Archive of SID

ارائه مدلی برای تخمین میزان برون‌گرایی اعضای شبکه اجتماعی با استفاده از اطلاعات ساختار گراف

ایمان گلکار امنیه[1*] و مرجان کائدی[2]

اطلاعات مقاله

واژگان کلیدی:
مدل‌سازی شخصیت،
شبکه اجتماعی،
برون‌گرایی،
برنامه‌نویسی ژنتیک،
رگرسیون M5.

چکیده

با آگاهی از شخصیت اعضای شبکه‌های اجتماعی می‌توان بسیاری از سرویس‌های ارائه شده به این افراد را بهبود بخشید و یا از این اطلاعات برای بهبود روابط اعضای شبکه اجتماعی با یکدیگر استفاده کرد. یک روش برای تخمین شخصیت اعضای شبکه‌های اجتماعی استفاده صریح از پرسشنامه‌های شخصیتی است. ولی بسیاری افراد این کار را نقض حریم شخصی خود می‌دانند و یا تمایلی به صرف وقت برای پرکردن پرسشنامه ندارند. به همین دلیل نیاز است ویژگی‌های شخصیتی اعضا به صورت غیرمستقیم تخمین زده شود. روش‌هایی که برای این منظور در تحقیقات پیشین ارائه شده‌اند همگی نیازمند دسترسی به پروفایل و اطلاعات محتوایی و متنی اعضا هستند. درحالیکه دسترسی به این اطلاعات همواره امکان‌پذیر نیست و می‌تواند موجب نقض حریم خصوصی افراد گردد. در این مقاله مدلی ارائه می‌شود که به صورت غیرمستقیم و بدون نقض حریم خصوصی افراد، تنها با در اختیار داشتن اطلاعاتی راجع به ساختار گرافی که حول هر عضو شبکه اجتماعی وجود دارد، می‌تواند میزان برون‌گرایی آن عضو را تخمین بزنند. برای ایجاد این مدل، ابتدا داده‌های مربوط به تعدادی از اعضای شبکه اجتماعی جمع‌آوری شده است. سپس با استفاده از دو روش برنامه‌نویسی ژنتیک و قوانین رگرسیون M5 بر روی این داده‌ها، روابطی استخراج شده‌اند که با دریافت سه ویژگی از ساختار گراف هر عضو شبکه اجتماعی می‌توانند میزان برون‌گرایی او را تخمین بزنند. نتایج نشان می‌دهد که این دو روش با دقت بالایی این تخمین را انجام می‌دهند. همچنین، مدل حاصل از برنامه‌نویسی ژنتیک در مقایسه با رگرسیون M5، دقت بالاتر و پیچیدگی محاسباتی کمتری دارد.

۱- مقدمه

شبکه‌های اجتماعی در سال‌های اخیر رشد فراوانی داشته‌اند. پژوهشی که در سال۲۰۰۵ بر روی وب‌سایت‌های شبکه‌های اجتماعی انجام شده است، نشان می‌دهد که حدود ۱۱۵میلیون کاربر عضو این وب‌سایت‌ها هستند[۱]و تنها پنج سال بعد، شبکه اجتماعی فیس‌بوک[2] به تنهایی بیش‌از ۵۰۰ میلیون عضو داشته است[۲]. شبکه اجتماعی یک سرویس مبتنی بر وب است که به افراد اجازه می‌دهد: ۱) یک پروفایل تحت چارچوب سازمان‌دهی شده ایجاد کنند؛ ۲) یک لیست از کاربران دیگر ایجاد

* پست الکترونیک نویسنده مسئول: marjan.kaedi@gmail.com
۱. دانشجوی کارشناسی ارشد،دانشکده مهندسی کامپیوتر، دانشگاه اصفهان
۲. استادیار، دانشکده مهندسی کامپیوتر، دانشگاه اصفهان

[2]Face book

www.SID.ir

Archive of SID

کنند که با آن‌ها در ارتباط هستند و ۳) در لیست ارتباطات خود حرکت کنند و اطلاعات آن‌ها را مشاهده کنند[۳]. در هنگام ایجاد پروفایل کاربری، اعضای شبکه اجتماعی اطلاعات بسیاری را در مورد خودشان ابراز می‌کنند.سپس، از طریق بروزرسانی پروفایل، قرار دادن عکس‌های جدید، اضافه کردن علاقمندی‌ها و ایجاد ارتباطات، بسیاری از جنبه‌های شخصیتی افراد وارد شبکه اجتماعی می‌شود و با دیگران به اشتراک گذاشته می‌شود[۴].

شخصیت را می‌توان به عنوان الگویی از صفات نسبتا دائمی و ویژگی‌های خاص دانست که به رفتار افراد ثبات می‌بخشد[۵]. مطالعات پیشین نشان داده است که پروفایل کاربری اعضای فیس‌بوک انعکاسی از شخصیت واقعی آن‌ها می‌باشد و نه شخصیت ایده‌آل آن‌ها[۶]؛ به بیان دیگر شخصیت واقعی افراد در پروفایل آن‌ها منعکس شده است.

مدل‌سازی و تخمین شخصیت اعضای شبکه‌های اجتماعی، به عنوان یک مساله در حوزه مهندسی کامپیوتر و فناوری اطلاعات شناخته شده است و تحقیقاتی مختلفی در این زمینه انجام شده که در بخش ۳ مرور می‌گردند. با تخمین شخصیت افراد در شبکه‌های اجتماعی می‌توان سرویس‌های اینترنتی ارائه شده در این شبکه‌ها را بهبود بخشید. نمونه‌هایی از سرویس‌های آنلاین شبکه‌های اجتماعی که با در اختیار داشتن شخصیت اعضا بهبود می‌یابند و به صورت واقع‌گرایانه‌تر انجام خواهند شد، عبارتند از:

- سیستم‌های پیشنهاددهنده[1]در شبکه‌های اجتماعی که به صورت خودکار،پیشنهاداتی را به هر کاربر پیشنهاد می‌دهند که به نظر می‌رسد مورد علاقه او قرار خواهند گرفت[۷, ۸].
- بازاریابی ویروسی آنلاین[2][۹-۱۲] که بر روی شبکه‌های اجتماعی پیاده‌سازی می‌شود و از طریق دهان به دهان گشتن اطلاعات مربوط محصول در بین اعضا، باعث ترویج آن محصول می‌شود.
- تشخیص و یافتن انجمن‌ها[3][۱۳, ۱۴]
- ارسال هدفمند پیام‌های تبلیغاتی[۱۵, ۱۶]

بدیهی است که تاثیر دادن شخصیت افراد در کاربردهای گفته شده، باعث واقع‌گرایانه‌تر شدن نتایج و بهبود آن‌ها خواهد شد.

یک روش برای تخمین شخصیت اعضای شبکه‌های اجتماعی این است که از کاربران خواسته شود که صریحا پرسشنامه‌های شخصیتی را پر کنند. ولی بسیاری افراد، این کار را نقض حریم شخصی[4]خود می‌دانند و یا تمایلی به صرف وقت برای پرکردن پرسشنامه ندارند. به همین دلیل، نیاز است که ویژگی‌های شخصیتی اعضا به صورت غیرمستقیم تخمین زده شود. تا کنون روش‌هایی برای مدل‌سازی و تحلیل شخصیت اعضای شبکه‌های اجتماعی به صورت غیرمستقیم ارائه شده است. این روش‌ها، برای تخمین شخصیت اعضا، از اطلاعات پروفایل کاربری اعضا و اطلاعات محتوایی و متنی آن‌ها استفاده می‌کنند. درحالیکه دسترسی به این اطلاعات همواره امکان‌پذیر نیست و ممکن است باعث نقض حریم شخصی افراد گردد. در بسیاری از موارد تنها گراف ارتباطات شبکه اجتماعی موجود می‌باشد و دسترسی به سایر اطلاعات اعضا دشوار خواهد بود. همچنین در هیچ‌یک از تحقیقات پیشین، معادله‌ای مدون برای تخمین شخصیت اعضای شبکه اجتماعی ارائه نشده است.

در این مقاله از بین ویژگی‌های مختلف شخصیتی، بر روی ویژگی شخصیتی برون‌گرایی[5]به عنوان یک ویژگی مهم تمرکز می‌شود و برای تخمین میزان برون‌گرایی اعضای شبکه اجتماعی مدلی ارائه می‌شود که تنها به اطلاعات قابل استخراج از گراف ارتباطات اعضا نیاز دارد. در ادامه،

[1] Recommender Systems
[2] Viral Marketing

[3] Community Identification
[4] Privacy
[5] Extroversion

www.SID.ir

Archive of SID

ابتدا در بخش۲ به معرفی ویژگی شخصیتی برون‌گرایی پرداخته می‌شود. در بخش۳ پژوهش‌های انجام شده پیشین برای تخمین ویژگی‌های شخصیتی اعضای شبکه-های اجتماعی مرور می‌گردند. در بخش۴ روش پیشنهادی این مقاله ارائه می گردد. در این بخش، ابتدا با انجام آزمون همبستگی پیرسون[1] بررسی می‌شود که کدامیک از متغیرهای قابل استخراج از شبکه اجتماعی با میزان برون‌گرایی اعضا همبستگی دارند. سپس دو روش برای تخمین شخصیت بر اساس این ویژگی‌های موثر، ارائه می‌گردد. در بخش۵ به ارزیابی و مقایسه نتایج پرداخته خواهد شد. نهایتا در بخش۶ مقاله نتیجه گیری می‌شود.

۲- برون‌گرایی

شخصیت، ترکیب خاصی از الگوهای واکنشی احساسی، رفتاری و گرایشی است که هر شخص از آن پیروی می‌کند. بر اساس نظریه شخصیتی پنج عامل بزرگ شخصیتی[2][۱۷]، شخصیت را می‌توان به پنج بعد اصلی تقسیم کرد. این پنج بعد عبارتند ازروان‌نژندی[3]، برون-گرایی، وظیفه‌شناسی[4]، تطابق‌پذیری[5] وگشودگی نسبت به تجربیات[6][۱۷, ۱۸].

