



ارائه شده توسط:

سایت ترجمه فا

مرجع جدیدترین مقالات ترجمه شده

از نشریات معتبر

سیستم پیشنهادی مبتنی بر تصویر داده کاوی پزشکی

چکیده

این مقاله یک الگوریتم مبتنی بر نمونه را ارائه می دهد که می تواند با ایجاد ارتباط دو طرفه بین بیماران، مجموعه داده های پزشکی بزرگ را کشف کند. در روش متریک آزاد ما، هر فرد در یک مجموعه داده، هر عضو از مجموعه داده را رده بندی می کند. با جمع آوری این صفوف، تجسم داده با توجه به افراد معمولی نشان دهنده زیر مجموعه های بیماران خویشاوند نزدیک، ممکن است. این مقاله همچنین یک ابزار تجسم برای بررسی پایگاه داده بیماران دیابتی را توصیف می کند. این نمونه اولیه از یک سیستم پیشنهادی، الگوریتم فوق را برای پر مایه کردن داده ها، ساختار بیماران، ایجاد ارتباط بین افراد و ارائه پیشنهاد ها، پیاده سازی می کند.

کلمات کلیدی: اکتشاف، توصیه، داده های پزشکی، یادگیری مبتنی بر نمونه

1. مقدمه

با توجه به اینکه پرونده های پزشکی الکترونیکی و سنسورهای پوشیدنی گسترده تر می شود، مجموعه داده های پزشکی با استفاده از روش های خاص برای اکتشاف به سمت بزرگتر شدن می روند. این مجموعه داده ها با مشکلات جدانشدنی همراه هستند: آنها حاوی داده های با ابعاد بزرگ که می توانند ناهمگن و بدون ساختار باشند، و اغلب شامل ارزش های از دست رفته هستند. با چنین مقدار داده های موجود، پزشکان برای پیدا کردن بیماران مربوطه، به کمک نیاز دارند. آنها بر روی سیستم های توصیه شده جهت مرور، کشف و دستکاری مدارک تکیه می کنند و این امر به آنها اجازه می دهد که ارتباط بین بیماران مشابه را پیدا کنند. با وجود استفاده از برنامه های کاربردی یادگیری ماشینی و نتایج نشان دهنده امیدوار کننده های پزشکی، استفاده از روش های سنتی دقیقاً مطابق با زمینه بالینی نیست. این روش ها تمایل به تعمیم دادن و حذف نابرابری ها دارند، الگوریتم ها باید آموزش داده شوند و بیشتر راه حل های "جعبه سیاه" هستند که باعث کاهش قابلیت تفسیر می شوند. در حالی که آنها در تلاش برای پیش بینی بیماری ها به خوبی عمل می کنند، آن الگوریتم ها هنگامی که سعی در تشخیص بیماری های مشابه و مطالعه بیماران غیرطبیعی را بدون توجه بر تخصص دامنه دارند، مناسب نیستند. هدف از این مقاله ارائه یک روشی است که قادر به ساختن عناصر یک مجموعه داده با ایجاد ارتباط بین

آنها است. ساختار به دست آمده با اندازه گیری نمایندگی، داده ها را غنی می کند و این راهی برای تجسم افراد معمولی یا بیماران مشابه در پایگاه داده پزشکی است. الگوریتم پیشنهادی به منظور جلوگیری از تعمیم افراطی مبتنی بر نمونه است. این نیز یک روش شبیه سازی استدلال پزشکان مواجهه با بیماران جدید است: آنها افراد جدید را با موارد مواجه شده گذشته مطابقت می کنند. مزیت دیگر این روش سهولت است که کاربر می تواند محدودیت ها در قواعد ارتباط بین عناصر را بیان کند و آن را یک ابزار قابل تنظیم سازد. جزء اصلی روش ما، یک الگوریتم انتخابی است که افراد نمره ای برای هر عضو از مجموعه داده اختصاص می دهند تا بتوانند معرف ترین افراد در سراسر جمعیت را تعیین کنند (این مفهوم نزدیک به شکل های اولیه دسته است، اما به هیچ گام اولیه خوشه بندی در مجموعه داده نیاز ندارد). با استفاده از توابع مختلف برای محاسبه نمره هر فرد، ساختار خروجی به راحتی می تواند تغییر کند.

