پیش‌بینی تیپ شخصیتی کاربران براساس روابط ضمنی در شبکه‌های اجتماعی چندگانه

فاطمه کریمی[1*]، استادیار، خیام صالحی[2]، استادیار

[1] گروه علوم کامپیوتر- دانشکده علوم ریاضی ـ دانشگاه کاشان ـ کاشان ـ ایران ـ fkarimi@kashanu.ac.ir

[2] گروه علوم کامپیوتر- دانشکده علوم ریاضی ـ دانشگاه شهرکرد- شهرکرد- ایران ـkh.salehi@sku.ac.ir

چکیده: امروزه تحلیل رفتار کاربران در شبکه‌های اجتماعی برای درک ویژگی‌های روانشناختی اهمیت فزاینده‌ای یافته است. باتوجه‌به حضور همزمان کاربران در شبکه‌های اجتماعی متنوع، ساختاردهی داده‌های استخراج شده از این پلتفرم‌ها در قالب یک شبکه‌ی چندگانه می‌توانند اطلاعات ارزشمندی را ارائه دهند. با این حال، رویکردهای مرسوم به دلیل اتکا بر روابط صریح که اغلب خصوصی هستند و همچنین استفاده از داده‌های تک‌منبعی، دارای محدودیت می‌باشند. بنابراین، در این مقاله یک چارچوب نوین دو مرحله‌ای برای پیش‌بینی تیپ شخصیتی کاربران ارائه شده است که نوآوری اصلی آن، استفاده از شبکه‌ی روابط ضمنی استخراج شده از داده‌های چندمنبعه و چندحالته شامل متن، تصویر و اطلاعات مکانی به جای تکیه بر روابط صریح دوستی کاربران است. در مرحله‌ی اول، ابتدا با استفاده از یک رویکرد تکاملی چندهدفه (MOEA/D-TS) ساختار جوامع در شبکه‌ی چندگانه شناسایی شده و سپس با استفاده از یک روش پیش‌بینی پیوند (CLPES)، شبکه روابط ضمنی میان کاربران استخراج می‌گردد. در مرحله‌ی دوم، از شبکه‌ی حاصل به عنوان یک منبع اطلاعاتی غنی در یک چارچوب مبتنی‌بر روش‌های یادگیری ماشین به‌منظور پیش‌بینی تیپ شخصیتی کاربران بهره‌گیری می‌شود. ارزیابی‌ها برتری قابل توجه چارچوب پیشنهادی را نشان می‌دهد به‌طوری که مدل نهایی با استفاده از جنگل تصادفی و همجوشی هر سه منبع داده، به میانگین امتیاز Macro-F1 برابر با ۰,۶۷۳ دست یافت که نشان‌دهنده بهبود عملکردی به طور میانگین به میزان ۲۳ درصد نسبت به مدل‌های تک‌منبعه است. این یافته، ارزش راهبرد تحلیل روابط ضمنی و همجوشی داده‌های چندمنبعه را برای درک عمیق‌تر شخصیت کاربران تأیید می‌کند.

واژه‌های کلیدی: شبکه‌های اجتماعی، پیش‌بینی شخصیت، روابط ضمنی، شبکه‌ی چندگانه، تشخیص جوامع، پیش‌بینی پیوند، رویکرد تکاملی و یادگیری ماشین

* فاطمه کریمی، **fkarimi@kashanu.ac.ir**

User Personality Type Prediction based on Implicit Relations in Multiplex Networks

Fatemeh Karimi [1*], Assistant professor, Khayyam Salehi [2], Assistant professor

[1] Dept. of Computer Science, Faculty of Mathematical Sciences, University of Kashan, Kashan, Iran, fkarimi@kashanu.ac.ir

[2] Dept. of Computer Science, Faculty of Mathematical Sciences, Sharekord University, Sharekord, Iran, kh.salehi@sku.ac.ir

Abstract: Analyzing user behavior on social networks to understand psychological traits is of growing importance. Structuring data extracted from users' concurrent activity across various platforms as a multiplex network can provide valuable insights. However, conventional approaches are limited by their reliance on explicit relationships, which are often private, and their use of single-source data. Therefore, this paper proposes a novel two-stage framework for personality type prediction. Its primary innovation is the use of an implicit relationship network, extracted from the fusion of multi-source, multi-modal data including text, image, and location, instead of relying on explicit friendship links. In the first stage, community structures in the multiplex network are identified using a multi-objective evolutionary approach (MOEA/D-TS), and then the implicit relationship network is constructed using a link prediction method (CLPES). In the second stage, the resulting network is leveraged as an information source within a machine learning framework to predict user personality types. Evaluations demonstrate the proposed framework's significant superiority. The final model, using a Random Forest classifier with the fusion of all three data sources, achieved an average Macro-F1 score of 0.673. This represents an average performance improvement of 23% compared to single-source models. This finding confirms the value of analyzing implicit relations and fusing multi-source data for a deeper understanding of user personality.

Keywords: *Social Networks, Personality Prediction, Implicit Relationships, Multiplex Network, Community Detection, Link Prediction, Evolutionary Approach and Machine Learning*

*** Corresponding author, Fatemeh Karimi, fkarimi@kashanu.ac.ir**

۱. مقدمه

فعالیت کاربران در پلتفرم‌های اجتماعی، حجم عظیمی از داده‌های پیچیده رفتاری را تولید می‌کند که بازتابی از ترجیحات، باورها و الگوهای ارتباطی افراد محسوب می‌شود. این داده‌های عظیم چالش‌های تحلیلی نوینی را در حوزه رفتارشناسی دیجیتال ایجاد کرده‌اند. تحلیل هوشمندانه‌ی محتوای تولید شده به‌وسیله‌ی کاربران و الگوهای تعاملی آنها می‌تواند به عنوان جایگزینی کارآمد برای روش‌های سنتی (مانند پرسش‌نامه‌های شخصیتی) در شناسایی ویژگی‌های روانشناختی افراد عمل کند. این ویژگی‌های شخصیتی تأثیر بسزایی در ترجیحات فرهنگی، هنری و سبک زندگی افراد دارند. در تحقیقات اخیر وجود رابطه بین شخصیت یک فرد و دوستان وی به اثبات رسیده است [۱]، [۲]. با این حال، محدودیت‌های دسترسی به این داده‌ها به دلایل مختلف مانند ملاحظات امنیتی و حفظ حریم خصوصی کاربران، تحلیل آنها را با دشواری مواجه ساخته است.

در فضای رسانه‌های اجتماعی، با حجم انبوهی از کاربران و تعاملات میان آنها، روابط آشکار میان کاربران صرفاً بخش محدودی از رفتارهای تعاملی را نشان می‌دهد، حال آنکه بسیاری از کاربران با علائق مشترک ممکن است هیچ ارتباط مستقیمی با یکدیگر نداشته باشند. در تحلیل این شبکه‌ها می‌توان دو نوع رابطه را متمایز نمود: روابط صریح[1] که شامل کنش‌های مستقیم و قابل مشاهده بین کاربران است و روابط ضمنی[2]، که از طریق شناسایی اشتراکات در زمینه‌های مختلفی چون الگوهای علایق، شباهت‌های فکری و یا رفتارهای تعاملی مشابه قابل استخراج هستند. این الگوهای رفتاری مشابه، پایه‌ای برای شناسایی روابط پنهان بین کاربران محسوب می‌شود [۳].

اگرچه پژوهش‌های موفقیت‌آمیزی جهت کشف روابط ضمنی بین کاربران صورت گرفته است، اما تحقیقات پیشین به‌طور عمده بر داده‌های تک‌منبعی از یک پلتفرم واحد متمرکز بوده‌اند که این امر دقت نتایج را تحت تأثیر قرار داده است [۴]. این در حالی است که کاربران امروزی اغلب در چندین پلتفرم مجازی به‌صورت همزمان فعالیت دارند و ادغام داده‌های چندمنبعی می‌تواند بینش‌های عمیق‌تری ارائه نماید. همچنین، اغلب پژوهش‌ها از روش‌های پیش‌بینی پیوند[3] شامل روش‌های مبتنی‌بر ساختار و خصیصه‌ی گره‌ها به‌منظور کشف روابط ضمنی بهره‌گیری کرده‌اند. با این حال، یک جنبه‌ی مهم که تاکنون توجه قرار نگرفته است مشابهت در الگوهای نقل مکان کاربران است.

از این رو، در پژوهش حاضر با هدف رفع این محدودیت‌ها، یک چارچوب نوین دو مرحله‌ای پیشنهاد شده است که مرحله‌ی اول با تلفیق روش‌های تشخیص جوامع[4] و پیش‌بینی پیوند[5] به کشف روابط ضمنی بین افراد می‌پردازد و در گام دوم سعی بر آن خواهد بود که وجود رابطه بین خصوصیات شخصیتی یک کاربر و دوستان ضمنی وی مورد تجزیه‌وتحلیل قرار گیرد.

براین اساس می‌توان دو فرضیه کلی را برای این پژوهش در نظر گرفت:

- فرضیه اول (H_1): پیش‌بینی تیپ شخصیتی کاربران براساس شبکه‌ی روابط ضمنی نسبت به حالت منفرد (مبتنی بر ویژگی‌های رفتاری هر فرد) با دقت بالاتری صورت خواهد گرفت.
- فرضیه دوم (H_2): بهره‌گیری از داده‌های چندمنبعه در مورد اطلاعات رفتاری کاربران در قالب یک شبکه‌ی چندگانه، دقت پیش‌بینی ویژگی‌های شخصیتی کاربران را افزایش دهد.

[1] explicit
[2] implicit
[3] link prediction
[4] community detection
[5] link prediction

سهم و نوآوری‌های اصلی این مقاله را می‌توان به شرح زیر خلاصه کرد:

- ارائه یک چارچوب نوین برای پیش‌بینی ویژگی‌های شخصیتی مبتنی‌بر روابط ضمنی: نتایج این تحقیق نشان می‌دهد که بدون دسترسی به گراف دوستی صریح کاربران، می‌توان با استخراج و تحلیل روابط ضمنی، به پیش‌بینی قابل قبولی از ویژگی‌های شخصیتی کاربران دست یافت.
- توسعه یک چارچوب همجوشی چندمنبعه: یک چارچوب جامع جهت تلفیق داده‌های ناهمگون متنی، تصویری و مکانی به منظور ساخت یک شبکه از روابط ضمنی غنی و معنادار بین کاربران ارائه شده است.
- اعتبارسنجی جامع تجربی: کارایی چارچوب پیشنهادی بر روی یک مجموعه داده واقعی و بزرگ ارزیابی شده و برتری آن نسبت به روش‌های متکی بر داده‌های فردی و تک‌منبعه از طریق تحلیل‌های آماری معناداری به اثبات رسیده است.

نتایج این پژوهش می‌تواند در حوزه‌هایی متنوعی نظیر «بازاریابی شخصی‌سازی‌شده[1]» برای پیشنهاد محصولات یا فعالیت‌های متناسب با شخصیت کاربر، «سامانه‌های توصیه‌گر» برای پیشنهاد محتوا یا دوستان جدید، «مدیریت منابع انسانی» به‌منظور تسهیل فرایند جذب و تشکیل تیم‌های کاری سازگار و حتی حوزه‌ی «سلامت روان» برای شناسایی الگوهای رفتاری مرتبط با اختلالات شخصیتی کاربردهای عملی گسترده‌ای داشته باشد.

ادامه مقاله به‌صورت زیر سازماندهی شده است. در بخش دوم به تعریف مفاهیم اولیه و الگوریتم‌های مورد استفاده پرداخته شده است. در بخش سوم به طور مختصر کارهای گذشته مرور می‌شود. در بخش چهارم مدلسازی مسئله‌ی موردنظر و راه‌کار پیشنهادی با جزییات شرح داده شده است و در بخش پنجم به ارزیابی آن پرداخته می‌شود. در نهایت در بخش پایانی نتیجه‌گیری ارائه می‌شود.

۲. مفاهیم پایه‌ای

در این بخش ابتدا کلیدی‌ترین اصطلاحات مورد استفاده در این پژوهش تشریح می‌شوند. سپس دو راهکار اساسی مورد استفاده جهت تشخیص جوامع و پیش‌بینی پیوند با جزئیات بیشتر مورد بررسی قرار می‌گیرند.

۱،۲. اصطلاحات

مطالعه‌ی شبکه‌های اجتماعی و بررسی روابط بین کاربران مستلزم اصطلاحاتی است که در این بخش به‌اختصار شرح داده شده‌اند.

شبکه چندلایه[2]: شبکه‌هایی که از چندین لایه‌ی مختلف تشکیل شده‌اند و هر نوع از تعامل‌های بین گره‌ها به‌وسیله‌ی یک لایه‌ی مجزا نمایش داده می‌شود.