برون‌گرایی در واقع به خوشه‌ای از صفات اطلاق می‌گردد که باعث می‌شود شخصی پرانرژی، معاشرتی، هیجان‌طلب و دارای اعتماد به نفس و احساسات مثبت باشد. افراد برون‌گرا به راحتی با افراد متفاوت کنار می‌آیند و نسبت به شغل خود احساس رضایت بیشتر دارند و عموما به سازمان و محیط پیرامونی خود احساس بهتری دارند. افراد برون-گرا از بودن با مردم لذت می برند، پر از انرژی هستند و اغلب هیجانات مثبت را تجربه می‌کنند. چنین افرادی مدام در حال حرف زدن و شور و هیجان بوده و در جمع با جرات صحبت می‌کنند و توجه را به خود جلب می‌کنند. برون‌گرایی،آمادگی فرد را برای تجربه حوادث لذت بخش افزایش می‌دهد. برعکس، درون‌گراها به کم تحرکی و بی-انرژی بودن، سکوت، کم‌حرفی، اندیشناکی و عدم وابستگی به دنیای بیرون مشهور هستند[۱۷-۲۰].

در نظریه پنج عامل بزرگ شخصیتی برای تخمین ابعاد شخصیتی اغلب از دو پرسشنامه استاندارد با نام‌های NEO-FFI و NEO-PI استفاده می‌شود[۲۱]. نسخه خلاصه شده‌ای برای پرسشنامه استاندارد NEO-FFI ارائه شده است که شامل ۶۰ سوال پنج گزینه ای است[۱۸, ۲۲]. این پنج گزینه در بازه کاملا مخالف تا کاملا موافق قرار دارند.

۳- مروری بر تحقیقات پیشین

مدل‌سازی شخصیت در بسیاری از شاخه‌های علوم و مهندسی کامپیوتر مورد توجه قرار گرفته و بکار رفته است. به عنوان نمونه می توان به مدل‌سازی شخصیت در علم رباتیک و فناوری عامل‌های هوشمند[7][۲۳, ۲۴]، سیستم‌های پیشنهاددهنده[۲۵]،مدیریت تعامل انسان و کامپیوتر[8](HCI)[۲۶]، مهندسی نرم‌افزار[۲۷]، آموزش الکترونیکی[9][۲۸]، طراحی واسط کاربر[10][۲۹]، تجارت الکترونیکی[۳۰]و همچنین شبکه‌های اجتماعی اشاره کرد. در ادامه این بخش، تحقیقات پیشین در زمینه مدل‌سازی شخصیت در شبکه‌های اجتماعی مرور خواهند شد.

در روانشناسی اجتماعی[11] این باور وجود دارد که رفتار افراد می‌تواند از طریق الگوهای شخصیتی توضیح داده شود. این الگوهای شخصیتی در طول زمان نسبتا ثابت هستند[۱۷]. در شبکه‌های اجتماعی نیز کاربران رفتار دارای ثباتی را از خودشان نشان می‌دهند که از روی آن‌ها

[1] Pearson Correlation
[2] Big-five model
[3] Neuroticism
[4] Conscientiousness
[5] Agreeableness
[6] Openness

[7] Intelligent Agents
[8] Human-Computer Interaction
[9] E-learning
[10] User Interface
[11] Social Psychology

www.SID.ir

Archive of SID

می‌توان شخصیت افراد را تحلیل نمود[۳۱]. اخیرا توجه ویژه ای به مطالعه شخصیت اعضای شبکه‌های اجتماعی معطوف شده است. در این بخش مقاله، تحقیقاتی که در این زمینه انجام شده است، دسته بندی و مرور می‌گردند.

در دیدگاه اول، تحقیقاتی که در این زمینه انجام شده است را بر اساس اطلاعاتی از اعضا که در این روش‌ها برای تحلیل شخصیت مدنظر قرار گرفته است، در سه دسته مرور می‌کنیم.

۱) در برخی تحقیقات، به تحلیل ارتباط بین شخصیت هر عضو شبکه اجتماعی و اطلاعات متنی که در پروفایل او موجود است و یا با دیگران به اشتراک گذاشته است (مانند سن، جنسیت، علاقمندی‌ها و پرسشنامه‌ها، یادداشت‌ها وبلاگ‌های اعضا)، پرداخته شده است[۴, ۳۲, ۳۳]. مشکل عمده ای که این روش‌ها دارند این است که دسترسی به اطلاعاتی مانند مانند سن، جنسیت، علاقمندی‌ها و یادداشت‌های اعضای شبکه‌های اجتماعی، در بسیاری از مواقع امکان‌پذیر نیست و باعث نقض حریم شخصی در شبکه‌های اجتماعی می‌گردد.

۲) در مقابل، در برخی تحقیقات، ارتباط بین شخصیت اعضای شبکه‌های اجتماعی و اطلاعات ساختاری گراف را بررسی کرده‌اند. منظور از اطلاعات ساختاری گراف، اتصالاتی است که از طریق یال‌های موجود در شبکه اجتماعی بین اعضا وجود دارد و بسته به نوع شبکه اجتماعی به عنوان ارتباط "دوست بودن"[1] و یا ارتباط "پیروی کردن"[2] تعبیر می‌گردد. این یال‌ها در مجموع، ساختار گراف شبکه اجتماعی را تشکیل می‌دهند. تحقیقاتی که تنها از اطلاعات ساختاری شبکه اجتماعی استفاده می‌کنند، اخیرا بیشتر مورد توجه قرار گرفته اند. به عنوان نمونه گلبک و همکاران[۲] رابطه مستقیمی بین تعداد ارتباطات افراد با میزان برون‌گرایی شخصیتی آن‌ها را گزارش کرده‌اند. الگین و همکاران[۳۴] در پژوهش خود نشان داده‌اند که برون‌گرایی و روان‌نژندی ارتباط مستقیمی با معیارهای درجه[3]، همبستگی[4] و ارتباطات بینابینی[5] دارد. همچنین مطالعاتی وجود دارد که نشان می‌دهد بین میزان برون‌گرایی و تعداد ارتباطات یک شخص رابطه وجود دارد هرچند که این رابطه چندان چشم گیر نباشد[۳۵]. به بیان دیگر برون‌گرا بودن یک شخص به تنهایی باعث نمی‌شود تعداد افراد در ارتباط با وی در شبکه اجتماعی زیاد باشند. لازم به ذکر است که در هیچ‌یک از تحقیقات ارائه شده در این زمینه، معادله‌ای مدون برای تخمین شخصیت اعضا ارائه نشده است بلکه تنها به تحلیل عوامل موثر پرداخته شده است.

۳) در تحقیقات دیگر بر روی شخصیت اعضای شبکه اجتماعی، محققان با تمرکز بر روی یک شبکه اجتماعی مشخص، از هر دو منبع اطلاعاتی ذکر شده یعنی اطلاعاتی که توسط خود افراد در شبکه اجتماعی به اشتراک گذاشته شده است و اطلاعات ساختاری گراف شبکه، در کنار هم برای تحلیل شخصیت افراد استفاده کرده‌اند[۳۶, ۳۷].

در دیدگاه دوم، تحقیقات انجام شده بر روی شخصیت اعضای شبکه اجتماعی را از نظر کیفی و کمّی بودن نتایج نیز به دو دسته تقسیم می‌کنیم.

۱) دسته اول، پژوهش‌هایی هستند که تنها به مطالعه کیفی ارتباط بین ویژگی‌های افراد در

[1] Friendship
[2] Following
[3] Degree
[4] Closeness
[5] Betweenness

www.SID.ir

Archive of SID

شبکه‌های اجتماعی و شخصیت آن‌ها پرداخته‌اند و ارتباطی عددی که بتوان از آن برای تخمین شخصیت افراد استفاده کرد، ارائه ننموده‌اند[۴, ۳۳, ۳۵, ۳۸, ۳۹]. با توجه به اینکه هیچ یک از آن‌ها به دنبال روشی برای تخمین شخصیت اعضای شبکه اجتماعی نبوده‌اند در این مقاله از بیان آن‌ها خودداری می‌کنیم.

در دسته دوم تحقیقات، ارتباطات بین اطلاعات موجود در شبکه‌های اجتماعی و شخصیت افراد، به صورت کمّی و عددی پرداخته شده است. تا جایی که می‌دانیم تنها دو نمونه تحقیق در این زمینه انجام شده است که در ادامه به بررسی آن‌ها می‌پردازیم. در پژوهش اول، گلبک وهمکاران[۲] به ارتباط کمّی بین پروفایل کاربران شبکه‌های اجتماعی و شخصیت آن‌ها پرداخته‌اند. در این پژوهش از مدل پنج عامل بزرگ شخصیتی استفاده شده است و به وسیله آزمون همبستگی پیرسون، روابط موجود بین پنج فاکتور شخصیتی اعضای شبکه اجتماعی و پارامترهای مختلف اطلاعات آن‌ها در شبکه اجتماعی مورد تحلیل قرار گرفته است. از بین این اطلاعات، ارتباط میان ساختار گراف شبکه و شخصیت برای ما حائز اهمیت است که نتایج آن در جدول۱ آورده شده است. در این جدول، مقادیر بزرگ‌تر از ۰/۱ به عنوان وجود همبستگی قوی تعبیر شده‌اند و به صورت پررنگ نمایش داده شده‌اند.

