

Silniki rekomendacyjne

Wprowadzenie

Systemy rekomendacji stały się ważnym obszarem badań od połowy lat 90. Zainteresowanie tym obszarem badawczym pozostaje wysokie, ponieważ z jednej strony takie aplikacje pomagają użytkownikom znaleźć odpowiednie przedmioty, a z drugiej strony systemy te są przydatne do wyszukiwania przedmiotów, do których nie można uzyskać dostępu, ponieważ użytkownicy nie wiedzą o ich istnieniu. Przykłady takich aplikacji obejmują rekomendacje filmów, wiadomości i książek oraz innych produktów na Amazon.com. Systemy rekomendacji zostały pierwotnie opracowane w celu sugerowania produktów lub usług, które zostały dostosowane do preferencji poszczególnych użytkowników. Istnieją jednak sytuacje, w których grupa użytkowników wspólnie uczestniczy w jednym działaniu, takim jak oglądanie filmu lub zwiedzanie miasta. W takich przypadkach potrzebujemy technik, które rozwiązują problem identyfikacji rekomendacji dla grupy użytkowników i zaspokojenia prawie wszystkich osób w grupie w największym możliwym stopniu. Jedną motywacją dla systemów rekomendacji grupowych; obszary związane z wyborem społecznym (zwanym także grupowym podejmowaniem decyzji) w ekonomii, polityce i socjologii, w których chcemy zdecydować, co jest najlepsze dla grupy, biorąc pod uwagę różne opinie jej członków, również ilustrują znaczenie rekomendacji grupowych. Problem rekomendacji grupowych został również zbadany w obszarach metasearch, oprogramowania pośredniego bazy danych, filtrowania grupowego i systemów wieloagentowych. Na przykład w metasearch listy rankingowe tworzone przez wiele wyszukiwarek muszą być połączone w jedną listę; ten problem jest znany jako problem agregacji rang. Oprogramowanie pośrednie bazy danych obejmuje porządkowanie obiektów, gdy każdemu obiektowi są przypisane wartości liczbowe związane z wieloma polami. W systemach wieloagentowych agenci muszą wykorzystywać decyzje, które są racjonalne nie tylko z punktu widzenia jednostki, ale także z punktu widzenia społecznego. Dlatego możemy powiedzieć, że rekomendacja grupy próbuje zidentyfikować przedmioty, które będą mile widziane przez grupę jako całość, a nie tylko przez poszczególnych członków grupy. Grupy zainteresowań mogą się różnić od ustalonych grup do grup losowych, które wymagają jedynie sporadycznych zaleceń. Większość badań dotyczących rekomendacji grupowych koncentruje się na badaniu podstawowych algorytmów używanych do generowania rekomendacji. Do tej pory zastosowano dwie główne strategie do generowania rekomendacji grupowych. Techniki te obejmują agregowanie indywidualnych prognoz w prognozy grupowe (agregowane prognozy) lub agregowanie indywidualnych modeli w modele grupowe (modele agregowane). Strategie te różnią się harmonogramem etapu agregacji danych. W strategii zagregowanej prognozy, indywidualne prognozy są generowane na podstawie indywidualnych modeli preferencji, a następnie są agregowane w predykcję grupową. Alternatywną strategią jest zbudowanie modelu preferencji grupowych przed wykonaniem rekomendacji przedmiotu. Innymi słowy, w strategii modelu zagregowanego poszczególne modele użytkowników są scalane w model oparty na grupie, a następnie generowane są rekomendacje przy użyciu zagregowanego modelu grupy. Jednak liczne badania wykazały, że rekomendacja grupowa jest zadaniem znacznie bardziej złożonym niż rekomendacja indywidualna, a gdy rozważane są grupy, pojawiają się różne wyzwania, które uniemożliwiają skuteczne działanie tradycyjnych technik rekomendacji dla rekomendacji pozycji dla grup. Systemy rekomendacji to narzędzia i techniki, które zapewniają sugestie dotyczące elementów, z których mogą korzystać użytkownicy. Systemy rekomendacji są zasadniczo ukierunkowane w kierunku pomagania użytkownikom w znajdowaniu przedmiotów, którymi mogą być zainteresowani spośród przeważającej liczby przedmiotów; takie systemy próbują przewidzieć najbardziej odpowiednie produkty lub usługi na podstawie preferencji i ograniczeń użytkowników. Ogólna klasyfikacja danych wykorzystywana przez systemy rekomendacji wykorzystuje trzy kategorie: przedmioty, użytkownicy i transakcje, to znaczy relacje między użytkownikami i przedmiotami.

1. Przedmioty to przedmioty zalecane. Przedmioty można scharakteryzować pod względem ich złożoności oraz ich wartości lub użyteczności.

