تابستان ۱۴۰۰، دوره ۶، شماره ۲، پیاپی ۲۱

DOR :20.1001.1.2538161.1400.6.2.1.1

دریافت مقاله: ۱۳۹۸/۱۲/۲۴

پذیرش مقاله: ۱۳۹۹/۰۷/۱۸

نوع مقاله: پژوهشی

توسعه سامانه‌های توصیه‌گر مشارکت‌محور با استفاده از شبکه توابع پایه شعاعی

مریم محمدی*

استادیار دانشکدهٔ علوم ریاضی و کامپیوتر، دانشگاه خوارزمی، تهران

پست الکترونیکی: m.mohammadi@khu.ac.ir

محمدعلی ناصری

دانشجوی کارشناسی ارشد دانشکدهٔ علوم ریاضی و کامپیوتر، دانشگاه خوارزمی، تهران

پست الکترونیکی: ma.naseri.uk@gmail.com

چکیده

حجم فراوان و رو به رشد اطلاعات بر روی وب و اینترنت، فرایند تصمیم‌گیری و انتخاب اطلاعات، داده یا محصولات مورد نیاز را برای بسیاری از کاربران وب دشوار کرده است. سامانه‌های توصیه‌گر با پیشنهادهایی متناسب با سلیقه کاربران، آن‌ها را در انتخاب محصولات مورد علاقه یاری می‌دهند. سامانه‌های توصیه‌گر مشارکت‌محور، دقیق‌ترین توصیه‌ها را با بررسی و کاوش کاربران قبلی از بانک اطلاعاتی، به کاربران جدید ارائه می‌کنند به‌طوری که با علایق آن‌ها مطابقت داشته باشد. چالش‌های بسیار چشم‌گیری که در سامانه‌های مشارکت‌محور وجود دارد عبارت هستند از: مقیاس‌پذیری، تنک‌بودن (خلوتی) و شروع‌سرد. در این مقاله به ارائه یک سامانه توصیه‌گر مشارکت‌محور با استفاده از شبکه توابع پایه شعاعی می‌پردازیم. سامانه پیشنهادی دارای دو مرحله برون‌خط و برخط می‌باشد. در مرحله برون‌خط، به تخمین مقادیر صفر یا امتیازات داده نشده کاربران در ماتریس امتیازات کاربران به محصولات، با استفاده از شبکه توابع پایه شعاعی می‌پردازیم. در این مرحله برای تعیین مراکز توابع پایه شعاعی از الگوریتم خوشه‌بندی کی-میانگین و برای یافتن وزن‌های آن‌ها از روش کاهش گرادیان استفاده می‌کنیم. در مرحله برخط، ابتدا کاربر فعالی وارد سامانه می‌گردد. سپس به کمک ماتریس کامل به‌دست آمده از مرحله برون‌خط، کاربران را، از طریق محاسبه مقدار تابع تشابه پیرسون بین کاربر فعال و سایر کاربران به دو خوشهٔ مثبت و منفی تقسیم می‌کنیم و با رویکرد همسایه نزدیک، پیشنهادهایی را به کاربر فعال وارد شده به سامانه بر اساس علایق او ارائه می‌دهیم. در این مقاله، مشکل مقیاس‌پذیری را با استفاده از خوشه‌بندی کاربران، مشکل شروع سرد را با پیشنهاد محصولات با بالاترین امتیاز و مشکل تنک بودن را با هموار نمودن ماتریس امتیازات به کمک شبکه توابع پایه شعاعی رفع می‌نماییم. نتایج آزمایش‌ها بر روی داده‌های MovieLens 100K بیانگر بالابودن دقت و کیفیت پیشنهادهای ارائه شده است.

واژه‌های کلیدی: سامانه توصیه‌گر، شبکه تابع پایه شعاعی، مشکل شروع سرد، مقیاس‌پذیری، ماتریس تنک.

* نویسندهٔ مسئول

۱ - مقدمه

طی سال‌های اخیر با بالا رفتن تعداد کاربران در اینترنت، پیشرفت‌های به‌وجود آمده در جمع‌آوری و ذخیره‌سازی داده‌ها و ایجاد بسترهای مناسب و وسیعی برای تجارت الکترونیکی از جمله: فروشگاه‌های اینترنتی، موتورهای جستجو برای یافتن کالاهای مورد نیاز و غیره رقابتی بین شرکت‌های مختلف در این حوزه ایجاد شده است. زمانی‌که کاربر وارد یک سامانهٔ فروش اینترنتی می‌شود، با توجه به سلیقه و نیاز خود به دنبال کالای مورد نیازش است. سامانهٔ توصیه‌گر با توجه به علایق و نظر کاربر فعال و همچنین با استفاده از بانک اطلاعاتی خود از امتیازات و نظرات کاربران گذشته تعدادی از محصولات را به کاربر توصیه می‌کند. کاربرد سامانه‌های توصیه‌گر در فیلم، موسیقی، اخبار، کتاب‌ها، مقالات، جستجوی پرسش‌ها و غیره می‌باشد که در پژوهش‌های متخصصان، شرکت‌ها، رستوران‌ها، خدمات مالی، بیمه عمر، شبکه‌های مجازی و غیره استفاده می‌شود. برای جمع‌آوری داده‌ها در بانک اطلاعاتی، راه‌های مختلفی وجود دارد، از جمله: درخواست از کاربر برای امتیازدهی به محصولات مورد علاقه‌اش، نمایش محصولات مختلف به کاربر برای مقایسه، تحلیل محصولات مشاهده شده، تحلیل شبکه‌های اجتماعی و غیره [٤, ١٠, ٢٠, ٢٤, ٢٨]. این نوع سامانه‌ها به سه دستهٔ مشارکت‌محور، محتوامحور و ترکیبی تقسیم می‌شوند:

• سامانه‌های توصیه‌گر مشارکت‌محور بر اساس سوابق و فعالیت‌های کاربر و بررسی این رفتارها با کاربران مشابه، سعی می‌کنند تا لیستی از محصولات مورد علاقه کاربر را به او توصیه کند. در این روش محتوای کالا اهمیتی ندارد و سامانه بر اساس سابقهٔ همه کاربران و انتخاب‌های خود کاربر، به او توصیه‌های جدیدی را ارائه می‌کند. این روش بر سلیقه‌ها و امتیازات مشترک بنا نهاده شده است.

• سامانه‌های توصیه‌گر محتوامحور بر اساس امتیازات موجود در محصولات انتخاب شده یا مورد علاقهٔ کاربر فعال توصیه‌هایی ارائه می‌کند. تکنیک‌هایی که در این سامانه استفاده می‌شوند، سعی بر پیشنهاد توصیه‌هایی دارند که شبیه به محصولاتی باشد که کاربر پیش از این امتیاز بالایی به آن‌ها داده یا جزو علایق او باشد. محصولات مورد بررسی با محصولاتی که قبلاً توسط کاربر امتیازبندی شده بودند، مورد مقایسه قرار گرفته و محصولاتی که بیشترین شباهت‌ها را با معیارهای کاربر داشته باشند، پیشنهاد می‌گردند.

• سامانه‌های توصیه‌گر ترکیبی که از ترکیب تکنیک‌های مختلف جهت ارائهٔ توصیه به‌دست می‌آید، به منظور افزایش کارایی و کاهش ضعف‌های سامانه‌های دیگر طراحی شده است. نتفلیکس از این نوع سامانه برای ارائه توصیه به کاربران استفاده می‌کند.

