نشریه علمی مطالعات مدیریت کسب و کار هوشمند- سال هشتم ـ شماره ۲۹ـ پاییز ۱۳۹۸
صفحات ۱۱۲- ۸۷

مدلی برای انتشار داده‌های شبکه‌های اجتماعی برخط با حفظ حریم خصوصی

روح‌الله کوثری لنگری *
سهیلا سردار **
سید عبدالله امین موسوی ***
رضا رادفر ****

چکیده

امروزه استفاده از شبکه‌های اجتماعی در میان اقشار مختلف جامعه جهانی، به‌صورت غیرقابل‌انکاری افزایش یافته است. پایگاه داده شبکه‌های اجتماعی؛ شامل منابع غنی و با ارزشی هستند که انتشار یا تحلیل آن‌ها برای مقاصد بازاریابی، تبلیغاتی، امنیت ملی، سلامت و ... می‌تواند برای محققان مؤسسات دولتی و خصوصی سودمند باشد؛ اما رعایت حریم خصوصی موجودیت‌هایی که اطلاعات آن‌ها در اختیار تحلیلگران داده‌کاوی قرار می‌گیرد، به‌عنوان یک پروتکل حقوقی ضروری است. در این مقاله، از طریق روش‌شناسی کیفی فراترکیب، کلیه ابعاد، شاخص‌ها و کدهای مربوطه استخراج و سپس میزان اهمیت و اولویت هر یک از عوامل، تعیین و متعاقباً مدل بهبودیافته گمنامی، به‌وسیله الگوریتم بهینه‌سازی کرم شب‌تاب و خوشه‌بندی فازی، ارائه شده است. نتایج شبیه‌سازی و ارزیابی‌های مدل پیشنهادی بر روی داده‌های چهار شبکه اجتماعی فیس‌بوک، یوتیوب، توییتر و گوگل پلاس، حاکی از حفظ حریم خصوصی داده‌ها با کمترین نسبت انحراف و بیشترین سودمندی است.

کلیدواژگان: شبکه اجتماعی، حریم خصوصی، گمنامی، الگوریتم کرم شب‌تاب، خوشه‌بندی فازی.

---

* دانشجوی دکتری، مدیریت فنّاوری اطلاعات، دانشکده مدیریت، دانشگاه آزاد اسلامی واحد تهران شمال، تهران.
** عضو هیئت‌علمی، گروه مدیریت صنعتی، دانشکده مدیریت، دانشگاه آزاد اسلامی واحد تهران شمال، تهران. (نویسنده مسئول)؛ s_sardar@iau-tnb.ac.ir
*** عضو هیئت‌علمی، گروه مدیریت فنّاوری اطلاعات، دانشکده مدیریت، دانشگاه آزاد اسلامی واحد علوم و تحقیقات، تهران.
**** عضو هیئت‌علمی، گروه مدیریت تکنولوژی، دانشکده مدیریت، دانشگاه آزاد اسلامی واحد علوم و تحقیقات، تهران، ایران.
تاریخ دریافت:۱۳۹۷/۱۲/۰۷ تاریخ پذیرش:۱۳۹۸/۰۶/۱۱

مقدمه

شبکه‌های اجتماعی به‌عنوان مجموعه‌ای از روابط تعریف می‌شوند که افرادی با علاقه‌مندی یکسان را به یکدیگر متصل می‌کند (الباراوی[1]، ۲۰۱۴). شبکه‌های اجتماعی آنلاین، روابط را به‌وسیله ساختار گراف یا جدول با استفاده از رئوس[2] و یال‌ها[3] مدل می‌کنند. رئوس (گره یا نود) نمایانگر فرد، موجودیت، گروه یا سازمان در یک شبکه هستند، درحالی‌که یال‌ها (لبه یا پیوند) نمایانگر روابط، تعاملات، دوستی‌ها یا همکاری‌های اعضاء می‌باشند. یکی از اهداف بهره‌برداری از پایگاه داده شبکه‌های اجتماعی آنلاین[4]، مسئله داده‌کاوی داده‌های حجیم[5] است که نگه‌دارندگان داده[6]، اقلام اطلاعاتی را برای پژوهش‌های علمی، بازاریابی و تبلیغاتی در اختیار تحلیلگران و نرم‌افزارهای داده‌کاوی قرار می‌دهند تا الگوها[7]، مدل‌ها و دانش مستتر مورد نیاز این سازمان‌ها از آن هجمه اطلاعات استخراج شود. با توسعه و گسترش شبکه‌های اجتماعی، مسئله نشست و افشای حریم خصوصی[8] داده‌های منتشرشده این شبکه‌ها، نگرانی مهمی تلقی می‌شود (تیواری و چودهاری[9]، ۲۰۱۷). نیاز به محرمانگی و پنهان ماندن اطلاعات حساس کاربران در شبکه‌های اجتماعی، مدل‌های مختلفی از گمنام سازی[10] را به وجود آورده است. این مدل‌ها؛ شامل دو هدف اساسی حفظ حریم خصوصی و میزان کیفیت و سودمندی داده‌های منتشر شده است (سرگلزاری و ازگمی، ۲۰۱۳). چالش مطرح‌شده در این فرایند، مسئله عدم توازن و تضاد این دو هدف است. برنامه‌های داده‌کاوی موثر حداقل باید تضمین‌های مناسبی برای حفظ حریم خصوصی بر روی داده‌های اساسی افراد تأمین کنند (تان و همکاران[11]، ۲۰۱۷). لذا، مفهوم حفظ حریم خصوصی داده‌کاوی[12]، مکانیسمی برای