2. Użytkownicy systemu rekomendacji mogą mieć bardzo różnorodne cele i cechy. Aby spersonalizować swoje rekomendacje, system rekomendacji próbuje wykorzystać informacje o swoich użytkownikach. Informacje te można uporządkować na wiele sposobów; wybór informacji do modelowania zależy od zastosowanej techniki rekomendacji.

3. Transakcja to zarejestrowana interakcja między użytkownikiem a systemem rekomendacji. Transakcje to dane przydatne w algorytmie generowania rekomendacji. Oceny to najpopularniejsza forma danych transakcyjnych gromadzonych przez system rekomendacji. Inne rodzaje danych transakcji mogą obejmować czytanie historii lub to, jak często użytkownik korzysta z rekomendacji, które przekazał. Oceny mogą przybierać różne formy, takie jak oceny liczbowe, oceny porządkowe i oceny binarne.

Systemy rekomendacji mogą być wykorzystywane do różnych celów, jak stwierdzono przy identyfikowaniu typowych zadań, w których system rekomendacji może pomóc, takich jak tematy wymienione poniżej.

* Znalezienie pożądaných przedmiotów: To zadanie polega na znalezieniu uszeregowanej listy przedmiotów dla użytkownika i przewidywaniu, jak bardzo użytkownik polubi te przedmioty (np. oceny w gwiazdkach).

* Znalezienie wszystkich pożądaných elementów: To zadanie polega na znalezieniu wszystkich, a nie tylko niektórych elementów, którymi użytkownik będzie zainteresowany. Czasami konieczne jest znalezienie wszystkich możliwych elementów dla użytkownika, a nie tylko niektórych z nich. Przykłady obejmują sytuacje medyczne, finansowe i prawne, w których fałszywe alarmy mogą potencjalnie mieć ogromne konsekwencje, jeśli są obecne.

* Dostarczanie adnotacji w kontekście: To zadanie polega na znalezieniu użytecznych elementów w bieżącym kontekście użytkownika. Na przykład system może przekazywać takie rekomendacje, jak „Klienci, którzy kupili ten produkt w tym konkretnym kontekście (Boże Narodzenie), kupili to” lub „Podobne filmy”, gdy użytkownik szuka filmu.

* Znalezienie zalecanej sekwencji: To zadanie polega na znalezieniu sekwencji elementów dla użytkownika, która jest „przyjemna jako całość”, zamiast dopasowywania preferencji użytkownika tylko w odniesieniu do poszczególnych elementów. Przykłady takich systemów rekomendacji obejmują generowanie list odtwarzania utworów i funkcję radia w Spotify.

* Proste przeglądanie: To zadanie polega na udzielaniu pomocy użytkownikom, którzy chcą przeglądać zestaw elementów, poprzez dostarczanie odpowiednich dla nich sugestii. Takie rekomendacje są oferowane przez różne usługi, w tym Amazon.com i eBay

* Znalezienie wiarygodnego polecającego: Niektórzy użytkownicy nie ufają systemom rekomendacji. Mogą eksperymentować z systemem i obserwować, jak zachowuje się on w różnych ustawieniach, na przykład zmieniając jego profil. Systemy rekomendacji mogą pomóc w potwierdzeniu ich wiarygodności, włączając również powód, dla którego dana pozycja była polecana w wynikach, na przykład „ta piosenka była polecana, ponieważ słuchałeś xy”.

* Poprawianie swojego profilu przez ocenianie elementów: Użytkownik może pomóc poprawić wynik algorytmu rekomendacji, umożliwiając algorytmowi lepsze zrozumienie rodzajów elementów, które lubi.

* Wyrażanie siebie: użytkownicy mogą również oceniać przedmioty, ponieważ lubią to robić, ponieważ lubią wyrażać siebie i czuć się z tym dobrze. Tutaj głównym celem nie jest poprawa wyników systemu rekomendacji, ale wyrażenie swoich uczuć na temat przedmiotu.

* Pomaganie innym: Głównym celem oceny przedmiotu może być również pomoc innym. Jeśli na przykład użytkownik miał złe doświadczenia z firmą turystyczną, opublikowanie odpowiedzi i oceny na stronie internetowej takiej jak tripadvisor.com w celu ostrzeżenia innych może być motywacją dla użytkownika do korzystania z systemu rekomendacji.

* Wpływanie na innych: systemy rekomendacji można wykorzystać do promocji firmy. Witryny, takie jak hotels.com i tripadvisor.com, mogą być używane do przypisywania ocen do hoteli, a także innych obiektów i usług. Nieetyczne podejścia, takie jak zachęcanie osób fizycznych (np. Pracowników) do przypisywania dużej ilości wysokich ocen do jednej firmy, przy jednoczesnym przypisywaniu niższych ocen do konkurentów, mogą być wykorzystane do promocji firmy.