در سامانه‌های مشارکت‌محور چالش‌هایی مانند مقیاس‌پذیری، تنک‌بودن(خلوتی) و شروع سرد مطرح هستند. دو رویکرد در سامانه‌های توصیه‌گر مشارکت‌محور وجود دارد. رویکرد اول به این صورت است که پیش‌بینی و توصیه بر اساس کاربران قبل در بانک اطلاعاتی صورت بگیرد و کاربر فعال با همه کاربران دیگر مقایسه شود. در واقعیت تعداد کاربران بسیار زیاد است و پیش‌رفتن با این رویکرد موجب مقیاس‌پذیری می‌شود. به این رویکرد، سامانهٔ توصیه‌گر مشارکت‌محور مبتنی بر حافظه می‌گویند. اساس کار رویکرد دوم خوشه‌بندی است. به این صورت که سامانه مدلی را بر اساس کاربران قبل طراحی کرده و سپس فرآیند پیش‌بینی توسط این مدل صورت می‌گیرد. این رویکرد، سامانهٔ توصیه‌گر مشارکت‌محور مبتنی بر مدل نام دارد. یکی از چالش‌هایی که بسیار مورد توجه است تنک بودن بانک اطلاعاتی است. بعضی از کاربران به دلایل مختلفی از جمله بی‌حوصلگی، از ارائهٔ امتیاز به محصولات مختلف سر باز می‌زنند که باعث ایجاد یک ماتریس تنک می‌شود[١٦, ١٧]. تکنیک‌های بسیاری برای غلبه بر مشکل تنک بودن ماتریس امتیاز ارائه شده است. از جمله آن‌ها،

رویکردی بود که سروار [۲۳] [1] و بیلسس [٤] [2] با استفاده از تجزیه مقدار منفرد، ماتریس تنک را به ماتریسی کامل با ابعاد پایین تبدیل می‌نمودند. این رویکرد در ماتریس امتیاز با ابعاد پایین، اطلاعات ارزشمندی را از دست می‌دهد. رویکرد دیگر پر کردن امتیازات خالی با استفاده از خوشه‌بندی کاربران مختلف و پالایش گروهی بود که توسط چانگ [٦] [3] بیان شد. همین رویکرد با این تفاوت که به جای کاربران، محصولات را خوشه‌بندی کنیم مورد توجه سروار قرار گرفت [۲۲]. مقاله [۲] با مدل‌سازی روابط اعتماد به صورت محلی یا سراسری به پیش‌بینی ترجیحات یک کاربر در سامانه‌های توصیه‌گر پرداخته است.

هدف از نگارش این مقاله، در وهلهٔ اول ایجاد یک ماتریس کامل به جای ماتریس تنک اولیه برای بانک اطلاعاتی با استفاده از شبکهٔ توابع پایه شعاعی و خوشه‌بندی کاربران و در وهلهٔ دوم ارائهٔ سامانهٔ توصیه‌گر مشارکت محور با استفاده از ماتریس کامل ایجاد شده برای کاربران، با توجه به امتیازات و علایقشان است. ابتدا در بخش ۲، پیش‌نیازها و مفاهیم مقدماتی را مرور می‌کنیم. سپس در بخش ۳ به معرفی و دسته‌بندی انواع سامانه‌های توصیه‌گر می‌پردازیم. سامانه توصیه‌گر پیشنهادی به همراه الگوریتم‌های مراحل برون‌خط و برخط را در بخش ٤ بیان می‌کنیم و در نهایت با ارائه نتایج و آزمایش‌های عددی به بررسی کارایی، دقت و کیفیت سامانه پیشنهادی می‌پردازیم.

۲- پیش‌نیازها

۱,۲ توابع پایه شعاعی [۳, ۹, ۲۷]

تعریف: تابع $\Phi: \mathbb{R}^d \to \mathbb{R}$ یک تابع پایه شعاعی است اگر تابع یک متغیره $\varphi: [0,\infty] \to \mathbb{R}$ موجود باشد به طوری که

$\Phi(x) = \varphi(r), r = ||x||$ که در آن، $||\cdot||$ نرم اقلیدوسی روی $\mathbb{R}^d$ است.

جدول ۱: انواع توابع پایه شعاعی

شرط	ضابطه	نوع تابع
ندارد	$\varphi(r) = r$	خطی (همانی)
ندارد	$\varphi(r,\sigma) = \exp\left(\frac{-r^2}{2\sigma^2}\right)$	گاوسین
$\beta \in \mathbb{R}_{\neq 0} \backslash 2\mathbb{N}$	$\varphi(r) = \left(1 + \frac{r^2}{\sigma^2}\right)^{\frac{\beta}{2}}$	چندربعی
$0 < \beta \neq 2N$	$\varphi(r) = r^\beta$	توانی
$n \in \mathbb{N}$	$\varphi(r) = r^{2n} \mathrm{Ln}(r)$	صفحه نازک اسپلاین

برخی از انواع مهم توابع پایه شعاعی در جدول ۱ آورده شده‌اند.

Φ در $\mathbb{R}^d$ را می‌توان به وسیلهٔ یک عدد ثابت مثبت مانند σ مقیاس نمود که به σ، پارامتر شکل می‌گویند. این عدد معمولا با استفاده از روش‌های عددی به گونه‌ای که جواب‌های عددی دقیق‌تر و عدد وضعیت ماتریس کوچکتر شود، تعیین می‌گردد[٥, ۱۱]. در پایه شعاعی گاوسین، σ^2 بیانگر واریانس در توزیع نرمال است.

۲,۲ شبکه توابع پایه شعاعی [۱۶, ۱۷, ۱۹]

روش شبکه توابع پایه شعاعی یک رویکرد شبکه عصبی با قابلیت طراحی در یک فضایی با بعد بالاست. فرایند آموزش، معادل یافتن یک تابع مناسب چندبعدی است. شبکه توابع پایه شعاعی از ۳ لایه مجزا تشکیل شده است: لایه ورودی که مجموعه‌ای از گره‌های ابتدایی است؛ لایه دوم که لایه‌ای پنهان با ابعاد بالا است و لایه خروجی که پاسخ شبکه را روی الگوهای فعال‌سازی به کار رفته در لایه ورودی ارائه می‌دهد. تبدیل فضای ورودی به فضای واحد پنهان، دارای رابطه‌ای غیرخطی است. قضیهٔ کاور توجیه ریاضی انتخاب این توابع را وجود فضای پنهان با بعد بالا در شبکهٔ توابع پایه شعاعی بیان کرده است. خروجی حاصل از مدل‌سازی عددی توسط شبکهٔ تابع پایه شعاعی، از رابطه زیر به دست می‌آید:

$$y = \sum_{j=1}^{K} w_j \phi(||x - c_j||), \qquad (۱)$$

که در آن c_jها به مراکز و w_jها به وزن‌های شبکه توابع پایه شعاعی معروف هستند.

1- Sarwar
2- Billsus
3- Cheung

در ایـــن مقاله از روش‌های خوشـــه‌بندی برای انتخاب مراکز، اســـتفاده کرده‌ایـــم. در حین آماده‌ســـازی داده‌ها، بردارهای امتیاز (برحســـب محصـــول) را وارد الگوریتم خوشه‌بندی مورد نظر می‌کنیم. خروجی الگوریتم علاوه بر خوشه‌های کاربران، مراکز خوشه برای هر داده را نیز به ما می‌دهد. زمانی که بردارهای امتیاز را وارد لایهٔ ورودی شبکه می‌کنیم، الگوریتم، در لایهٔ پنهان با توجه به نوع تابع پایه شـــعاعی و مراکزی که هر داده در خوشـــهٔ مورد نظر خود دارد

$$\phi(||x_i - c_k||),\ i=1,2,\cdots,n.$$

را محاســـبه می‌کند. متغیر k بین عـــدد یک تا حداکثر تعداد خوشـــه‌ها، برای مجموعه داده‌ها، می‌باشـــد. روش اعمال شده در این مقاله، برای انتخاب وزن‌ها تکنیک کاهش گرادیان می‌باشد.

۳,۲ تکنیک کاهش گرادیان [۱۵, ۱۸]

کاهش گرادیان یک تکنیک بهینه‌سازی برای پیدا کردن کمینهٔ یک تابع اســـت. در این روش، با هدف مینیمم کردن f، بایـــد جهت u را بـــه گونه‌ای بیابیم کـــه f در آن جهت، بیشترین کاهش را داشته باشد. کافیست مسئلهٔ بهینه‌سازی زیر را حل کنیم:

$$\min_u u^T\nabla f(x_1,x_2,\cdots,x_n) = \min_u ||\nabla f(x_1,x_2,\cdots,x_n)||_2\cos(\theta) \quad (۳)$$
$$= \nabla f(x_1,x_2,\cdots,x_n)\min_u \cos(\theta).$$

که در آن، θ زاویهٔ بین بردار گرادیان و بردار u است.