جدول۱– بخشی از نتایج ارائه شده توسط گلبک و همکاران[۲]

مشخصات ساختاری گراف	گشودگی	وظیفه‌شناسی	برون‌گرایی	تطابق‌پذیری	روان‌نژندی
درجه هر گره	- ۰/۰۹۴	- ۰/۰۷۸	**۰/۱۸۶**	۰/۰۱۳	- ۰/۰۶۹
چگالی گره-محور	**- ۰/۱۵۲**	۰/۰۵۰	**- ۰/۲۲۴**	۰/۰۵۹	۰/۰۳۲

سه نتیجه کلی از جدول۱ استخراج شده است: الف) رابطه مستقیمی بین درجه هر گره و میزان برون‌گرایی هر فرد وجود دارد؛ ب) رابطه معکوسی بین چگالی گره-محور هر گره و میزان برون‌گرایی هر فرد موجود می‌باشد و ج) رابطه معکوس بین چگالی گره-محور هر گره و میزان گشودگی هر فرد نیز برقرار است. در این پژوهش در ادامه از الگوریتم M5 برای تخمین شخصیت اعضا استفاده شده است، با این وجود تنها خطای تخمین شخصیت‌ها ارائه شده و روابطی در این خصوص گزارش نشده است. در پژوهش دوم، کوئرشا و همکاران[۳۱] نیز به بیان ارتباطات کمّی موجود بین پروفایل کاربری اعضای شبکه‌های اجتماعی و شخصیت آن‌ها پرداخته‌اند. در این پژوهش، اعضای شبکه اجتماعی بر اساس عملکردشان به گروه‌های مختلفی تقسیم‌بندی شده‌اند (اعضای سرشناس، اعضای شنونده وغیره) و روابط همبستگی آن‌ها با هریک از پنج بعد شخصیتی مدل پنج عامل بزرگ شخصیتی بیان شده است. همچنین در این پژوهش نیز از الگوریتم M5 برای تخمین شخصیت اعضای شبکه‌های اجتماعی استفاده کرده‌اند که مجددا تنها به بیان درصد خطای تخمین‌ها بسنده شده است و در این پژوهش نیز روابط فرموله شده‌ای برای تخمین شخصیت افراد ارائه نشده است.

از مرور مطالعات قبلی مشخص می‌شود که تا کنون هیچ رابطه ای به صورت فرموله شده، ارائه نشده است که بتوان با استفاده از آن ویژگی‌های شخصیتی اعضای شبکه اجتماعی را تخمین زد. در این مقاله ما به دنبال ارائه روشی هستیم که با استفاده از اطلاعات گراف ارتباطات شبکه اجتماعی، شخصیت اعضای شبکه اجتماعی را تخمین بزنند. بنابراین روشی که در این مقاله، ارائه خواهد شد، در دسته روش‌های کمّی قرار می‌گیرد. با توجه به اینکه دسترسی به پروفایل و اطلاعاتی مانند سن، جنسیت، علاقمندی‌ها و یادداشت‌های اعضای شبکه‌های اجتماعی همواره امکان‌پذیر نیست و ممکن است باعث نقض حریم خصوصی اعضا گردد، در این مقاله تمرکز ما تنها بر روی استفاده از اطلاعات ساختاری گراف شبکه اجتماعی خواهد بود. همچنین از بین ویژگی‌های مختلف شخصیتی بر روی برون‌گرایی تمرکز شده است. بنابراین

www.SID.ir

Archive of SID

در این مقاله روشی ارائه خواهد شد که تنها بر اساس اطلاعات گراف ارتباطات شبکه اجتماعی، روابط کمّی مناسبی را در قالب معادله ریاضی برای تخمین شخصیت اعضای شبکه اجتماعی ارائه می‌کند. این روش پیشنهادی در بخش بعد معرفی می‌گردد.

۴- روش پیشنهادی

برای طراحی مدل پیشنهادی، نیاز است که تعدادی از اعضای شبکه‌های اجتماعی به عنوان جامعه آماری در نظر گرفته شوند و اطلاعات مربوط به ساختار گراف هر یک از آن‌ها به همراه درجه برون‌گرایی هر یک، فراهم باشد. چگونگی جمع‌آوری این داده‌ها و همچنین استخراج متغیرهای مرتبط، در بخش ۴-۱ معرفی می‌شوند. سپس در بخش‌های۴-۲ و ۴-۳، دو روش برای طراحی مدل مورد نظر ارائه می‌گردد.

۴-۱- داده‌های جامعه آماری و آزمون پیرسون

برای جمع‌آوری اطلاعات پرسشنامه ای طراحی شده است و با توزیع کردن این پرسشنامه به صورت الکترونیکی، اطلاعات مورد نظر در مورد اعضای شبکه اجتماعی فیس‌بوک گردآوری شده است. در پرسشنامه طراحی شده، سه دسته سوال موجود است.

- دسته اول، شامل سوالاتی است که با استفاده از آن‌ها می‌توان درجه برون‌گرایی شخص را تخمین زد. برای این منظور از نسخه خلاصه شده پرسشنامه استاندارد NEO-FFI به زبان فارسی[۴۰] استفاده شده که شامل ۶۰ سوال پنج گزینه ای است.
- دسته دوم سوالات، راجع به اطلاعاتی هستند که می‌توانیم از گراف ارتباطات اعضا استخراج کنیم. برای این منظور، دو ویژگی برای گراف هر عضو درنظر گرفته شده است که عبارتند از تعداد دوستان هر عضو (درجه گره مربوطه در گراف) وتعداد دوستان مشترک هر عضو[1]. سپس با استفاده از این دو ویژگی، ویژگی دیگری به نام چگالی گره-محور[2][۴۱] نیز برای هر عضو از طریق رابطه (۱) محاسبه می‌شود.

$$EgoCentric\ Density = M \Big/ \frac{D*(D+1)}{2} - D \quad (۱)$$

که در این رابطه M نشان دهنده تعداد دوستان مشترک است (در واقع، M نشان می‌دهد چند نفر از دوستان یک شخص با یکدیگر نیز دوست هستند)و D نشان دهنده تعداد دوستان هر عضو می‌باشد. در نظریه گراف،چگالی گره محور، میزان تراکم ارتباطات حول هر گره (که در اینجا معادل هر عضو شبکه اجتماعی است) را نشان می‌دهد[۴۱]. در اینجا، این رابطه نشان می‌دهد که تعداد ارتباطات موجود بین دوستان هر عضو (یعنی صورت کسر) چه نسبتی با حداکثر تعداد رابطه‌های ممکن بین آن دوستان (یعنی مخرج کسر) دارد.

- دسته سوم سوالات، مربوط به جنسیت، سن و مدت عضویت در شبکه اجتماعی هستند. این اطلاعات برای ارزیابی‌های بعدی، در پرسشنامه قرار داده شده‌اند هرچند که به گراف ارتباطات شبکه اجتماعی مربوط نیستند.

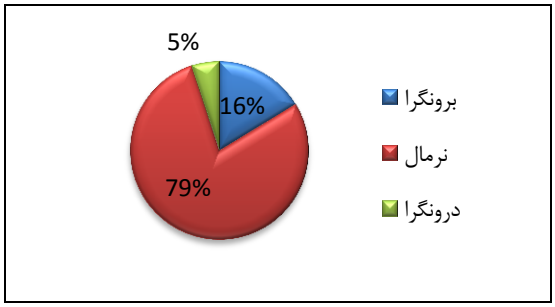
این پرسش‌نامه با استفاده از سامانه ساخت پرسش‌نامه گوگل[3]، به صورت یک فرم طراحی شد و لینک فرم در بین دوستان نویسنده در شبکه اجتماعی فیسبوک قرار داده شد. پس از توزیع پرسشنامه به صورت الکترونیکی، بیش از ۱۴۰ نفر از اعضای شبکه اجتماعی فیس‌بوک، در پر کردن پرسشنامه مشارکت داشته‌اند که پس از پردازش اطلاعات و تشخیص و حذف رکوردهای نامعتبر، نهایتا در مجموع ۱۴۰ رکورد اطلاعاتی صحیح مربوط به ۱۴۰ نفر از ۹ شهر مختلف کشور به دست آمده است. این ۱۴۰ نفر از محدوده‌های جغرافیایی مختلف هستند و سطح تحصیلات

[1] Mutual Friends
[2] Egocentric Density
[3] docs.google.com

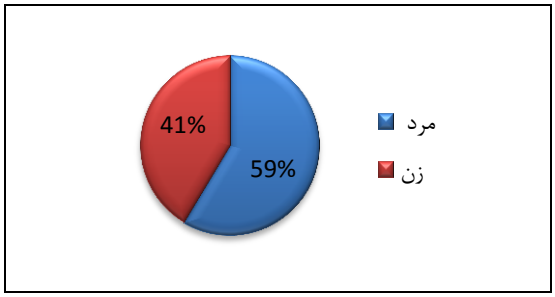
www.SID.ir

Archive of SID

متفاوتی دارند. کمترین سن و بیشترین سن آنها به ترتیب ۱۵ سال و ۵۶ سال است. شکل ۱ توزیع برون‌گرایی و جنسیت را در بین شرکت‌کنندگان نشان می‌دهد.



الف) توزیع برون‌گرایی



ب) توزیع جنسیت

شکل ۱ – توزیع برون‌گرایی و جنسیت در مجموعه داده

همانطور که نمودار شکل ۱ نشان می دهد، ۱۶درصد افراد شرکت کننده، از برون‌گرایی بالای متوسط (یعنی بیشتر از ۴۸) برخوردار هستند و ۵ درصد افراد از برون‌گرایی زیر متوسط (یعنی کمتر از ۲۴) برخوردار هستند و ۷۹ درصد از لحاظ برون‌گرایی نرمال محسوب می‌شوند. همچنین ۸۲ نفر از شرکت کنندگان (۵۹درصد) مرد و ۵۸ نفر (۴۱درصد) زن هستند.

