

ACREA TEXT MINING

Podporujeme český a slovenský jazyk

Jaké jsou výhody zapojení analýzy textu do marketingových a obchodních analýz?

**ZÍSKEJTE
KONZULTACI
ZDARMA**



Data o zákaznících, jejich chování a interakce s nimi velmi výrazně:

- zpřesní všechny dosavadní prediktivní modely napříč organizací,
- zvýší úspěšnost marketingových kampaní,
- odhalí vzorce v nákupním chování zákazníků, které využijete v prodeji,
- zvýší hodnotu všech zákazníků díky cross/up-sellu,
- zpřesní retenční kampaně,
- zrychlí hledání příčin chyb v reklamačních řízeních,
- zefektivní a zrychlí práci call center, help desků a front office managerů.

Strojové učení a Text mining

- **Text mining** dokáže transformovat nestrukturovaná data na strukturovaná.
- Jsme schopni analyzovat texty v českém i slovenském jazyce – mailly, poznámky, komentáře, přepisy hovorů z call centra, otevřené otázky v dotaznících, reklamační protokoly, servisní zprávy a mnoho dalších.

Modul ACREA Text Mining zapojený v PS Clementine dokáže:

- analyzovat sentiment komentářů, mailů, přepisů telefonních hovorů,
- roztřídit dokumenty podle klíčových slov nebo témat (stížnost, spokojenost, reklamace, konkrétní produkt, jméno),
- roztřídit dokumenty nebo mailly do předem známých kategorií,
- doplnit diakritiku,
- rozpoznat jazyk textu,
- kombinovat strukturovaná data s nestrukturovanými a vytěžit z dat ještě více než doposud.

Text miningové úlohy

Analýza a zpracování textových dokumentů

Textminingové procedury umožňují z textových dokumentů extrahovat informativní atributy a zpracovat volné texty podobným způsobem jako strukturovaná tabulková data.

Textové dokumenty

emaily, revizní nebo lékařské zprávy, odborné články, přepsané telefonní hovory, odpovědi na otevřené otázky ve výzkumu, webové stránky, výpovědi svědků, požadavky zákazníků, recenze produktů, novinové zprávy a mnoho dalších.

Abychom mohli snadno analyzovat textové dokumenty, uspořádáme je do tabulky, kde řádky reprezentují jednotlivé dokumenty a sloupce vlastnosti extrahované z volného textu dokumentů. Tabulková reprezentace umožní dokumenty klasifikovat do předem stanovených tříd, seskupovat je do předem neznámých skupin, vyhledávat dokumenty podobné dotazu či jinému dokumentu nebo jen analyzovat jejich obsah případně změny obsahu v čase.

Klasifikace dokumentů

Jádrem většiny textminingových řešení je klasifikace dokumentů. Textové dokumenty jsou rozřazovány do předem známých tříd. Klasifikační model je možné přirovnat k lidskému čtenáři, jenž byl vyškolen na rozřídění velkého množství textových dokumentů.

Binární klasifikace

Dokumenty nejčastěji rozřazujeme do dvou skupin. Zřejmě nejznámějším příkladem je třídění příchozích emailů spamovým filtrem. Strukturovaná reprezentace emailů však umožňuje jejich téměř libovolné filtrování. Například můžeme emaily vybrat a seřadit podle relevance na ty, které je třeba přečíst a oddělit je od těch nepodstatných.

Podle relevance se též klasifikují a filtrují novinové články, tiskové zprávy, prohlášení mluvčích, přepsané telefonní hovory či projevy politiků.

Relevantní dokumenty vybírané binárním klasifikátorem mohou být zacíleny na sledování teroristických hrozeb, prevenci kriminality, sledování konkurence, monitorování vybrané kauzy apod.

Nominální klasifikace

Je-li skupin více, hovoříme o klasifikaci nominální.

V organizacích je třeba, aby se požadavky zákazníků dostaly k osobě kompetentní k jejich vyřízení. Manuální rozdělování došlé korespondence je časově náročné a málo efektivní. Klasifikaci příchozích dokumentů do skupin podle obsahu a přání zákazníků je vhodné automatizovat pomocí textminingového klasifikačního modelu. Model rozřadí dokumenty do předem definovaných kategorií. Kategorie odpovídají kompetencím a schopnostem pracovníků zodpovědných za vyřizování požadavků. Potřeba manuálního čtení před rozříděním dokumentů se minimalizuje pouze na dokumenty, jež klasifikátor nedokáže na základě strojového učení automaticky přiřadit do jedné z kategorií.