این بخش به بخش های زیر تقسیم می شود: ابتدا برخی از زمینه های مربوط به کار ما، مشکلاتی که با آنها برخورد می کنیم و داده هایی که ما برای نشان دادن این مقاله استفاده می کنیم را به اشتراک می گذاریم. بخش دوم تمرکز بر مفاهیم و الگوریتم هایی است که داده ها را ارزشمند می کنند و پیشنهادها را به وجود می آورند. بخش زیر جزئیات دقیق ابزار تجسم و اکتشافی را که ما با استفاده از الگوریتم های فوق الذکر انجام دادیم شرح می دهد. این مقاله با بحث و ارائه آثار مربوط به آینده به پایان می رسد.

2. مسئله و مجموعه داده

همانطور که قبلاً ذکر کردیم، یادگیری عمیق و روشهای سنتی معیار ما را به طور کامل برآورده نمی کند. اولاً پیش بینی وقوع یک بیماری لزوماً به شناسایی منشا آن کمک نمی کند. با تطبیق با بیمار مطالعه شده با سایر افراد مشابه و خویشاوند نزدیک، برای یک پزشک ترسیم سیر تکاملی وضعیت زندگی بیمار، ساده تر می شود. علاوه بر این، هدف ما این است که تا آنجا که ممکن است نزدیک به راهی که پزشکان پرونده های پزشکی را تحلیل می کنند باشیم: آنها به تجربه خود برای ارتباط بیماران مشابه متکی هستند و دانش خود در مورد این بیماران برای درمان یک فرد جدید استفاده می کنند. هدف از روش پیشنهادی این است که به پزشکان کمک کند تا دانش خود را به طور خودکار با بیماران مشابه پیشنهادی افزایش دهند و اجازه دهد همه اطلاعات جمع آوری شده در مورد آن بیماران مشابه را تجسم نمایند. هدف ما استفاده از روش کادر پزشکی است که سوابق

بیماران را تجزیه و تحلیل می کند و آنها را به بیمارستان هایی که مجموعه های بزرگ بالینی جمع آوری می کنند میفرستند. برای انجام این کار، ابزارهایی که ایجاد می کنیم، باید تمایل به ظهور فرضیه های جدید و استخراج منشا و ارتباطات و الگوهای مرتبط داشته باشند. مجموعه داده ای که در حال کار بر روی آن هستیم از پرونده پزشکی الکترونیکی به اشتراک گذاشته شده بیماران دیابتی استخراج می شود. این اطلاعات گمنام از پایگاه داده یک شبکه بهداشت منطقه ای می آید: *CARéDIAB*. با استفاده از این سوابق، مستلزم تعداد زیادی پیش پردازش، از قبیل ادغام پرونده های اضافی، ارزیابی مقادیر از دست رفته و اصلاح مقادیر نادرست است. علاوه بر این، به دلیل تمرکز مطالعه بر عوارض مربوط به دیابت، مجموعه داده ها نیاز به فیلتر شدن و سرپرستی دارد. پس از انجام این کار، مجموعه داده های حاصل شامل اطلاعات در مورد 28073 روش پزشکی مرتبط با 93 آزمایش پزشکی 1437 بیمار است.

