aby bol čo najrelevantnejší a najzaujímavejší. Tento personalizovaný prístup sa nezameriava len na zvýšenie užívateľskej spokojnosti a zapojenia, ale je tiež základom pre efektívnu cieľovú reklamu, kde sú reklamy prispôbené konkrétnym záujmom a správaniu používateľov.

Súčasťou procesu je aj predpovedanie správania. Algoritmy sociálnych médií sa pokúšajú predvídať, čo by mohlo používateľov zaujímať v budúcnosti, a prispôbujú svoje stratégie, aby bol obsah, ktorý zobrazujú, vždy krok vpred. Vďaka neustálej spätnej väzbe a iterácii sa algoritmy neustále učia a zdokonaľujú, aby boli ešte lepšie v predvídaní a napĺňaní očakávaní užívateľov.

Názorný príklad pre Facebook:

The Facebook algorithm



[?]

Obr. 1: Truncation of SVD for LSI.

Výsledkom je, že algoritmy sociálnych médií tvoria dynamický a neustále sa vyvíjajúci systém, ktorý je mimoriadne citlivý na potreby a chovanie používateľov,

a to všetko v úsilí udržať ich zapojenie a poskytnúť im maximálne prispôsobený zážitok. Ekonomický model sociálnych sietí, založený predovšetkým na reklamných príjmoch, je priamo ovplyvnený úspešnosťou týchto algoritmov v predpovedaní a ovplyvňovaní používateľského správania.

4 Metodológia

V rámci tejto štúdie sme sa zameriavali na analýzu algoritmov sociálnych médií a ich vplyvu na správanie používateľov. Metodológia našej práce zahŕňala niekoľko kľúčových krokov. Začali sme výberom relevantných dátových sád zo sociálnych médií, pričom sme sa zamerali na obdobie posledných dvoch rokov. Tieto dáta zahŕňali používateľské interakcie, ako sú lajky, komentáre a zdieľanie obsahu. Následne sme analyzovali špecifické algoritmy používané týmito platformami, vrátane NaiveBayes a Adaboost, pričom sme hodnotili ich efektivitu v personalizácii obsahu a vplyve na používateľské správanie. Použili sme techniky spracovania prirodzeného jazyka a analýzy sentimentu na spracovanie týchto dát. Počas celého procesu sme si boli vedomí etických zvážení spojených so zberom a analýzou používateľských dát. Na vyhodnotenie účinnosti rôznych algoritmov sme použili špecifické metriky, ako sú miera zapojenia a čas strávený používateľmi na platforme. Táto metodológia nám umožnila poskytnúť hĺbkový pohľad na to, ako algoritmy sociálnych médií ovplyvňujú používateľské správanie a súkromie.

5 Analýza

Porovnavane algoritmy:

Algorithm	Advantages	Disadvantages	Best Use Case
Adaboost	Improves weak models, adaptable	Sensitive to noise, prone to overfitting	Classification with 'difficult' cases
NaiveBayes	Simple, fast, effective for text data	Assumes independence of features, can be less accurate	Text classification, spam filtering
GBM	High accuracy, effective for complex patterns	Prone to overfitting, computationally intensive	Classification and regression with complex patterns
Random Forest	Resistant to overfitting, effective for large datasets	Less interpretable, computationally intensive	Classification and regression, large datasets

Tabuľka 1: Comparison of Machine Learning Algorithms

5.1 NaiveBayes[4]

5.2 Adaboost[1]

5.3 Random Forest[2]

5.4 Gradient Boosting Machines (GBM)[3]

6 Výsledok

7 Záver

PREDBEZNY ZÁVER:

V našej štúdii sme skúmali, ako algoritmy sociálnych médií získavajú a využívajú informácie od používateľov. Naše zistenia ukazujú, že tieto algoritmy hrajú kľúčovú úlohu v personalizácii používateľského obsahu a v ovplyvňovaní správania používateľov. Avšak spolu s týmto prispôbením prichádza aj zvýšené riziko zneužitia používateľských údajov a narušenia súkromia. Preto je nevyhnutné, aby sa pri vývoji a implementácii týchto algoritmov kládol dôraz na etické zváženia a ochranu súkromia. Naše výskumy naznačujú potrebu väčšej transparentnosti a regulácie v oblasti algoritmov sociálnych médií, aby sa zaistila ochrana používateľov a ich údajov.

8 Reakcia na prednášky

8.1 Kreatívne písanie (Prednáška 8)

Kreatívne písanie je umenie, ktoré umožňuje autorom vyjadriť svoje myšlienky, emócie a predstavy prostredníctvom slov. Je to proces, ktorý spočíva v tvorbe originálnych textov, či už ide o poéziu, prózu, drámu alebo iné literárne formy. Pri kreatívnom písaní je dôležitá nielen schopnosť gramaticky správneho a zrozumiteľného vyjadrovania, ale predovšetkým originalita a kreativita. Autor by mal byť schopný preniesť čitateľa do svojho sveta, ukázať mu nové perspektívy a vyvolať v ňom emócie.

Kľúčom k úspechu v kreatívnom písaní je pravidelná prax a ochota experimentovať so slovami a štýlom. Dobrý autor by mal byť otvorený novým nápadom a neprestajne rozvíjať svoje písacie zručnosti. Kreatívne písanie nie je len o výroku umeleckých diel, ale aj o sebaujadrnení a osobnom rozvoji. Táto forma písania môže byť terapeutická a umožňuje autorom vyrovnáť sa s osobnými záležitosťami alebo spracovať svoje myšlienky a pocity. Preto je dôležité vytvárať prostredie, kde sa autori cítia inšpirovaní a majú priestor na kreativitu.

8.2 Plagiátorstvo (Prednáška 9)

Plagiátorstvo je téma, o ktorej by mal byť každý tvorca článku, eseje, prezentácie alebo autor akéhokoľvek verejného prejavu oboznámený. Dá sa to nazvať krádežou, ale niekedy sa môže človek dopustiť neúmyselnému plagiátorstvu. Spôsob, akým sa mu vyhnúť je jednoduchý, a to odkazovať sa na použité zdroje.

8.3 Technológia a ľudia: Scrum. doc. Valentino Vranić (Prednáška 11)

Niaky text pojdem na prednasku

Literatúra

- [1] S. Desai and S. Patil. Efficient regression algorithms for classification of social media data. In *2015 International Conference on Pervasive Computing (ICPC)*, pages 1–5, 2015.
- [2] P. D. Reddy and L. R. Parvathy. Prediction analysis using random forest algorithms to forecast the air pollution level in a particular location. In *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, pages 1585–1589, 2022.
- [3] P. V. Reddy and S. M. Kumar. A method for determining the accuracy of stock prices using gradient boosting and the support vector machines algorithm. In *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, pages 1596–1599, 2022.
- [4] L. Zhang, H. Yuan, and R. Y. K. Lau. Predicting and visualizing consumer sentiments in online social media. In *2016 IEEE 13th International Conference on e-Business Engineering (ICEBE)*, pages 92–99, 2016.