1. ElBarawy
2. Nodes
3. Edges
4. Online Social Networks
5. Big Data
6. Data Holders
7. Patterns
8. Privacy Preserving
9. Tiwari & Choudhary
10. Anonymity
11. Tan et al.
12. Privacy Preserving Data Mining(PPDM)

ماسک کردن یا پاک کردن داده‌های برای جهت تضمین نتایج داده‌کاوی معتبر (سودمند) و پاسخ به نگرانی حفظ حریم خصوصی داده‌ها و تعادل در دقت داده‌ها است (سوئینی[1]، ۲۰۰۲). با توجه به موارد پیش گفته، این پژوهش به دنبال آن است تا ضمن رعایت محرمانگی داده‌ها، از طریق اصول الگوریتم بهینه‌سازی کرم شب‌تاب[2] و تکنیک خوشه‌بندی میانگین فازی[3]، یک مدل جامع، فارغ از نوع ساختار شبکه‌های اجتماعی ارائه نماید.

پیشینه پژوهش

شبکه‌های اجتماعی، ساختاری مبتنی بر گراف بدون جهت G(V,S,E) که در آن V مجموعه گره[4]، E مجموعه لبه[5] و S نمایانگر ویژگی حساس متشکل از گره‌هایی (عموماً فردی یا سازمانی) حاوی هویت‌ها و جریان اطلاعاتی با الگوهای پنهان و معتبر هستند؛ بنابراین، پیش-نیاز انتشار داده‌ها، اعمال فرایند حیاتی گمنام سازی است تا ضمن رعایت حفظ سودمندی داده، مانع از نشت و افشای هویت کاربران شود (تورا و ناوارو[6]، ۲۰۱۵). تاکنون تکنیک‌های متنوعی برای تضمین محرمانگی شبکه‌های اجتماعی مطرح شد است اما، یا در مقابل افشای صفت و هویت ضعف داشته یا نسبت انحراف[7] ایجادشده در داده‌ها و میزان اطلاعات ازدست‌رفته بالا بوده است (لی و همکاران[8]، ۲۰۱۷). مدل K-گمنامی[9] به‌عنوان پایه همه مدل-های گمنامی، نخستین بار توسط سوئیتی ارائه شد. ایده اصلی در K–گمنامی این است که در یک گراف، تعداد K رأس یا یال با خاصیت یکسان وجود دارد که مهاجم قادر به تشخیص هویت داده اصلی نشود. در پژوهشی با استفاده از مفهوم رمزگذاری، راهکاری جدید جهت افزایش حریم خصوصی کاربران شبکه اجتماعی فیس‌بوک و گوگل پلاس ارائه شد (اشلیگل

1. Sweeney
2. Firefly Algorithm(FFA)
3. Fuzzy C-Means Clustering(FCM)
4. Node
5. Edge
6. Torra & Navarro
7. Distortion Ratio
8. Li et al.
9. K-Anonymity(KA)