Techniki systemu rekomendacji

Techniki rekomendacji mogą wykorzystywać kilka różnych klasyfikacji. Klasyfikacje te oparte są na danych, które są wykorzystywane jako dane wejściowe do rekomendacji. W następujących rozdziałach przedstawimy taką klasyfikację, która jest proponowana w literaturze i zapewni przegląd różnych rodzajów systemów rekomendujących. Systemy rekomendujące są zazwyczaj podzielone na następujące kategorie na podstawie sposobu formułowania rekomendacji:

Zalecenia oparte na treści

Podejście oparte na treści opiera się na wyszukiwaniu informacji i opiera się na założeniu, że użytkownicy będą preferować przedmioty, które są podobnymi przedmiotami, z których wcześniej wyrazili przyjemność. Systemy rekomendacji oparte na treści analizują opisy elementów w celu zidentyfikowania elementów szczególnie interesujących dla użytkownika i koncentrują się na ich cechach. Podobieństwo elementów określa się, mierząc podobieństwo ich cech. Rzeczywiście, systemy rekomendacji próbują polecać produkty, które są podobne do przedmiotów, które dany użytkownik lubił w przeszłości. Taki system analizuje zestaw funkcji elementów, które zostały wcześniej ocenione przez użytkownika i konstruuje model lub profil zainteresowań użytkownika na podstawie cech elementów ocenianych przez tego użytkownika. Taki model można na przykład zastosować do filtrowania wyników wyszukiwania, decydując, czy użytkownik jest zainteresowany określoną stroną internetową. Przykładem takiej rekomendacji może być rekomendacja konkretnego hotelu na podstawie liczby gwiazdek lub poziomu cen lub tego, czy oferuje on usługi na zewnątrz. Algorytm filtrowania treści dla rekomendacji hotelowych można opisać w następujący sposób: System zbiera zestaw funkcji w połączeniu z prawdopodobieństwem, że każdemu użytkownikowi spodoba się każdy element. Aby określić hotel, który użytkownik najprawdopodobniej preferuje, system przeprowadza porównanie tego zestawu funkcji z różnymi hotelami w systemie. Porównanie może obejmować heurystykę lub po prostu obliczenie podobieństwa, takie jak obliczenie oparte na cosinusie. Następnie system zwraca hotele, które najlepiej pasują do preferencji użytkownika. Filtrowanie na podstawie treści może być bardzo przydatne, gdy dla elementów dostępna jest znaczna ilość danych opisowych. Dane te mogą zostać skodyfikowane i wewnątrznie przedstawione jako ontologia funkcji, zapewniając w ten sposób szczegółowy opis domeny, który ma zastosowanie do obliczania podobieństwa do preferencji użytkownika. Niemniej jednak, aby był użyteczny, schemat oparty na treści musi mieć taką ontologię dostępną oprócz możliwości klasyfikowania przedmiotów według ontologii. Dlatego filtrowanie na podstawie treści może być problematyczne, gdy informacje dostępne dla każdego elementu są rzadkie lub trudne do interpretacji przy użyciu algorytmów komputerowych. Ponadto rekomendacje oparte na treści mogą opierać się wyłącznie na dostępnych wyraźnych informacjach

dotyczących określonego elementu. Kilka elementów, takich jak reprezentacje elementów, profile użytkowników i zdolność uczenia się modelu użytkownika, jest obecnych w architekturze wysokiego poziomu filtrowania opartego na treści; elementy te zostaną szczegółowo wyjaśnione poniżej.

Reprezentacje przedmiotów

Elementy, które można polecić użytkownikom, są reprezentowane przez zestaw funkcji. Na przykład każdy film jest powiązany z pewnymi funkcjami, takimi jak gatunki, aktorzy, tematy i daty. Gdy każdy element jest opisany przez zestaw funkcji, a wartości tych funkcji są znane, elementy mogą być reprezentowane przez ustrukturyzowane dane. Gdy dane nie mają struktury, jak ma to miejsce w przypadku tekstu, wymagany jest etap wstępnego przetwarzania w celu wyodrębnienia odpowiednich informacji ustrukturyzowanych. Podstawowym zadaniem takiego komponentu jest reprezentowanie zawartości pozycji, na przykład dokumentów, stron internetowych, wiadomości i opisów produktów, w formie odpowiedniej do użycia na kolejnych etapach przetwarzania. Elementy danych są analizowane przy użyciu technik ekstrakcji cech, aby przenieść reprezentacje elementów z oryginalnej przestrzeni informacyjnej na przestrzeń docelową. Na przykład strony internetowe mogą być reprezentowane jako wektory słów kluczowych