در ادامـــه به یافتن وزن‌های w_j حاصل از شـــبکه تابع پایه شعاعی در رابطه (۱) می‌پردازیم. فرض کنید y جواب تقریبی حاصل از شـــبکه توابع پایه شعاعی و $\tilde{y}$ نیز جواب دقیق این شـــبکه به ازای مجموعـــه‌ای از داده‌های از پیش تعیین شده است. هدف ما یافتن وزن‌های w_j با استفاده از تکنیک کاهش گرادیان اســـت به گونه‌ای که تابع هزینه زیر مینیمم گردد:

$$E = \frac{1}{2}\sum(\tilde{y}-y)^2.$$

بنابراین

$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial y}\times\frac{\partial y}{\partial w_i} = -(\tilde{y}-y)\phi_i.$$

حال با توجـــه به روش کاهش گرادیان، مینیمم زمانی حاصل می‌گردد که در خلاف جهت گرادیان حرکتی کنیم یا به عبارتی $\cos(\theta) = -1$. بنابراین با انتخاب یک بردار وزن اولیه با اســـتفاده از رابطه تکراری زیر، وزن‌های W_j را به صورت زیر به‌روزرسانی می‌کنیم.

$$w_j(t+1) = w_j(t) + \eta(\tilde{y}-y)\phi_j,$$

که در آن η نرخ یادگیری است.

۴,۲ خوشه‌بندی کی-میانگین [۱,۸,۱۲]

خوشـــه‌بندی کی-میانگیـــن، یکی از انـــواع روش‌های خوشـــه‌بندی برای گروه‌بنـــدی مجموعـــه‌ای از داده‌ها و قرار دادن آن‌ها در طبقاتی از نمونه‌های مشـــابه اســـت. به منظـــور ارزیابی تشـــابه، نیاز به یک معیار تشـــابه داریم. معیارهای ســـنجش شباهت بســـیار متعدد هستند و نحوه انتخـــاب و محاســـبهٔ آن‌ها در خوشـــه‌بندی بســـیار مهم است.

دو نمونه از انواع توابع تشابه عبارت هستند از:

تابع تشابه همبســـتگی پیرسون، $Psim: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$، برای دو بردار $X, Y \in \mathbb{R}^n$ به صورت زیر تعریف می‌شود:

$$Psim\ (X,Y) = \frac{(X-\bar{X},Y-\bar{Y})}{||X-\bar{X}||_2||Y-\bar{Y}||_2} = \frac{\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^n (x_i-\bar{x})^2}\sqrt{\sum_{i=1}^n (y_i-\bar{y})^2}}$$

تعریف: تابع تشابه کسینوسی، بین دو بردار غیر صفر $X,Y \in \mathbb{R}^n$ را به صورت زیر بیان می‌کنیم.

$$Csim\ (X,Y) = \frac{\langle X,Y\rangle}{||X||_2||Y||_2}$$

در خوشه‌بندی کی-میانگین، هدف افراز n نقطه d بعدی زیر

$$X = \{x_1,x_2,\ldots,x_n\},$$

به تعداد $K \le n$ خوشه با مراکز

$$C = \{c_1,c_2,\ldots,c_k\},$$

اســـت. الگوریتم کی-میانگین به دنبـــال یافتن افرازی اســـت که تابع هدف زیر که همان مربعات خطا میان مرکز هر خوشـــه (همان میانگین داده‌های خوشه) و تمامی نقاط آن خوشه است، مینیمم گردد.

$$J(C) = \sum_{i=1}^{n} \sum_{k=1}^{K} \| x_i - c_k \|^2 .$$

شبه کد الگوریتم کی-میانگین در مرجع [۱] ارائه شده است.

۳- سامانه‌های توصیه‌گر

برای نگارش این بخش، از مراجع [۸, ۱۶, ۱۷, ۲۶, ۲۷] استفاده شده است. تعاریف متفاوتی برای سامانه‌های توصیه‌گر ارائه شده است. از جملهٔ آن‌ها، تعریف لیانگ [۲۴] می‌باشد که سامانه‌های توصیه‌گر را زیرمجموعه‌ای از سامانه پشتیبانی تصمیم[4] می‌داند و آن‌ها را سامانهٔ اطلاعاتی[5] نامگذاری می‌کند. سامانه‌های توصیه‌گر اساسا به این منظور ایجاد می‌شوند تا در فرایند جستجو یا انتخاب، گزینه‌ای را به کاربر پیشنهاد دهند و بدین وسیله، وی را در انتخاب گزینه بهتر یاری رسانند، به گونه‌ای که این فرایند برای همان کاربر، شخصی‌سازی شده باشد. تکنیک‌هایی که در سامانه‌های توصیه‌گر به‌کار برده می‌شود، در اصل برای پردازش داده‌ها در بانک اطلاعاتی است. تلاش سامانه‌های توصیه‌گر بر این است که با حدس زدن طرز تفکر کاربر، به کمک اطلاعاتی که از نحوه رفتار او با کاربران مشابه وی و نظرات آن‌ها دارد، مناسب‌ترین و نزدیک‌ترین کالا به سلیقه کاربر را شناسایی و پیشنهاد کنند.

۱,۳ اصطلاحات

قبل از این‌که انواع سامانه‌های توصیه‌گر را بیان کنیم، با چند اصطلاح آشنا می‌شویم:

۱. کاربر فعال: در سامانه‌های توصیه‌گر به کاربری که توصیهٔ جاری در سامانه، برای او در حال پردازش و آماده شدن است، کاربر فعال می‌گویند.

۲. ماتریس امتیازات: ماتریسی است که درایه‌های آن همان امتیاز یا میزان رضایت کاربران $\{U_1,\ldots,U_M\}$ به محصولات $\{I_1,\ldots,I_N\}$ است:

$$[R]_{M\times N} = \{r_{ij} | 1 \le i \le M, 1 \le j \le N\}.$$

۳. از فعل مصرف کردن در سامانه‌های توصیه‌گر، زمانی استفاده می‌کنند که کاربر، توصیه ارائه شده را می‌پذیرد. به عبارتی وقتی کاربری پیشنهادی که توسط سامانه به وی شده را می‌پذیرد، می‌گوییم کاربر آن پیشنهاد را مصرف کرده است. این پذیرش می‌تواند به شکل‌های مختلفی باشد، مثلاً کاربر، کتاب پیشنهادی را می‌خرد، وبگاه پیشنهادی را مرور می‌کند یا به شرکت خدماتی که به او پیشنهاد شده مراجعه می‌کند.

۲,۳ سامانه‌های توصیه‌گر مشارکت‌محور

در سامانه‌های توصیه‌گر مشارکت‌محور، بر اساس سوابق و فعالیت‌های کاربر و بررسی این رفتارها با کاربران مشابه، سامانه سعی می‌کند تا لیستی از محصولات مورد علاقه کاربر را به او توصیه کند. به این صورت که به کاربر فعال محصولاتی توصیه خواهد شد که کاربران دیگر با امتیازات و ترجیحات مشابه او این محصولات را پسندیده‌اند. بر اساس رابطهٔ بین کاربران و کالاها، اقلام جدید به کاربر توصیه می‌شود. در این روش، خود کالا اهمیتی ندارد و بر اساس امتیازات کاربران دیگر و همچنین امتیازات گذشته خود کاربر، به او توصیه‌های جدیدی ارائه می‌شود. اساس این روش، سلایق مشترک است. الگوریتم‌های سامانهٔ توصیه‌گر مشارکت‌محور به دو دسته تقسیم‌بندی می‌شوند:

- مبتنی بر حافظه: در این دسته، پیش‌بینی با توجه به امتیازات کاربران قبل صورت می‌گیرد و کاربر فعال با کاربران موجود مقایسه می‌شود. الگوریتم‌های مبتنی بر حافظه از کل بانک اطلاعاتی کاربران برای پیش‌بینی و همچنین از تکنیک‌های آماری برای یافتن مجموعه‌ای از کاربران، معروف به همسایگی‌ها استفاده می‌کنند. پس از تشکیل همسایگی، کاربران خوشه‌بندی می‌شوند. این تکنیک از الگوریتم‌های مختلفی برای اولویت‌بندی همسایه‌ها استفاده می‌کند تا یک پیش‌بینی یا تعدادی از بالاترین توصیه‌ها را به کاربر فعال ارائه کند. این تکنیک‌ها، همچنین به عنوان نزدیکترین همسایه یا مبتنی بر کاربر

4- Decision Support System
5- Information system

شناخته می‌شـوند و در عمل بسـیار مورد استفاده قرار می‌گیرند. مهم‌ترین مشـکلی که در این روش وجود دارد، مقیاس‌پذیری اسـت. به این صورت کـه وقتی حجم بانک اطلاعاتی بالا باشد، استفاده از این روش موجب کند شدن الگوریتم و افزایش زمان پردازش می‌شود.