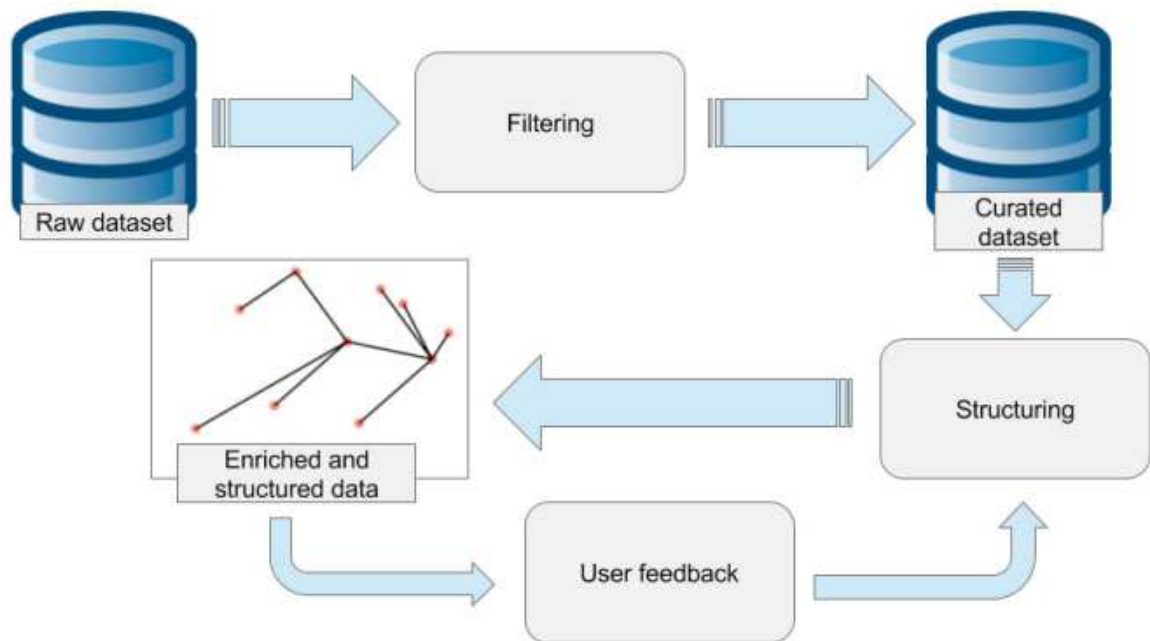
3. بررسی داده ها از طریق تجسم و نظریه

بر اساس نظریه انتخاب اجتماعی، الگوریتم پیشنهادی ترکیبی از اولویت های فردی در "نمره معرف" داده شده به هر نهاد است. سپس بیماران محله خود را بررسی و همسایه ای با بالاترین امتیاز را به عنوان معرف خود را انتخاب خواهند کرد. بنابراین افراد متعدد با معرف یکسان، بیماران خویشاوند نزدیک خواهند بود. مفهوم معرف همانطور که در این مقاله استفاده می شود، در یک مقاله قبل بیان شده است. به منظور نزدیک بودن به چگونگی استدلال پزشکان، روش ارائه شده در این بخش هیچ فرضیه ای در مورد تعداد طبقاتی که در آن بیماران تقسیم شده اند، ارائه نمی دهد و بیماران را بر اساس فاصله اقلیدسی خود در رابطه با بقیه مجموعه داده ها تنبیه یا محروم نمی کند.

1.3. گردش کار

به منظور کمک به پزشکان در تشخیص بیماران خود، هدف ما ارائه کامل یک چارچوب برای یک سیستم پیشنهادی گرایش دار به سمت داده های بالینی است. گردش کار این ابزار در شکل 1 توضیح داده شده است. این مقاله بر روی گام ساختاری تمرکز دارد و الگوریتم در زیر بخش بعدی، همراه با نمونه های مورد استفاده از مجموعه داده های مصنوعی و واقعی، به تفصیل شرح داده می شود.

مرحله فیلترینگ فرایند در یک کار قبل طی شد. اغلب فیلترینگ متشکل از همگن سازی داده های ورودی به یک ساختار مشابه برای یک پایگاه داده ارتباطی است. برای رسیدن به این هدف، چندین عکس از وضعیت سلامت بیمار در چند سال گذشته جمع آوری شده و در مجموعه های زمانی مختلف با هم ادغام می شوند.