Profile użytkowników

Ten komponent tworzy model oparty na zainteresowaniach użytkowników wywnioskowanych z poprzednich ocen pozytywnych i negatywnych. Zazwyczaj strategia generalizacji jest wdrażana przy użyciu technik uczenia maszynowego, które są w stanie wywnioskować model zainteresowań użytkowników na podstawie elementów, które w przeszłości lubiły lub nie. Na przykład w przypadku rekomendacji wiadomości opartych na historii czytania użytkownika można włączyć różne właściwości artykułów prasowych, takie jak treść wiadomości, wzorce dostępu i nazwane podmioty, aby utworzyć profil użytkownika. Ten profil jest następnie używany jako przedstawienie elementu do użycia w celu identyfikacji elementów odpowiednich dla użytkownika docelowego. Ta reprezentacja zazwyczaj składa się z wektora ważonych elementów: im większa waga, tym bardziej użytkownik lubi tę cechę. Rozmiar wektora odpowiada liczbie funkcji. Istnieje kilka podejść, które można wykorzystać do tworzenia profili użytkowników z profili produktów. Możliwe jest obliczenie każdego profilu użytkownika jako średniego wektora na podstawie profili elementów ocenianych przez użytkownika.

Nauka modeli użytkowników

Ten moduł wykorzystuje profil użytkownika do sugerowania odpowiednich elementów poprzez dopasowanie reprezentacji profilu do tych, które mają być polecane. Wynik może być oparty na binarnej lub ciągłej ocenie trafności przy użyciu różnych wskaźników podobieństwa.

Zalecenia dotyczące współpracy

W tym podejściu, zamiast używać funkcji elementów do określania ich podobieństwa, nacisk kładziony jest na podobieństwo ocen użytkowników dla elementów. Na przykład w zaleceniach filmowych, aby przewidzieć ocenę filmu u przez użytkownika, technika filtrowania grupowego próbuje zidentyfikować użytkowników podobnych do Ciebie, to znaczy innych użytkowników o podobnych gustach w filmach (tzn. Oceniają te same filmy podobnie). Następnie technika agreguje oceny wyrażone przez użytkowników o podobnych wzorcach oceny i oblicza przewidywaną ocenę dla danego użytkownika. Filtrowanie kooperacyjne jest zdecydowanie najbardziej rozpowszechnioną techniką rekomendacji ze względu na jego moc i prostotę. Największą siłą technik współpracy jest to, że są one całkowicie niezależne od każdej możliwej do odczytu maszynowej reprezentacji zalecanych obiektów i mogą być skutecznie stosowane do złożonych obiektów, takich jak muzyka i filmy, które użytkownik może polubić

lub niechęć z wielu powodów. Na przykład użytkownik może polubić określony gatunek filmów, określonego reżysera, określonego aktora lub aktorkę itp. Filtrowanie grupowe jest często określane jako filtrowanie społecznościowe, ponieważ technika rozpoznaje podobieństwa między użytkownikami na podstawie ich ocen i generuje nowe rekomendacje oparte na porównaniach między użytkownikami. Ten proces naśladuje społeczny proces polecania przedmiotów znajomym, ponieważ system zapewnia użytkownikom rekomendacje na podstawie elementów, które podobały się osobom, które są podobne do użytkownika pod względem ocen. Algorytmy stosowane we współpracujących systemach filtrujących można pogrupować w dwie ogólne klasy: algorytmy oparte na pamięci (lub oparte na heurystyce) i algorytmy oparte na modelu. Algorytmy oparte na pamięci zapewniają przewidywania ocen dla użytkowników na podstawie ich poprzednich ocen. Podejścia oparte na pamięci stosują heurystykę, która wykorzystuje wszystkie lub próbkę danych elementu użytkownika do wygenerowania prognozy. Stosowane są dwa ogólne podejścia do opartego na pamięci wspólnego filtrowania (CF): oparte na użytkownikach wspólne filtrowanie i oparte na elementach wspólne filtrowanie. Oparte na użytkownikach wspólne filtrowanie tworzy prognozy dla danego użytkownika, najpierw identyfikując użytkowników, którzy są podobni (najbliżsi sąsiedzi) do danego użytkownika, a następnie obliczając najczęściej oceniane elementy, których dany użytkownik nie widział. Na przykład wspólny system rekomendacji filtrowania książek może sugerować książkę, która nie została przeczytana przez danego użytkownika, ale przeczytana przez użytkowników o podobnym smaku. Wspólne filtrowanie oparte na elementach tworzy prognozy dla danego użytkownika, najpierw obliczając najbardziej podobne elementy dla każdego elementu zgodnie z ich podobieństwami. Algorytmy oparte na modelach modelują użytkowników na podstawie ich wcześniejszych ocen, a systemy rekomendujące używają tych modeli do przewidywania ocen nieocenionych pozycji. W tej metodzie, gdy algorytm uczy się modelu, nie potrzebuje już dostępu do każdej oceny i może wykorzystać parametry modelu do obliczenia swoich prognoz. Podejścia oparte na modelach (takie jak algorytmy uczenia maszynowego i eksploracji danych) są wykorzystywane do projektowania i opracowywania modeli, które umożliwiają systemowi uczenie się złożonych wzorców na podstawie danych szkoleniowych, a następnie dostarczanie prognoz opartych na tak wyuczonych modelach. Zaletą tego podejścia jest to, że po nauczaniu się modelu system nie wymaga już dostępu do każdej oceny i parametrów modelu mogą być używane do obliczania prognoz; jednak faza uczenia się może być czasochłonna. Natomiast podejścia oparte na pamięci są wolniejsze podczas fazy prognozowania, ponieważ muszą heurystycznie zbadać pełny zestaw ocen, podczas gdy podejścia oparte na modelu są dość szybkie, ponieważ wykorzystują tylko parametry modelu do obliczenia przewidywanej oceny.