• مبتنـی بر مـدل: در این رویکرد، طراحـی مدل‌ها با اسـتفاده از داده‌کاوی و الگوریتم‌هـای پیش‌بینی یادگیری ماشین، صورت می‌گیرد. الگوریتم‌های بسیاری در این زمینه وجود دارد. از جمله می‌توان به شـبکه‌های بیزی، مدل‌های خوشه‌بندی، تجزیه مقدار تکین و غیره نام برد. الگوریتم‌های مبتنی بر مدل سـعی دارند بانک اطلاعاتی عظیمی را در یک مدل فشـرده کنند. در این روش‌ها علاوه بر افزایش سرعت پـردازش، زمان انجام محاسـبات نیز کاهـش می‌یابد. این روش‌ها از خطای پایین‌تری در پیش‌بینی برخوردار هستند و معمولا اساس کار آن‌ها خوشه‌بندی است.

سـامانه‌های توصیه‌گـر مشـارکت‌محور بـه عنوان موفق‌ترین روش‌های توصیه شـناخته شـده‌اند. با وجود اسـتفادهٔ گسـترده از این سامانه‌ها، سـه چالش در آن‌ها وجود دارد که عبارت هستند از:

۱. مقیاس‌پذیری: در سامانه‌های مشارکت‌محور، با تکنیک همسایه یا همسایگی نزدیک، همسایگی با رویکرد مبتنی بر حافظه یا مبتنی بر مدل انتخاب می‌شود. در رویکرد مبتنی بر حافظه، کاربر فعال با تمامی کاربران بانک اطلاعاتی مقایسه می‌شـود. با افزایش تعـداد کاربران، پیچیدگی محاسـباتی افزایش می‌یابد. این چالش به عنوان مقیاس‌پذیری تعریف شده اسـت. در رویکرد مبتنی بر مدل، کاربران موجود بر اساس شباهتشان خوشه‌بندی می‌شوند و مقایسه بر روی کاربران درون خوشه ها صورت می‌گیرد. بنابراین در رویکرد مبتنی بر مدل، چالش مقیاس‌پذیری آنچنان تاثیرگذار نیست.

۲. تنک بودن: بعد از ورود کاربر، سـامانه لیسـتی از محصولات را برای امتیازدهی به کاربر فعال ارائه می‌کند. در واقـع، کاربر فعال تعداد بسـیار کمی از محصولات را امتیازدهـی و خریداری می‌کند کـه منجر به ماتریس تنک می‌شـود. درجهٔ تنک‌بـودن این ماتریس، از نسـبت تعداد درایه‌های صفر (امتیازات داده نشـده توسط کاربران) به تعداد درایه‌های کل ماتریس محاسـبه می‌شود. پیامد این چالش، فقدان همسایگی اسـت. در این مقاله با استفاده از شـبکهٔ توابع پایه شـعاعی، ماتریس امتیاز تنک را به یک ماتریس کامل تبدیل خواهیم کرد.

۳. مشکل شروع سرد: این چالش زمانی رخ می‌دهد که کاربر فعال، هیچ امتیازی به محصولات ندهد. ندادن امتیاز دلایل مختلفی دارد از جمله: نداشتن آگاهی از محصولات، تردید در امتیازدهی، بی‌حوصلگی و غیره. یکی از پیامدهای این چالش، ایجاد مشکل در جستجوی همسایگی است.

تذکر ۱: تمرکز این مقاله روی سـامانه‌های توصیه‌گر مشارکت‌محور مبتنی بر مدل است.

۴ – سامانه توصیه‌گر پیشنهادی

در این بخش به معرفی سامانه توصیه‌گر پیشنهادی که یک سامانه توصیه‌گر مشارکت‌محور توسعه‌یافته با استفاده از شـبکه توابع پایه شـعاعی با وزن‌های کاهش گرادیان است، می‌پردازیم. ماتریس امتیازات $R = [r_{ij}]_{M\times N}$ را در نظر بگیرید. که در آن درایهٔ r_{ij} امتیازی اسـت که توسط کاربر i–ام برای کالای j–ام ثبت شده است. هدف رفع چالش‌هایی است که در سـامانه‌های توصیه‌گر مشارکت‌محور وجود دارد. سـامانهٔ توصیه‌گر مورد استفاده در این مقاله به دو مرحلهٔ برون‌خط و برخط بخش‌بندی شده است. در مرحله برون‌خط در صدد پر نمودن ماتریس امتیازات یا به عبارتی حل مشکل تنک‌بودن هستیم. این مرحله به پیش‌بینی تقریبا ضعیفی از درایه‌های صفـر ماتریس R منجر می‌گردد. ما ابتدا برای یافتن ارتباط بین کاربران، از تکنیک خوشه‌بندی کی–میانگین اسـتفاده شده است. در ادامه مراکز خوشه‌ها را به‌عنوان مراکز توابع پایه شعاعی در نظر می‌گیریم و در نهایت ماتریس R را به یـک ماتریس کامل تبدیل می‌کنیم. این تکنیک اصطلاحا هموارسـازی نامیده می‌شود که هم همبستگی بین کاربران را افزایش می‌دهد و هم چالش تنک

بودن ماتریس امتیـــاز R را رفع می‌کند. در مرحله برخط، پـــس از ورود کاربـــری فعال به ســـامانه و دریافت بردار امتیازات او، ســـامانه به پیش‌بینی امتیازات داده نشده وی و همچنین توصیه T محصول بـــه او می‌پردازد. از طرفی به تخمینی دقیق‌تر از ماتریس امتیازات نهایی می‌رسیم. در این مرحله، اگر کاربر فعال، هیچ‌گونه سابقه و امتیازدهی به کالاهای ارائه شده از جانب سامانه را نداشته باشد (مشکل شروع سرد)، ممکن است شباهتی بین خوشه‌های موجود برقرار نشـــود. در این حالت، مدل پیشنهادی، محصولاتی با بالاترین امتیاز داده شـــده توسط کاربران قبل را، از هر خوشه، به عنوان توصیهٔ پیش‌فرض به کاربر ارائه می‌دهد که راه‌حلی برای مشکل شروع سرد به شمار می‌آید. چالش دیگری که پیش رو داریم، چالش مقیاس‌پذیری است. چون محاسبات بر اساس تعداد خوشه‌ها انجام می‌شود، چالش مقیاس‌پذیری تا حدی کاهش می‌یابد.

تذکر ۲: نوع امتیازهای ماتریـــس R در این مقاله، عدد می‌باشـــد. امتیازها، شـــامل اعداد ۱، ۲، ۳، ٤ و ٥ می‌باشند. امتیاز ۱ به مفهوم بســـیار بد و امتیاز ٥ به معنای عالی از جانب کاربر می‌باشد.

تذکر ۳: تمام الگوریتم‌های ارائه شـــده در این مقاله بر اساس دستورات نرم‌افزار MATLAB نگارش شده‌اند.

۱,٤ الگوریتم مرحله برون‌خط

Algorithm Offline
Input: $R = [r_{ij}, 1 \le i \le m, 1 \le j \le n]$: sparse user-item rating matrix.

ϵ: termination tolerance. η: weight learning rate.
Output: $\widehat{R} = [\hat{r}_{ij}, 1 \le i \le m, 1 \le j \le n]$: smooth user-item rating matrix.