و همکاران[1]، ۲۰۱۷). در تحقیقی دیگر راهکاری نوین داده‌کاوی با حفظ حریم خصوصی مبتنی بر الگوریتم ژنتیک[2]؛ شامل یک تابع چند هدفه مبتنی بر قطعه‌بندی، پیشنهاد شد (سوزان و کریستوفر[3]، ۲۰۱۶). همچنین مدل جدید حریم خصوصی نیز برای امن سازی فرایند داده-کاوی شبکه‌های اجتماعی مطرح شد (بلواورگاز و همکاران[4]، ۲۰۱۶). پژوهشی دیگر، از الگوریتم بهینه‌سازی فاخته برای حفظ حریم خصوصی و پنهان‌سازی قوانین ارتباطی حساس[5] در داده‌های کلان بهره برد. عمل پنهان‌سازی، با استفاده از روش اعوجاج سبب افزایش مدل در گریز از هر جواب بهینه محلی[6] گردید (افشاری و همکاران[7]، ۲۰۱۶). در پژوهشی با استفاده از تعمیم برچسب گره و افزایش لبه، همسایگی گره را K-گمنام تا با دانش زمینه‌ای مهاجم و افشای هویت مقابله کند (لی و همکاران[8]، ۲۰۱۶). همچنین با استفاده از رویکرد جدیدی به نام MR-Cube بر مشکلات محرمانگیِ داده‌های توزیع‌شده شبکه‌های اجتماعی غلبه شد (آکشایا و آمریت[9]، ۲۰۱۶). در تحقیقی دیگر مدلی به نام TLK3L ارائه شد که ضمن بهبود روش K-گمنامی و توزیع متغیرهای حساس، از افشای صفت در حملات تشابه جلوگیری به عمل آمد (رحیمی و همکاران[10]، ۲۰۱۶). در پژوهشی دیگر با تغییر گراف، به ارائه الگوریتمی حریصانه برای مقابله با افشای پیوند، بهبود سودمندی داده‌ها و حفظ حریم خصوصی داده‌های منتشرشده پرداخته شد (سرگلزاری و همکاران، ۱۳۹۲). همچنین مدل بهینه K-هم‌ریختی با رویکرد مقابله با افشای پیوند معرفی شد. در این مدل آن‌قدر لبه به شبکه اضافه می‌شود تا هر گره با حداقل k-1 گره دیگر یک ریخت شود (چن و همکاران[11]، ۲۰۱۴). مدلی دیگر با بهره‌گیری از بهینه‌سازی گروهی و الگوریتم اجتماع ذرات[12]، روش داده‌کاوی با

1. Schlegel et al.
2. Genetic Algorithm
3. Susan & Christopher
4. Bello-Orgaz et al.
5. COA4ARH
6. Stepwise Optimal
7. Afshari et al.
8. Li et al.
9. Akshaya & Amrit
10. Rahimi et al.
11. Chen et al.
12. Particle Swarm Optimization(PSO)

حفظ حریم خصوصی را توسعه داد. (مانداپاتی و همکاران[۱]، ۲۰۱۳). در تحقیقی دیگر از طریق الگوریتم ژنتیک، روش K-گمنامی بهبود داده شد. K مقدار آستانه گمنام سازی ساختار و تعداد خوشه‌هایی است که نودها می‌توانند دسته‌بندی شوند (سیهاک و همکاران[۲]، ۲۰۱۲). با استفاده از تکنیک فازی و خوشه‌بندی مدلی به نام K-Member Anonymity(KFKA) به‌منظور بهبود عملکرد الگوریتم K-گمنامی ارائه شد (هوندا و همکاران[۳]، ۲۰۱۴). همچنین در مدلی دیگر با استفاده از تکنیک درخت تصمیم، الگوریتم K-گمنامی توسعه و بهبود داده شد (بینچگام و هوهوآ[۴]، ۲۰۱۱). در پژوهشی از روش حذف و اضافه نمودن تصادفی لبه برای رسیدن به حد مشخصی از آشفتگی و با در نظر گرفتن معیارهای مشابهت گره به‌عنوان دانش زمینه‌ای مهاجم، استفاده کرده است (یانگ[۵]، ۲۰۱۱). همچنین به‌منظور بهبود مدل K-گمنامی، روشی با نام K-Support Anonymity ارائه شد که حفاظت از حریم خصوصی را تا حدودی نسبت به روش‌های قبلی افزایش داد (تای و همکاران[۶]، ۲۰۱۰). به‌منظور رفع مشکل تنوع کم در مقادیر متغیرهای حساس که منجر به ضعف داده در برابر حملات افشاء صفت می‌شود مدلی دیگر به نام P-Sensitive K-anonymity(PKA) ارائه شد (تروتا و وینای[۷]، ۲۰۰۶). همچنین به‌منظور توسعه مدل K-گمنامی و رفع محدودیت افشاء صفت، مدل L-تنوع[۸] مطرح شد؛ اما این روش توزیع متغیرهای حساس را به‌طور کامل رعایت نکرد (ماچاناواجالا و همکاران[۹]، ۲۰۰۶). در پژوهشی دیگر برای بهبود مدل K-گمنامی و رفع محدودیت میزان توزیع‌شدگی ویژگی‌های حساس داده‌ها، مدل T-همسایگی[۱۰] ارائه شد اما همچنان در برابر افشای صفت و حملات تشابه، آسیب‌پذیر است (لی و همکاران، ۲۰۰۷).