شکل 1: گردش کاری مراحل مختلف شامل نمونه اولیه سیستم پیشنهادی ما.

البته این قدم، بسته به چگونه داده های ورودی ساختار یافته است، متفاوت خواهد بود، اما هدف آن ذخیره دانش در یک فرم عملی است. توجه به بازخورد کاربر نسبت به گزارش، در حالی که هنوز در نمونه اولیه ما پیاده سازی نشده است، یک عنصر کلیدی برای فرآیند است. اگر کاربران می توانند ارتباط بین بیماران را به عنوان "معنی دار" یا "نادرست" ارزیابی کنند، این می تواند برای هدایت فرآیند ساختار به منظور ارائه پیشنهادهای بهتر استفاده شود. بازخورد می تواند ساختار بندی را به روش های مختلفی، مانند وزن کردن هر میزان برای دادن اهمیت بیشتر به معیار خاص، یا استفاده از رای مخالف دادن برای مشخص کردن ابعاد خاصی که نباید در هنگام تطبیق افراد مورد توجه قرار گیرد، تحت تأثیر قرار دهد. پیاده سازی بازخورد کاربر در رابطه با پیشنهادهای گفته شده به کاربران اجازه می دهد تا سیستم را آموزش دهند و به نتایج و دقیق تر بهتر و بهبود یافته تری هدایت کنند.

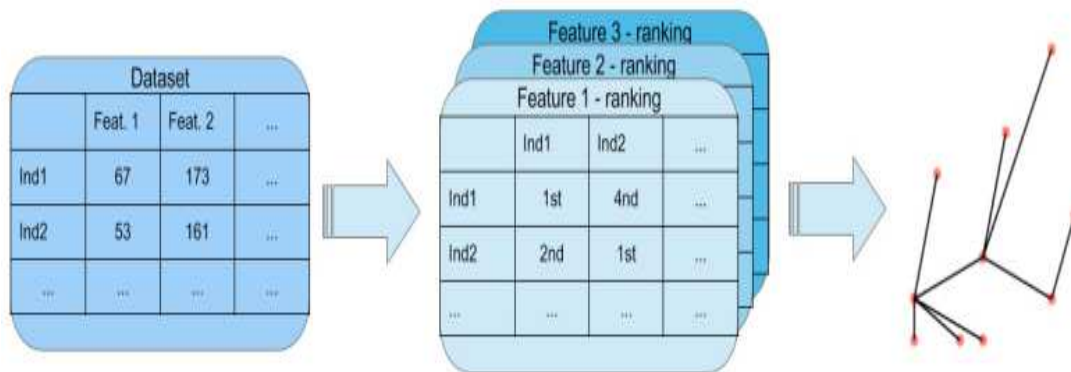
2.3. الگوریتم

تمام مجموعه داده ها براساس درجه معرف (DOR) ساختار یافته است. درجه معرف به عنوان یک مقیاس اهمیت برای هر چیز در مجموعه عمل می کند و تعیین می کند تا چه میزان این چیز می تواند یک نمونه باشد. درجه معرف با جمع کردن نمرات داده شده توسط هر عضو از مجموعه داده، در هر بعد محاسبه می شود. الگوریتم دقیق زیر به وسیله شکل 2 نشان داده شده است.

به Ω اجازه دهید یک مجموعه از n عنصر در یک فضای چند بعدی باشد. این n عنصر، شی نامیده می شوند، و شامل ویژگی های کیفی یا کمی D هستند. اجازه دهید که صفات مورد نیاز برای محاسبه درجه معرف هر شی را تعریف کنیم. برای هر یک از ابعاد D ، ما یک ماتریس متمایز برای بعد داده شده را محاسبه میکنیم. سپس هر شی با توجه به شباهت خود هر هدف را به ترتیب می کند، که نتیجه آن یک ماتریس سلسله ای است. این کار برای هر یک از ابعاد D انجام می شود. اجازه دهید این صف ها را به نمرات تبدیل کنیم. اجازه دهید X یک هدف از Ω باشد: X یک نمره نسبی S_{CXD} را به همه اهداف Ω برای ابعاد D اختصاص می دهد. نمره S_{CXD} نسبت به X می تواند هر تابع دلخواه باشد، اما در این مقاله آن تعریف خواهد شد توسط:

$$\forall y \in \Omega, S_{CXD}(y) = n - Rk_{xD}(y) \quad (1)$$

که $Rk_{xD}(y)$ درجه ای از یک هدف y نسبت به X در ابعاد D است. محاسبه تمام نمرات نسبی، هر هدف X نمرات $n * D$ دریافت می کند، که مربوط به رای همه n هدف از Ω در تمام ابعاد D است. تمام این نمرات نسبی $n * D$ برای تعریف درجه معرف X جمع می شوند. نمره کل به صورت زیر تعریف می شود:



شکل 2: داده های ساختاری بر اساس اولویت های فردی، بر اساس درجه معرف

$$DoR : \Omega \rightarrow \mathbb{R}^+ \quad (2)$$

$$x \mapsto \text{Aggreg}_{y \in \Omega} (S c_{yD}(x))$$

در این مقاله، تابع تراکم، تابع مجموع است. اجازه دهید K را به عنوان عامل مقیاسی که بتواند تعداد نمونه ها را کنترل کند، تعریف کنیم. با استفاده از k مجاورت بر اساس فاصله اقلیدسی، هر هدف با همسایه ای که دارای بالاترین درجه معرف است ارتباط دارد. با $N_k(x)$ مجموعه k نزدیک ترین اهداف از x ، لینک ها تعریف شده اند با :

$$\forall x \in \Omega, \quad x \mapsto y = \text{argmax}_{z \in N_k(x)} DoR(z) \quad (3)$$

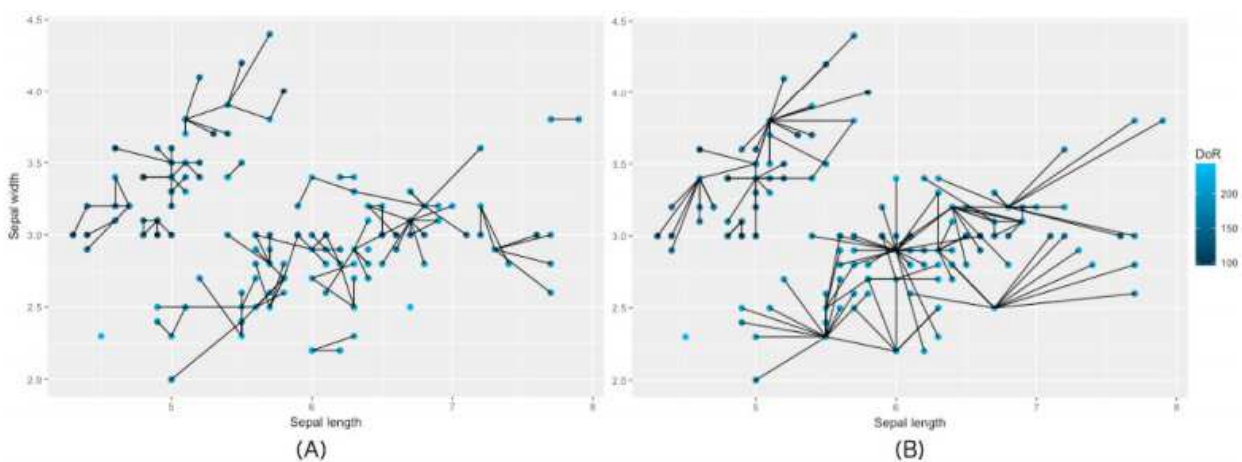
هر شی به معرف ترین هدف در مجاورت خود مرتبط است، و عامل مقیاس k اندازه مجاورت مورد نظر را تعیین می کند، بنابراین چند نماینده بسیار معرف (با مقدار بالا برای k) و یا نمونه های زیادی که به صورت نزدیک اهدافشان هماهنگ می شوند (با مقدار کم برای k)، ایجاد می شود.