1. Set $\widehat{R} = R$.
2. Set range=(max_rating-min_rating)+1.
3. Set number of clusters $\tilde{k} = \min k$ subject to $\left\lfloor \frac{\text{range}}{k} \right\rfloor \le 3$.
4. **For all** $j = 1, \cdots, n$ do
5. Set $X = R(:,j)$, call Algorithm 2.5 and partition the users into $\tilde{k}$ clusters.
6. Set cluster indices $I(:,j) = \mathcal{I}$.
7. Set cluster centers $C(:,j) = \mathcal{C}$.
8. **End for**
9. Set initial weight matrix $[W_0]_{m \times n}$ with randomly entries in the range $0.5 - 1$.
10. **For all** $j = 1, \cdots, n$ do
11. **For all** $i = 1, \cdots, \tilde{k}$ do
12. Set $f =$ find $(I(:,j) == i)$.
13. Set $Q(f,j) = \phi(\|R(f,j) - C(i,j)\|_2)$, where ϕ is one of the RBFs in Table 1.
14. Set $F(i,j) = \langle Q(f,j), W_0(f,j) \rangle$ according to Eq. 1.
15. **End for**
16. **End for**
17. Round entries of F to the nearest integers.
18. Set $F(F < \min_rating) = \min_rating$.
19. Set $F(F > \max_rating) = \max_rating$.
20. Set $[S,T] =$ find $(R == 0)$.
21. Set $\widehat{R}(S,T) = F(I(S,T),T)$.
22. **repeat**
23. Set $W = W_0 + \eta\left((\widehat{R} - R) .* Q\right)$.
24. Set $E = \frac{\|W - W_0\|_1}{\|W\|_1}$.
25. Set $W_0 = W$.
26. **untill** $E \le \epsilon$.
27. Repeat Steps 10-21 with $W_0 = W$.
28. **Return** $\widehat{R}$.

ورودی‌های الگوریتم ۱,٤ به ترتیب عبارت هستند از: R به عنوان ماتریس امتیاز (بانک اطلاعاتی امتیازات کاربران)، ϵ برای شـــرط توقف در بروزرسانی وزن‌های شبکهٔ تابع پایه شعاعی و η به عنوان پارامتر یادگیری این وزن‌ها. در نهایت، الگوریتم، ماتریس کامل $\widehat{R}$ را به عنوان خروجی ارائه می‌دهد. هدف از ارائـــهٔ الگوریتم در مرحلهٔ برون‌خط، رفع چالش تنک بودن ماتریس امتیاز اســـت. فرآیند الگوریتم با قرار دادن ماتریس R در یک ماتریسی بنام $\widehat{R}$ آغاز می‌شود. برای پیدا کردن ارتباط بین کاربران، از تکنیک خوشه‌بندی بهره می‌بریم. تابع $Q(f,j)$ در شـــبکه تابع پایه شعاعی، همان $\phi(\| R(f,j) - C(i,j) \|_2)$ و متغیر f، شمارهٔ زیرنویس هر خوشه در ماتریس $[I]_{m \times n}$ برای هر ســـتون می‌باشد. با توجه به این‌که بردار F، از طریق وزن‌های اولیهٔ $[W_0]_{m \times n}$ به‌دســـت آمده اســـت، برای بروزرســـانی وزن‌هـــا، از روش کاهش گردایان اســـتفاده می‌کنیم. در حلقهٔ بروزرســـانی وزن‌ها، رابطه‌ای را برای توقف حلقه به صورت

$$E = \frac{\|W - W_0\|_1}{\|W\|_1}$$

تعریف می‌کنیم که شرط توقف حلقه با توجه به رابطهٔ بالا

$$E \le \epsilon$$

می‌باشـــد. ســـامانه با توجه به وزن‌های بروزرسانی شده، گام ۱۰ تا ۲۱ الگوریتم را دوباره اجرا خواهد کرد.

۲,٤ الگوریتم مرحله برخط

Algorithm Online
Input: $R = [r_{ij}, 1 \le i \le m, 1 \le j \le n]$: sparse user-item rating matrix. $\widehat{R} = [\hat{r}_{ij}, 1 \le i \le m, 1 \le j \le n]$: smooth user-item rating matrix. u: active user's ratings to the items. T: number of items to be recommended.
Output: $\tilde{R} = [\tilde{r}_{ij}, 1 \le i \le m, 1 \le j \le n]$: predicted user-item rating matrix. $\tilde{u}$: predicted ratings for the active user. Z: item recommendations for the active user.

1. Set $\tilde{u} = u$.
2. **if** the active user is a new user and the rating vector is NULL (cold start problem) **then**
3. Set $L = [l_1, \cdots, l_n]$, $l_j = \sum_{i=1}^{m} \widehat{R}_{ij}$.
4. Set $Z =$ argmax$_j$(L,T) to be T top rated items.
5. **else**
6. Set v to be the Pearson similarity vector between u and all users corresponding to $\widehat{R}$:

$$v_l = \frac{\sum_{t \in r_l \cap u} (\hat{r}_{lt} - \overline{\hat{r}_l})(u_t - \overline{u})}{\sqrt{\sum_{t \in r_l \cap u} (\hat{r}_{lt} - \overline{\hat{r}_l})^2} \sqrt{\sum_{t \in r_l \cap u} (u_t - \overline{u})^2}}, (l = 1, \cdots, m)$$

7. Set $P =$ find $(v >= 0)$ to be the positive neighbouring cluster.
8. Set $N =$ find $(v < 0)$ to be the negative neighbouring cluster.
9. Set $J =$ find $(u == 0)$ to be the zero rated items for the active user.
10. Set s to be the Cosine similarity vector between u and all users corresponding to R:

$$s_l = \frac{\sum_{t\in r_l \cap u} r_{lt}u_t}{\sqrt{\sum_{t\in r_l \cap u} (r_{lt})^2}\sqrt{\sum_{t\in r_l \cap u} (u_t)^2}}, (l=1,\cdots,m)$$

11. Predict the rating of the set of items J for the active user as the weighted average of deviations from the positively correlated neighbours' mean

$$\tilde{u}_j = \overline{r_j} + \frac{\sum_{i\in P} (\hat{r}_{ij}-\overline{\hat{r}_i})s_i}{\sum_{i\in P} s_i}, (j \in J)$$

12. For each j ∈ J, i∈[i=1,2,⋯,m|r_{ij}=0,1≤j≤n], set $\tilde{r}_{ij} = \tilde{u}_j$.
13. For each j ∉ J, i∈[1,2,⋯,m], set $\tilde{r}_{ij} = \hat{r}_{ij}$.
14. Set X = find ($\tilde{u} \geq 3$) to be the set of recommended items based on liking.
15. Predict the rating of the set of items J for the active user based on negatively correlated neighbours

$$\tilde{v}_j = \overline{r_j} + \frac{\sum_{i\in N} (\hat{r}_{ij}-\overline{\hat{r}_i})s_i}{\sum_{i\in N} s_i}, (j \in J)$$

16. Set Y = find ($\tilde{v} \geq 3$) to be the set of recommended items based on disliking.
17. Set Z = X − Y.
18. **if** Z ==NULL then
19. Set L = [$l_1,\cdots,l_n$], $l_j = \sum_{i=1}^{m} \hat{R}_{ij}$.
20. Set Z = argmax_j(L,T) to be T top rated items.
21. **else if** length (Z) ≥ T then
22. Z = Z(1: T).
23. **else**
24. Set L = [$l_1,\cdots,l_n$], $l_j = \sum_{i=1}^{m} \hat{R}_{ij}$.
25. Set $\tilde{Z} = \text{argmax}_j$(L,T − length (Z)) to be T − length (Z) top rated items.
26. Z = [Z,$\tilde{Z}$] .
27. **end if**
28. **end if**
29. **return** $\tilde{R}$, $\tilde{u}$ and Z.