به منظور افزایش دقت اعتبارسنجی روش‌های ارائه شده و روایی آن‌ها، تعداد رکوردهای موجود را با اضافه کردن نویز به داده‌ها افزایش داده‌ایم. برای این منظور از روش نویز گوسی[1] با توزیع ۰/۰۵ استفاده کرده‌ایم و مجموعا ۴۰۰ رکورد اطلاعاتی را بدست آورده‌ایم. رکورد اطلاعاتی هر عضو شبکه اجتماعی شامل فیلدهای زیر است:

- درجه برون‌گرایی آن عضو که عددی بین ۱۲ تا ۶۰ است و از پاسخ‌های کاربر به ۶۰ سوال پرسشنامه ذکر شده محاسبه شده است.
- مشخصات مربوط به گراف آن عضو که شامل موارد زیر هستند:
  - تعداد دوستان آنعضو (درجه گره مربوطه در گراف)
  - تعداد دوستان مشترک آن عضو
  - چگالی گره-محور آن عضو
- جنسیت
- سن
- مدت عضویت در شبکه اجتماعی فیس‌بوک

سپس آزمون پیرسون[2] با در نظر گرفتن برون‌گرایی به عنوان متغیر وابسته و بقیه اطلاعات به عنوان متغیرهای مستقل، بر روی داده‌های جمع‌آوری شده انجام شده است تا متغیرهایی که با برون‌گرایی اشخاص همبستگی قوی دارند، تشخیص داده شوند. نتایج این آزمون در جدول۲ آورده شده‌اند. در این جدول، مقادیر بزرگ‌تر از ۰/۱ به عنوان ارتباطات قوی درنظر گرفته شده‌اند و به صورت پررنگ نمایش داده شده‌اند.

جدول۲ - نتایج آزمون همبستگی پیرسون بر روی داده‌های جمع‌آوری شده

مشخصات	برون‌گرایی
درجه هر گره	**۰/۱۵۸**
تعداد دوستان مشترک	**- ۰/۲۹۸**
چگالی گره-محور	**- ۰/۲۵۲**
جنسیت	۰/۰۵
سن	**- ۰/۱۷۶**
مدت عضویت	**۰/۱۸۹**

همانگونه که در جدول۲ دیده می‌شود، شباهت‌هایی بین نتایج به دست آمده و نتایج ارائه شده توسط گلبک و همکاران[۲] وجود دارد: هر سه نتیجه قابل استخراج از مقاله گلبک و همکاران(ارائه شده در بخش۳) در اینجا نیز

[1] Gaussian noise

[2] Pearson Correlation

www.SID.ir

Archive of SID

تکرار می‌شوند با این تفاوت که شدت همبستگی‌ها متفاوت است. این سه نتیجه عبارتند از: الف) وجود رابطه مستقیم بین درجه و میزان برون‌گرایی هر فرد؛ ب) وجود رابطه معکوس بین تعداد دوستان مشترک و میزان برون‌گرایی هر فرد و ج) وجود رابطه معکوس بین چگالی گره-محور و میزان برون‌گرایی هر فرد.

۴-۲- روش اول: استفاده از قوانین M5 برای تخمین میزان برون‌گرایی اعضا

همانطورکه در بخش ۳ اشاره شد، تنها در دو مورد از تحقیقات پیشین، به صورت کمّی به تحلیل شخصیت کاربران شبکه‌های اجتماعی پرداخته شده است. در هر دوی این تحقیقات[۲, ۳۱]، از روش M5 برای این منظور استفاده شده است. به همین دلیل، در اینجا نیز در روش پیشنهادی اول ما، الگوریتم M5 برای استخراج مجموعه ای از قوانین برای تخمین برون‌گرایی افراد استفاده می‌شود. تفاوت این روش پیشنهادی ما با دو روش قبلی، در ویژگی‌هایی از هر عضو است که برای تخمین شخصیت او مورد استفاده قرار می‌گیرند. این ویژگی‌ها عبارتند از درجه گره مربوطه در گراف، تعداد دوستان مشترک و چگالی گره-محور آن عضو.

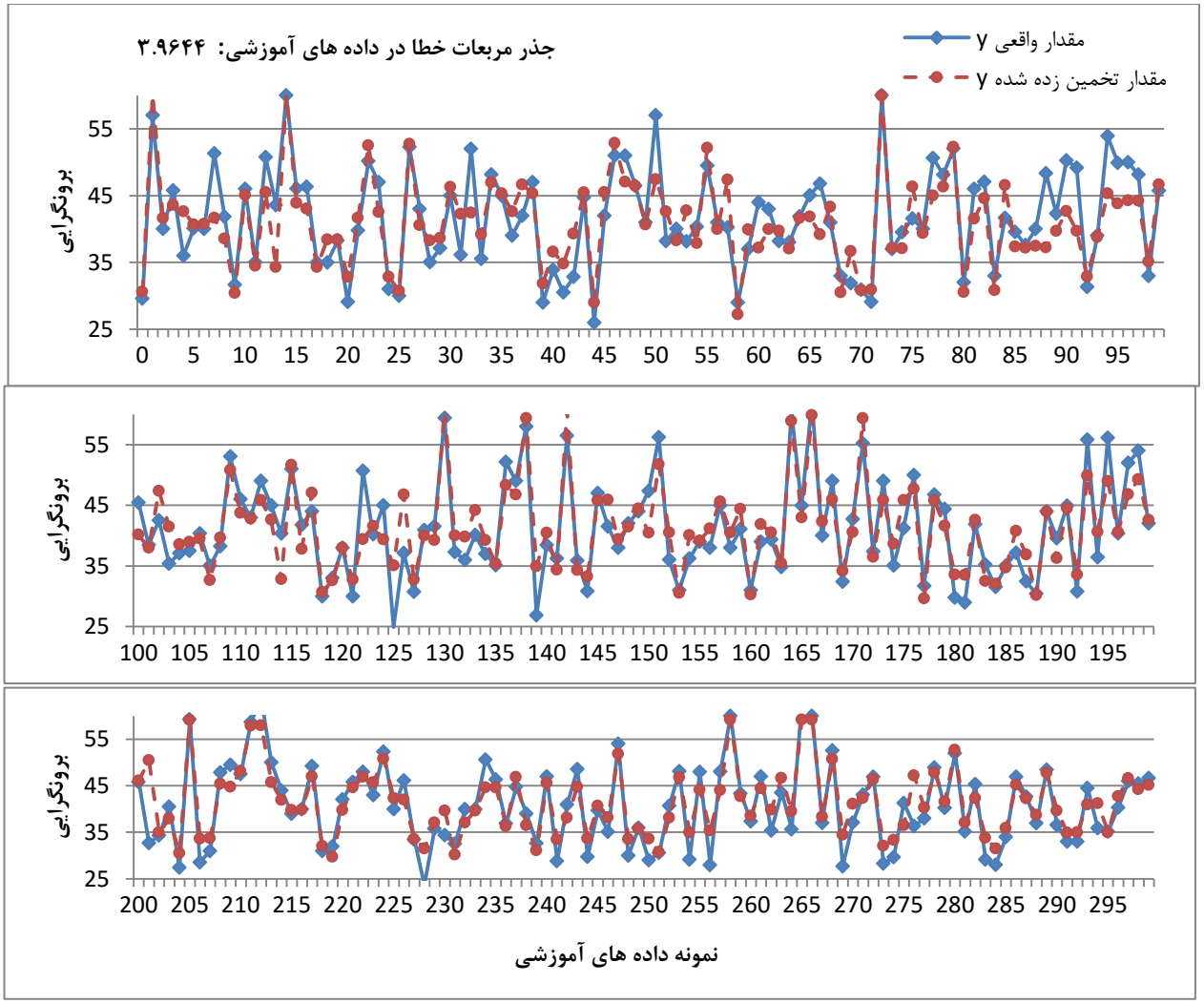
الگوریتم قوانین M5 از جمله روش‌های پیش‌بینی است که با استفاده از روش تفرقه و حکومت[1] لیستی از قوانین تصمیم‌گیری را برای مسائل رگرسیونی تولید می‌کند. در هر تکرار از الگوریتم، مدل درختی از مسئله ساخته می‌شود که راه‌حل‌ها در برگ‌های درخت وجود دارند و بهترین برگ به قانون تبدیل می‌شود. با انجام تکرارهای مختلف، مجموعه‌ای از قوانین ایجاد می‌شود که خلاصه‌تر از مدل درختی ساخته شده بر روی کل مجموعه داده هستند[۴۲]. برای پیاده سازی الگوریتم قوانین M5 از نرم افزار وکا[2][۴۳] استفاده شده است. از بین ۴۰۰ داده موجود، ۳۰۰تا به عنوان داده آموزشی و ۱۰۰تا به عنوان داده تست استفاده شده اند و با استفاده از روش M5، قوانینی طراحی شده است که از روی سه ویژگی گراف هر عضو شبکه اجتماعی (یعنی درجه گره مربوطه در گراف، تعداد دوستان مشترک و چگالی گره-محور آن عضو) می‌تواند میزان برون‌گرایی او را تخمین بزند. شکل ۲ نتایج پیاده‌سازی این روش را برای تخمین برون‌گرایی بر روی داده‌های آموزشی و داده‌های تست نشان می‌دهد. در این شکل، محور افقی، نشان‌دهنده تک تک افرد جامعه آماری (داده‌های آموزشی و داده‌های تست) است و محور عمودی میزان برون‌گرایی (واقعی و تخمین زده شده) را نشان می‌دهد. همانگونه که از شکل ۲ مشخص است، قوانین M5 نتایج بسیار خوبی را بر روی داده‌های آموزشی و تست بدست می‌آورد. همانطورکه در شکل ۲ ذکر شده است جذر مربعات خطا[3] بر روی داده‌های تست ۵/۳۸ می‌باشد. بایدتوجه داشت که این مقادیر در مقیاس ۶۰ هستند، در نتیجه خطای موجود در تخمین‌ها ۸/۹۶ درصد است.