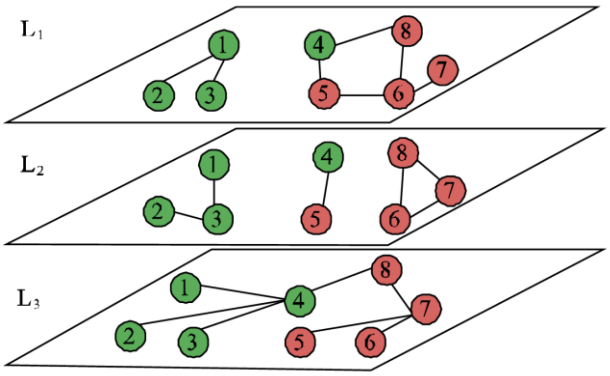
شبکه چندگانه: زیرمجموعه‌ای از شبکه‌های چندلایه که در آن یک مجموعه ثابت از گره‌ها در تمام لایه‌ها حضور دارند، هر لایه نوع خاصی از ارتباط بین همان گره‌ها را نمایش می‌دهد و روابط در لایه‌های مختلف می‌توانند ماهیت کاملاً متفاوتی داشته باشند. شکل *۱*: **یک نمونه از شبکه‌ی چندگانه با سه لایه**۱ نمونه‌ای از یک شبکه‌ی چندگانه با سه لایه را نمایش می‌دهد. لازم به ذکر است که در این مقاله تمرکز اصلی بر شبکه‌های چندگانه می‌باشد.

روابط صریح: به کنش‌های مستقیم و قابل مشاهده در پلتفرم‌های رسانه‌های اجتماعی، مانند برقراری ارتباط دوستی در فیسبوک[3] با ارسال مستقیم «درخواست[4]» اشاره دارد [۳].

روابط ضمنی: این روابط که بر خلاف روابط صریح از تعاملات مستقیم نشأت نمی‌گیرند، از طریق تحلیل شباهت‌ها در

[1] Personalized Marketing
[2] multi-layer network
[3] Facebook
[4] friend request

رفتار، علایق یا الگوهای فکری میان کاربران استخراج می‌شوند و به صورت مستقیم قابل مشاهده نیستند [۳].



شکل ۱: یک نمونه از شبکه‌ی چندگانه با سه لایه

تیپ شخصیتی: برای ارزیابی شخصیت، چندین مقیاس معتبر توسط جامعه علمی ارائه شده که از مهم‌ترین آنها می‌توان به مدل پنج عاملی شخصیت (FFM[۱]) و نظریه‌ی تیپ‌های شخصیتی مایرز-بریگز (MBTI[۲]) اشاره کرد. پژوهش حاضر از مدل MBTI بهره می‌برد که شخصیت افراد را بر اساس چهار بعد دوگانه و شانزده تیپ مختلف دسته‌بندی می‌کند. براساس این مدل، تیپ شخصیتی افراد با چهار حرف نشان داده می‌شود که بیانگر موارد زیر است:

- شیوه دریافت انرژی فرد (درون‌گرا[۳] I، برون‌گرا[۴] E): این بعد بر مبنای چگونگی برخورد و تعامل افراد با دنیا و اینکه انرژی خود را متوجه کجا می‌کنند و از کجا انرژی دریافت می‌کنند، تعریف شده است.
- انواع اطلاعاتی را که افراد بیشتر متوجه می شوند (حسی[۵] S، شهودی[۶] N): نوع اطلاعاتی که اغلب مورد توجه افراد قرار می‌گیرد و همچنین چگونگی به‌دست آوردن آنها این بعد را مشخص می‌کند.
- چگونگی تصمیم‌گیری فرد (فکری[۷] T، احساسی[۸] F): این معیار بر چگونگی پردازش اطلاعات و نوع تصمیم‌گیری هر شخص دلالت دارد.
- شیوه‌ی نگرش به دنیای بیرون (ملاحظه‌کننده[۹] P، داوری‌کننده[۱۰] J): این امر که افراد ترجیح می‌دهند در شرایط سازمان یافته و با ساختار فعالیت کنند یا استقلال بیشتری داشته باشند، تعیین کننده‌ی این بعد خواهد بود.

همان‌طور که در شکل ۲ مشاهده می‌شود، هر یک از این تیپ‌های شخصیتی دارای ویژگی‌های متمایز و منحصربه‌فردی هستند.

INTJ مستقل، نوآور، تحلیلگر	ISTJ مسئولیت‌پذیر، منطقی، منظم	INFJ خلاق، آرمان‌گرا، دلسوز	ISFJ وظیفه‌شناس، عملگرا، حامی
INTP جستجوگر، خلاق، متفکر	INFP آرمان‌گرا، منعطف، خردمند	ISFP صبور، واقع‌گرا، سازگار	ISTP بی‌طرف، منعطف، نتیجه‌گرا
ENTP جسور، اجتماعی، خلاق	ENFP خوش‌بین، نوآور، دلسوز	ESFP خوش‌مشرب،سازگار، تکانشی	ESTP اهل عمل، پرانرژی، تکانشی
ENTJ رهبر، قاطع، برنامه‌ریز	ENFJ خوش‌اخلاق، حامی، آرمان‌گرا	ESFJ سازمان‌یافته، وفادار، حامی	ESTJ ساختارگرا، اجتماعی، عملگرا

شکل ۲: شانزده تیپ شخصیتی بر اساس مدل MBTI [۵]

تحقیقات علوم رفتاری نشان می‌دهد پلتفرم‌های رسانه‌های اجتماعی تأثیرات عمیقی بر الگوهای ارتباطی بین فردی و تعامل با محیط اجتماعی برجای گذاشته‌اند [۶] که حاکی از آن است که نوع شناسی[۱۱] MBTI به دلیل ماهیت چندبعدی خود، چارچوب مناسبی برای تحلیل رفتارهای دیجیتال کاربران ارائه می‌دهد و می‌تواند هنگام اختصاص دادن برچسب‌های شخصیتی به داده‌های کاربر، برمبنای رفتارها و فعالیت‌ها در رسانه‌های اجتماعی، مورد استفاده قرار گیرد. علیرغم مناقشات درباره اعتبار علمی مقایسه‌ای

[1] Five-Factor Model
[2] Myers-Briggs Type Indicator
[3] introverted
[4] extroverted
[5] sensing
[6] intuitive
[7] Thinking
[8] feeling
[9] perceiving
[10] judging
[11] typology

مدل‌های FFM و MBTI، شواهد تجربی حاکی از برتری کاربردی MBTI در زمینه‌های صنعتی و ارزیابی‌های شخصیتی است [۷].

۲،۲. تشخیص جوامع با روش MCDMT[1]

جوامع، خوشه‌ها[2] یا پیمانه‌ها[3] به گروه‌هایی از رأس‌ها اطلاق می‌شوند که تمایل بالاتری برای تشکیل ساختارهای همگن در شبکه دارند. فرآیند شناسایی جوامع در واقع یک مسئله بهینه‌سازی است که هدف آن یافتن چینش مطلوبی از گره‌هاست؛ به گونه‌ای که گره‌های با ویژگی‌های مشابه در یک گروه قرار گیرند و تعامل بین جوامع مختلف به میزان کمینه برسد.

در سال‌های اخیر، مطالعات گسترده‌ای در حوزه کشف جوامع صورت گرفته است [۸]، [۹] که بسیاری از آنها بر تحلیل شبکه‌های تک‌لایه متمرکز بوده‌اند. ساختار جوامع در این نوع شبکه‌ها، از مجموعه‌ی مشابهی از گره‌ها تشکیل شده است با این حال، در شبکه‌های چندلایه، تعریف یکپارچه‌ای برای جوامع وجود ندارد، چرا که به طور معمول الگوهای متفاوتی از جوامع در لایه‌های مختلف قابل مشاهده است [۱۰]. ازاین‌رو، نتیجه تشخیص جامعه در این ساختارها نمی‌تواند یک ساختار دقیق از جوامع را در شبکه‌های چندلایه تعیین کند. علاوه‌براین، برخی رویکردهای متداول مانند ادغام لایه‌ها و خوشه‌بندی تلفیقی نیز با محدودیت‌هایی مواجهند. این روش‌ها عموماً به تعریف سراسری جوامع تجمیعی[4] در شبکه‌های چندلایه توجه ندارند، روابط بین جوامع را در نظر نمی‌گیرند و تحلیل کلی ساختار چندلایه را نادیده می‌گیرند [۱۰].

در روش MCDMT، الگوریتم تکاملی MOEA/D-TS که تلفیقی از جستجوی سراسری و محلی را به کار می‌گیرد، به عنوان راهکار مناسبی برای مسئله شناسایی جوامع در شبکه‌های چندلایه مورد استفاده قرار گرفته است [۱۱]. انگیزه اصلی استفاده از MOEA/D-TS برای حل این مسئله، پیچیدگی محاسباتی پایین و قابلیت تحسین‌برانگیز MOEA/D برای تولید راه‌حل‌های نامغلوب[5] در مسائل بهینه‌سازی چندهدفه به‌همراه جستجوی محلی جامع الگوریتم IDTS[6] [۱۲] برای کاهش مناطق غیرقابل کشف در فضای مسئله و سرعت بخشیدن به روند همگرایی است.

با داشتن یک شبکه‌ی چندگانه G_L با l لایه به‌صورت $L = \{L_1, L_2, \dots, L_l\}$، مسئله‌ی موردنظر، چگونگی شناسایی ساختارهای مشترک جوامع در شبکه با بهینه‌سازی یک تابع هدف است. در اینجا مسئله تشخیص جوامع در شبکه‌های چندگانه در قالب یک چارچوب بهینه‌سازی چندهدفه ارائه گردیده که ترکیبی از دو معیار ارزیابی را به کار می‌گیرد: معیار پیمانگی[7] (Q) برای هر لایه‌ی L_i و همچنین معیار NMI[8] برای ساختار جامعه به‌دست‌آمده در هر زوج لایه L_i و L_{i-1}. سپس از MOEA/D-TS بهبود یافته به‌منظور بهینه‌سازی تابع هدف موردنظر در یک روند تکراری بهره‌گیری شده است. بنابراین، می‌توان مسئله‌ی موردنظر را به‌صورت بیشینه کردن پیمانگی Q برای هر لایه L_i و NMI برای هر زوج لایه L_i و L_{i-1} در شبکه به شرح زیر فرموله کرد:

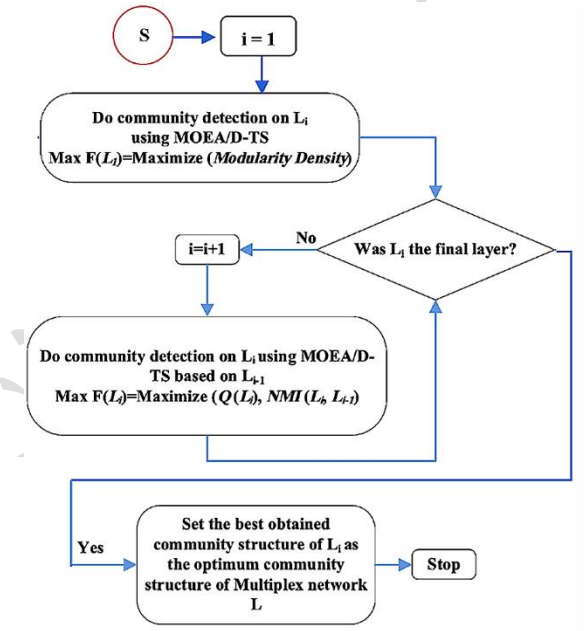
$$MaxF(L_i) = Maximize\big(Q(L_i), NMI(L_i, L_{i-1})\big) \qquad (۱)$$

به بیان دیگر، ابتدا جوامع موجود در لایه اول شبکه شناسایی می‌شوند و سپس با بیشینه‌سازی پیمانگی Q برای هر لایه و NMI بین لایه‌ی فعلی و قبلی، روند تشخیص جوامع برای لایه‌های دیگر ادامه می‌یابد. در نهایت، ساختار جوامع شناسایی‌شده در آخرین لایه (L_l)، به عنوان الگوی جامع و مشترک برای کل شبکه چندلایه در نظر گرفته می‌شود.

---

[1] Multiplex Community Detection based on MOEA/D-TS
[2] clusters
[3] modules
[4] collective communities
[5] non-dominated
[6] Improved Diversificator Tabu Search
[7] modularity
[8] Normalized Mutual Information

شکل ۳ مراحل کلی روش MCDMT را برای شناسایی جوامع در شبکه‌های چندگانه را نمایش می‌دهد. شرح کامل جزئیات فنی الگوریتم MOEA/D-TS و تحلیل جامع عملکرد آن خارج از محدوده این مقاله است. خوانندگان علاقه‌مند برای بررسی دقیق‌تر این الگوریتم و معادلات آن می‌توانند به پژوهش پیشین ما در مرجع [۱۱] مراجعه نمایند.



شکل (۳): کارنمای روش MCDMT برای تشخیص جوامع [۱۱]

۳،۲. پیش‌بینی پیوند با روش CLPES

پیش‌بینی پیوند شامل شناسایی پیوندهای ناموجود، پیش‌بینی پیوندهای آتی یا تشخیص پیوندهای جعلی[1] در یک شبکه می‌باشد [۱۳]. در اغلب روش‌های موجود، پیش‌بینی پیوند صرفاً بر اساس داده‌های استخراج‌شده از گره‌ها یا پیوندها صورت می‌گیرد. بااین‌حال، در سامانه‌های واقعی نظیر شبکه‌های اجتماعی، جریان اطلاعات در قالب تعاملات گروهی در مجموعه‌هایی متراکم (جوامع) نمود می‌یابد، به‌گونه‌ای که این روابط جمعی جایگزین ارتباطات فردی می‌شوند. ازاین‌رو، اطلاعات مربوط به پیوندها به‌همراه اطلاعات جوامع می‌بایست در الگوریتم‌های پیش‌بینی پیوند لحاظ شوند. بنابراین، الگوریتم‌های پیش‌بینی پیوند باید علاوه بر داده‌های پیوندها، اطلاعات مربوط به ساختار جوامع را نیز در نظر بگیرند. این امتیازها برای تمامی پیوندهای مفقود[2] در شبکه از طریق روش موردنظر ارزیابی می‌شوند.