ورودی‌های الگوریتم برخط سامانهٔ توصیه‌گر مشارکت‌محور عبارت هستند از: R به عنوان ماتریس امتیاز، $\widehat{R}$ به عنوان ماتریس هموار شده (کامل شده) در مرحلهٔ برون‌خط، بردار امتیاز u که توسط کاربر فعال به سامانه تحویل داده شده است و T که تعداد توصیه‌های سامانه را مشخص می‌کند. بردار u می‌تواند بردار صفر باشد که در این صورت مشکل شروع سرد برای سامانه رخ خواهد داد. با پیشامد حاصل، سامانه از طریق ماتریس $\widehat{R}$، کالاهایی که بیشترین امتیاز را برای هر خوشه دارا می‌باشند، به تعداد T، به کاربر فعال، به عنوان توصیه، ارائه می‌کند. در نتیجه با استفاده از خروجی الگوریتم برون‌خط در ۱,٤، مشکل شروع سرد برطرف خواهد شد. هنگامی که کاربر فعال، حداقل به یک کالا امتیاز دهد روند الگوریتم ۲,٤ در گام ٦ شروع می‌شود. رویکردی که در پیش‌بینی امتیازات صفر کاربر فعال به کار گرفته شده است، رویکرد «همسایهٔ نزدیک» است. به این شکل که کاربران را، از طریق تابع تشابه پیرسون (٥) به دو خوشهٔ مثبت و منفی تقسیم می‌کنیم. برد این تابع $[-1,1]$ می‌باشد و کاربران با تابع تشابه نامنفی در خوشهٔ مثبت و کاربران با تابع تشابه منفی در خوشهٔ منفی قرار می‌گیرند. حال برای خوشهٔ مثبت از تابع تشابه کسینوسی (٦) برای یافتن نزدیک‌ترین همسایه بهره می‌بریم. با استفاده از تابع پیش‌بینی گام ۱۱ که همان انحراف از میانگین همسایگی وزنی است، امتیازهای صفر کاربر فعال را پیش‌بینی می‌کنیم. در این رابطه مقادیر تابع تشابه کسینوسی، همان وزن‌های تابع پیش‌بینی هستند. ماتریس پیش‌بینی را با استفاده از امتیازات پیش‌بینی شدهٔ کاربر فعال به‌دست می‌آوریم. به این صورت که مقادیر پیش‌بینی شده توسط تابع پیش‌بینی به جای امتیاز صفر در ماتریس R، برای زیرنویس کالاهایی با امتیاز صفر در بردار u جایگذاری می‌شوند. دیگر امتیازات صفر در ماتریس امتیاز R، با مقادیر ماتریس هموار $\widehat{R}$ پر خواهد شد. زیرنویس‌هایی با مقادیر قابل قبول نزدیک‌ترین همسایه برای بردار پیش‌بینی شده در خوشهٔ مثبت را در مجموعهٔ X قرار داده و همین روند محاسبه را برای خوشهٔ منفی تکرار و Y را به عنوان مجموعهٔ زیرنویس‌هایی با مقادیر قابل قبول نزدیک‌ترین همسایه برای بردار پیش‌بینی شده در خوشهٔ منفی، تعریف می‌کنیم. مقادیر قابل قبول، از قبل برای هر سامانه‌ای معین می‌شود. در این مقاله این مقدار، بزرگ‌تر یا مساوی ۳ می‌باشد. با محاسبهٔ $Z = X\backslash Y$، کالاهایی توسط سامانه انتخاب می‌شود که با استفاده از محاسبات قبل، به نظر و سلیقهٔ کاربر نزدیک‌تر هستند. در انتها خروجی‌های الگوریتم عبارت هستند از: ماتریس پیش‌بینی $\tilde{R}$، بردار پیش‌بینی $\tilde{u}$ برای کاربر فعال و کالاهای توصیه شده توسط سامانه با اولویت مجموعهٔ Z.

۵ – ارزیابی عددی

در این قسمت به ارزیابی عددی الگوریتم‌های ذکر شده می‌پردازیم. داده‌های مورد استفاده در این مقاله، مربوط به داده‌های Movielens [٦] است که حاوی 100000 امتیاز ناصفر متعلق به 943 کاربر در 1682 فیلم مختلف می‌باشد. همچنین از معیارهای زیر برای ارزیابی دقت و کیفیت سامانه توصیه‌گر پیشنهادی استفاده شده است [٢٦].

معیار MAE یا میانگین فدرمطلق خطا که در واقع میانگین قدرمطلق تفاضل، بین امتیاز پیش‌بینی شده توسط سامانهٔ توصیه‌گر مشارکت‌محور و امتیاز واقعی است.

6-https://grouplens.org/datasets/movielens

$$MAE = \frac{1}{mn}\sum_{i,j} |\tilde{r}_{ij} - r_{ij}|.$$

معیار RMSE یا میانگین مربع خطا که از رابطه زیر محاسبه می‌شود.

$$RMSE = \sqrt{\frac{1}{mn}\sum_{i,j} (\tilde{r}_{ij} - r_{ij})^2}$$

معیار دقت Precision که از تقسیم «تعداد امتیازهای واقعی قابل قبول و امتیازهای پیش‌بینی شدهٔ قابل قبول» بر «تعداد امتیازهای پیش‌بینی شدهٔ قابل قبول» به صورت زیر و با در نظرگرفتن پارامترهای جدول ۲ به صورت زیر محاسبه می‌گردد.

$$Precision = \frac{TP}{TP + FP}.$$

معیار بازیابی Recall که از تقسیم «تعداد امتیازهای واقعی قابل قبول و امتیازهای پیش‌بینی شدهٔ قابل قبول» بر «تعداد امتیازهای واقعی قابل قبول» با استفاده از رابطه زیر قابل محاسبه است.

$$Recall = \frac{TP}{TP + TN}.$$

مقادیر معیارهای Precision و Recall عددی بین ۰ تا ۱ است. هرچه این معیار به عدد ۱ نزدیک باشد، دقت الگوریتم مطلوب‌تر است.

معیار F-Measure که یک معیار مناسب برای ارزیابی دقت یک الگوریتم است، با استفاده از رابطه زیر محاسبه می‌گردد.

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

تذکر ٤. برای محاسبهٔ معیارهای بالا، باید مشخص کنیم که چه امتیازی (واقعی و پیش‌بینی شده) قابل قبول می‌باشد[۱۰]. در این مقاله برای محاسبهٔ معیارهای Precision و Recall، امتیازات بزرگ‌تر یا مساوی ۳ قابل قبول هستند.

۱,۵ مدل ارزیابی k-fold Cross-Validation

اعتبارسنجی متقابل یا Cross-Validation، یک مدل ارزیابی است که مشخص می‌کند که نتایج یک تحلیل آماری بر روی یک مجموعه داده تا چه اندازه قابل تعمیم است. این روش به‌طور ویژه در کاربردهای پیش‌بینی مورد استفاده

جدول ۲: جدول توصیف

	Recommended items	Not Recommended items
Used items	TP	TN
Not used items	FP	FN

قرار می‌گیرد تا مشخص شود مدل موردنظر تا چه اندازه در عمل مفید خواهد بود. در اعتبارسنجی متقابل $k-fold$، داده‌ها به طور تصادفی، به k زیرمجموعهٔ $X_1, X_2, \cdots, X_k$ افراز می‌شوند. از این k زیرمجموعه، هر بار یکی برای اعتبارسنجی و $k-1$ تای دیگر برای مدل‌سازی به کار می‌روند. این روال k بار تکرار می‌شود و همه داده‌ها دقیقا یک بار برای مدل‌سازی و یک بار برای اعتبارسنجی به کار می‌روند. در نهایت میانگین نتیجهٔ این k بار اعتبارسنجی به عنوان یک تخمین نهایی برگزیده می‌شود. جفت foldهای ترکیب شدهٔ (اعتبارسنجی، مدل‌سازی) به صورت زیر تعریف می‌شوند [۲۵].

$$(T_1,V_1),(T_2,V_2),\cdots,(T_k,V_k).$$

به طوری که

$$\begin{aligned} V_1 &= X_1, \quad T_1 = X_2 \cup X_3 \cup \cdots \cup X_k, \\ V_2 &= X_2, \quad T_2 = X_1 \cup X_3 \cup \cdots \cup X_k, \\ &\vdots \qquad\quad \vdots \\ V_k &= X_k, \quad T_2 = X_1 \cup X_2 \cup \cdots \cup X_{k-1}. \end{aligned}$$

عملکرد کلی مدل اعتبارسنجی متقابل $k-fold\ Cross-Validation$. برای معیارهای خطای MAE و RMSE همان‌طور که در روابط (٤) و (٥) نشان داده شده است، محاسبه می‌شود.

$$MAE = \frac{1}{k}\sum_{L=1}^{k} \left(\frac{1}{|V_L|}\sum_{(i,j)\in V_L} |\tilde{r}_{ij} - r_{ij}|\right). \qquad (٤)$$

$$RMSE = \frac{1}{k}\sum_{L=1}^{k} \left(\sqrt{\frac{1}{|V_L|}\sum_{(i,j)\in V_L} (\tilde{r}_{ij} - r_{ij})^2}\right). \qquad (٥)$$