Podejścia hybrydowe

Inną popularną metodą formułowania rekomendacji jest zastosowanie podejścia hybrydowego. Można albo oddzielnie wdrożyć systemy oparte na współpracy i treści, a wyniki połączyć, w ten sposób łącząc cechy jednej metody z drugą metodą, lub stworzyć model ujednociający, który ma cechy obu podejść. Zastosowanie takiej techniki pomaga złagodzić pewne ograniczenia systemów opartych na treści i współpracy. Wdrażanie metod współpracy i opartych na treści osobno i łączenie ich prognoz Włączanie niektórych cech opartych na treści w podejście oparte na współpracy Włączanie niektórych cech opartych na współpracy w podejście oparte na treści Konstruowanie ogólnego modelu unifikacyjnego obejmującego zarówno cechy oparte na treści, jak i cechy oparte na współpracy

Zastosowania silników rekomendacyjnych w biznesie

Jeśli korzystałeś z usługi przesyłania strumieniowego Netflix, kupiłeś coś od Amazon lub związałeś się z „osobami, które możesz znać” na LinkedIn lub Facebooku, to użyłeś silnika rekomendacji. Mechanizmy rekomendacji dopasowują cechy dotyczące Ciebie do rzeczy, które mogą Cię zainteresować. Na przykład film ma rok premiery, gatunek, aktorów i wyniki kasowe. Masz funkcje. Masz preferencje,

wiek i być może ukończyłeś ankietę wyrażającą niektóre twoje postawy wobec niektórych filmów. Być może oceniłeś niektóre oglądane filmy. Dzięki ustaleniu, które zestawy filmów mają być pokazywane, oraz odpowiedzi na te zalecenia, maszyna uczy się z czasem, aby tworzyć lepsze sugestie. Jeśli obejrzałeś kilka filmów science fiction i oceniłeś je wysoko, silnik nauczy się pokazywać więcej filmów science fiction, a także, dla odmiany, filmy, które podobają się innym osobom, takim jak ty, lubiącym filmy science fiction. Systemy rekomendacji zostały wdrożone w wielu branżach i kontekstach, zwłaszcza w ramach witryn zakupów internetowych, wyjaśnił Thompson. Zazwyczaj są używane w:

- * Supermarkety: Rejestry, które generują niestandardowe kupony na następne zakupy, przeglądają wcześniejsze zakupy i kupony na miarę, aby uwzględnić przedmioty, które prawdopodobnie Cię zainteresują.

- * Sklepy z książkami i muzyką: Firmy mogą wysyłać klientom e-maile po zakupie i polecać nowe książki lub albumy - lub w przypadku Amazon, Pandora i podobnych witryn, dostarczać rekomendacje produktów w czasie rzeczywistym w oparciu o to, co klient ogląda lub słucha do.

- * Firmy inwestycyjne: systemy rekomendacji mogą analizować, które akcje prawdopodobnie byłyby zainteresowane na podstawie tego, co wybrali podobni klienci.

- * Usługi telewizyjne i filmowe: firmy takie jak Netflix analizują wcześniejsze wybory treści każdego klienta i na ich podstawie formułują rekomendacje, a także proaktywnie polecają produkty w czasie rzeczywistym na podstawie historii przeglądania.