روش پیشنهادی ما در [۱۴]، موسوم به CLPES، یک چارچوب نوین برای پیش‌بینی پیوند در شبکه‌های چندلایه ارائه می‌دهد که در آن از ویژگی‌های داخلی و شباهت خارجی در کنار ساختار جوامع در شبکه‌ها بهره‌گیری می‌کند. این روش قادر است با بهره‌گیری از اطلاعات بین لایه‌ای و الگوهای ارتباطی درون جوامع، دقت پیش‌بینی پیوندها را در هر لایه بهبود بخشد. در گام نخست از روش CLPES، ساختار جوامع شبکه با استفاده الگوریتم MOEA/D-TS شناسایی می‌شود. پس از آن، احتمال تشکیل پیوند با استفاده از ویژگی‌های داخلی لایه‌ها محاسبه می‌شود. سرانجام، با تلفیق این ویژگی‌ها و معیار جدید شباهت خارجی[3]، مقادیر نهایی احتمال تشکیل پیوند تعیین می‌شود.

برای گراف $G(V,E)$ با ساختار جوامع مشخص، به پیوندهای بین هر جفت گره (i,j) وزن‌های مختلفی تخصیص داده می‌شود. این یال‌ها شامل پیوندهای موجود $e_{ij} \in E$ و پیوندهای بالقوه $e_{ij} \in U - E$ هستند که در آن U مجموعه سراسری یال‌ها برای گراف G است [۱۳]. براین‌اساس، تمام پیوندها بین گره‌های متعلق به یک جامعه، وزن‌های مثبت دریافت می‌کنند، درحالی‌که سایر پیوندها وزن منفی دریافت می‌کنند. برای هر پیوند $e_{ij} \in U$، میزان وزن‌دهی مبتنی بر جامعه (CEW) به‌صورت زیر محاسبه می‌شود:

$$CWE(e_{ij}) = \begin{cases} +\frac{\gamma_i}{n}, & if\ C(i) = C(j) \\ -\frac{\gamma_i}{n}, & otherwise \end{cases} \quad (۲)$$

که در آن، n تعداد گره‌ها در G یعنی $|V|$ است و γ_i اندازه جامعه‌ای که شامل گره i است را نشان می‌دهد. همچنین، $C(i)$ و $C(j)$ نیز

[1] fictitious links
[2] missing links
[3] External Similarity (ExSim)

برچسب جوامعی که گره‌های i و j به آنها تخصیص یافته‌اند را مشخص می‌کنند. این رابطه، یک وزن اولیه بر اساس ساختار جامعه به هر پیوند بالقوه اختصاص می‌دهد؛ به طوری که پیوندهای درون یک جامعه منسجم، امتیاز بالاتری دریافت می‌کنند. در مواردی که شبکه مورد مطالعه دارای ساختار وزن‌دار باشد، وزن نهایی هر یال از ترکیب خطی مقدار حاصل برای CEW و وزن قبلی محاسبه می‌شود. یک شبه کد از روش CLPES برای پیش‌بینی پیوند چندگانه در الگوریتم ۱ ارائه شده است [۱۴].

در یک شبکه چندگانه G_L با l لایه $L = \{L_1, L_2, \dots, L_l\}$ احتمال کلی تشکیل پیوند بین دو گره i و j در لایه L_t به شکل زیر محاسبه می‌شود:

$$S_{L_t}(i,j) = (1-\varphi).IS_{L_t}(i,j) + \varphi.ES_{L_t}(i,j) \qquad (۳)$$

الگوریتم (۱): چارچوب کلی روش CLPES

Input: adjacency matrix *A* for multiplex network G_L
Output: score *S* of links in each layer
a_{ij} : The corresponding element in adjacency matrix A(i,j)
Step 1: multiplex community detection
- Determine the community structure of the network using MOEA/D-TS
- CEW ← compute the community-based weight for all edge $e_{ij} \in U$

Step 2: for any layer L_k in the multiplex network:
- *IS* ← Compute likelihood score of a link using internal features

Step 3: for *t = 1* to *l*
for any pair of nodes (i, j)
$$ES_{L_t}(i,j) = \sum_{k=1,k\neq t}^{l} CEW(e_{ij}) \times (\tilde{a}_{ij} - IS_{L_t}(i,j)) \times (\tilde{a}_{ij} - S_{L_tL_k}(i,j))$$
Where $\tilde{a}_{ij} = a_{ij}^{L_t}$
$$S_{L_t}(i,j) = (1-\varphi).IS_{L_t}(i,j) + \varphi.ES_{L_t}(i,j)$$
end
end

مقادیر به‌دست‌آمده برای $S_{L_t}(i,j)$ امتیاز احتمال وجود پیوند بین جفت گره (i,j) در لایه L_t را با تکیه بر ویژگی‌های داخلی لایه، نشان می‌دهد. این امتیاز در گام ۲ الگوریتم با استفاده از معیارهای ساختاری مختلف برای یافتن شباهت، محاسبه می‌شود. در این رابطه، $ES_{L_t}(i,j)$ امتیاز وجود پیوند بین جفت گره (i,j) در لایه L_t را نشان می‌دهد که براساس اطلاعات خارجی لایه‌ها به‌دست می‌آید. پارامتر φ همچنین اثر هر قسمت از رابطه را کنترل می‌کند. پس از مشخص کردن معیار شباهت جفت گره‌ها، نتیجه نهایی احتمال وجود یک پیوند با استفاده از یک روش برچسب زدن آستانه [1] مشخص می‌شود که از رابطه زیر به‌دست می‌آید:

$$ES_{L_t}(i,j) = \sum_{k=1,k\neq t}^{l} CEW(e_{ij}) \times (\tilde{a}_{ij} - IS_{L_t}(i,j)) \times (\tilde{a}_{ij} - S_{L_tL_k}(i,j)) \qquad (۴)$$

در این رابطه، $CEW(e_{ij})$ امتیاز احتمال شکل‌گیری یک پیوند بین دو گره i و j را تعیین می‌کند و $S_{L_tL_k}(i,j)$ شباهت جفت گره (i,j) بین دو لایه L_t و L_k را محاسبه می‌کند. در رابطه فوق $\tilde{a}_{ij}$ متمم عنصر مربوطه در ماتریس مجاورت $A(i,j)$ در لایه L_t است. چنانچه یک پیوند گره‌های i و j را در این لایه متصل کند، $\tilde{a}_{ij} = 0$ و در غیر این‌صورت $\tilde{a}_{ij} = 1$ خواهد بود. این رابطه به تحلیل نقش تأثیر ساختار جامعه در کنار معیارهای شباهت خارجی و داخلی بر صحت روش پیش‌بینی پیوند می‌پردازد. به بیان دیگر، این رابطه با در نظر گرفتن همزمان سه عامل کلیدی، یعنی تعلق به یک جامعه مشترک، شباهت درونی کاربران در همان لایه و شباهت خارجی آنها که از رفتارشان در لایه‌های دیگر نشأت می‌گیرد، به تحلیل صحت و احتمال وجود یک پیوند می‌پردازد. این رویکرد یک دید جامع و چندبعدی برای کشف روابط فراهم می‌کند. خوانندگان علاقه‌مند برای بررسی دقیق‌تر این چارچوب، به مرجع [۱۴] ارجاع داده می‌شوند.

۳. پیشینه‌ی پژوهش

در سال‌های اخیر و با توجه به فراگیر شدن شبکه‌های اجتماعی مطالعات وسیعی در راستای بررسی رابطه بین فعالیت‌های کاربران در پلتفرم‌های مختلف و ارتباط آن با ویژگی‌های شخصیتی کاربران

[1] threshold labeling

صورت گرفته است[۱۵]، [۱۶]، [۱۷] که در ادامه به بررسی مهم‌ترین پژوهش‌ها در این حوزه می‌پردازیم.

در پژوهش پیشگامانه‌ای که توسط گولدبرگ و همکاران [۱۸] انجام شد، برای اولین بار امکان پیش‌بینی ویژگی‌های شخصیتی افراد از طریق تحلیل فعالیت‌های آنها در رسانه‌های اجتماعی مورد بررسی قرار گرفت. این مطالعه نشان داد که بین محتوای پست‌های کاربران و خصوصیات شخصیتی آنان ارتباط معناداری وجود دارد. در این تحقیق، با تحلیل دو دسته ویژگی شامل ویژگی‌های زبانی[۱] و ویژگی‌های اجتماعی (شامل تعداد دوستان، وضعیت رابطه) امتیازهای شخصیتی ۲۷۹ کاربر فیسبوک پیش‌بینی شد. در این تحقیق یک برنامه‌ی کاربردی مبتنی‌بر فیسبوک ایجاد شده است که یک آزمون کوتاه شخصیتی برای کاربران اجرا کرده و با بهره‌گیری از تکنیک‌های پیشرفته‌ی پردازش داده، امکان تحلیل همزمان نتایج آزمون شخصیتی و اطلاعات پروفایل را برای پژوهشگران فراهم ساخته است.

پژوهش انجام شده توسط باچ‌راچ و همکاران [۱۹] ارتباط معناداری بین فعالیت‌های کاربران در فیسبوک و ویژگی‌های شخصیتی آنها بر اساس مدل پنج عامله (FFM) را آشکار ساخت. در این مطالعه، محققان با بهره‌گیری از مجموعه‌ی داده‌ای موسوم به myPersonality [۱۹]، همبستگی میان تیپ‌های شخصیتی و شاخص‌های پروفایل فیسبوک را مورد بررسی قرار دادند. این ویژگی‌ها شامل اندازه و چگالی شبکه‌ی دوستان، تعداد تصویرهای بارگذاری شده، تعداد رخدادهای دنبال شده، تعداد عضویت در گروه‌ها و تعداد دفعاتی که کاربر در تصویرها ضمیمه شده، تعیین می‌گردد.

ورهوون و همکاران [۲۰] در مطالعه‌ی خود به بررسی امکان پیش‌بینی شاخص‌های شخصیتی MBTI از طریق تحلیل محتوای توییت‌ها پرداختند. آنها نشان دادند که ابعاد بعضی ویژگی‌های MBTI مانند «درونگرایی/برونگرایی» و «احساسی/فکری» با دقت قابل قبولی از طریق سبک نوشتاری کاربران قابل پیش‌بینی هستند درحالی‌که پیش‌بینی دو دسته دیگر چالش‌برانگیزتر است. محققان یک مجموعه داده‌ای به نام TWISTY از توییت‌های کاربران ارائه دادند که با اطلاعات تیپ شخصیتی کاربران براساس مدل MBTI و اطلاعات درباره جنسیت کاربران حاشیه‌نویسی[۲] شده است. روش مورد استفاده جهت پیش‌بینی شامل ماشین بردار پشتیبان طبقه‌بند خطی[۳] و رگرسیون لجستیک[۴] است.

پژوهش [۲۱] رویکردی نوین برای پیش‌بینی ویژگی‌های شخصیتی انسان از طریق پلتفرم‌های شبکه‌های اجتماعی (فیسبوک، توییتر و اینستاگرام) با استفاده از مدل پنج‌عاملی شخصیت ارائه می‌دهد. در این مطالعه، یک مدل Transformer با تقسیم‌بندی دوتایی (BPT[۵]) در ترکیب با روش فراوانی واژه و لحظه جاذبه معکوس (TF-IGM) برای استخراج و تحلیل ویژگی‌های شخصیتی از داده‌های متنی شبکه‌های اجتماعی معرفی شده است. این روش‌شناسی چندین تکنیک استخراج ویژگی از جمله ابزارهای تحلیل زبان‌شناختی را به کار گرفته و از همبستگی پیرسون برای انتخاب ویژگی‌ها استفاده می‌کند، سپس با طبقه‌بندی ماکزیمم آنتروپی به پیش‌بینی شخصیت می‌پردازد.

مورایس و همکاران [۲۲] مجموعه‌ای از رویکردهای شناخته شده‌ی یادگیری ماشین را برای پیش‌بینی دسته‌بندی‌های MBTI براساس ویژگی‌های استخراج شده از داده‌های متنی توسط تکنیک‌های پردازش زبان طبیعی (NLP[۶]) به‌کار بردند. نتایج این تحقیق حاکی از این است که متن‌های منتشر شده در شبکه‌های اجتماعی و تعامل کاربران با یکدیگر در قالب کامنت تا حد قابل‌قبولی بیانگر ویژگی‌های شخصیتی افراد می‌باشد.

رانی و همکاران [۲۳] یک رویکرد ماشین بردار پشتیبان با حداقل مربعات مبتنی بر هسته مخلوط (MK-LSSVM) را در

[1] linguistic features
[2] annotate
[3] Linear SVC
[4] Logistic Regression
[5] Binary-Partitioning Transformer
[6] Natural Language Processing

پیش‌بینی ویژگی‌های شخصیتی انسان از پلتفرم‌های رسانه‌های اجتماعی پیشنهاد دادند. این مطالعه چالش تعیین مؤثر ویژگی‌های شخصیتی از داده‌های رفتاری را با استفاده از هسته‌های مخلوط در SVM برای مدل‌سازی روابط پیچیده داده‌ها و شناسایی ویژگی‌های ورودی متنوع مورد بررسی قرار می‌دهد، و بدین ترتیب دقت پیش‌بینی را روی مجموعه داده‌های استاندارد شامل مقاله و شاخص تیپ‌شناسی MBTI افزایش می‌دهد از روش‌های موجود مانند الگوریتم کت‌بوست پیشی می‌گیرد.