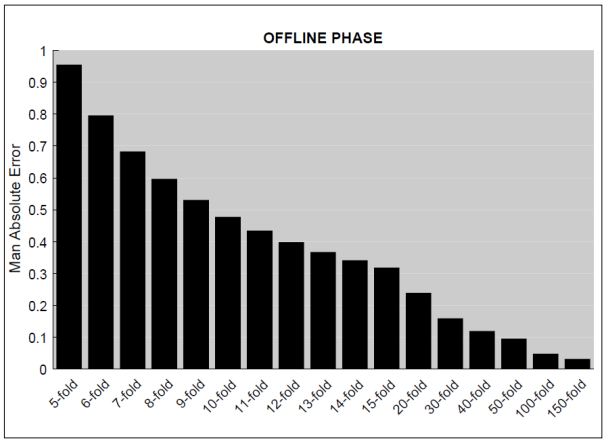
۲,۵ – نمایش خطا در قالب نمودار

در این قسمت به تحلیل دقت الگوریتم‌های سامانهٔ توصیه‌گر مشارکت‌محور توسعه‌یافته با استفاده از وزن‌های کاهش گرادیان در قالب نمودار برای داده‌های MovieLens می‌پردازیم. قبل از ارائهٔ ارزیابی، ورودی الگوریتم‌های برون‌خط و برخط را در جدول ۳، برای

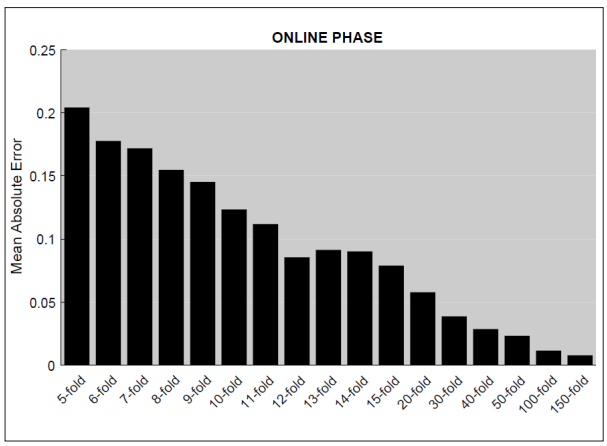
جدول ۳. پارامترهای وروری برای ارزیابی سامانه با داده‌های MovieLens

پارامترهای ورودی	
نوع پارامتر	مقدار پارامتر
Φ	$\exp\left(\frac{-\|R-C\|_2^2}{2\sigma^2}\right)$
σ	1
η	0.1
ϵ	0.9

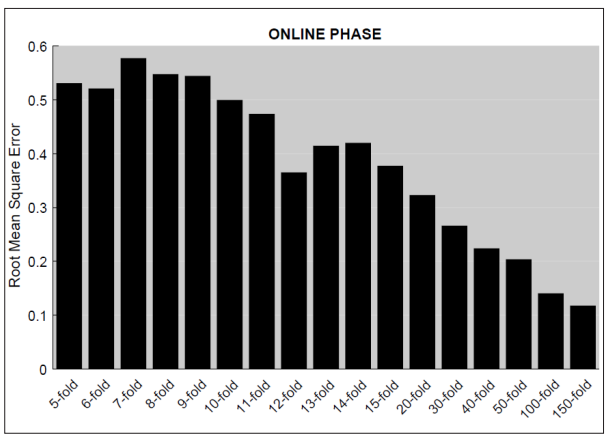
ارزیابی مقداردهی کرده‌ایم. با توجه به این‌که از تابع پایه شعاعی مقیاس‌نشده استفاده نموده‌ایم مقدار σ را برابر ۱ در نظر گرفته‌ایم. همچنین مقدار نرخ یادگیری η در تکنیک کاهش گرادیان را مقداری کوچک‌تر از ۱ و نزدیک ۰ اختیار می‌کنند که ما در این مقاله مقدار ۰,۱ را در نظر گرفتیم. مقدار ϵ نیز کنترل‌کننده میزان خطای نسبی وزن‌ها است که برای جلوگیری از افزایش زمان اجرای برنامه مقدار نه چندان کوچک ۰,۹ را اختیار نموده‌ایم. نمودارهای میله‌ای مربوط به معیار MAE برای الگوریتم برون‌خط ۱,٤، در شکل ۱ نمایش داده شده است. بالاترین میزان خطای الگوریتم پیشنهادی در شکل ۱ برابر با ۰,۹۵ است در حالی که در مقاله [۱۳] برابر با ۱,۱۹۳۷، در مقاله [۲۱] برابر با ۱,۰۰۸٦ و درمقاله [۱٤] برابر با ۱,۰۱۹۳ گزارش شده است. نمودارهای میله‌ای مربوط به معیارهای MAE و RMSE به ترتیب در شکل‌های ۲ و۳ و نمودارهای میله‌ای مربوط به معیارهای Recall، Precision و F-Measure متعلق به الگوریتم برخط ۲,٤، در شکل ٤ رسم شده‌اند. بالاترین میزان خطای الگوریتم پیشنهادی در شکل ۲ برابر با ۰,۲۱ است در حالی که در مقاله [۱۳] برابر با ۰,۹۲۰۱، در مقاله [۲۱] برابر با ۰,۷۰۹٤ و در مقاله [۱٤] برابر با ۰,٦۷٦۰ گزارش شده است. بالاترین میزان خطای الگوریتم پیشنهادی در شکل ۳ برابر با ۰,٥٦ است در حالی که در مقاله [۱۳] برابر با ۰,۹۳۱۸، در مقاله [۲۱] برابر با ۰,۹۰۹٦ و درمقاله [۱٤] برابر با ۰,۹ گزارش شده است. همان‌طور که انتظار می‌رود،



شکل ۱: نمودار میله‌ای مربوط به معیار MAE در مرحلهٔ برون‌خط

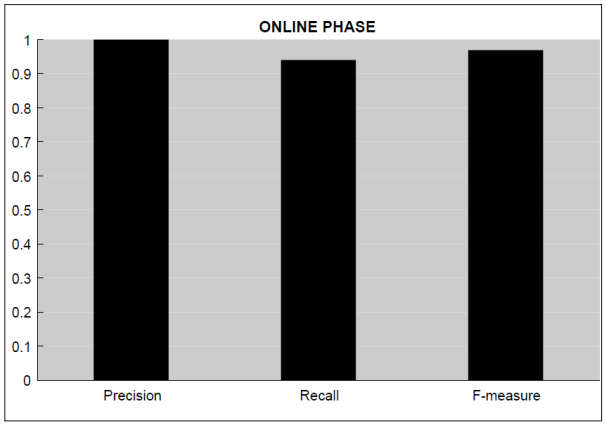


شکل ۲: نمودار میله‌ای مربوط به معیار MAE در مرحلهٔ برخط



شکل ۳: نمودار میله‌ای مربوط به معیار RMSE در مرحلهٔ برخط

عملکرد سامانهٔ توصیه‌گر مشارکت‌محور با استفاده از توابع پایه شعاعی، با توجه به نتایج حاصل شده بسیار مطلوب است و توصیه‌های حاصل از این سامانه، قابل اعتماد می‌باشند.



شکل ۴: نمودار میله‌ای مربوط به معیارهاPrecision، Recall و F-measure در مرحلهٔ برخط

۶- نتیجه‌گیری

سامانه‌های توصیه‌گر، ابزارهای موثر نرم‌افزاری هستند که پس از تجزیه و تحلیل داده‌های موجود، پیشنهادهایی مناسب و نزدیک به واقعیت را در اختیار مخاطب قرار می‌دهند. هر چه این پیشنهادها به نظرات کاربران و علایق آن‌ها نزدیک‌تر باشد، سامانهٔ توصیه‌گر دقیق‌تر است. در این مقاله، به توسعه سامانه‌های توصیه‌گر مشارکت‌محور با استفاده از شبکه توابع پایه شعاعی پرداختیم. ابتدا در مرحله برون‌خط، مقادیر امتیاز داده نشده در ماتریس امتیازات را با استفاده از توابع پایه شعاعی تقریب زدیم. مراکز این توابع را با اعمال خوشه‌بندی کی-میانگین بر روی ستون محصولات و وزن‌های این توابع را به کمک روش کاهش گرادیان به‌دست آوردیم. سپس به کمک ماتریس کامل به‌دست آمده از مرحله برون‌خط، کاربران را، از طریق محاسبه مقدار تابع تشابه پیرسون بین کاربر فعال و سایر کاربران به دو خوشهٔ مثبت و منفی تقسیم می‌کنیم و با رویکرد همسایه نزدیک، پیشنهادهایی را به کاربر فعال وارد شده به سامانه بر اساس علایق او ارائه می‌دهیم. در این مقاله، مشکل مقیاس‌پذیری را با استفاده از خوشه‌بندی کاربران، مشکل شروع سرد را با پیشنهاد محصولات با بالاترین امتیاز و مشکل تنک بودن را با هموار نمودن ماتریس امتیازات به کمک شبکه توابع پایه شعاعی رفع می‌نماییم. نتایج آزمایش‌ها بر روی داده‌های MovieLens 100K بیانگر بالابودن دقت و کیفیت پیشنهادهای ارائه شده است. استفاده از روش‌های خوشه‌بندی فازی که توانایی تشخیص داده‌های پرت و تخمین اولیه و نادقیق مقادیر صفر در ماتریس امتیازات را دارند و همچنین تکنیک بهینه‌سازی مقید برای یافتن مراکز توابع پایه شعاعی را به تحقیقات آتی می‌سپاریم.