Vícečetná klasifikace

Vícečetné klasifikátory nachází uplatnění tam, kde každý textový dokument můžeme zařadit do několika kategorií zároveň. Například textové odpovědi na otevřené otázky ve výzkumech, mohou poukazovat na více důvodů nespokojenosti se službami či výrobky nebo odborné články se věnují více tématům. V takových situacích se v řešení kombinují binární klasifikátory pro jednotlivé kategorie. Opět platí, že kvalitu každého z klasifikátorů determinuje dostatečný počet trénovacích příkladů, tj. předem správně označených dokumentů kategoriemi, do nichž náleží.

Klastrování dokumentů

Ačkoli u většiny úloh, kdy textové dokumenty rozdělujeme do skupin, kategorie známe předem, můžeme se setkat se situacemi, kdy buď kategorie neznáme, známe je jen částečně nebo nemáme dostatečný počet dokumentů pro vytvoření textminingového klasifikátoru. Například dokumenty kategorizujeme podle kauz či osob v nich obsažených, které se neustále

Při detekci nových skupin dokumentů nám pomohou textminingové klastrovací modely. Ty na základě témat extrahovaných z textu navrhnou nové skupiny vzájemně si podobných dokumentů.

Pomocí klastrovacích modelů můžeme roztřídit webové stránky, tiskové zprávy či obchodní sdělení. Při zpracování odpovědí na výzkumné otevřené otázky nám detekované klastry pomohou navrhnout výsledné kategorie pro třídění odpovědí.

Klastrování dokumentů najde své uplatnění i v úlohách, kdy kategorie známe předem, ale jsou málo podrobné a dokumenty v rámci jedné kategorie potřebujeme ještě rozdělit do homogennějších skupin. Ze záznamů o servisních zásazích tak můžeme identifikovat podrobnější příčiny poruch, příchozí poštu rozdělit do více hierarchických složek nebo najít konkrétní důvody nespojenosti s produktem.

Rozpoznání sentimentu

Sentimentem výroku rozumíme pozitivní nebo negativní postoj autora. Stanovení sentimentu není jednoduchou klasifikační úlohou. Jeho rozpoznání vyžaduje extrakci speciálních atributů vyjadřujících postoje. Velmi obtížné je rozpoznat sentiment v textech obsahujících dvojité záporny, idiomy, nadsázky, sarkasmus, ironii či cynismus. Úlohu komplikuje i fakt, že v mnoha dokumentech sentiment není obsažen vůbec, nebo naopak se v nich vyskytují pozitivní i negativní postoje zároveň.

Postoje jsou klíčové při řízení vztahu se zákazníky a sledování veřejného mínění. Rozpoznání sentimentu se využívá pro detekci problémů s produkty nebo problémů při vyřizování požadavků a reklamací. Především posílení negativního sentimentu signalizuje potřebu zásahu do procesu.

Postoje zákazníků jsou spjaté s loajalitou zákazníků. Tu lze měřit pomocí Net Promoter Score (NPS). U zákazníků, kde nebyla NPS měřena a máme k dispozici jejich komentáře, doporučení, stížnosti apod., můžeme NPS odhadnout modelem nebo přímo předpovídat jejich loajalitu z analýzy sentimentu jejich výpovědí.

Analýzu sentimentu využijeme všude tam, kde zpracováváme hodnocení a recenze ve formě volného textu. Hodnotitelé mohou vyjadřovat své názory na filmy, politiky, státní správu, lékaře, umělce, právníky, řemeslníky atd

Analýza obsahu

Analýza obsahu umožňuje optimalizovat procesy sledování tiskových zpráv, webových stránek, vyšetřovacích zpráv apod., kdy chceme nahradit lidské čtenáře, kteří nás upozorní na zajímavá témata a trendy ve sledovaných dokumentech.

Pravidelné sledování tiskových zpráv o konkurenčních subjektech umožní monitorovat jejich aktivitu včas reagovat na nová zjištění.

Podobným způsobem jako monitorování konkurence je možné monitorovat i spokojenost zákazníků s vlastními produkty službami. Vstupem zde budou zákaznické recenze, diskuse a tiskové zprávy týkající se vlastní firmy. Díky analytice navázané na extrahované atributy ze vstupních textů lze bez zdlouhavého čtení dokumentů sledovat spokojenost zákazníků a řídit PR aktivity.