مطالعه‌ی سیوکومار و همکاران [۲۴] یک روش‌شناسی نوآورانه برای استنباط نوع شخصیت را با ادغام داده‌های چندین پلتفرم رسانه‌های اجتماعی با امتیازات شاخص تیپ شخصیتی MBTI ارائه می‌دهد. این پژوهش یک چارچوب جامع را ایجاد می‌کند که شامل فرآیندهای آماده‌سازی داده با کیفیت، اعتبارسنجی دقت داده‌ها و توسعه مدل دقیق با استفاده از تکنیک‌های یادگیری ماشین مانند رگرسیون لجستیک، ماشین بردار پشتیبان و شبکه‌های عصبی می‌شود.

این پژوهش به بررسی پیش‌بینی شخصیت از رفتار کاربران فیس‌بوک با استفاده از شبکه‌های عصبی مصنوعی در ترکیب با تکنیک‌های پردازش زبان طبیعی می‌پردازد. این مطالعه داده‌های متنی از پست‌ها و نظرات فیس‌بوک را تحلیل می‌کند تا بینش‌های معنادار از تعاملات کاربران استخراج کند و مدل‌های دقیق پیش‌بینی شخصیت را توسعه دهد. با بهره‌گیری از ترکیب طبقه‌بندی‌های NLP و ANN[1]، این پژوهش الگوها و ظرایف زبانی در محتوای تولید شده توسط کاربران را شناسایی می‌کند تا پیش‌بینی ویژگی‌های شخصیتی را تسهیل کند. این کار پتانسیل رویکردهای یادگیری ماشین برای درک رفتار انسان در شبکه‌های اجتماعی را نشان می‌دهد. این مطالعه بینش‌های ارزشمندی در کاربرد الگوریتم‌های پیشرفته برای تحلیل شخصیت در بسترهای رسانه‌های اجتماعی ارائه می‌دهد.

به تازگی، رویکردهای مبتنی بر گراف دانش در پیش‌بینی شخصیت محبوبیت پیدا کرده‌اند. رمضانی و همکاران [۲۵] رویکردی را پیشنهاد دادند که برای هر متن ورودی یک گراف دانش می‌سازد. این گراف با اطلاعات هستی‌شناسی، و ویژگی‌های روان‌شناختی زبانی غنی‌تر شده که منجر به یک بازنمایی جامع و غنی از نظر معنایی از متن می‌گردد. این گراف دانش به عنوان ورودی برای طبقه‌بندی‌کننده‌های یادگیری عمیق مختلف برای پیش‌بینی چندبرچسبی پنج ویژگی بزرگ شخصیت استفاده می‌شوند. آنان همچنین در [۲۶] از شبکه توجه گراف دانش در بهبود دقت طبقه‌بندی بهره بردند.

ژو و همکاران [۲۷] رویکردی نوآورانه برای تشخیص شخصیت از متن تولید شده توسط کاربر را معرفی می‌کند که با ترکیب دانش روان‌شناختی زبانی و تکنیک‌های پیشرفته شبکه‌های عصبی گراف همراه است. نویسندگان یک گراف ناهمگن واژگانی شخصیت برای هر کاربر با ترکیب معناشناسی متن، لکسیکون‌های روان‌شناختی زبانی و یک گراف دانش شخصیت می‌سازند. با استفاده از این گراف، آنها یک شبکه انتقال پیام شخصیت ناهمگن را پیشنهاد می‌دهند که تعبیه واژگان را بامکانیزم‌های توجه دوگانه پالایش می‌کند و به مدل اجازه می‌دهد تا روابط معنایی و مبتنی بر دانش محلی و سراسری را در متن دریافت کند.

اگرچه پژوهش‌های مرور شده سهم قابل توجهی در زمینه‌ی تشخیص پروفایل شخصیتی کاربران ایفا نموده‌اند، اما اتکای آنها به یک یا حداکثر دو منبع داده‌ای، کارایی این سیستم‌ها را در شرایط عملی و پیچیده‌ی دنیای واقعی به شدت محدود ساخته است.

مطالعه‌ی بورایا و همکاران [۲۸] نخستین تلاش نظام‌مند در زمینه تشخیص ویژگی‌های شخصیتی با استفاده ترکیبی از داده‌های چندین شبکه اجتماعی محسوب می‌شود. هسته اصلی این پژوهش بر مدلسازی روابط بین فردی متمرکز است که به‌صورت ذاتی با

[1] Artificial Neural Networks

مکانیزم‌های تشخیص شخصیت در پیوند است. مسئله‌ی اصلی در این مطالعه بررسی امکان افزایش کارآیی پیش‌بینی شخصیت با بهره‌گیری از داده‌های شبکه‌های اجتماعی مختلف است.

پژوهش فارسیو و همکاران [۲۹] نخستین مطالعه جامع در حوزه یادگیری پروفایل کاربر[1] از طریق تلفیق چندین منبع داده متفاوت را معرفی نموده است. در ابتدا داده‌های کاربران از سه پلتفرم توییتر[2]، فوراسکوئر[3] و اینستاگرام[4] جمع‌آوری شده و یک مجموعه‌ی داده‌ای وسیع و قوی به نام NUS-MSS[5] برای سه ناحیه‌ی جغرافیایی مختلف با بیشترین فعالیت کاربران فوراسکوئر شامل سنگاپور، لندن و نیویورک ارائه شده است [۲۹]. در مرحله بعد، پژوهشگران مدل پیشنهادی خود را بر روی مجموعه داده‌ای NUS-MSS پیاده‌سازی کردند تا ویژگی‌های جمعیت‌شناختی کاربران شامل سن و جنسیت را پیش‌بینی نمایند. نتایج نشان می‌دهند که منابع داده‌ای مختلف به‌صورت متقابل، مکمل یکدیگر هستند و ترکیب مناسب آنها عملکرد پروفایل‌یابی کاربر را تقویت خواهد کرد.

۴. چارچوب پیشنهادی برای پیش‌بینی تیپ شخصیتی (PTPIR)

در این پژوهش، ما یک چارچوب جامع دومرحله‌ای با عنوان PTPIR برای پیش‌بینی تیپ شخصیتی کاربران ارائه می‌دهیم. در ابتدا یک شبکه روابط ضمنی غنی از طریق تحلیل داده‌های چندوجهی ساخته می‌شود. این مرحله خود متکی بر دو تکنیک قدرتمند است: ۱) شناسایی جوامع در شبکه‌ی چندگانه با الگوریتم بهینه‌سازی چندهدفه MOEA/D-TS [۱۱] و ۲) بهره‌گیری از چارچوب پیش‌بینی پیوند CLPES [۱۴] به منظور استخراج نهایی روابط ضمنی بین کاربران.

مرحله دوم که هسته اصلی نوآوری مقاله حاضر را تشکیل می‌دهد، به کاربرد این شبکه ضمنی در مسئله پیش‌بینی شخصیت اختصاص دارد. در این مرحله، ما نشان می‌دهیم که چگونه می‌توان از توپولوژی این گراف استخراج‌شده به عنوان یک منبع اطلاعاتی قدرتمند برای آموزش مدل‌های یادگیری ماشین و پیش‌بینی دقیق ویژگی‌های شخصیتی کاربران بهره برد. در ادامه‌ی این بخش، ابتدا مسئله به صورت رسمی تعریف شده، سپس هر یک از مراحل اصلی چارچوب، شامل ساخت شبکه روابط ضمنی و پیش‌بینی شخصیت مبتنی بر آن، شرح داده می‌شود.

۱،۴. بیان مسئله

فرض کنید برای کاربران u_1 u_2 ... وu_m، مجموعه‌ی $SN = \{sn_1, sn_2, \dots, sn_k\}$ نشان‌دهنده‌ی k شبکه‌ی اجتماعی مختلف باشد که این افراد دارای حساب کاربری در تمامی آنها هستند. فرض بر این است که داده‌های فعالیت برای تمام کاربران در تمام پلتفرم‌ها قابل دسترس است؛ اما هیچ اطلاعاتی درباره روابط دوستی یا دنبال کردن موجود نیست. بنابراین در فاز اول، شبکه‌ی روابط ضمنی بین تمامی کاربران براساس فعالیت‌های آنها در k شبکه‌ی اجتماعی براساس راهکار CLPES به‌دست می‌آید. در فاز دوم، با بهره‌گیری از الگوریتم‌های یادگیری ماشین تشخیص تیپ شخصیتی هر کاربر براساس شبکه‌ی ارتباطات ضمنی وی پیش‌بینی می‌شود. در این مطالعه سه شبکه‌ی اجتماعی توییتر، اینستاگرام و فوراسکوئر به‌عنوان منابع اصلی داده انتخاب شده‌اند؛ که هر یک نماینده‌ای از یک نوع محتوای خاص محسوب می‌شوند [۳۰].

۲،۴. مرحله‌ی اول: ساخت شبکه‌ی روابط ضمنی

[1] user profile learning
[2] Twitter
[3] Foursquare
[4] Instagram
[5] National University of Singapore Multi-Source data-Set

در این بخش، یک چارچوب دو مرحله‌ای جهت پیش‌بینی تیپ شخصیتی کاربران براساس شبکه‌ی روابط ضمنی آنها ارائه می‌شود. در ادامه، به بررسی گام‌های مختلف چارچوب پیشنهادی در هر مرحله پرداخته می‌شود.

۱،۲،۴. شناسایی کاربران در شبکه‌های اجتماعی مختلف و استخراج روابط فعال

به‌منظور یک‌پارچه کردن اطلاعاتی که کاربران در پلتفرم‌های مختلف تولید می‌کنند، لازم است تا افراد در میان رسانه‌های اجتماعی گوناگون شناسایی شوند. بسیاری از شبکه‌های اجتماعی امروزی مانند توییتر، گوگل‌پلاس، اینستاگرام، فوراسکوئر، پینترست[1]، تردز[2] و غیره، قابلیت جدید اتصال متقابل را معرفی نموده‌اند که امکانی را فراهم می‌کند تا کاربران حساب کاربری خود در یک شبکه‌ی اجتماعی را با سایر حساب‌های کاربری خود در شبکه‌های اجتماعی دیگر پیوند دهد [۳۱]. طبق مطالعه صورت گرفته در [۳۱] بیش از ۷۰ درصد از کاربران فوراسکوئر از قابلیت اتصال متقابل استفاده می‌کنند. لازم به ذکر است که این مرحله در حوزه‌ی مطالعه‌ی این مقاله نیست و به‌منظور ارزیابی روش پیشنهادی از مجموعه‌های داده‌ای استخراج شده با روش فوق به‌وسیله‌ی سایر پژوهشگران [۲۹] بهره‌گیری شده است.

در ابتدا مسئله به‌صورت یک شبکه‌ی چندگانه از ارتباطات بین کاربران در شبکه‌های اجتماعی موردنظر (در اینجا توییتر، اینستاگرام و فوراسکوئر) درنظر گرفته شود. این مقاله خلاف روش‌های قبلی، تنها به روابط «غیرفعال[3]» دوستی و دنبال کردن اکتفا نمی‌کند و بر این فرض استوار است که تحلیل «فعال[4]» بین کاربران (تعاملات مستمر و معنادار) نسبت به روابط اسمی (اتصالات بدون تعامل)، تصویر واقع‌بینانه‌تری از ماهیت پویای ارتباطات کاربران ارائه می‌دهد. این انتخاب روش‌شناختی از آن جهت حائز اهمیت است که در شبکه‌های اجتماعی، بسیاری از ارتباطات دوستی صرفاً صوری بوده و فاقد هرگونه تعامل محتوایی هستند. تمرکز بر روابط فعال این امکان را فراهم می‌سازد که به جای توجه صرف به تعدد ارتباطات، کیفیت و عمق تعاملات اجتماعی کاربران مورد سنجش قرار گیرد [۴]. برای تحقق این هدف، پس از شناسایی کاربران در پلتفرم‌های مختلف از طریق روش اتصال متقابل، یک شبکه‌ی چندگانه از تعاملات فعال کاربران در شبکه‌های اجتماعی مورد مطالعه ایجاد می‌شود.

ازاین‌رو، گرافی از روابط فعال بین کاربران مبتنی‌بر عملیات «پاسخ دادن[5]» در شبکه‌ی توییتر در نظر گرفته می‌شود که برای شبکه‌های اینستاگرام و فوراسکوئر مبتنی‌بر عملیات «ذکر کردن[6]» خواهد بود. گراف روابط به‌صورت وزن‌دار طراحی شده است که در آن وزن هر یال نشان‌دهنده شدت و کیفیت رابطه بین دو کاربر می‌باشد.

۲،۲،۴. کشف روابط ضمنی

هدف اصلی در این مرحله، بهره‌گیری از روش پیش‌بینی پیوند چندگانه CLPES جهت شناسایی روابط ضمنی بین کاربران در شبکه‌ی چندگانه‌ی حاصل از مرحله‌ی قبل است. این روابط در مرحله آتی جهت تشخیص ویژگی‌های شخصیتی کاربران مورد استفاده قرار خواهند گرفت. این فرآیند در دو مرحله انجام می‌شود: ابتدا جوامع موجود در شبکه با الگوریتم CLPES تشخیص داده می‌شوند، سپس به هر ارتباط در لایه‌های مختلف شبکه، وزن‌های متناسب با ساختار جوامع تخصیص می‌یابد. در گام بعدی می‌بایست امتیاز شکل‌گیری پیوند بین گره‌ها در هر لایه براساس ویژگی‌های داخلی مشخص شود.