مراجع

[۱] اسماعیلی، مهدی. انبار داده‌ها و داده‌کاوی. چاپ دوم. انتشارات دیباگران تهران، تهران، ۱۳۹۵.

[۲] گوهری، فائزه‌السادات. شمس، فریدون. حقیقی، حسن. ارائه یک سیستم توصیه گر اثربخش با به کارگیری روابط اعتماد محلی و سراسری آگاه از زمینه، علوم رایانشی، صفحات: ۱۴-۳، ۱۳۹۸.

[۳] میرزایی، داود. آنالیز عددی پیشرفته. چاپ دوم. انتشارات دانشگاه اصفهان، اصفهان، ۱۳۹۳.

[4] D. Billsus, M.J. Pazzani. Learning collaborative information filters. 15th International Conference on Machine Learning. (1998)46-54.

[5] M. Bozzini. L. Lenarduzzi. M. Rossini. R. Schaback. Interpolation by basis functions of different scales and shape. Calcolo. 41(2004)77-87.

[6] K. Cheung, K. Tsui, J. Liu. Extended latent class models for collaborative recommendation. IEEE Trans. Systems, Man, and Cybernetics—Part A: Systems and Humans. 34(2004)143-148.

[7] T.M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE transactions on electronic computers. 3(1965)326-334.

[8] A. De Carvalho, M.M. Brizzotti. Combining RBF networks trained by different clustering techniques. Neural Processing Letters. 14(2001)227-240.

[9] S. De Marchi. Four lectures on Radial Basis Functions. Department of Mathematics, University of Padua. (2013).

[10] M.D. Ekstrand, J.T Riedl, J.A. Konstan. Collaborative filtering recommender systems. Foundations and Trends® in Human–Computer Interaction. 4(2011)81-173.

[11] G. Fasshauer. J. Zhang. On choosing “optimal” shape parameters for RBF approximation. Numerical Algorithms. 45(2007)345-368.

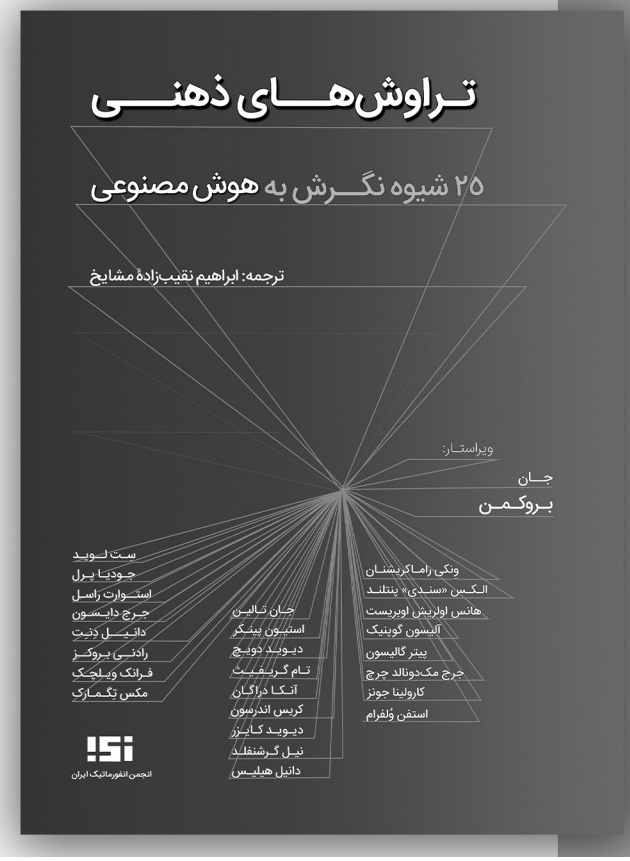
[12] E.W. Forgy. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. biometrics. 21(1965)768-769. ACM computing surveys (CSUR). 31(1999):264-323.

[13] X. Guan, C. Li, Y. GUAN. Matrix Factorization With Rating Completion: An Enhanced SVD Model for Collaborative Filtering Recommender Systems. IEEE Access. 2017, pp:27668-27678.

[14] A. Hernando, J. Bobadillal, F. Ortega. A nonnegative matrix factorization for collaborative filtering recommender

based collaborative filtering recommendation. algorithms. 10th International World Wide Web Conference. (2001)285–295.
[23] B. Sarwar, G. Karypis, J.A. Konstan, T.J. Riedl. Application of dimensionality reduction in recommender systems: A case study. ACM WebKDD Workshop, New York. (2000).
[24] L. Ting-Peng, H. Lai, Y. Ku. Personalized content recommendation and user satisfaction: Theoretical synthesis and empirical findings. Journal of Management Information Systems. 23(2006)45-70.
[25] J. Wit. Evaluating recommender systems: an evaluation framework to predict user satisfaction for recommender systems in an electronic programme guide context. MS thesis. University of Twente. (2008).
[26] G. Xue, C. Lin, Q. Yang, W. Xi, H. Zeng, Y. Yu, Z. Chen. Scalable collaborative filtering using clusterbased smoothing. 28th Annual International ACMSIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil. (2005)114–121.
[27] S. Zahra, M.A. Ghazanfar, A. Khalid, M.A. Azam, U. Naeem, A. Prugel-Bennett. Novel centroid selection approaches for KMeans-clustering based recommende systems. sciences. 320(2015)156-189.
[28] Y. Zhao, X. Feng, J. Li, B. Liu. Shared collaborative filtering. Proceedings of the fifth ACM conference on Recommender systems. (2011)29-36.
systems based on a Bayesian probabilistic model. Knowledge-Based Systems. 97(2016)188-202.
[15] N.B. Karayiannis. Reformulated radial basis neural networks trained by gradient descent. IEEE transactions on neural networks. 10(1999)657-671.
[16] M.K. Kavitha devi, P. Venkatesh. Smoothening approach to alleviate the meager rating problem in collaborative Recommender System. Future generation computer systems. 29(2013)262-270.
[17] M.K. Kavitha devi, P. Venkatesh. Kernel based collaborative filtering system for E-purchasing. Sadhana: Academy Proceedings in Engineering Sciences. 35(2010)513–524.
[18] D. Kim, J.A. Fessler. Optimized first-order methods for smooth convex minimization. Mathematical programming. 159(2016)81-107.
[19] M. J. Orr. Introduction to radial basis function networks. http://anc.ed.ac.uk/rbf/papers/intro.ps.gz. (1996).
[20] C. Perlich, S. Rosset. Identifying bundles of product options using mutual information clustering. Proceedings of the 2007 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics. (2007).
[21] M. Ranjbar, P. Moradi, M. Azami, M.~Jalili. An imputation-based matrix factorization method for improving accuracy of collaborative filtering systems. Engineering Applications of Artificial Intelligence. 46 (2015)58-66.
[22] B. Sarwar, G. Karypis, J.A. Konstan, T.J. Riedl. Item-

جدیدترین کتاب
از انتشارات انجمن انفورماتیک ایران
منتشر شد!



تراوش‌های ذهنی

تهیهٔ کتاب از دفتر انجمن انفورماتیک ایران
(۶۶۴۱۲۸۶۱) و فروشگاه اینترنتی چاره
www.chare.ir

قیمت ۴۰/۰۰۰ تومان