برای نشان دادن الگوریتم فوق الذکر، می توانیم از Iris، یک مجموعه داده شناخته شده استفاده کنیم. شکل 3 (a) ساختار مجموعه داده ها را برای یک عامل مقیاس $k = 5$ نشان می دهد که نمونه های بسیاری را به همراه می آورد. هر یک از نمونه ها بسیار شبیه به افرادی است که نشان می دهد: این پیکربندی است که در سیستم پیشنهادی استفاده خواهد شد. همانطور که می توانیم در شکل 3 (b) ببینیم، تعداد نمونه ها با افزایش فاکتور مقیاس ($k = 15$) کاهش می یابد. این پیکربندی برای خلاصه کردن یک جمعیت به فقط تعداد انگشت شماری از افراد کلیدی قابل استفاده است و هنگام کاوش در یک مجموعه داده ناشناخته ثابت می کند که چیز با ارزشی است. برای یک عامل مقیاس داده شده، بعضی از کدها به عنوان یک نمونه به یکدیگر پیوند دارند: زمانی که فاکتور مقیاس افزایش می یابد آنها کمتر می شوند، و در نهایت تنها یک کد وابسته به خود برای به یک عامل مقیاس، برابر با اندازه جمعیت ($k = n$) می شود. در حالیکه فاکتور مقیاس از 1 به n افزایش می یابد، تکرارهای بیشتری که فرد به خودش پیوند می دهد، غیر معمول تر است. هنگامی که مجموعه داده های CARÉDIAB در متن یک سیستم پیشنهادی درخواست می شود، روش ما ساختار نشان داده شده در شکل 4 را برای بدست آوردن این نتیجه اراده می دهد، برای هر بیمار، اطلاعات مربوط به عوارض پزشکی مرتبط با دیابت و سری زمانی

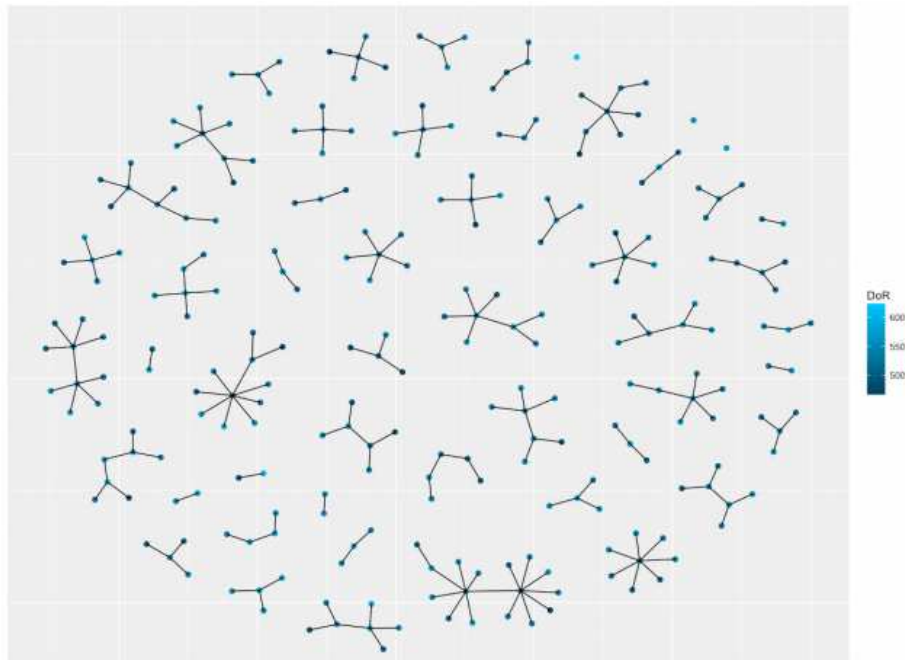
نشان دهنده تکامل سطح هموگلوبین گلیکوز شده (HbA1c) را استخراج کردیم. سپس ده شاخص کمی از این سری زمانی ایجاد کردیم: مجموعه داده حاصل شامل 256 فرد و 10 ویژگی است.