- * Witryny sieci społecznościowych: witryny takie jak LinkedIn i Facebook używają systemów rekomendacji do sugerowania dodatkowych połączeń lub znajomych w oparciu o istniejącą sieć danej osoby.

Wiele algorytmów używanych w silnikach rekomendacji i uczenie maszynowe nie jest wcale takie nowe. Regresja, drzewa decyzyjne, k-najbliższy sąsiad, maszyny wektorów nośnych (SVM), sieci neuronowe i naiwne Bayes to ustalone metody o dobrze znanych ograniczeniach i odpowiednich zastosowaniach. Wiele z tych metod wykorzystywano od dawna do wspierania decyzji biznesowych opartych na danych. Typowy silnik rekomendacji przetwarza dane przez cztery następujące fazy, a mianowicie zbieranie, przechowywanie, analizę i filtrowanie.

Zbieranie danych

Pierwszym krokiem w tworzeniu silnika rekomendacji jest zebranie danych. Dane mogą być danymi jawnymi lub niejawnymi. Jawne dane składałyby się z danych wprowadzanych przez użytkowników, takich jak oceny i komentarze do produktów. Dane niejawne to historia zamówień / historia zwrotów, zdarzenia koszyka, odsłony, kliknięcia i dziennik wyszukiwania. Ten zestaw danych zostanie utworzony dla każdego użytkownika odwiedzającego witrynę. Dane o zachowaniu są łatwe do zebrania, ponieważ możesz przechowywać dziennik działań użytkowników w swojej witrynie. Zbieranie tych danych jest również proste, ponieważ nie jest potrzebne wszelkie dodatkowe działania ze strony użytkownika; już korzystają z aplikacji. Minusem tego podejścia jest to, że trudniej jest analizować dane. Na przykład filtrowanie potrzebnych dzienników z mniej potrzebnych może być kłopotliwe. Ponieważ każdy użytkownik ma inne upodobania dotyczące produktu, jego zestawy danych będą różne. Z biegiem czasu, gdy „karmisz” silnik większą ilością danych, staje się on coraz mądrzejszy dzięki swoim zaleceniom, dzięki czemu Twój subskrybenci i klienci są bardziej skłonni do angażowania się, klikania i kupowania. Podobnie jak silnik rekomendacji Amazon współpracuje z kartami „Często kupowane razem” i „Polecane dla Ciebie”.

Przechowywanie danych

Im więcej danych możesz udostępnić algorytmom, tym lepsze rekomendacje będą. Oznacza to, że każdy projekt rekomendacji może szybko zmienić się w projekt dużych zbiorów danych. Typ danych, których używasz do tworzenia rekomendacji, może pomóc ci zdecydować, jakiego rodzaju pamięci należy użyć. Możesz wybrać użycie bazy danych NoSQL, standardowa baza danych SQL lub nawet jakiś obiektowy obiekt pamięci. Każda z tych opcji jest wykonalna w zależności od tego, czy przechwytyjesz dane wejściowe lub zachowanie użytkownika, a także od czynników, takich jak łatwość implementacji, ilość danych, którymi może zarządzać magazyn, integracja z resztą środowiska i przenośność. Podczas zapisywania ocen lub komentarzy użytkowników skalowalna i zarządzana baza danych minimalizuje liczbę wymaganych zadań i pomaga skupić się na rekomendacji.

Analiza danych

Aby znaleźć elementy, które mają podobne dane dotyczące zaangażowania użytkowników, filtrujemy je, stosując różne metody analizy. Jeśli chcesz udzielić natychmiastowych rekomendacji użytkownikowi przeglądającemu produkt, potrzebujesz analizy typu zwinnego. Oto niektóre z metod analizy danych: Systemy w czasie rzeczywistym mogą przetwarzać dane podczas ich tworzenia. Ten typ systemu zwykle obejmuje narzędzia, które mogą przetwarzać i analizować strumienie zdarzeń. Wymagany byłby system działający w czasie rzeczywistym, który dawałby aktualne rekomendacje. Analiza wsadowa wymaga okresowego przetwarzania danych. Podejście to sugeruje, że należy stworzyć wystarczającą ilość danych, aby analiza była odpowiednia, na przykład dzienna wielkość sprzedaży. System wsadowy może działać poprawnie, aby wysłać wiadomość e-mail w późniejszym terminie. Analiza w czasie rzeczywistym pozwala szybko zbierać dane, dzięki czemu można odświeżyć dane analityka co kilka minut lub sekund. System zbliżony do czasu rzeczywistego działa najlepiej w celu dostarczania rekomendacji podczas tej samej sesji przeglądania.