روش‌های مبتنی‌بر شباهت از کاربردی‌ترین راهبردهای پیش‌بینی پیوند به شمار می‌روند و به سه دسته‌بندی کلی، شامل روش‌های مبتنی‌بر ساختار، مبتنی‌بر خصیصه‌ی گره و روش‌های مبتنی‌بر مکان تقسیم می‌شوند. به‌منظور تعیین ویژگی‌های داخلی

[1] Pinterest
[2] Threads
[3] passive
[4] active
[5] reply
[6] mention

در هر لایه از روش‌های بالا استفاده می‌شود که در ادامه به طور خلاصه معرفی خواهند شد.

به‌منظور سنجش شباهت کاربران در شبکه‌ی توییتر براساس ویژگی‌های داخلی، از یک معیار شباهت مبتنی‌بر خصیصه‌های فردی با عنوان «شباهت کلمات[1]» که بر اساس تحلیل لغوی و معنایی محتوای تولیدی کاربران محاسبه می‌گردد [۳۲]. این معیار براساس ۵۰ دسته‌بندی از عناوین متداول مورد استفاده در توییتر است و شباهت واژه‌های مورد استفاده‌ی کاربر u و v با استفاده از یک فاصله‌ی همینگ[2] اصلاح شده به‌صورت زیر محاسبه می‌گردد:

$$W_S(u,v) = 1 - \frac{1}{2}\sum_{n=2}^{50} |f_{u,n} - f_{v,n}| \qquad (۵)$$

که در آن $f_{u,n}$ بسامد نرمال شده‌ی استفاده از کلمه‌ی موجود در nامین دسته‌بندی به‌وسیله‌ی کاربر u است. همچنین، شباهت کاربران در شبکه‌ی فوراسکوئر براساس معیار «جکارد مکان‌ها[3]» به شرح زیر محاسبه می‌شود:

$$JP_S(u,v) = \frac{|\Phi_L(u,v)|}{|\Phi_L(u) \cup \Phi_L(v)|} \qquad (۶)$$

در این رابطه، $\Phi_L(u)$ مکان‌های بازدید شده به‌وسیله‌ی کاربر u است و $\Phi_L(u,v) = \Phi_L(u) \cap \Phi_L(v)$ مجموعه‌ی دسته‌بندی‌های مکانی مشترک بازدید شده به‌وسیله‌ی دو کاربر u و v را نشان می‌دهد. درنهایت معیار شباهت برای شبکه‌ی اینستاگرام نیز بر پایه‌ی معیار جکارد براساس ۱۰۰ ویژگی تصویری پرکاربرد استخراج شده برای کاربران است. ازاین‌رو، معیار «جکارد تصاویر[4]» شباهت دو کاربر u و v باتوجه‌به تصاویر منتشر شده‌ی آنها را به شکل زیر محاسبه می‌کند:

$$JI_S(u,v) = \frac{|Y_L(u,v)|}{|Y_L(u) \cup Y_L(v)|} \qquad (۷)$$

که در آن، $Y_L(u)$ ویژگی‌های تصویری استخراج شده از تصاویر کاربر u است و مجموعه‌ی ویژگی‌های بین دو کاربر u و v به شکل $Y_L(u,v) = Y_L(u) \cap Y_L(v)$ نمایش داده شده است.

در مرحله‌ی نهایی از CLPES، پس از محاسبه‌ی ویژگی‌های داخلی و شباهت خارجی بین هر لایه، امتیاز احتمال شکل‌گیری پیوند بین هر جفت گره در لایه‌های مختلف محاسبه می‌گردد. به عبارت دقیق‌تر، این روش با تحلیل امتیازهای محاسبه‌شده، شبکه پنهان روابط کاربران را در لایه‌های مختلف یک سیستم چندبعدی آشکار می‌سازد. این شبکه استنتاج شده، به عنوان پایه و اساس مرحله‌ی بعدی، یعنی پیش‌بینی ویژگی‌های شخصیتی کاربران، مورد استفاده قرار می‌گیرد.

۳،۴. مرحله‌ی دوم: پیش‌بینی تیپ شخصیتی

در این بخش، حل مسئله‌ی پیش‌بینی شخصیت کاربران با در نظر گرفتن شبکه‌ی روابط ضمنی بین افراد و بهره‌گیری از روش‌های یادگیری ماشین مورد بررسی قرار می‌گیرد. ایده اصلی این است که مسئله به‌صورت یک مدل طبقه‌بندی دودویی در نظر گرفته شود که برچسب‌ها در دسته‌بندی‌های مختلف MBTI را مشخص می‌کند و مدل‌های مختلف براساس ویژگی‌های همسایگان ضمنی هر کاربر آموزش ببینند.

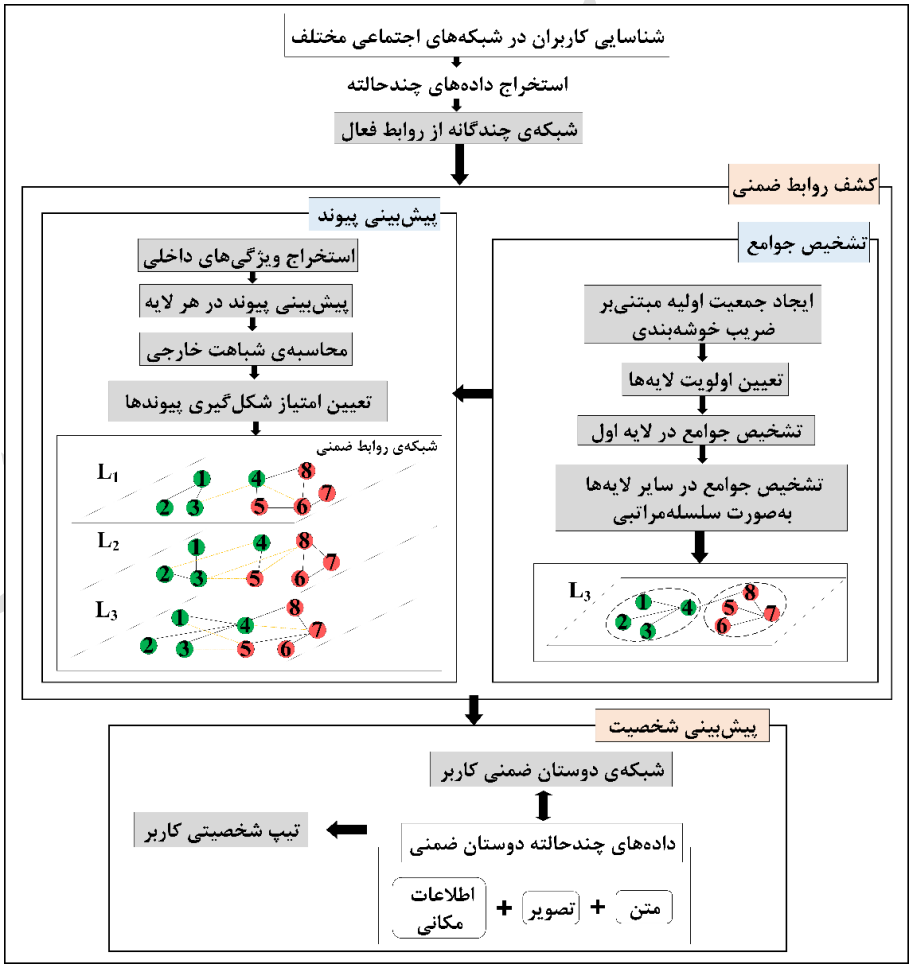
چنانچه ذکر شد، استفاده از مدل MBTI با بهره‌گیری از تعاملات و محتوای منتشرشده توسط کاربران در توییتر، نیاز به مشارکت فعال افراد را کاهش داده و فرآیند تحلیل شخصیت را تسهیل می‌کند. از طرف دیگر، شمای اصلی MBTI فرض می‌کند که هر دسته‌بندی MBTI نمایانگر جنبه‌های مختلفی از یک شخصیت انسانی است. لذا، با توجه به اهمیت کاربردی ابعاد مختلف MBTI در حوزه‌های عملی مانند کارآفرینی، رویکردی جزءنگر اتخاذ شده است. به‌طور مشخص، مدل‌سازی

[1] word similarity
[2] Hamming distance
[3] Jaccard of locations
[4] Jaccard of image

به‌صورت طبقه‌بندهای دودویی طراحی شده که هر یک از مؤلفه‌های چهارگانه MBTI (برون‌گرایی E/درون‌گرایی I، حسی S/شهودی N، منطقی F/احساسی T، قضاوتی J/ادراکی P) به‌طور مستقل پیش‌بینی و در نهایت تجمیع [1] این چهار بعد، تیپ شخصیتی نهایی (از میان ۱۶ حالت ممکن) را تعیین می‌نماید. این امر مزیت‌های شامل موارد زیر را به‌همراه دارد:

- با تبدیل مسئله به چهار طبقه‌بندی دودویی مجزا، مرزهای تصمیم‌گیری ساده‌تر و قابل‌تمایزتر می‌شوند و دقت به‌طرز چشمگیری بهبود می‌یابد.
- با تقسیم کلاس‌ها به حالت دودویی (به‌عنوان مثال E/I) در مقایسه با ۱۶ حالت مختلف، داده‌های آموزشی بیشتری برای هر کلاس وجود دارد.
- آموزش چهار طبقه‌بند مختلف، هر یک به‌جای داشتن یک مدل کلی برای تمام ویژگی‌ها، می‌توانند به‌طور جداگانه به بهترین شکل باتوجه‌به هدف متناسب خود بهینه‌سازی شوند.

همچنین، مطالعات نشان می‌دهند که اگرچه شخصیت افراد در طول زندگی تکامل می‌یابد، این تغییرات عمدتاً در بازه‌های زمانی بلندمدت (چندین سال) رخ می‌دهد. با توجه به اینکه مجموعه داده‌ی مورد استفاده در این پژوهش مربوط به بازه‌ی ۱۶۳ روزه است، می‌توان با اطمینان فرض کرد که تغییرات شخصیتی کاربران در این مدت ناچیز بوده و تأثیر محسوسی بر نتایج ندارد. شکل ۴ یک نمای کلی از روش پیشنهادی برای پیش‌بینی تیپ شخصیتی کاربران را نمایش می‌دهد.



شکل ۴: شمای کلی از چارچوب پیشنهادی جهت پیش‌بینی تیپ شخصیتی کاربران

[1] aggregate

۵. نتایج عملی

در این بخش با برر سی جامع رفتارهای دیجیتال کاربران در سه شـبکه‌ی اجتماعی توییتر، اینسـتاگرام و فوراسـکوئر به ارزیابی راهکار پیشنهادی جهت تشخیص ویژگی‌های شخصیتی کاربران پرداخته می‌شود. برای بررسی نتایج از طبقه‌بندهای مختلفی شامل گرادیان افزایشی[1]، رگرسیون لجستیک[2]، بیز ساده[3] AdaBoost، XGBoost، فرایندهای گوسی[4] و پرسپترون چندلایه (MLP[5]) و جنگل تصادفی[6] استفاده شده است.

۱،۵. جزییات پیاده‌سازی و تنظیمات آزمایش‌ها

در این زیربخش، جزئیات مربوط به مجموعه داده مورد استفاده، مراحل پیش‌پردازش، معیارهای ارزیابی و پارامترهای پیاده‌سازی تشریح می‌گردد.

۱،۱،۵. مجموعه داده‌ای و پیش‌پردازش

باتوجه‌به عدم وجود مجموعه داده‌ای پروفایل‌یابی شخصیتی متنوع به‌صورت چند منبعه، در این مطالعه از بزرگترین مجموعه داده‌ای موجود یعنی NUS-MSS [۲۹] استفاده شده است که شامل داده‌های چندحالته جمع‌آوری شده از سه شبکه اجتماعی توییتر، اینستاگرام و فوراسکوئر در کشور سنگاپور، به‌همراه یک ملاک حقیقت‌مبنا[7] در رابطه با وضعیت تأهل و تیپ شخصیتی کاربران است[8]. داده‌های مورد نیاز با استفاده از راهبرد اتصال متقابل پروفایل‌های کاربران در شبکه‌های اجتماعی مختلف جمع‌آوری شده‌اند به‌طوری‌که شبکه توییتر به عنوان پایگاه اصلی انتخاب شده است و انتشار مجدد پست‌های اینستاگرام و فوراسکوئر به‌همراه توییت‌های شبکه توییتر را در یک کانال اطلاعات ذخیره می‌کند. برای به‌دست آوردن ملاک صحت مربوط به شخصیت، از API جستجوی توییتر برای انجام جستجوی نتایج آزمون MBTI معتبر برخط، مانند 16Personalities[9]، Jung Typology Test[10] و MBTI Online[11] استفاده شده است. پس از جمع‌آوری توییت‌ها با نتایج آزمون MBTI، برچسب‌های ملاک صحت MBTI از آنها استخراج شده است. جزئیات داده‌های جمع‌آوری شده در جدول ۱ نشان داده شده است.