با وجود اینکه در روش پیشنهادی اول، همانند دو روش قبلی از قوانین M5 استفاده شده است ولی نتایج روش پیشنهادی ما در مقایسه با نتایج ارائه شده در مطالعات کوئرشا و همکاران[۳۱] که جذر مربعات خطا را در مقیاس ۵ برابر با ۰/۸۸ (یعنی ۱۷/۶ درصد) گزارش داده اند، بهبود قابل توجهی را به همراه دارد. همچنین گلبک وهمکاران[۲] خطای اعمال الگوریتم M5 بر روی ۷۴ ویژگی استخراجی از شبکه اجتماعی افراد را برابر با ۱۱درصد گزارش کرده‌اند. این در حالی است که در روش پیشنهادی اولِ ما، با حفظ حریم خصوصی اعضا و تنها با اعمال الگوریتم M5 بر روی داده‌های استخراجی از گراف شبکه، تخمین دقیق‌تری برای شخصیت افراد بدست آمده است.

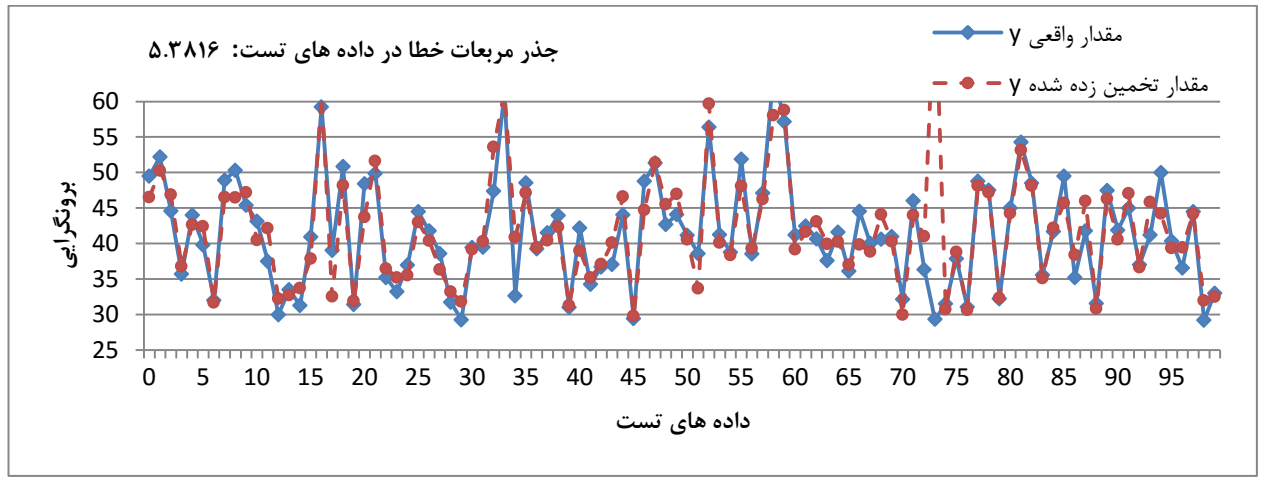
[1] Separate-And-Conquer
[2] WEKA
[3] Root Mean Square Error

www.SID.ir

Archive of SID



شکل ۲–الف- نتایج پیاده سازی روش M5 بر روی داده‌های آموزشی (به علت طول زیاد محور افقی، نمودار در سه سطر ارائه شده است)



شکل ۲–ب - نتایج پیاده سازی روش M5 بر روی داده‌های تست

۴-۳- روش دوم: استفاده از برنامه‌نویسی ژنتیک برای تخمین میزان برون‌گرایی اعضا

در روش پیشنهادی اول که در بخش ۴-۲ معرفی شد، از روش M5 برای تخمین برون‌گرایی اعضا استفاده شد. روش M5 تعداد زیادی از قوانین پیچیده را استخراج می‌کند که با داده‌های آموزشی منطبق هستند. این قوانین زیاد به راحتی برای کاربردهای آینده قابل اعمال نیستند. در اکثر سرویس‌هایی از شبکه اجتماعی که در

www.SID.ir

Archive of SID

آن‌هامی‌توان از شخصیت اعضا استفاده کرد، نیاز است که به صورت آنلاین و سریع، شخصیت تعدادی زیادی از اعضا تخمین زده شود و برای ارائه سرویس‌ها مورد استفاده قرار بگیرد. استفاده از مجموعه قوانین پیچیده بدست آمده از روش M5، نیاز به زمان محاسباتی زیادی دارد و عملا در چنین کاربردهایی قابل استفاده نیست. در روش پیشنهادی دوم، دو هدف مدنظر است. هدف اول این است که ارتباط بین سه متغیر مورد نظر (یعنی درجه، تعداد دوستان مشترک و چگالی گره-محور) و میزان برون‌گرایی شخص، به صورت خلاصه در قالب یک رابطه ریاضی منفرد، مدل‌سازی گردد تا بتوان به راحتی و بدون نیاز به زمان محاسباتی زیاد از آن برای تخمین شخصیت هر یک از اعضا استفاده کرد. هدف دوم، افزایش دقت تخمین است. برای برآورده کردن این دو هدف، در روش پیشنهادی دوم از برنامه‌نویسی ژنتیک استفاده می‌شود. تا جایی که می‌دانیم تا کنون از برنامه‌نویسی ژنتیک برای تخمین شخصیت بر روی داده‌های مستخرج از گراف شبکه اجتماعی استفاده نشده است.

برنامه‌نویسی ژنتیک توسعه ای از الگوریتم ژنتیک است که در آن هر یک از کروموزوم‌ها، یک معادله یا رابطه بین متغیرها را نشان می‌دهد. این رابطه‌ها که اغلب در قالب درخت نمایش داده می‌شوند، برنامه نامیده می‌شوند. در این روش، با تولید جمعیتی از برنامه‌ها و سپس تکامل دادن آن‌ها به صورت تکرار شونده و نسل به نسل، به دنیال یافتن برنامه ای هستیم که با داده‌های مورد نظر مساله، به خوبی منطبق باشد[۴۴].

برای اعمال برنامه‌نویسی ژنتیک بر روی داده‌های جمع‌آوری شده از اعضای شبکه اجتماعی و تخمین میزان برون‌گرایی آن‌ها، جعبه ابزار GPTips[۴۵] در نرم افزار MATLAB استفاده شده است. هدف این است که برنامه‌نویسی ژنتیک با توجه به داده‌ها، معادله ریاضی را بدست آورد که دارای قالب کلی زیر است.

$$Extraversion = F(d, m, e) \qquad (۲)$$

در رابطه (۲)، d، m و e به ترتیب نشان دهنده درجه یک گره (یعنی تعداد دوستان آن عضو شبکه اجتماعی)، تعداد دوستان مشترک آن عضوو چگالی گره-محور او هستند. هدف این است که با استفاده از برنامه‌نویسی ژنتیک، رابطه F بر روی سه متغیر معرفی شده به گونه ای طراحی شود که میزان برون‌گرایی تخمینی برای هر فرد (یعنی *Extraversion*)، به مقدار واقعی برون‌گرایی اوتا حد ممکن نزدیک باشد. اختلاف بین مقدار تخمینی برون-گرایی و مقدار واقعی آن بر اساس جذر مربعات خطا توصیف می‌شود.

پیاده سازی برنامه‌نویسی ژنتیک را می‌توان به دو صورت ژن یکتا[1] و ژن چندگانه[2] انجام داد[۴۵]. استفاده از ژن چندگانه جذر مربعات خطا را به میزان قابل توجه‌ای کاهش می‌دهد.در ادامه نتایج به دست آمده بوسیله روش ژن چندگانه را بررسی می‌کنیم.

در پیاده‌سازی برنامه‌نویسی ژنتیک حداکثر عمق درخت برابر ۷ و جمعیت اولیه برابر با ۵۰ در نظر گرفته شده است و الگوریتم به تعداد۱۰۰ نسل تکرار شده است.روش مسابقه[3] با اندازه ۱۰ به عنوان روش انتخاب[4]در برنامه-نویسی ژنتیک استفاده شده است و پارامتر نخبه-گرایی[5]برابر با ۰/۱ در نظر گرفته شده است. همچنین چهار عمل اصلی جمع، تفریق، ضرب و تقسیم برای قرار دادن در معادله‌ها منظور شده اند. در اینجا نیز همانند روش پیشنهادی اول، از بین ۴۰۰ داده موجود،۳۰۰تا به عنوان داده آموزشی و ۱۰۰تا به عنوان داده تست استفاده شده است. پس از اجرای برنامه‌نویسی ژنتیک بر روی داده‌ها، مدل ارائه شده در رابطه (۳) برای تخمین برون‌گرایی افراد بدست آمده است.