Algorytm rekomendacji produktu

Filtrowanie danych: Aby uzyskać odpowiednie dane niezbędne do przedstawienia rekomendacji dla użytkownika. Musimy wybrać algorytm, który lepiej pasowałby do silnika rekomendacji.

Oparte na treści: popularny, polecany produkt ma cechy podobne do tego, co użytkownik ogląda lub lubi.

Klaster: polecane produkty pasują do siebie, bez względu na to, co zrobili inni użytkownicy.

Współpraca: inni użytkownicy, którzy lubią te same produkty, co widok lub polubienie innego użytkownika, również polubią zalecany produkt. Filtrowanie oparte na współpracy umożliwia tworzenie teoretycznych atrybutów produktu i przewidywanie na podstawie gustów użytkowników.

Wyniki tego filtrowania oparte są na założeniu, że dwóch użytkowników, którzy lubili te same produkty w przeszłości, prawdopodobnie polubią te same teraz lub w przyszłości

Możesz reprezentować dane o ocenach lub interakcjach jako zbiór macierzy, a produkty i użytkownicy jako wymiary. Załóżmy, że następujące dwie macierze są podobne, ale następnie odejmujemy drugą od pierwszej, zastępując istniejące oceny liczbą pierwszą, a brakujące oceny liczbą zero. Powstała macierz jest tabelą prawdy, w której liczba pierwsza reprezentuje interakcję użytkowników z produktem. Ostatecznie wynik uzyskany po odfiltrowaniu i użyciu algorytmu zalecenia są przekazywane użytkownikowi na podstawie aktualności typu zalecenia, niezależnie od tego, czy jest to rekomendacja w czasie rzeczywistym, czy też wysłana później wiadomość e-mail. Ponieważ silnik rekomendacji produktu działa głównie na danych. Firma może nie mieć wystarczającej pojemności do przechowywania tej ogromnej ilości danych od odwiedzających witrynę. Można używać platform internetowych takich jak Hadoop i Spark, które umożliwiają przechowywanie danych na wielu

urządzeniach w celu zmniejszenia niezawodności na jednym komputerze. Hadoop używa HDFS do dzielenia plików na duże bloki i dystrybuje je między węzłami w klastrze. Dzięki temu zestaw danych może być przetwarzany szybciej i wydajniej niż w bardziej konwencjonalnej architekturze superkomputera, która opiera się na równoległym systemie plików, w którym obliczenia i dane są dystrybuowane za pośrednictwem szybkich sieci. Wreszcie przetwarzamy zestawy dużych danych przy użyciu modelu programowania MapReduce. Dzięki temu możemy jednocześnie uruchomić algorytm w rozproszonym systemie plików i wybrać najbardziej podobny klaster. W ten sposób każda firma może opracować własną architekturę silnika rekomendacji za pomocą narzędzi open source, a my możemy pomóc im we wdrożeniu silnika, korzystając z naszej wiedzy technicznej.

Biznesowy przypadek użycia

Silniki rekomendacji odgrywają kluczową rolę w zaangażowaniu i utrzymywaniu klientów w mediach internetowych i przemyśle rozrywkowym. Przy wykładniczej objętości danych medialnych silniki rekomendacji z dużymi danymi pokazują nowoczesne, zorientowane na użytkownika podejście do dostarczania mediów poprzez wydajne przetwarzanie danych, uczenie maszynowe i analizy predykcyjne. Silnik rekomendacji medialnych można zbudować dla filmów lub teledysków, książek lub dowolnych produktów. Taki przypadek użycia filmu (dużych) danych koncentruje się na architekturze silnika rekomendacji dla klientów korzystających z dekodera (STB), które:

- * Wykorzystuje architekturę Hadoop typu open source jako podstawę dużych zbiorów danych
- * Zbiera nieprzetworzone dane użytkownika z filmów na żądanie, set-top dzienników aktywności, zaplanowanych nagrań i kilku katalogów mediów
- * Przetwarza i analizuje dane dziennika użytkownika w ramach Big Data Hadoop
- * Dostarcza wyniki do wyszukiwarki, która następnie dostarcza unikalne rekomendacje za pośrednictwem interfejsu przeglądarki skierowanego do użytkownika

Na poziomie szczegółowym indywidualne zachowania użytkowników, takie jak oglądane filmy, kliknięte katalogi, programy zaplanowane do nagrywania, średni czas oglądania wideo, są dokładnie analizowane zgodnie z metodologią analizy dzienników dużych zbiorów danych. Na wysokim poziomie te ogromne dane mogą namalować obraz „tego, co jest na topie” dla konkretnego użytkownika lub wśród grup użytkowników o podobnych gustach. Dla analityków danych ta architektura silnika rekomendacji wykracza poza uruchamianie zapytań SQL względem hurtowni danych w celu przewidywania trendów i preferencji.