4. برنامه کاربردی

به یک ابزار تجسم ما یک ابزار نمونه اولیه را برای اجرای مراحل توصیف شده در شکل 1 طراحی کردیم. ما در حال حاضر می توانیم هر دو فیلتر سازی و ساختار یک مجموعه داده را به صورت خودکار انجام دهیم، که منجر به یک سیستم توصیه کننده پزشکی می شود. نقطه عطف بعدی ما جمع آوری بازخورد کاربر برای هر پیشنهاد پیشنهاد شده است و به طور خودکار از آن برای اصلاح نمودار شباهت استفاده می شود که منجر به توافق مناسب تر و قابل تنظیم بین اشخاص می شود. نمونه اولیه نهایی برای آزمایش در یک محیط بالینی به کار گرفته می شود. هنگامی که توسط پزشکان و کارشناسان استفاده می شود، استفاده از ابزار و کیفیت نتایج به منظور بهبود الگوریتم و بهینه سازی نمونه اولیه مورد بررسی قرار می گیرد.



شکل 3



شکل 4



شکل 5

5. نتیجه گیری و کارهای آتی

نظریه انتخاب اجتماعی می تواند به عنوان یک راه موثر برای غنی سازی داده ها، به منظور ایجاد یک سیستم توصیه شود. الگوریتم استفاده شده علاوه بر مقیاس پذیر بودن برای بیماران با توجه به شباهت آنها ساده و موثر است. این یک چارچوب شفاف و قابل فهم برای کارشناسان پزشکی برای دستکاری و کشف پرونده های پزشکی فراهم می کند و ثابت می کند که یک ابزار قابل تنظیم است: هر کاربر می تواند بر ویژگی های خاص و معیارهای تشابه که در طول مرحله ساختاری مورد توجه قرار گیرد، تاکید کند. ما یک نمونه اولیه را ارائه دادیم

که در حال حاضر توسط پزشکان آزمایش و بررسی می شود. این برنامه وب با استفاده از الگوریتم فوق به عنوان راهی برای مرتبط کردن بیماران مرتبط با یکدیگر، ارائه پرونده های پزشکی مشابه به کاربر، در نتیجه ایجاد چارچوبی برای دستکاری و کشف مجموعه های داده های بزرگ با هزاران نفر از افراد است. آثار آینده راه را به طور خودکار جمع آوری بازخورد کاربر در مورد توصیه ها به تدریج بهبود وزن ویژگی ها، تصفیه توصیه ها به عنوان کاربر دستکاری مجموعه داده ها را در نظر بگیرند. برای ایجاد یک تجربه یکپارچه، کاربران با استفاده از دکمه های "مانند" و "دوست نداشتن" برای هر یک از توصیه ها به روش بی قاعده درخواست خواهند شد. بازخورد آنها به عنوان یک ورودی برای الگوریتم برای انطباق نفوذ هر ویژگی در طی مرحله ساختار استفاده می شود.



این مقاله، از سری مقالات ترجمه شده رایگان سایت ترجمه فا میباشد که با فرمت PDF در اختیار شما عزیزان قرار گرفته است. در صورت تمایل میتوانید با کلیک بر روی دکمه های زیر از سایر مقالات نیز استفاده نمایید:

لیست مقالات ترجمه شده ✓

لیست مقالات ترجمه شده رایگان ✓

لیست جدیدترین مقالات انگلیسی ISI ✓

سایت ترجمه فا ؛ مرجع جدیدترین مقالات ترجمه شده از نشریات معتبر خارجی