1. Mandapati et al.
2. Sihag et al.
3. Honda et al.
4. Bingchun & Guohua
5. Yang
6. Tai et al
7. Truta & Vinay
8. L-Diversity
9. Machanavajjhala et al
10. T-Closeness

روش‌شناسی پژوهش

این پژوهش از نظر هدف کاربردی و از لحاظ راهبرد تحقیق از نوع تحقیق آمیخته (کیفی و کمی) محسوب می‌شود. در فاز کیفی، داده‌های کیفی سایر تحقیقات و مدل‌های پیشین، گردآوری و با روش‌شناسی فراترکیب[1] در تدوین چارچوب مدل پیشنهادی استفاده خواهد شد. چراکه فراترکیب به مطالعهٔ کیفی اطلاعات و یافته‌های استخراج‌شده از سایر مطالعات مرتبط با موضوع پرداخته و با فراهم کردن نگرش سامانمند، به کشف راهکارهایی جدید کمک می‌کند (باروسو و ساندلوسکی[2]، ۲۰۰۶). به منظور تعیین روایی[3] از روش ارزیابی حیاتی[4] و برای تعیین پایائی[5] فاز کیفی از طریق محاسبه ضریب کاپا-کوهن[6] (توافق دو کدگذار) اقدام شده است. اخذ نمونه آماری پژوهش، از طریق روش گردآوری کتابخانه‌ای؛ شامل دیتا گرافی با ۱۳۸۳ گره از چهار شبکه اجتماعی فیس‌بوک، یوتیوب، توییتر و گوگل پلاس از موسسه علم و فناوری دانشگاه UTM مالزی؛ شامل آخرین وضعیت چهار شبکه موصوف در سال ۲۰۱۷ است. نهایتاً در فاز کمی، ساختار مدل پیشنهادی به همراه چهار مدل منتخب از پیشینه پژوهش از طریق برنامه‌نویسی در محیط نرم‌افزاری متلب پیاده‌سازی شده و نتایج مدل‌ها از طریق اعمال نمونه آماری، مورد مقایسه قرار خواهند گرفت.

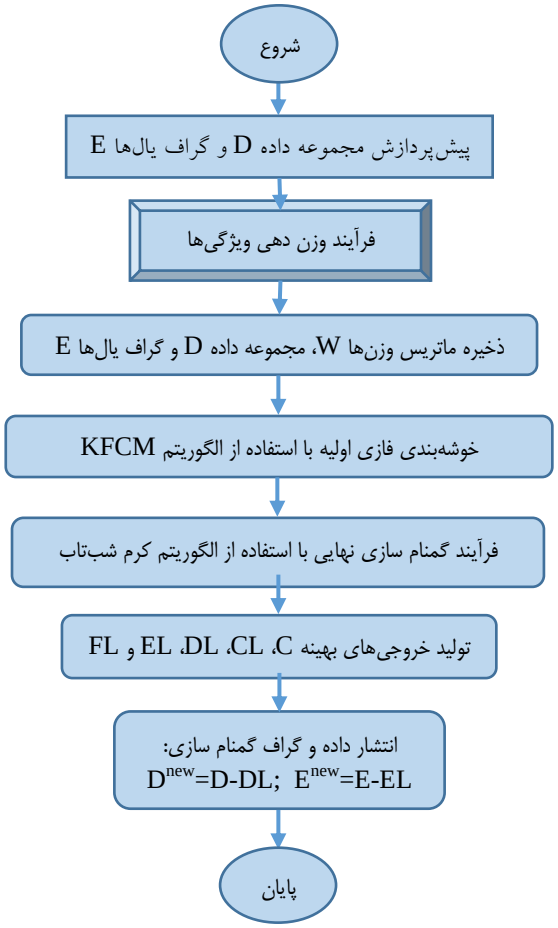
مدل مفهومی پژوهش

همه مدل‌های گمنامی درصدد هستند تا میزان محرمانگی داده را بیشینه و میزان تخریب داده را کمینه کنند. این تحقیق یک مسأله خوشه‌بندی بسیار پیچیده[7] است که باید قیدها و معیارهای مختلف گمنام‌سازی را در کنار جامعیت داده‌ها رعایت نماید؛ بنابراین، الگوریتم‌های فرا ابتکاری همچون کرم شب‌تاب مناسب است. ابتدا، مسأله را به‌صورت مدل بهینه‌سازی چندهدفه مقید فرموله می‌کنیم. ساختار بهینه‌سازی الگوریتم کرم شب‌تاب، یک خوشه‌بندی