Nakvalitu vlastních produktů mohou též upozornit textové poznámky o opravách a revizích. Namísto čtení tisíců zpráv můžeme analyzovat tabulková data s příčinami problémů a nespojenosti uživatelů. Data umožní analyzovat kvalitu a spokojenost podle jednotlivých produktů, produktových skupin, způsobu užívání či typu uživatelů. Důležitou součástí monitorování je i analýza časových změn, která může upozornit na nově vzniklé problémy.

Velmi bohatý informační potenciál nabízí textové dokumenty pocházející ze zpravodajství a publicistiky. Na základě jejich strukturované reprezentace získané textminingovými postupy budeme schopni monitorovat politickou orientaci a náladu voličů, identifikovat nové klíčové události či analyzovat vyváženost prezentovaných témat. V záplavě tiskových zpráv můžeme upozornit na ty důležité a doporučit je k přečtení.

Využití strukturované reprezentace dokumentů

Hlavním přínosem textminingu je možnost obohacení běžných strukturovaných dat o atributy extrahované z textových dokumentů. Predikční modely budované nad kombinací strukturovaných a nestrukturovaných dat pak využijí přidané informace ukryté ve volných textech ke zpřesnění individuálních předpovědí.

Klasický **retenční model** odhadující pravděpodobnost odchodu zákazníka lze zpřesnit díky dostupnosti nových atributů pocházejících z textových dokumentů jako jsou **otevřené otázky ve výzkumu spokojenosti, poznámky z jednání se zákazníkem nebo přepsané hovory z telefonního centra.**

Volné texty s informacemi o marketingových a obchodních **interakcích se zákazníky** se ideálně hodí pro **doporučování produktů a prodejní kampaně.** Na základě potřeb vyjádřených v textu můžeme zákazníka vhodně oslovit a nabídnout mu produkt na míru. Přesnost pravděpodobnosti úspěšného obchodu zde zajisté zvýší **analýza sentimentu zákaznických interakcí.**

Policie při **vyšetřování trestné činnosti** zaznamenává **popis případu, úkony a pozorování** především v textové podobě, jen malá část informací je dostupná ve formě strukturovaných dat. Atributy extrahované z **výpovědí a popisů událostí** mohou nejen významně zpřesnit a rozšířit sledované policejní statistiky ale i pomoci při samotném vyšetřování případů. Díky novým strukturovaným atributům extrahovaným z textů lze lépe **kategorizovat trestnou činnost, hledat podobné modus operandi** a sestavovat **sociální sítě osob, firem a dalších entit.**



Textové poznámky nalezneme téměř ve všech **zprávách z revizí a kontrol.** Ve spojení se strukturovanými daty umožní lepší **posouzení rizikivosti daňového subjektu** po kontrole z finančního úřadu, **zpřesnit odhad životnosti stroje či součástky** po servisní zásahu nebo vhodně modifikovat zařazení zařízení do rizikové třídy po revizní kontrole. U servisních zásahů a revizních kontrol mohou textové poznámky napomoci i výrobcům při **identifikaci problémových součástek a materiálů.**

Díky volným textům můžeme také **zpřesnit znalecké odhady tržních cen nemovitého i movitého majetku.** Slovní **popis lokality nemovitosti či stavu zařízení** nabízí přesnější informace než standardizované odhadní koeficienty. Z textů je možné extrahovat i nové vlastnosti ovlivňující odhadní cenu. Následné regresní modely predikující cenu, tyto vlastnosti kvantifikují a nabídnou přesnější odhady.

Při **detekci pojistných podvodů** jsme schopni podobným způsobem vylepšit model poskytující rizikové skóre podvodu při hlášení pojistné události. Opět můžeme využít **přepsaný záznam telefonního hlášení** nebo text elektronicky zasláného **popisu události.**

ACREA CR | Váš partner v oblasti analýzy dat

Společnost ACREA nabízí pomoc při zpracování textových dokumentů, zapojení dat do standardních modelů nebo konzultace ke stávajícím projektům. Nabízíme i širokou škálu kurzů zaměřených na metody, které využijete při tvorbě a spravování modelů.

Zaujaly Vás tyto informace?

Neváhejte nás kontaktovat na emailu info@acrea.cz.
Rádi zodpovíme Vaše dotazy a navštívíme Vás.