جدول (۱): جزئیات آماری مجموعه داده‌ای مورد استفاده

	Twitter	Instagram	Foursquare
#users	15788	10254	3090
	tweets	**images**	**check-ins**
#posts	122,584,534	4,789,519	420,603

به منظور آماده‌سازی داده‌ها برای مدل‌سازی، مجموعه‌ای از عملیات پیش‌پردازش و استخراج ویژگی بر روی داده‌های متنی، مکانی و تصویری به شرح زیر انجام شد.

- **ویژگی‌های متنی:** در ابتدا پیش از استخراج ویژگی، مراحل پیش‌پردازش زیر بر روی متن توییت‌ها اعمال گردید:
  - **فیلتر کردن بازتوییت‌ها:** داده‌های مربوط به بازتوییت‌ها در آزمایش‌های ما لحاظ نشدند، زیرا اطلاعات جمعیت‌شناختی جدیدی از خود کاربر ارائه نمی‌دهند.
  - **تبدیل کلمات عامیانه:** با استفاده از یک واژه‌نامه خارجی، تمام کلمات و اصطلاحات عامیانه به مترادف‌های استاندارد خود تبدیل شدند تا از تعداد واژگان ممکن در توییت‌ها کاسته شود.
  - **حذف تگ‌ها، منشن‌ها و هشتگ‌ها:** منشن‌های کاربران، تگ‌های مکانی و هشتگ‌ها از متن توییت‌ها حذف گردیدند.

پس از فرآیند پاک‌سازی، ویژگی‌های متنی زیر استخراج و نرمال‌سازی شدند:

[1] gradient boosting
[2] logistic regression
[3] naïve bayes
[4] Gaussian processes
[5] MultiLayer Perceptron
[6] random forest
[7] ground truth

[8] این مجموعه داده‌ای طی یک توافق‌نامه با شرکت SoMin در سنگاپور در اختیار ما قرار گرفته است.

[9] http://16personalities.com
[10] http://humanmetrics.com
[11] http://mbtionline.com

- *ویژگی‌های مبتنی بر LDA*[1]: تمام توییت‌های هر کاربر در یک سند واحد ادغام شدند. سپس با اعمال روش *LDA*، این اسناد به یک فضای موضوعی پنهان نگاشت شده است. در نتیجه، برای هر کاربر ۵۰ ویژگی *LDA* استخراج شده است.

- *ویژگی‌های زبان‌شناختی*: در این بخش ۶۴ ویژگی برای هر کاربر استخراج شده است که روش قدرتمندی برای اهداف پیش‌بینی شخصیت، سن و جنسیت هستند [۲۳]. این ویژگی‌ها شامل ویژگی‌های LIWC[2] [۳۳] استخراج شده برای کاربران هستند.

- *ویژگی‌های اکتشافی*[3] *و لغوی*[4]: ابتدا تعداد نشانی‌های وب، تعداد هشتگ‌ها و تعداد ذکرکردن‌های کاربر محاسبه شده است. در مرحله دوم، تعداد کلمات عامیانه، کلمات عاطفی، تعداد شکلک‌ها و میانگین امتیاز احساسات محاسبه شده است. این ویژگی‌ها می‌توانند سیگنال‌های خوبی از ویژگی‌های شخصیتی کاربر باشند که به نوبه خود وابسته به جنس و سن هستند. در بخش سوم، ویژگی‌های سبک زبانی[5] مانند «تعداد کاراکترهای تکراری» در کلمات، «تعداد غلط‌های املایی» و «تعداد کلمات ناشناخته» برای بررسی کننده‌ی کلمات محاسبه شده است. درمجموع، ۵۳ ویژگی برای این دسته استخراج گردیده است.

– **ویژگی‌های تصویری**: برای استخراج ویژگی‌های تصویری، هر عکس اینستاگرام با استفاده از یک مدل از قبل آموزش دیده‌ی GoogleNet به‌طور خودکار به ۱۰۰۰ مفهوم تصویری (مانند گل، توپ، صندلی و غیره) ایجاد شده ازImageNet [۳۴] نگاشت شده است. سپس احتمال وقوع مفهوم پیش‌بینی شده برای هر کاربر را خلاصه شده و بردار به‌دست‌آمده به تعداد تصاویر ارسال شده به‌وسیله‌ی این کاربر تقسیم شده است. در نهایت، از روش تحلیل مؤلفه‌های اصلی (PCA[6]) برای کاهش ابعاد فضای ویژگی تصاویر به ۱۰۰ ویژگی تصویری برای هر کاربر استفاده شد.

- **ویژگی‌های مکانی**: برای محاسبه ویژگی‌های مکانی از ۸۸۶ دسته محل[7] فوراسکوئر (مانند هنر و سرگرمی، موزه، دانشگاه، پارک، فروشگاه، رستوران و غیره) بهره‌گیری شده است. برای هر کاربر، تعداد کل ورودهای فوراسکوئر وی در محل‌های مختلف از هر دسته‌بندی محل شمارش می‌شود. سپس تعداد کل ورودهای کاربر در هر دسته‌بندی محل بر تعداد کل ورودهای وی تقسیم شده است.

۲،۱،۵. معیارهای ارزیابی

اگرچه روش طبقه‌بندی دودویی بر روی هر برچسب در هر دسته‌بندی اعمال شده است، اما نمی‌توان برچسب را در هر گروه اولویت‌بندی کرد؛ زیرا دقیقا معکوس یکدیگر نیستند. به‌همین دلیل، جهت ارزیابی آزمایشات از معیار Macro-F1 استفاده شده است. این معیار نشانگر میانگین معیارهای دقت[8]، خوانش[9] و معیار $F1$ برای هر برچسب است. این معیار ارزیابی به‌دلیل توانایی آن در رفتار یکسان با تمام برچسب‌های داده در مجموعه داده‌های نامتوازن[10] با هدف جلوگیری از ارزش‌گذاری ناصحیح عملکرد کلی مدل، انتخاب شده است [۲۹]. فرض کنید P_i، R_i و $F1_i$ به‌ترتیب مشخص کننده‌ی دقت، بازیابی و معیار $F1$ باتوجه‌به کلاس i باشند؛ آنگاه:

$$F1_i = \frac{2P_iR_i}{P_i + R_i} \qquad (۸)$$

درنتیجه، معیار Macro-F1 باتوجه‌به میانگین حسابی مقادیر $F1$ برای هر کلاس به شکل زیر محاسبه می‌شود:

$$MacroF1 = \frac{1}{n}\sum_x F1_x = \frac{1}{n}\sum_x \frac{2P_xR_x}{P_x + R_x} \qquad (۹)$$

[1] Latent Dirichlet Allocation
[2] Linguistic Inquiry and Word Count
[3] heuristically-inferred
[4] lexicon
[5] linguistic style features

[6] Principal Component Analysis
[7] venue category
[8] precision
[9] recall
[10] imbalanced

دلیل اصلی انتخاب معیار Macro-F1، ماهیت نامتوازن مجموعه داده‌های شخصیتی است که در آن، فراوانی برخی تیپ‌ها بسیار بیشتر از بقیه است. این معیار با محاسبه میانگین ساده و بدون وزن‌دهی از امتیاز F1 هر کلاس، از سوگیری به سمت کلاس‌های پرجمعیت جلوگیری کرده و تضمین می‌کند که به عملکرد مدل در تمام کلاس‌ها، چه نادر و چه متداول، اهمیت یکسانی داده می‌شود.

۳،۱،۵. جزئیات پیاده‌سازی و تنظیمات پارامترها

چارچوب پیشنهادی برای عملکرد صحیح، نیازمند تنظیم دقیق پارامترهای دو الگوریتم زیربنایی خود، یعنی CLPES و MOEA/D است. این تنظیمات به شرح زیر در تمام آزمایش‌ها اعمال شده‌اند. تنها پارامتر روش CLPES ضریب φ می‌باشد که مقدار آن برای تمام آزمایش‌ها ۰،۵ در نظر گرفته شده است. همچنین بر اساس پژوهش پیشین ما [۱۱]، برای الگوریتم MOEA/D، اندازه جمعیت یا تعداد زیرمسئله‌ها $N = 200$، احتمال برش $P_c = 0.9$، احتمال جهش $P_m = 0.2$ و تعداد نسل‌ها $gen_{max} = 200$ در نظر گرفته شده است. در الگوریتم جستجوی ممنوعه[1]، پارامتر همسایگی $T = 20$ (معادل ۱۰ درصد اندازه جمعیت)، اندازه لیست ممنوعه $T' = 100$ و معیار توقف الگوریتم، ۱۰۰ تکرار بدون هیچ‌گونه بهبودی در نتایج، در نظر گرفته شده است. برای ارزیابی در تمامی آزمایش‌ها، مجموعه داده‌ها به مجموعه‌های آموزشی(۸۰ درصد) و آزمون (۲۰ درصد) تقسیم شده است و برای تنظیم ابرپارمترها از اعتبارسنجی متقاطع ۵-بخشی[2] استفاده شده است.

۲،۵. ارزیابی و مقایسه

در این بخش ابتدا به ارزیابی راهکار پیشنهادی جهت پیش‌بینی تیپ شخصیتی کاربران براساس روابط ضمنی پرداخته می‌شود و برای این منظور نتایج حاصل به‌وسیله‌ی مدل‌های آموزش داده شده با ترکیب منابع داده‌ای مختلف با یکدیگر مقایسه خواهند شد. ارزیابی نهایی برای پیش‌بینی شخصیت مبتنی‌بر ۲۹۲۴ کاربری است که تیپ شخصیتی آنها مشخص بوده و در هر سه منبع داده‌ای موردنظر فعالیت داشته‌اند و لذا می‌توان آن را به‌عنوان ملاک صحت قابل اعتماد در نظر گرفت.

جدول ۲ نتایج ارزیابی هریک از روش‌ها جهت پیش‌بینی دسته‌بندی‌های مختلف شخصیتی را باتوجه‌به طبقه‌بندهای مختلف نمایش می‌دهد. در این جدول نماد *Ir* نشان دهنده‌ی روش پیشنهادی جهت پیش‌بینی ویژگی‌های شخصیتی، مبتنی‌بر شبکه‌ی روابط ضمنی کاربران است و *Ind* مشخص کننده سناریوی پیش‌بینی براساس ویژگی‌های فردی هریک از کاربران است. طبق نتایج جدول ۲ مشاهده می‌شود که روش پیشنهادی در سه مورد از چهار دسته‌بندی ویژگی‌های شخصیتی نسبت به حالت مبتنی‌بر فعالیت‌های انفرادی کاربر در شبکه‌های اجتماعی نتایج بهتری را به‌دست آورده است. براین‌اساس، به‌منظور پیش‌بینی ویژگی شخصیتی «درون‌گرایی/برون‌گرایی»، PTPIR[3] با طبقه‌بند گرادیان تقویتی به بهترین نتیجه دست یافته است. همچنین، بهترین نتیجه‌ی پیش‌بینی ویژگی «حسی/شهودی» و «فکری/احساسی» متعلق به PTPIR با روش جنگل تصادفی است. تنها در مورد پیش‌بینی ویژگی «ملاحظه کننده/داوری کننده» تکیه بر فعالیت‌های انفرادی هر فرد در شبکه‌های اجتماعی و بهره‌گیری از طبقه‌بند گرادیان تقویتی نتیجه‌ی برتر را به خود اختصاص داده‌است. این نتایج با تعاریف مایرز سازگار است که بیان می‌کند ویژگی «ملاحظه کننده/داوری کننده» در شخصیت افراد واکنش‌های آنها به تغییرات مختلف زندگی را توصیف می‌کنند [۳۵]؛ و لذا بررسی رفتار فردی کاربران در شبکه‌های اجتماعی نمود بهتری از این ویژگی را آشکار خواهد کرد. درعین‌حال، سایر دسته‌های ویژگی شخصیتی بیشتر مربوط به احساسات و ادراک از زندگی هستند که می‌توانند نقاط مشترک بیشتری با دوستان را به نمایش بگذارند.

[1] Tabu Search
[2] 5-fold cross validation
[3] Personality Type Prediction based on Implicit Relations

۲،۳،۵. بررسی مدل با ترکیبات مختلف منابع

باتوجه‌به عملکرد به‌نسبت بالای ویژگی‌ها در منابع تکی، در این بخش به بررسی عملکرد راهکار پیشنهادی با در نظر گرفتن ترکیبات مختلف منابع پرداخته خواهد شد.

برای این منظور عملکرد طبقه‌بندهای مختلف در مدل‌های تک‌منبعه و ترکیبات مختلف منابع مورد بررسی قرار گرفته است. به‌منظور بررسی عملکرد مدل با در نظر گرفتن ترکیبات مختلف منابع، راهکار همجوشی ثانویه برای آموزش طبقه‌بندهای مختلف از ترکیبات مختلف مجموعه ویژگی‌ها استفاده می‌شود. برای این منظور، ویژگی‌های هم‌نوع (متنی، تصویری، مکانی) متناظر با مجموعه همسایگی ضمنی هر کاربر در کنار یکدیگر قرار می‌گیرند و برای هر نوع از ویژگی‌ها، یک طبقه‌بند مجزا آموزش خواهد دید.