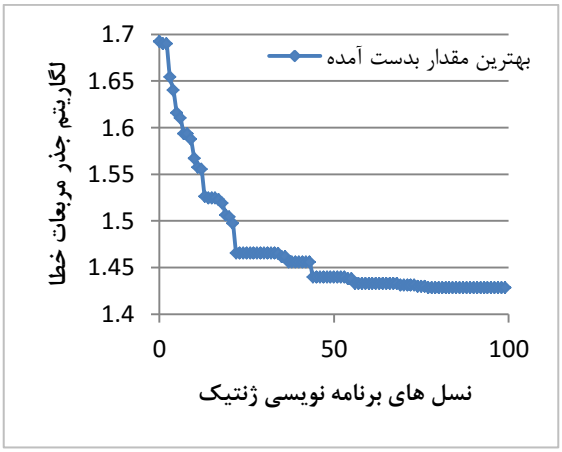
[1] Single Gene
[2] Multi Gene
[3] Tournament
[4] Selection
[5] Elitism

www.SID.ir

Archive of SID

$$Extraversion = 0.0007\,m - 0.0008\,d - \frac{39.58}{d\ e} - \frac{0.95\,m}{d + 6.03} + \frac{0.0001\,d}{e} + \frac{7105}{m^2} + 49 \quad (۳)$$

روند بهبود نتایج در نسل‌های مختلف برنامه‌نویسی ژنتیک در شکل ۳ نشان داده شده است. همانگونه که مشخص است بهترین نتیجه در نسل ۷۷ بدست آمده است و پس از آن الگوریتم همگرا شده است و تغییری در نتایج حاصل نشده است. در اجراهای مختلف برنامه‌نویسی ژنتیک، نتایج کمی با یکدیگر متفاوت هستند و معمولا میانگین تعدادی اجرای مختلف را می‌توان گزارش نمود ولی در اینجا هدف اصلی بدست آوردن معادله ای برای تخمین شخصیت می‌باشد که نمی‌توان آن را در اجراهای مختلف ترکیب نمود؛ به همین دلیل در ادامه مقاله تنها بر روی نتایج یکی از معادله های بدست آمده تمرکز شده است. قابل ذکر است که اختلاف نتایج معادلات مختلف چندان زیاد نیست و میزان تغییرات جذر مربعات خطا در شرایطیکسان حداکثر به میزان ۰/۲ مشاهده شده است.



شکل ۳– روند بهبود در برنامه‌نویسی ژنتیک

شکل ۴ نتایج اعمال رابطه (۳) بر روی داده‌های آموزشی و داده‌های تست را نشان می‌دهد.همانطور که در شکل ۴ ذکر شده است،جذر مربعات خطا در داده‌های تست۳/۸۱ می‌باشد. با توجه به اینکه این مقادیر در مقیاس ۶۰ هستند، خطای موجود در تخمین‌ها برابر با ۶/۳۵درصد است.

با توجه به ماهیت روش برنامه‌نویسی ژنتیک، با وجود اینکه مدل به دست آمده دقت بالایی دارد، ولی این مدل به راحتی قابل تحلیل نیست. همانطورکه انتظار می‌رود جملاتی به صورت ترکیب خطی از متغیرها در فرمول وجود دارند (جملات اول و دوم) و همچنین تاثیرات مثبت و منفی متغیرها با قرار گرفتن آن‌ها در صورت و مخرج کسرها مشخص شده است (جملات سوم تا ششم) و نهایتا جمله آخر مدل باعث می شود که مقدار خروجی مدل در بازه اصلی برون‌گرایی (یعنی بین ۱۲ تا ۶۰) قرار بگیرد. جمله پنجم را می‌توان با توجه تحقیق انجام شده توسط گلبک و همکاران[۲] توجیه کرد؛ گلبک و همکاران[۲]در تحقیق خود تلویحا به ارتباط مستقیم متغیر "درجه" با میزان برون‌گرایی فرد و همچنین ارتباط معکوس متغیر "چگالی گره-محور"با میزان برون‌گرایی فرد اشاره کرده‌اند. این موضوع به نوعی در جمله پنجم مدل بدست آمده بازتاب شده است.

۴-۳-۱- مثالی از کاربرد رابطه ارائه شده

برای استفاده از رابطه استخراج شده، کافی است موارد خواسته شده در رابطه (۳)، یعنی درجه، تعداد دوستان مشترک و چگالی گره-محور هر فرد را در آن جایگذاری کنیم و میزان برونگرایی شخص را بدست آوریم. بعنوان مثال اگر تعداد دوستان (درجه) یک فرد در شبکه اجتماعی برابر ۱۱۸، تعداد دوستان مشترک او برابر ۴۲۶ و چگالی گره-محورش برابر ۰/۰۶ باشد، فرمول ارائه شده به صورت رابطه (۴) در خواهد آمد:

$$0.0007 * \mathbf{426} - 0.0008 * \mathbf{118} - \frac{39.58}{\mathbf{118}*0.06} - \frac{0.95*\mathbf{426}}{\mathbf{118} + 6.03} + \frac{0.0001*\mathbf{118}}{0.06} + \frac{7105}{\mathbf{426}^2} + 49 \quad (۴)$$

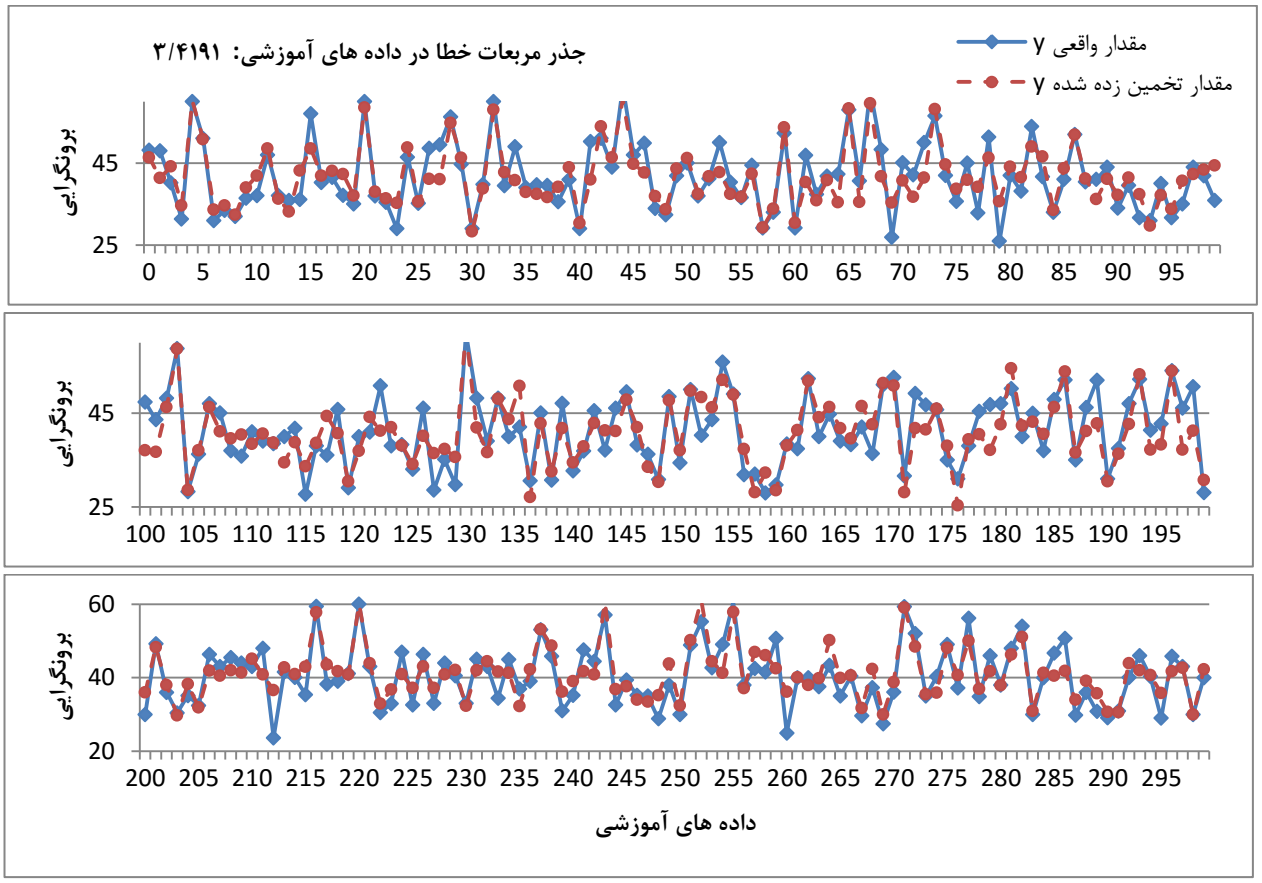
نتیجه رابطه (۴) برابر ۴۰/۵۸ می‌باشد. این مقادیر مربوط به نمونه اطلاعاتی ۶۴ام در شکل ۴ است. مقدار واقعی برونگرایی برای این شخص ۴۱,۸۵ بوده است. اختلاف

www.SID.ir

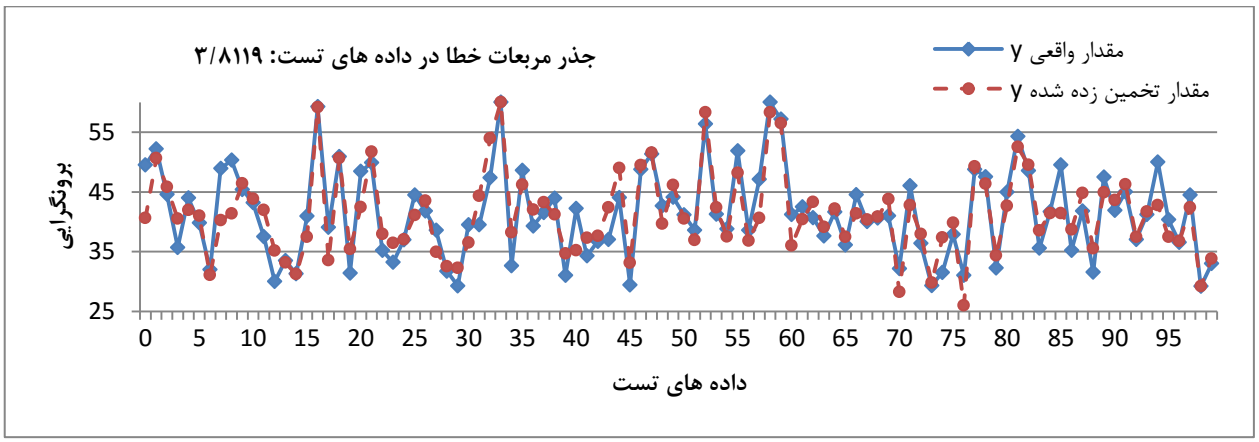
Archive of SID

۱/۲۷ بین مقدار تخمین زده شده و مقدار واقعی نشان دهنده میزان خطای ۲/۱۱ درصد در این نمونه اطلاعاتی است. تمامی نمونه‌های شکل۴ به همین صورت محاسبه و ارائه شده‌اند.