1. Meta Synthesis
2. Sandelowski, & Barroso
3. Validity
4. CASP
5. Reliability
6. Cohen's KAPPA
7. NP-Hard

بهینه از مجموعه نمونه‌ها ارائه کرده و همزمان به دنبال حداقل کردن نسبت انحراف داده و معیارCAVG[1] است. همچنین، برآورده شدن K- گمنامی، L- تنوع و T- همسایگی به‌عنوان قیدهای مسئله، در قالب یک تابع جریمه به تابع هدف مسأله افزوده می‌شود. به منظور بهبود کارایی روش پیشنهادی، به‌جای استفاده از خوشه‌بندی اولیه تصادفی، از خوشه‌بندی فازی متوازن به‌منظور ایجاد خوشه‌های اولیه برای الگوریتم کرم شب‌تاب استفاده می‌شود. توابع هدف مورد استفاده به‌منظور حداقل کردن نسبت فواصل درون خوشه‌ای به فواصل بین خوشه‌ای، حداقل کردن میانگین نسبت انحراف داده و گراف و حداقل کردن معیار CAVG در نظر گرفته شده است. هدف از مدل پیشنهادی مبتنی بر خوشه‌بندی فازی و الگوریتم کرم شب‌تاب، خوشه‌بندی نمونه‌ها به C دسته مختلف است که تعداد نمونه‌های موجود در هر دسته حداقل K عضو و معیارهای L-تنوع و T-همسایگی نیز به‌طور هم‌زمان برای همه خوشه‌ها برقرار باشد. فرایند مدل بدین صورت است که ابتدا، پیش‌پردازش داده و تشکیل گراف از مجموعه داده خام اولیه انجام می‌شود. بدین منظور، ماتریسی باینری یال E به ابعاد $N\times N$ ایجاد می‌شود که در آن $E_{ij}=1$ بیانگر وجود یال بین نمونه i و j است. همچنین، ماتریس داده D به ابعاد $N\times M$ ایجاد کرده که در آن، M تعداد ویژگی‌های هر نمونه است. به‌عبارت‌دیگر، D_{ik} ویژگی k از نمونه i را نشان می‌دهد. اگر ماتریس E ماتریس گراف اولیه و D ماتریس داده اولیه باشد، پس از گمنام سازی، ماتریس گراف و داده منتشر شونده به ترتیب با E^{new}، D^{new} نشان داده می‌شوند. به‌طورکلی متغیرهای بهینه‌سازی در الگوریتم پیشنهادی؛ شامل تعداد خوشه‌های بهینه C، نتایج خوشه‌بندی CL که تخصیص نمونه‌های مختلف به خوشه‌های مختلف را بیان می‌کند، ماتریس انتخاب ویژگی FL که نشان‌دهنده حضور یا حذف ویژگی‌های مختلف داده اصلاح شده است، ماتریس تغییرات گراف EL که بیان‌کننده تغییرات در بعضی یال‌های گراف است و ماتریس تغییرات داده DL که بیانگر تغییرات در داده است. ماتریس‌های داده و گراف منتشر شونده به‌راحتی با استفاده از $D^{new}=D-DL$ و $E^{new}=E-EL$ قابل محاسبه هستند.

1. Cernability Average Groups(CAVG)



شکل ۱: مراحل کلی آماده‌سازی اطلاعات و اجرای مدل پیشنهادی

الگوریتم خوشه‌بندی پیشنهادی KFCM[1]

به منظور تولید C و CLA برای هر کرم شب‌تاب اولیه، فرایند خوشه‌بندی با استفاده از الگوریتم KFCM به این صورت است که ابتدا مقداری برای C به نام $C_{initial}$ در بازه مجاز