جدول (۲): مقایسه نتایج به دست آمده جهت پیش‌بینی ویژگی‌های شخصیتی با روش PTPIR و راهکار منفرد براساس معیار Macro-F1

	E/I		*S/N*		*T/F*		*J/P*	
Method	***Ir***	***Ind***	***Ir***	***Ind***	***Ir***	***Ind***	***Ir***	***Ind***
Gradient Boosting	**0.741**	0.725	0.524	0.495	0.667	**0.670**	0.561	0.510
Logistic Regression	0.572	0.600	0.451	0.425	0.473	0.435	0.424	0.394
Naive Bayes	0.472	0.440	0.403	0.330	0.421	0.440	0.467	0.470
Random Forest	0.563	0.483	**0.735**	0.562	0.664	0.506	**0.737**	0.692

چنانچه مشاهده می‌شود، این چارچوب از طبقه‌بندی کننده‌های درخت تصادفی چند برچسبه[1] تشکیل شده است که بر روی انواع مختلف ویژگی، آموزش دیده‌اند. در این بخش، مدل طبقه‌بندی جنگل تصادفی به‌دلیل توانایی شناخته شده‌اش در انجام طبقه‌بندی در فضا با ابعاد بالا به‌همراه رویکرد انتخاب ویژگی تصادفی انتخاب شده است [۲۹]. همچنین، نتایج سایر تحقیقات نشان داده است که طبقه‌بند جنگل تصادفی به‌طور معمول عملکرد بهتری برای پیش‌بینی سن و جنسیت نسبت به سایر روش‌های یادگیری ماشین مانند ماشین بردار پشتیبان یا بیز ساده نشان می‌دهد. تعداد بهینه درختان تصادفی برای هر طبقه‌بند با اعتبارسنجی متقاطع ۱۰-بخشی مشخص شده است. براین‌اساس، تعداد درختان برای ویژگی‌های «متنی»، شامل ویژگی‌های اکتشافی، ویژگی‌های زبانی LIWC و LDA برابر ۱۵۰، برای «ویژگی‌های تصویری» برابر ۱۱۰ و برای «ویژگی‌های مکانی» برابر با ۱۳۵ در نظر گرفته شده است.

با ترکیب سه طبقه‌بند جنگل تصادفی در مدل یادگیری جمعی، درنهایت امتیاز طبقه‌بندی نهایی برای هر برچسب به شکل زیر محاسبه می‌شود:

$$SC(l) = \sum_{i=0}^{n} \frac{Pl_i \times d_i \times a_i \times dw_i}{n} \qquad (۱۰)$$

در این رابطه، p_{l_i} مشخص کننده‌ی اطمینان از پیش‌بینی مثبت[2] برای برچسب l از iامین مدل طبقه‌بندی است. همچنین d_i تعداد رکوردهای داده‌ای مربوط به کاربر فعلی برای آموزش مدل طبقه‌بندی iام تقسیم بر تعداد متوسط رکوردهای داده‌ای از نوع داده شده است که برای تمامی کاربران جمع‌آوری شده است. در نهایت، دقت مدل طبقه‌بندی iام و وزن متناظر هر مدل به‌ترتیب با a_i و w_i مشخص شده است. لازم به ذکر است که دقت a_i برابر با سطح صحت کلان[3] هر طبقه‌بند در نظر گرفته شده است که پس از انجام اعتبارسنجی متقاطع ۱۰-بخشی در هر مجموعه از ویژگی‌های متناظر تخمین زده می‌شود. به‌منظور یادگیری وزن w_i

[1] multi-label random forest
[2] positive prediction confidence
[3] macro accuracy

برای داده‌ها با ماهیت‌های مختلف از روش تپه‌نوردی با شروع مجدد تصادفی (SHCR[1]) استفاده شده است که قادر به دستیابی بهینه محلی برای مسائل غیرمحدب[2] است و بنابراین می‌تواند وزن مناسب جمعی را تولید کند. پس از اینکه هریک از طبقه‌بندها با روش فوق وزن‌دهی شدند، امتیاز خروجی با میانگین‌گیری مشخص می‌شود و نتیجه، امتیاز طبقه‌بندی جمعی که به برچسب l در تکرار فعلی n تخصیص داده شده است را نشان می‌دهد. این روند تا زمان پوشش SHCR با تغییر پارامترهای w_{i_n} در هر تکرار با هدف بیشینه کردن دقت کلی تکرار می‌شود. برای پیش‌بینی مشخصات کامل MBTI، چهار مدل مجزا برای چهار دسته MBTI آموزش داده می‌شود که هریک از آنها شامل دو برچسب هستند؛ به‌طوری‌که هر مدل یک برچسب نهایی را برای هر دسته‌بندی MBTI پیش‌بینی می‌کند. در این آزمایش، هدف پیش‌بینی تیپ کامل شخصیتی MBTI برای کاربران است. برای این منظور، ابتدا هر چهار دسته MBTI به‌طور جداگانه پیش‌بینی می‌شوند و سپس دسته‌بندی‌های پیش‌بینی شده در یک نمایه کامل با یکدیگر ادغام می‌شوند.

جدول (۳): مقایسه‌ی نتایج حاصل به وسیله‌ی چارچوب PTPIR برای پیش‌بینی ویژگی‌های شخصیتی، با ترکیب منابع داده‌ای مختلف (براساس Macro-F1)

Method	Data sources	E/I	S/N	T/F	J/P
SVM	Text	0.337	0.417	0.436	0.398
	Image	0.318	0.327	0.42	0.387
	Location	0.494	0.381	0.416	0.438
Naïve Bayes	Text	0.433	0.438	0.473	0.453
	Image	0.325	0.332	0.321	0.427
	Location	0.486	0.429	0.442	0.436
Logistic Regression	Text	0.346	**0.452**	0.367	0.493
	Image	0.328	**0.421**	**0.514**	**0.484**
	Location	0.497	**0.477**	**0.480**	**0.505**
Adaboost	Text	0.549	0.407	0.471	0.361
	Image	0.532	0.418	0.492	0.417
	Location	0.359	0.435	0.399	0.369
Gaussian processes	Text	0.534	0.419	0.476	0.361
	Image	0.338	0.407	0.377	0.361
	Location	0.353	0.407	0.283	0.360
Gradient boosting	Text	0.572	0.407	0.484	0.383
	Image	0.546	0.412	0.459	0.433
	Location	0.374	0.421	0.397	0.401
XGBoost	Text	**0.554**	0.423	0.485	0.365
	Image	**0.556**	0.407	0.461	0.390
	Location	**0.510**	0.413	0.406	0.418
Random Forest	Text	0.545	0.440	**0.519**	**0.450**
	Image	0.532	0.405	0.472	0.442
	Location	0.470	0.422	0.458	0.449
MLP	Text	0.551	0.440	0.496	0.388
	Image	0.534	0.418	0.419	0.406
	Location	0.486	0.407	0.391	0.371
Random Forest (late fusion)	Text + Image	0.411	0.615	0.517	0.526
	Image+ Location	0.563	0.668	0.55	0.683
	Text + Location	0.570	0.683	0.608	0.679
	Text + Image + Location	**0.581**	**0.746**	**0.613**	**0.752**

[1] Hill climbing with Random Restart

[2] non-convex

نتایج پیش‌بینی هر یک از دسته‌بندی‌های مدل MBTI باتوجه‌به ترکیب منابع داده‌ای مختلف، براساس معیار Macro-F1 در جدول ۳ نمایش داده شده است. ارزیابی منابع داده به صورت منفرد، یک تعامل پیچیده بین نوع داده و الگوریتم طبقه‌بند را آشکار ساخت. در حالی که داده‌های تصویری به طور مداوم کمترین سهم را در پیش‌بینی داشتند، برتری نسبی بین داده‌های متنی و مکانی، تابعی از خانواده الگوریتم مورد استفاده بود. به طور مشخص، طبقه‌بندهای خطی (مانند رگرسیون لجستیک) با داده‌های مکانی به عملکرد بهینه‌ای دست یافتند، در حالی که مدل‌های پیچیده‌تر مبتنی بر یادگیری گروهی (Ensemble) مانند جنگل تصادفی و XGBoost، قادر به استخراج الگوهای غنی‌تری از داده‌های متنی بودند. این مشاهدات حاکی از آن است که قابلیت استخراج ویژگی‌های شخصیتی از منابع داده ناهمگون، تنها به خود داده بستگی ندارد، بلکه به تناسب ساختاری بین مدل یادگیری و پیچیدگی ذاتی آن منبع داده نیز وابسته است.

با در نظر گرفتن ویژگی‌های متنی به‌همراه ویژگی‌های مکانی دوستان ضمنی و یادگیرنده‌ی پایه‌ی جنگل تصادفی، بهترین ترکیب برای پیش‌بینی تیپ شخصیتی در بین کلیه مجموعه‌های دو منبعه به‌دست‌آمده است. دلیل احتمالی این است که در انتخاب‌های ترجیحی مکان‌های مختلف و سبک نوشتن در بین شخصیت‌های مشابه، شباهت‌های مشخصی وجود دارد. به‌عنوان‌مثال، افراد برون‌گرا برای تفریح مکان‌هایی مانند شهربازی، کنسرت و مراکز خرید را انتخاب کنند و در متن‌های نوشته شده نظرات خود در مورد رویدادهای روزانه را با کلمات صریح بیان کنند. اما ممکن است تصاویر منتشر شده به‌وسیله‌ی افراد با ویژگی‌های شخصیتی مشابه به‌طور الزامی اطلاعات نزدیکی را فراهم نکند. همچنین، این مطابق با نتایج حاصل از آموزش مدل‌ها بر روی داده‌های تک حالته است که در آن متن و مکان به‌عنوان بهترین منابع جهت دستیابی به نتایج بیشینه دقت مشخص شدند.

تحلیل نهایی، مدل جنگل تصادفی با استراتژی همجوشی ثانویه بر روی هر سه منبع داده را به عنوان چارچوب برتر معرفی می‌کند. جالب توجه است که قدرت پیش‌بینی این مدل بر اساس ابعاد شخصیتی متغیر است. دقت بالای آن در ابعاد حسی/شهودی (S/N) و ملاحظه‌کننده/داوری‌کننده (J/P) نشان می‌دهد که چارچوب ما به ویژه در تشخیص ویژگی‌های مرتبط با ادراک و سبک زندگی مؤثر است. عملکرد نسبتاً پایین‌تر (هرچند همچنان برتر) در بعد تفکری/احساسی (T/F)، با یافته آماری ما (در تحلیل فرضیه H_1) که در ادامه آورده شده است همخوانی دارد و نشان می‌دهد این ویژگی ممکن است کمتر به داده‌های موجود در شبکه اجتماعی وابسته بوده و بیشتر در فرآیندهای شناختی فردی منعکس شود.

۳،۵. تحلیل آماری فرضیه‌ها

به منظور اعتبارسنجی علمی چارچوب پیشنهادی و پاسخ به سوالات تحقیق، فرضیه‌های اصلی پژوهش با استفاده از آزمون‌های آماری معناداری مورد ارزیابی قرار گرفتند. به منظور اطمینان از پایداری نتایج و کاهش تأثیر عوامل تصادفی، هر یک از آزمایش‌های طبقه‌بندی ۲۰ بار به صورت مستقل با مقداردهی اولیه‌ی متفاوت اجرا شدند. تحلیل‌های آماری گزارش‌شده در ادامه، بر روی مجموع نتایج حاصل از این اجراهای متعدد انجام شده است. برای این تحلیل، از آزمون ناپارامتریک رتبه علامت‌دار ویلکاکسون[1] با سطح معناداری $\alpha = 0.05$ استفاده شده است. در ادامه، نتایج آزمون برای هر فرضیه به تفکیک ارائه می‌گردد.

برای آزمون فرضیه اول (H_1)، که بر برتری رویکرد مبتنی بر روابط ضمنی (*Ir*) در مقابل رویکرد متکی بر اطلاعات فردی (*Ind*)تأکید دارد، یک تحلیل آماری دقیق به تفکیک برای هر یک از چهار بعد شخصیتی انجام شد. نتایج آزمون ویلکاکسون که در جدول ۴ خلاصه شده است، تصویر دقیقی از عملکرد روش پیشنهادی ما ارائه می‌دهد. همانطور که در جدول ۴ مشاهده می‌شود، روش مبتنی بر روابط ضمنی در پیش‌بینی سه بعد از چهار

[1] Wilcoxon signed-rank test

بعد شخصیتی به طور معناداری بهتر عمل کرده است ($p<0.001$). این یافته، قدرت اطلاعات شبکه اجتماعی در درک جنبه‌های تعاملی و سبک زندگی افراد را به وضوح نشان می‌دهد.

جدول (۴): نتایج آزمون ویلکاکسون برای اعتبارسنجی فرضیه H1

بعد شخصیتی	*p-value*
E/I	<0.001
S/N	<0.001
T/F	0.254
J/P	<0.001

نکته قابل توجه، عدم وجود تفاوت آماری معنادار بین دو روش برای بعد تفکری/احساسی (T/F) است ($p=0.254$). این نتیجه نشان می‌دهد که اگرچه اطلاعات شبکه اجتماعی برای پیش‌بینی رفتارها و ادراکات کلی بسیار کارآمد است، اما ممکن است برای درک فرآیندهای تصمیم‌گیری درونی افراد (که توسط بعد T/F سنجیده می‌شود) ارزش افزوده کمتری داشته باشد. این بعد از شخصیت احتمالاً بیشتر در سبک نگارش و محتوای تولیدی خود فرد منعکس می‌شود تا در ویژگی‌های شبکه دوستان ضمنی او. در مجموع، این نتایج فرضیه H_1 را به طور گسترده تأیید می‌کنند و نشان می‌دهند که رویکرد مبتنی بر روابط ضمنی، به ویژه برای ابعاد اجتماعی و ادراکی شخصیت، یک راهبرد برتر و مؤثر است.