در بخش بعد به مقایسه دو روش پیشنهادی می‌پردازیم.



شکل ۴–الف- مقایسه مقادیر واقعی و تخمین زده شده برون‌گرایی در برنامه‌نویسی ژنتیک بر روی داده‌های آموزشی



شکل ۴–ب- مقایسه مقادیر واقعی و تخمین زده شده برون‌گرایی در برنامه‌نویسی ژنتیک بر روی داده‌های تست

۵- مقایسه روش‌ها

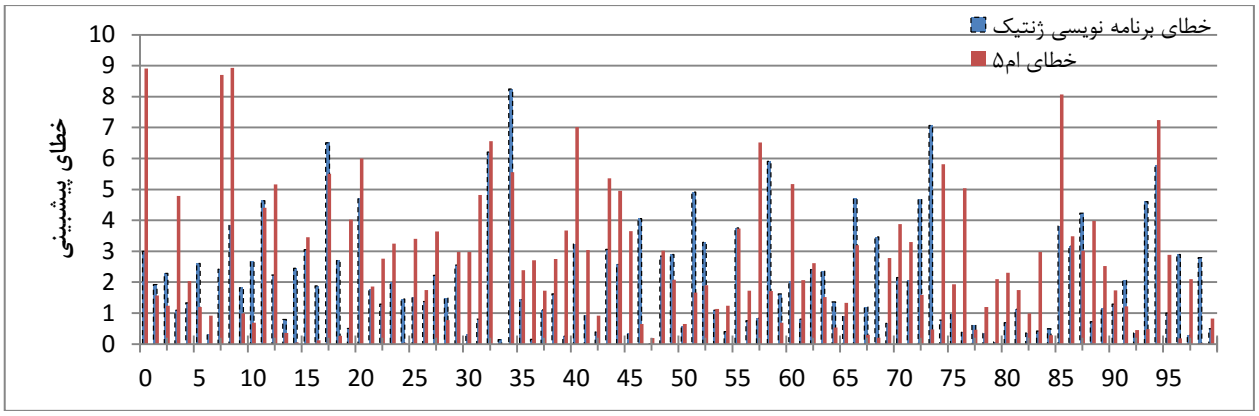
در این بخش به مقایسه دو روش پیشنهادی در این مقاله پرداخته می‌شود. برای این منظور ابتدا این دو روش در شرایط فعلی یعنی تنها استفاده از سه متغیر ساختاری گراف، با یکدیگر مقایسه می‌شوند. سپس فرض می‌شود که دو روش، علاوه بر سه متغیر مذکور به اطلاعات جانبی مانند جنسیت و سن افراد و مدت عضویت آن‌ها در شبکه

www.SID.ir

Archive of SID

اجتماعی هم دسترسی دارند. نتایج این دو روش در این شرایط نیز با یکدیگر مقایسه می‌گردد.

شکل ۵، میزان خطای تخمین بر روی ۱۰۰تا داده تست را برای دو روش قوانین M5 و برنامه‌نویسی ژنتیک نشان می‌دهد. در این نمودار برای تخمین برون‌گرایی هر یک از این ۱۰۰ نفر، تنها از سه متغیر ساختاری گراف آن‌ها استفاده شده است. با بررسی شکل ۵ مشخص می‌شود که در مجموع روش برنامه‌نویسی ژنتیک بهتر از قوانین M5عمل می‌کند. برای بیش از ۸۰ درصد افراد، برنامه‌نویسی ژنتیک با خطایی کمتر از ۳ تخمین برون‌گرایی را انجام داده است. همچنین جذر مربعات خطا که در شکل‌های ۲ و ۴ ارائه شد نیز برای روش قوانین M5 برابر با ۵/۳۸ است در حالیکه برای برنامه‌نویسی ژنتیک ۳/۸۱ است که نشان دهنده نتایج بهتر توسط برنامه‌نویسی ژنتیک است. علاوه بر اینکه روش برنامه-نویسی ژنتیک خطای کمتری در تخمین برون‌گرایی افراد دارد، دارای قابلیت استفاده بیشتری است؛



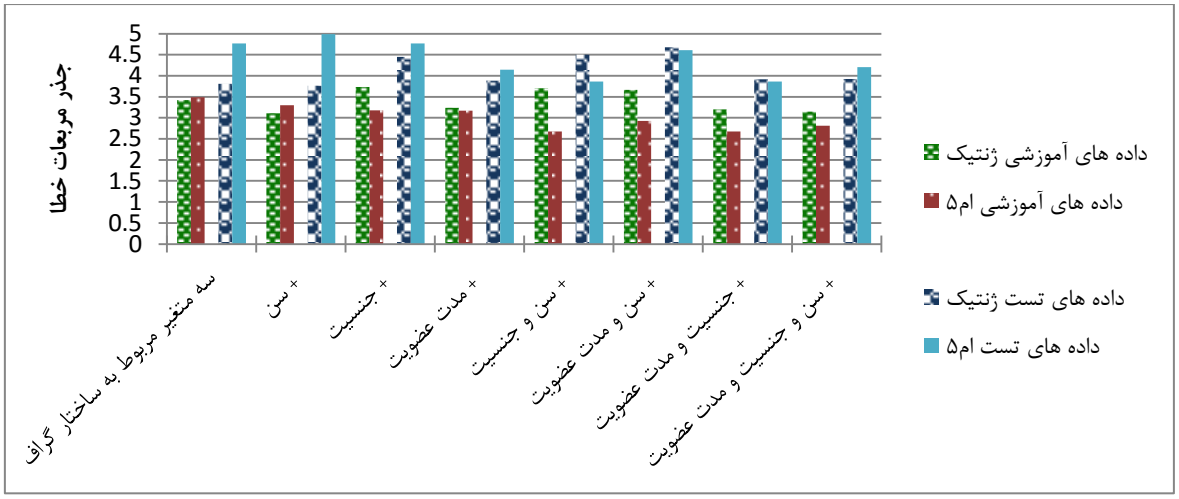
شکل ۵– مقایسه خطای نتایج تخمین بوسیله روش‌های M5 و برنامه‌نویسی ژنتیک بر روی داده‌های تست

چراکه همانطورکه در بخش ۴-۲ ذکر شد،این روش معادله ریاضی را ارائه می‌کند که به سادگی و به سرعت شخصیت افراد را تخمین می‌زند. این معادله برای کاربردهای آنلاین که نیاز به سرعت عمل زیادی دارند، در مقایسه با استفاده از تعداد زیادی قوانین به دست آمده از M5 بسیار کاراتر عمل می‌کند. بنابراین می‌توان گفت که هر دو هدف ذکر شده در بخش ۴-۳ برآورده شده اند.

همانگونه که مشاهده شد، تنها با استفاده از اطلاعات ساختاری گراف شبکه می‌توان شخصیت اعضای شبکه اجتماعی را با دقت بالایی تخمین زد. با این وجود، برای بررسی اینکه اضافه کردن اطلاعات غیر ساختاری، چه تاثیری بر روی تخمین شخصیت دارد، آزمایش‌های دیگری نیز انجام شد. در این آزمایش‌ها دو روش برنامه‌نویسی ژنتیک و قوانین M5 در شرایطی که اطلاعات دیگری نیز به عنوان ورودی موجود هستند، استفاده شدند و خطای آن‌ها با یکدیگر مقایسه شد. شکل ۶، جذر مربعات خطا را به ازای درنظر گرفتن ورودی‌های دیگر برای دو روش پیشنهادی نشان می‌دهد.بررسی شکل ۶ نشان می‌دهد که در نظر گرفتن اطلاعات بیشتر لزوما باعث بهبود نتایج تخمین شخصیت و کاهش خطای تخمین‌ها نمی‌شود. به عنوان مثال در روش برنامه‌نویسی ژنتیک، استفاده از اطلاعات ساختار گراف در مقایسه با استفاده از ساختار "گراف + سن و جنسیت و مدت عضویت" خطای کمتری بر روی داده‌های تست داشته است.همچنین از شکل ۶ می‌توان نتیجه گرفت که استفاده ازاطلاعات غیرساختاری، تاثیر چندانی در کاهش خطای تخمین‌ها نداشته است. بیشترین اختلاف مشاهده شده ۱/۱۳ است که مربوط به روش M5 و بین شرایط "+ سن" در مقایسه با شرایط"+ جنسیت و مدت عضویت" می‌باشد.

www.SID.ir

Archive of SID



شکل ۶– مقایسه خطا با در نظر گرفتن اطلاعات غیر ساختاری گراف شبکه برای هر دو روش ارائه شده (علامت + بدین معنی است که متغیرهای ذکر شده در کنار سه متغیر ساختاری گراف استفاده شده اند)

۶- نتیجه گیری

در این مقاله روشی برای فرموله‌سازی تخمین برون‌گرایی اعضای شبکه اجتماعی، با استفاده از اطلاعات ساختاری گراف شبکه اجتماعی ارائه شد. در مطالعات پیشین به وجود ارتباطاتی بین مشخصات ساختاری گراف شبکه اجتماعی و شخصیت اعضای آن پرداخته شده بود. با این وجود تاکنون معادله‌های مدونی برای تخمین شخصیت اعضا ارائه نشده بود. در این مقاله با استفاده از برنامه-نویسی ژنتیک معادله ریاضی برای تخمین برون‌گرایی اعضای شبکه اجتماعی ارائه شد که با حفظ حریم خصوصی اعضا، می‌تواند با دقت بالا و زمان محاسباتی کمی این کار را انجام دهد. برای این منظور، ۱۴۰ رکورد اطلاعاتی از اعضای شبکه اجتماعی فیس‌بوک جمع‌آوری شد و روش مذکور بر روی آن‌ها اجرا گردید. سپس، یکبار نیز تخمین برون‌گرایی این اعضا با استفاده از الگوریتم M5 نیز تکرار شد و دقت حاصل از برنامه‌نویسی ژنتیک با دقت حاصل از الگوریتم M5 بر روی داده‌های جمع-آوری شده با یکدیگر مقایسه شد. با توجه به اینکه مدل‌های بدست آمده بر روی نمونه داده‌های مربوط به افراد جامعه ایرانی توسعه یافته‌اند، می‌توان این مدل‌ها را برای تخمین شخصیت کاربران ایرانی شبکه‌های اجتماعی استفاده کرد و برای بهبود سرویس‌های شبکه‌های اجتماعی نظیر سیستم‌های پیشنهاددهنده، بازاریابی ویروسی آنلاین، یافتن انجمن‌ها و ارسال هدفمند پیام‌های تبلیغاتی در جامعه افراد ایرانی به کار برد.