1. K-anonymity Fuzzy C-Means Clustering

C_{min} تا C_{max} انتخاب می‌شود. سپس، خوشه‌بندی فازی بر روی تمام N نمونه اجرا می‌شود. خوشه‌هایی که دارای تعداد اعضای بیشتر یا مساوی مقدار K هستند را به‌عنوان مجموعه S1 و خوشه‌هایی که تعداد اعضایشان کمتر از K باشد را به‌عنوان مجموعه S2 در نظر می‌گیریم. در مرحله بعد، ادغام و تفکیک خوشه‌ها به‌گونه‌ای انجام می‌شود که تمام خوشه‌ها متوازن بوده و شرط K را برقرار کنند. برای این منظور، تک‌تک خوشه‌های مجموعه S2 به‌صورت جداگانه بررسی می‌شوند. برای هر خوشه از مجموعه S2(خوشه c) که دارای m عضو است (m<K)، نزدیک‌ترین خوشه از بین کل خوشه‌های موجود انتخاب می‌شود (خوشه k). اگر خوشه k از مجموعه S2 باشد، دو خوشه c و k با هم ادغام می‌شوند و درصورتی‌که هنوز شرط K برقرار نباشد، تا زمانی که تعداد اعضای خوشه به بالای K عضو برسد این کار تکرار می‌شود. از طرفی، چنانچه خوشه k از مجموعه S1 باشد، اگر تعداد اعضای خوشه k حداقل به‌اندازه K-m عضو بیشتر از K عضو باشد، آنگاه تعداد K-m عضو از نزدیک‌ترین اعضای خوشه k به مرکز خوشه c انتخاب شده و از خوشه k به خوشه c انتقال می‌یابند. در غیراینصورت دومین خوشه نزدیک به خوشه c برای ادغام یا تفکیک در نظر گرفته می‌شود. پس از اصلاح، خوشه c از مجموعه S2 خارج شده و به مجموعه S1 تعلق می‌یابد و مراکز خوشه مجدداً به‌روزرسانی می‌شوند. این فرآیند تا زمانی که هیچ خوشه‌ای درون مجموعه S2 باقی نمانده باشد، ادامه می‌یابد. پس از اجرای الگوریتم KFCM پیشنهادی، خوشه‌های متوازن حاصل می‌شود. شبه کد روش خوشه‌بندی KFCM در الگوریتم ۱ شکل ۲ و فلوچارت آن در شکل ۳ نشان داده شده است.

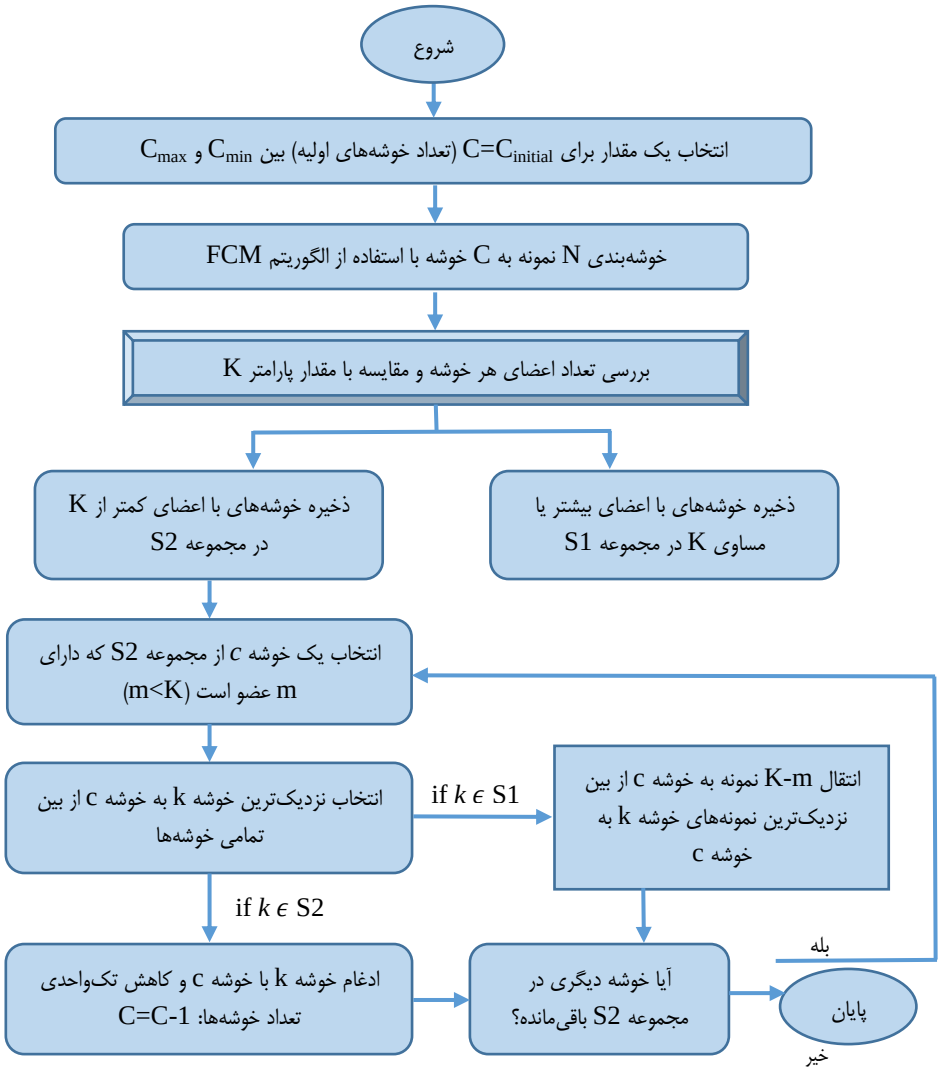
Algorithm 1. Proposed Hybrid K-anonymity and FCM Clustering (KFCM)