به منظور اعتبارسنجی فرضیه دوم (H_2)، که بر برتری راهبرد همجوشی داده‌های چندمنبعه تأکید دارد، یک تحلیل آماری دقیق انجام شد. برای آزمون این فرضیه، عملکرد مدل نهایی ما (جنگل تصادفی با همجوشی ثانویه) با هر سه منبع داده‌ای با عملکرد مدل‌هایی که از زیرمجموعه‌ای از این داده‌ها استفاده می‌کنند، مقایسه شد. به منظور افزایش دقت و عمق تحلیل، این مقایسه‌ها با استفاده از آزمون ویلکاکسون به صورت مجزا برای هر یک از چهار خصیصه دوگانه شخصیتی (I/E, S/N, T/F, J/P) انجام شد.

نتایج به دست آمده به طور قاطع و یکپارچه نشان داد که در تمام مقایسه‌ها و برای تمام ابعاد شخصیتی، مدل همجوشی کامل عملکردی برتر و به لحاظ آماری کاملاً معنادار داشته است ($p<0.001$). این یافته، فرضیه H_2 را با قدرت بسیار بالایی تأیید می‌کند و نشان می‌دهد که راهبرد همجوشی داده‌های چندمنبعه، یک روش مؤثر و پایدار برای بهبود پیش‌بینی در تمام جنبه‌های شخصیتی مورد بررسی است.

۶. نتیجه‌گیری و کارهای آتی

در این مقاله، یک چارچوب نوین با عنوان PTPIR برای پیش‌بینی تیپ شخصیتی کاربران در شبکه‌های اجتماعی ارائه گردید که با درنظر گرفتن ساختار جوامع و پیش‌بینی پیوند در آن‌ها، تیپ شخصیت کاربران را تحلیل می‌نماید. نتایج پیش‌بینی شخصیت براساس روش پیشنهادی که بر روی ترکیب منابع دوتایی آموزش دیده‌اند نشان می‌دهد که بهترین عملکرد با ترکیب داده‌های متنی و مکانی با بهبود بین ۱۱ تا ۳۸ درصد صورت گرفته است. این نتایج برای PTPIR که براساس هر سه منبع داده‌ای آموزش دیده است، از سایر روش‌ها پیشی گرفته و در مقایسه با استفاده از اطلاعات تک‌منبعه مربوط به دوستان ضمنی به‌طور میانگین، ۲۳ درصد، بهبود عملکرد را گزارش می‌دهد. این یافته، ارزش حیاتی تحلیل روابط ضمنی و یکپارچه‌سازی داده‌های چندوجهی را برای درک عمیق‌تر شخصیت در فضای مجازی به اثبات می‌رساند.

علیرغم نتایج امیدوارکننده، این پژوهش با محدودیت‌هایی نیز مواجه است. از جمله مهم‌ترین آنها می‌توان به وابستگی عملکرد مدل به کیفیت و در دسترس بودن داده‌ها و همچنین قابلیت تعمیم‌پذیری نتایج به جوامع کاربری با ویژگی‌های فرهنگی و جغرافیایی متفاوت اشاره کرد. اگرچه مجموعه داده NUS-MSS یک محک استاندارد و معتبر است، اما به کاربران مناطق جغرافیایی و فرهنگ‌های خاصی محدود می‌شود. از این رو، ارزیابی عملکرد مدل بر روی جوامع کاربری متنوع‌تر می‌تواند به درک عمیق‌تری از پایداری آن منجر شود.

به عنوان کارهای آتی، چندین مسیر تحقیقاتی مختلف قابل تصور است. از یک سو، می‌توان با بهره‌گیری از مدل‌های دینامیک مانند شبکه‌های عصبی گرافی زمانی[1]، به تحلیل چگونگی تکامل شخصیت و روابط کاربران در طول زمان پرداخت. همچنین، با الهام از پژوهش‌های اخیر که در بخش پیشینه مرور شد، می‌توان از گراف‌های دانش برای غنی‌سازی معنایی ویژگی‌های کاربران

[1] Temporal GNNs

بهره برد؛ ساخت یک گراف دانش برای هر کاربر و تلفیق آن با اطلاعات روانشناختی، می‌تواند بازنمایی بسیار قدرتمندتری برای مدل‌های یادگیری عمیق فراهم کند. در نهایت، در نظر گرفتن جوامع دارای همپوشانی و بررسی منابع داده‌ای جدید مانند داده‌های موسیقایی (اسپاتیفای) یا حرفه‌ای (لینکداین)، چشم‌اندازهای روشنی برای دستیابی به پروفایل‌های شخصیتی جامع‌تر و دقیق‌تر فراهم می‌کند.

مراجع

[1] W. Youyou, D. Stillwell, H. A. Schwartz, and M. Kosinski, "Birds of a feather do flock together: Behavior-based personality-assessment method reveals personality similarity among couples and friends," *Psychol Sci*, vol. 28, no. 3, pp. 276–284, 2017, doi: 10.1177/0956797616678187.

[2] M. Rasouli and V. Kiani, "A review of deep learning methods for emotion classification in text: advances, challenges, and opportunities," *Soft Comput. J.*, vol. 13, no. 1, pp. 40-57, 2024, doi: 10.22052/scj.2023.248812.1126 [In Persian].

[3] C. C. Yang, X. Tang, Q. Dai, H. Yang, and L. Jiang, "Identifying implicit and explicit relationships through user activities in social media," *International Journal of Electronic Commerce*, vol. 18, no. 2, pp. 73–96, 2013, doi: 10.2753/JEC1086-4415180203.

[4] S. M. Taheri, H. Mahyar, M. Firouzi, E. Ghalebi K, R. Grosu, and A. Movaghar, "Extracting implicit social relation for social recommendation techniques in user rating prediction," in *Proceedings of the 26th International Conference on World Wide Web Companion*, 2017, pp. 1343–1351, doi: 10.1145/3041021.3051153

[5] Ü. Atecs, "Inference of personality using social media profiles," Middle East Technical University, 2014.

[6] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of social media," *Bus Horiz*, vol. 53, no. 1, pp. 59–68, 2010, doi: 10.1016/j.bushor.2009.09.003.

[7] J. E. Barbuto Jr., "A critique of the Myers-Briggs Type Indicator and its operationalization of Carl Jung's psychological types," *Psychol Rep*, vol. 80, no. 2, pp. 611–625, 1997, doi: 10.2466/pr0.1997.80.2.611.

[8] G. Sperlì, "Community detection in Multimedia Social Networks using an attributed graph model," *Online Soc Netw Media*, vol. 46, p. 100312, May 2025, doi: 10.1016/J.OSNEM.2025.100312.

[9] D. Jin *et al.*, "A Survey of Community Detection Approaches: From Statistical Modeling to Deep Learning," *IEEE Trans Knowl Data Eng*, vol. 35, no. 2, pp. 1149–1170, 2023, doi: 10.1109/TKDE.2021.3104155.

[10] L. Ma, M. Gong, J. Yan, W. Liu, and S. Wang, "Detecting composite communities in multiplex networks: A multilevel memetic algorithm," *Swarm Evol Comput*, vol. 39, pp. 177–191, 2018, doi: /10.1016/j.swevo.2017.09.012.

[11] F. Karimi, S. Lotfi, and H. Izadkhah, "Multiplex community detection in complex networks using an evolutionary approach," *Expert Syst Appl*, vol. 146, 2020, doi: 10.1016/j.eswa.2020.113184.

[12] S. Lotfi and F. Karimi, "A Hybrid MOEA/D-TS for Solving Multi-Objective Problems," *Journal of AI and Data Mining*, vol. 5, no. 2, pp. 183–195, 2017, doi: 10.22044/jadm.2017.886.

[13] A. Biswas and B. Biswas, "Community-based link prediction," *Multimed Tools Appl*, vol. 76, no. 18, pp. 18619–18639, 2017, doi: 10.1007/s11042-016-4270-9.

[14] F. Karimi, S. Lotfi, and H. Izadkhah, "Community-guided link prediction in multiplex networks," *J Informetrics*, 2021, doi: 10.1016/j.joi.2021.101178.

[15] J. Serrano-Guerrero, B. Alshouha, M. Bani-Doumi, F. Chiclana, F. P. Romero, and J. A. Olivas, "Combining machine learning algorithms for personality trait prediction," *Egyptian Informatics Journal*, vol. 25, p. 100439, Mar. 2024, doi: 10.1016/J.EIJ.2024.100439.

[16] M. Eftekharian and A. Nodehi, "Breast cancer diagnosis and classification improvement based on deep learning and image processing," *Soft Comput. J*, vol. 12, no. 1, pp. 22–26, 2023, doi: 10.22052/scj.2023.246416.1067.

[17] Z. Sharifi Mehrjerd, H. Momeni and H. Adabi Ardakani, "A review of machine learning algorithms for diagnosing autism using EEG signals," *Soft Comput. J.*, vol. 13, no. 1, pp. 2-19, 2024, doi: 10.22052/scj.2023.248522.1110 [In Persian].

[18] L. R. Goldberg, "An alternative" description of personality": the big-five factor structure.," *J Pers Soc Psychol*, vol. 59, no. 6, p. 1216, 1990, doi: 10.1037//0022-3514.59.6.1216.

[19] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell, "Personality and patterns of Facebook usage," in *Proceedings of the 4th annual ACM web science conference*, 2012, pp. 24–32, doi: 10.1145/2380718.2380722.

[20] B. Verhoeven, W. Daelemans, and B. Plank, "Twisty: A multilingual twitter stylometry corpus for gender and personality profiling," in *Proceedings of the 10th Annual Conference on Language Resources and Evaluation (LREC 2016*, 2016, pp. 1–6, doi: 10.5281/zenodo.4638948

[21] M. D. Kamalesh and B. Bharathi, "Personality prediction model for social media using machine learning Technique," *Computers and Electrical Engineering*, vol. 100, p. 107852, 2022, doi: g/10.1016/j.compeleceng.2022.107852.

[22] I. Morais-Quilez, M. Graña, and J. de Lope, "Machine Learning for Personality Type

Classification on Textual Data," 2024, pp. 261–269. doi: 10.1007/978-3-031-61140-7_26, doi: 10.1007/978-3-031-61140-7_26.

[23] M. J. Rani, M. M. Adnan, N. N. Saranya, G. Karuna, and G. S. Margarat, "Mixture Kernel based Least Square Support Vector Machine for Personality Prediction on Social Media," in *2024 First International Conference on Software, Systems and Information Technology (SSITCON)*, 2024, pp. 1–5, doi: 10.1109/SSITCON62437.2024.10795938.

[24] S. Sivakumar, S. Priyanka, and P. Selvam, "Enhancing Personality Type Prediction with Ensemble Models: A Robust Predictive Approach," in *2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, 2023, pp. 1–7, doi: 10.1109/ICSES60034.2023.10465452.

[25] M. Ramezani, M.-R. Feizi-Derakhshi, M.-A. Balafar, and T. Ni, "Knowledge Graph-Enabled Text-Based Automatic Personality Prediction," *Intell. Neuroscience*, vol. 2022, Jan. 2022, doi: 10.1155/2022/3732351.

[26] M. Ramezani, M.-R. Feizi-Derakhshi, and M.-A. Balafar, "Text-based automatic personality prediction using KGrAt-Net: a knowledge graph attention network classifier," *Sci Rep*, vol. 12, no. 1, p. 21453, 2022, doi: 10.1038/s41598-022-25955-z.

[27] Y. Zhu, L. Hu, N. Ning, W. Zhang, and B. Wu, "A lexical psycholinguistic knowledge-guided graph neural network for interpretable personality detection," *Knowl Based Syst*, vol. 249, p. 108952, 2022, doi: https://doi.org/10.1016/j.knosys.2022.108952.

[28] K. Buraya, A. Farseev, A. Filchenkov, and T.-S. Chua, "Towards user personality profiling from multiple social networks," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[29] A. Farseev, L. Nie, M. Akbari, and T.-S. Chua, "Harvesting multiple sources for user profile learning: a big data study," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, 2015, pp. 235–242, doi: 10.1145/2671188.274938.

[30] A. Farseev, M. Akbari, I. Samborskii, and T.-S. Chua, "360 user profiling: past, future, and applications," *ACM SIGWEB Newsletter*, no. Summer, pp. 1–11, 2016, doi: 10.1145/2956573.2956577.

[31] Y. Chen, C. Zhuang, Q. Cao, and P. Hui, "Understanding cross-site linking in online social networks," in *Proceedings of the 8th Workshop on Social Network Mining and Analysis*, 2014, pp. 1–9, doi: 10.1145/3213898.

[32] C. A. Bliss, I. M. Kloumann, K. D. Harris, C. M. Danforth, and P. S. Dodds, "Twitter reciprocal reply networks exhibit assortativity with respect to happiness," *J Comput Sci*, vol. 3, no. 5, pp. 388–397, 2012, doi: 10.1016/j.jocs.2012.05.001.

[33] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: LIWC 2001," *Mahway: Lawrence Erlbaum Associates*, vol. 71, no. 2001, p. 2001, 2001.

[34] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[35] I. B. Myers, M. H. McCaulley, and R. Most, *Manual, a guide to the development and use of the Myers-Briggs type indicator*. consulting psychologists press, 1985.