در این مقاله تنها یکی از فاکتورهای شخصیتی مدل پنج عامل بزرگ شخصیتی، یعنی برون‌گرایی مورد بررسی قرار گرفت. در صورت وجود همبستگی بین دیگر فاکتورهای شخصیتی و ویژگی‌های اعضای شبکه اجتماعی، می‌توان دیگر فاکتورهای شخصیتی را نیز به روش مشابه فرموله-سازی کرد. همچنین با توجه به اینکه در اختیار داشتن شخصیت اعضای شبکه اجتماعی می‌تواند نقش مهمی در بهبود سرویس‌های ارائه شده در شبکه اجتماعی داشته باشد، می‌توان کاربرد مدل ارائه شده در این مقاله را برای بهبود نتایج سرویس‌های مختلف شبکه های اجتماعی مورد بررسی قرار داد. همچنین در این مقاله جامعه محدودی از افراد برای ارزیابی مدل ارائه شده جمع‌آوری و استفاده شده است که از محدودیت‌های مدل به شمار می‌آید و می‌توان در کارهای آینده تعداد افراد بیشتری را مورد ارزیابی قرار دارد.

www.SID.ir

Archive of SID

۷- مراجع

[1] J. A. Golbeck, "Computing and applying trust in web-based social networks," PhD, College Park, University of Maryland, 2005.

[2] J. Golbeck, C. Robles, and K. Turner, "Predicting Personality with Social Media," *ACM 978-1-4503-0268*, pp. 253-262, 2011.

[3] N. B. Ellison, "Social network sites: Definition, history, and scholarship," *Journal of Computer Mediated Communication*, vol. 13, pp. 210-230, 2007.

[4] Y. Amichai-Hamburger and G. Vinitzky, "Social network use and personality," *Computers in Human Behavior*, vol. 26, pp. 1289-1295, 2010.

[5] J. Feist, G. J. Feist, and T.-A. Roberts, *Theories of personality*: McGraw-Hill New York, 2006.

[6] M. D. Back, J. M. Stopfer, S. Vazire, S. Gaddis, S. C. Schmukle, B. Egloff, and S. D. Gosling, "Facebook profiles reflect actual personality, not self-idealization," *Psychological Science*, vol. 21, pp. 372-374, 2010.

[7] P. Bonhard and M. Sasse, "'Knowing me, knowing you'—Using profiles and social networking to improve recommender systems," *BT Technology Journal*, vol. 24, pp. 84-98, 2006.

[8] M. Goeksel and C. P. Lam, "System and method for utilizing social networks for collaborative filtering," ed: Google Patents, 2010.

[9] Y.-M. Li, C.-Y. Lai, and C.-H. Lin, "Discovering Influential Nodes for Viral Marketing," in *the 42nd Hawaii International Conference on System Sciences - IEEE*, Hawaii, USA, 2009.

[10] Y.-M. Li, C.-H. Lin, and C.-Y. Lai, "Identifying influential reviewers for word-of-mouth marketing," *Electronic Commerce Research and Applications*, pp. 294–304, 2010.

[11] S. Ahmed and C. Ezeife, "Discovering influential nodes from trust network," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, 2013, pp. 121-128.

[12] Z. Zhu, "Discovering the influential users oriented to viral marketing based on online social network," *Physica A: Statistical Mechanics and its Applications*, 2013.

[13] K. Wakita and T. Tsurumi, "Finding community structure in mega-scale social networks:[extended abstract]," in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 1275-1276.

[14] Y. Zhao, E. Levina, and J. Zhu, "Community extraction for social networks," *Proceedings of the National Academy of Sciences*, vol. 108, pp. 7321-7326, 2011.

[15] P. P. Zubcsek and M. Sarvary, "Advertising to a social network," *Quantitative Marketing and Economics*, vol. 9, pp. 71-107, 2011.

[16] P.-T. Chen and H.-P. Hsieh, "Personalized mobile advertising: Its key attributes, trends, and social impact," *Technological Forecasting and Social Change*, vol. 79, pp. 543-557, 2012.

[17] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *Journal of personality*,vol. 60, pp. 175-215, 1992.

[18] C. Sand, A Guide to the Big Five Personality Traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism: Webster's Digital Services, 2012.

[19] R. R. McCrae and P. T. Costa, "Validation of the five-factor model of personality across instruments and observers," *Journal of personality and social psychology*, vol. 52, pp. 81-90, 1987.

[20] T. Oren and N. Ghasem-Aghaee, "Personality representation processable in fuzzy logic for human behavior simulation," in *Summer Computer Simulation Conference*, 2003, pp. 11-18.

[21] P. T. Costa and R. R. MacCrae, Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO FFI): Professional Manual: Psychological Assessment Resources, 1992.

[22] B. E. de Raad and M. E. Perugini, *Big five assessment*: Hogrefe & Huber Publishers, 2002.

[23] A. Aly and A. Tapus, "A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction," in *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, 2013, pp. 325-332.

[24] A. Zoumpoulaki, N. Avradinis, and S. Vosinakis, "A multi-agent simulation framework for emergency evacuations incorporating personality and emotions," in *Artificial Intelligence: Theories, Models and Applications*, ed: Springer, 2010, pp. 423-428.

[25] M. A. S. Nunes and R. Hu, "Personality-based recommender systems: an overview," in *Proceedings of the sixth ACM conference on Recommender systems*, 2012, pp. 5-6.

[26] M. D. Ekstrand, J. T. Riedl, and J. A. Konstan, "Collaborative filtering recommender systems," *Foundations and Trends in Human-Computer Interaction*, vol. 4, pp. 81-173, 2011.

[27] L. F. Capretz and F. Ahmed, "Making sense of software development and personality types," *IT professional*, vol. 12, pp. 6-13, 2010.

www.SID.ir

Archive of SID

[28] H. Keller and S. J. Karau, "The importance of personality in students' perceptions of the online learning experience," *Computers in Human Behavior*, vol. 29, pp. 2494-2500, 2013.

[29] O. Nov, O. Arazy, C. López, and P. Brusilovsky, "Exploring personality-targeted UI design in online social participation systems," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 361-370.

[30] T. Zhou and Y. Lu, "The effects of personality traits on user acceptance of mobile commerce," *Intl. Journal of Human–Computer Interaction*, vol. 27, pp. 545-561, 2011.

[31] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft, "Our Twitter profiles, our selves: Predicting personality with Twitter," in *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, 2011, pp. 180-185.

[32] Y. Kalish and G. Robins, "Psychological predispositions and network structure: The relationship between individual predispositions, structural holes and network closure," *Social Networks*, vol. 28, pp. 56-84, 2006.

[33] L. Qiu, H. Lin, J. Ramsay, and F. Yang, "You are what you tweet: Personality expression and perception on twitter," *Journal of Research in Personality*, vol. 46, pp. 710-718, 2012.

[34] D. O. Olguın, P. A. Gloor, and A. S. Pentland, "Capturing individual and group behavior with wearable sensors," in *Proceedings of the 2009 aaai spring symposium on human behavior modeling, SSS*, 2009.

[35] S. G. Roberts, R. Wilson, P. Fedurek, and R. Dunbar, "Individual differences and personal social network size and structure," *Personality and Individual Differences*, vol. 44, pp. 954-964, 2008.

[36] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from twitter," in Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom), 2011, pp. 149-156.

[37] D. Quercia, R. Lambiotte, D. Stillwell, M. Kosinski, and J. Crowcroft, "The personality of popular facebook users," in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, 2012, pp. 955-964.

[38] J. Staiano, B. Lepri, N. Aharony, F. Pianesi, N. Sebe, and A. Pentland, "Friends don't lie: inferring personality traits from social network structure," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 2012, pp. 321-330.

[39] L. Li, A. Li, B. Hao, Z. Guan, and T. Zhu, "Predicting Active Users' Personality Based on Micro-Blogging Behaviors," *PloS one*, vol. 9, p. e84997, 2014.

[40] م. گ. فرشی, رویکردی نوین در ارزیابی شخصیت. تبریز: نشر جامعه پژوه, ۱۳۸۰.

[41] J. Scott, *Social network analysis*: SAGE Publications Limited, 2012.

[42] G. Holmes, M. Hall, and E. Prank, *Generating rule sets from model trees*: Springer, 1999.

[43] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, pp. 10-18, 2009.

[44] R. Poli and J. Koza, *Genetic Programming*: Springer, 2014.

[45] D. Searson, "GPTIPS: Genetic programming & symbolic regression for MATLAB," *Available from gptips. sourceforge. net,[Accessed 10 February, 2010]*, 2009.

www.SID.ir