Input: Initial Table T containing Dataset D & Graph E
Parameters: K & weights of features W
Output: C & CLA
Begin

1 Construct input matrix for clustering: MAT=E+W×D.
2 Calculate a value for $C=C_{initial}$ in the range of $[C_{min},C_{max}]$.
3 Clustering MAT into C clusters using FCM.
4 Consider all clusters with at least K members as S1 and the others in S2.
5 **For** c = 1: Number of Clusters in S2
6 **For** k = 1: C ($k \neq c$)
7 Calculate distance between the centroid of the two clusters: Dist(c,k).
8 **End For**
9 Find the nearest cluster k to cluster c.
10 **If** cluster k belongs to S2
11 Merge the two clusters k and c as a single cluster c.
12 Update the number of clusters: C=C-1.
13 **If** the merged clusters have at least K members
14 Eliminate the merged cluster from S2 and add it into S1.
15 **End If**
16 **Else**
17 Find the K-m nearest members of cluster k and transfer them into cluster c.
18 Eliminate cluster c from S2 and add it into S1.
19 **End If**
20 **End For**

End

شکل ۲: شبه کد الگوریتم خوشه‌بندی KFCM



شکل ۳: فلوچارت الگوریتم خوشه‌بندی متوازن پیشنهادی KFCM

تابع چندهدفه مقید پیشنهادی

در این پژوهش، به‌منظور دستیابی هم‌زمان به بالاترین دقت گمنام سازی، کمترین خطای

خوشه‌بندی و خطای گمنام سازی ، رعایت شرایط K-گمنامی، L- تنوع و T- همسایگی، مدل را به‌صورت یک مسئله بهینه‌سازی فرموله می‌کنیم. بنابراین، یک تابع چندهدفه مقید؛ شامل یک تابع خطای سه هدفه و یک تابع جریمه؛ شامل سه قید مجزا تعریف می‌شود. اهداف تابع هدف؛ شامل حداقل کردن نسبت میانگین فواصل درون خوشه‌ای به میانگین فواصل بین خوشه‌ای، حداقل کردن معیار CAVG که منجر به بهبود پراکندگی خوشه‌ها و تفکیک بهتر خوشه‌ها می‌شود و حداقل کردن میانگین نسبت انحراف (DR) جدول اصلاح‌شده نسبت به اطلاعات اولیه (اتلاف اطلاعات)، می‌باشند. در روش پیشنهادی، تغییرات و اصلاحات باید بر روی گراف، داده و ویژگی‌های داده اعمال شوند، لذا میانگین نسبت انحراف (DR) از میانگین تغییرات گراف، داده و ویژگی‌ها حاصل می‌شود. سه قید اصلی مسئله؛ شامل رعایت K-گمنامی، L-تنوع و T-همسایگی برای تک‌تک خوشه‌ها است. برای اطمینان از برقراری این شرایط، برای همه خوشه‌ها یک تابع جریمه مناسب بیان می‌شود که بر اساس تعداد شروطی که ارضا نشده‌اند، قابل محاسبه است. تابع چندهدفه مقید پیشنهادی (ObjFun)؛ شامل تابع خطای سه هدفه (Cost) و تابع جریمه (Penalty) برای محاسبه خطای هر کرم شب‌تاب به‌صورت رابطه ۱ بیان می‌شود.

$$\mathrm{ObjFun}=\mathrm{Cost}\times(1+\mathrm{Penalty}) \quad \text{رابطه (۱)}$$

$$\mathrm{Cost}=w_1\times F_1+w_2\times F_2+w_3\times F_3 \quad \text{رابطه (۲)}$$

$$F_1=\frac{\mathrm{IntraDist}}{\mathrm{InterDist}}=\frac{\frac{1}{N}\sum_{i=1}^{N}d\left(node_i,center_{CLA(i)}\right)}{\frac{1}{C\times(C-1)}\sum_{j=1}^{C}\sum_{k=1}^{C}d\left(center_j,center_k\right)} \quad \text{رابطه (۳)}$$

$$F_2=\mathrm{CAVG}=\frac{\left(\frac{N}{C}\right)}{K} \quad \text{رابطه (۴)}$$

$$F_3=\mathrm{DR}=\frac{\mathrm{DR_F}+\mathrm{DR_D}+\mathrm{DR_E}}{3} \quad \text{رابطه (۵)}$$

$$\mathrm{DR_F} = \frac{\sum_{k=1}^{M} FL(k)}{M}$$ رابطه (۶)

$$\mathrm{DR_D} = \frac{\sum_{i=1}^{N}\sum_{k=1}^{M} DL(i,k)}{N \times M}$$ رابطه (۷)

$$\mathrm{DR_E} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} EL(i,j)}{N \times N}$$ رابطه (۸)

$$FL(k) = \begin{cases} 1 & \text{if feature } k \text{ has been eliminated} \\ 0 & \text{otherwise} \end{cases}$$ رابطه (۹)