Big Data pozwala im być bardziej wydajnym dzięki przetwarzaniu ogromnych ilości danych użytkownika w ułamku czasu w porównaniu do tradycyjnego SQL. Spersonalizowane wyniki zostały załadowane do wyszukiwarki i wyświetlone we wbudowanej aplikacji internetowej. Dla marketerów jest to doskonałe narzędzie do wyszukiwania osobowości użytkowników i dostarczania tego, czego chcą użytkownicy. Dla użytkowników końcowych spersonalizowane rekomendacje oszczędzają im ręcznego przeglądania ogromnej bazy danych filmów. Uczenie maszynowe i analizy predykcyjne umożliwiają silnikowi rekomendacji bardziej precyzyjne przewidywanie preferencji użytkowników, tj. Zwiększanie zadowolenia i retencji użytkowników. Duże zbiory danych, ze skalowalnością i mocą do przetwarzania ogromnych ilości zarówno ustrukturyzowanych (np. Tytułów wideo, których szukają użytkownicy, preferowanego gatunku muzyki), jak i nieustrukturyzowanych danych (np. Schematów przeglądania / odsłuchu użytkowników), mogą umożliwić firmom analizę miliardów kliknięć i przeglądania dane od Ciebie i innych użytkowników takich jak Ty w celu uzyskania najlepszych rekomendacji.

Chociaż silniki rekomendacji były szeroko stosowane i badane przez ponad dekadę, pozostaje kilka kluczowych wyzwań. Skalowalność: Ilość danych wykorzystywanych jako dane wejściowe do mechanizmów rekomendacji rośnie szybko wraz z dodawaniem większej liczby użytkowników i elementów. Na przykład w popularnej witrynie rozmiar przechowywanych danych o zachowaniach użytkowników może łatwo osiągnąć terabajty dziennie. Pomimo dużej ilości danych większość mechanizmów rekomendacji stara się reagować interaktywnie w czasie krótszym niż sekunda, aby utrzymać zaangażowanie użytkowników. Głównym wyzwaniem jest tutaj zaprojektowanie wydajnych algorytmów uczenia się, które mogłyby obsługiwać tak duże zbiory danych.

Prywatność: rozumiejąc wartość danych użytkownika, większość stron internetowych gromadzi jak najwięcej danych użytkownika. Takie podejście budzi obawy dotyczące prywatności, ponieważ dane mogą zawierać poufne informacje, które użytkownicy chcą zachować prywatność, np. adresy użytkowników i historia płatności. Mimo że użytkownikom przedstawiono politykę prywatności dotyczącą wykorzystania danych, zwykle nie mają oni wyraźnej kontroli nad danymi.

Sparsity: Sparsity to problem braku informacji. W sklepach internetowych, które mają ogromną liczbę użytkowników i przedmiotów, prawie zawsze są użytkownicy, którzy ocenili tylko kilka produktów. Za pomocą mechanizmów rekomendacji opartych na współpracy i innych podejściach można zasadniczo tworzyć dzielnice użytkowników korzystających z ich profili. Jeśli użytkownik ocenił tylko kilka przedmiotów, określenie jego gustu jest dość trudne i może być związane z niewłaściwym sąsiedztwem.

Ustrukturyzowane rekomendacje: obecne silniki rekomendacji przewidują poszczególne elementy, których mogą chcieć użytkownicy. Fascynującym rozszerzeniem jest przewidywanie preferencji dla zestawów przedmiotów. Na przykład, jeśli system odkryje, że użytkownik wybiera się po raz pierwszy na sporty zimowe, może polecić parę butów, kasku, okularów i sukienki narciarskiej, które mają pasujące kolory i poziomy cenowe. W ten sposób użytkownik może uzyskać wszystko, czego potrzebuje do jazdy na nartach przy jednym zakupie. Istnieją dwa wyzwania w tego rodzaju uporządkowanych zaleceniach. Po pierwsze, liczba możliwych zestawów rośnie wykładniczo wraz z rozmiarem grupy.

Biorąc pod uwagę, że liczba elementów jest już bardzo duża, wydajność algorytmów uczenia się może stanowić problem. Po drugie, w przeciwieństwie do poszczególnych pozycji, nie jest jasne, jak wybrać odpowiednią funkcję punktacji dla zestawów.

Zaufanie: głosy osób z krótką historią mogą nie być tak istotne jak głosy osób, które mają bogatą historię w swoich profilach. Problem zaufania powstaje w związku z ocenami określonego klienta. Problem można rozwiązać poprzez dystrybucję priorytetów wśród użytkowników.