رابطه (۱۰)

$$DL(i,k) = \begin{cases} 1 & \text{if feature } k \text{ of node } i \text{ has been changed} \\ 0 & \text{otherwise} \end{cases}$$

رابطه (۱۱)

$$EL(i,j) = \begin{cases} 1 & \text{if edge between node } i \text{ and } j \text{ has been ch} \\ 0 & \text{otherwise} \end{cases}$$

در رابطه ۲، w_1، w_2 و w_3 سه ضریب وزنی ثابت هستند که به ترتیب نسبت تأثیر سه تابع خطای F_1، F_2 و F_3 را بر تابع هدف کلی (Cost) تعیین می‌کنند. مجموع این ضرایب باید برابر با یک باشد. هر چه ضریب یک تابع خطا بزرگ‌تر باشد، ارزش تابع خطای متناظر در تابع هدف بیشتر می‌شود. F_1 در رابطه ۳ به‌صورت نسبت میانگین فواصل درون خوشه‌ای به میانگین فواصل بین خوشه‌ای تعریف شده است که $d(\mathrm{node}_i,\mathrm{center}_{CLA(i)})$ فاصله اقلیدسی نمونه i تا مرکز خوشه مربوطه و $d(\mathrm{center}_j,\mathrm{center}_k)$ فاصله مرکز خوشه j و مرکز خوشه k *را* بیان می‌کند. F_2 در رابطه ۴ معیار CAVG را تعریف می‌کند که با توجه به شرط K-گمنامی همواره عددی بزرگ‌تر یا مساوی ۱ است و بهترین مقدار آن، عدد ۱ است. همچنین در رابطه ۵، میانگین نسبت انحراف (DR) جدول اصلاح‌شده نسبت به جدول اولیه آورده

شده است که به‌صورت میانگین تغییرات ایجادشده در ویژگی‌ها (DR_F)، تغییرات ایجادشده در داده (DR_D) و تغییرات ایجادشده در گراف یال‌ها (DR_E) به ترتیب بر اساس روابط ۶ تا ۸ محاسبه می‌شود. لازم به ذکر است که با توجه به میانگین‌گیری در روابط بالا، هر سه تابع خطای F_1، F_2 و F_3 در محدوده عدد ۱ قرار دارند و نیازی به نرمال‌سازی ندارند. تابع جریمه (Penalty) باید به‌گونه‌ای باشد که درصورتی‌که تمام قیود مسئله برقرار باشند مقدار آن برابر با یک شود. در غیر این صورت حتی اگر یکی از قیدها برقرار نباشد، Penalty باید مقدار بزرگی داشته باشد و هر چه تعداد قیود بیشتری برقرار نباشند، مقدار آن به‌مراتب بزرگ‌تر شود. تابع جریمه به‌صورت رابطه ۱۲ بیان شده است که در آن P_K، P_L و P_T به ترتیب تعداد عدم ارضای قیود K-گمنامی، L-تنوع و T-همسایگی هستند. به‌عبارت‌دیگر اگر شرط K-گمنامی برای خوشه i برقرار باشد K_not(i)=0 و در غیر این صورت K_not(i)=1 است. به همین ترتیب اگر شرط L-تنوع برای خوشه i برقرار نباشد L_not(i)=1 و اگر شرط T-همسایگی برای خوشه i برقرار نباشد T_not(i)=1 است.

$$\text{Penalty} = P_K + P_L + P_T$$ رابطه (۱۲)

$$P_K = \sum_{i=1}^{C} K\_not(i)$$ رابطه (۱۳)

$$P_L = \sum_{i=1}^{C} L\_not(i)$$ رابطه (۱۴)

$$P_T = \sum_{i=1}^{C} T\_not(i)$$ رابطه (۱۵)

اجرای الگوریتم کرم شب‌تاب

الگوریتم کرم شب‌تاب مبتنی بر جمعیت بوده و استراتژی جستجوی سراسری داشته و در فرآیندی تکراری به راه‌حل بهینه[1] (راه‌حل نزدیک به بهینه) دست می‌یابد.

1. Optimal Solution

تولید جمعیت اولیه تصادفی

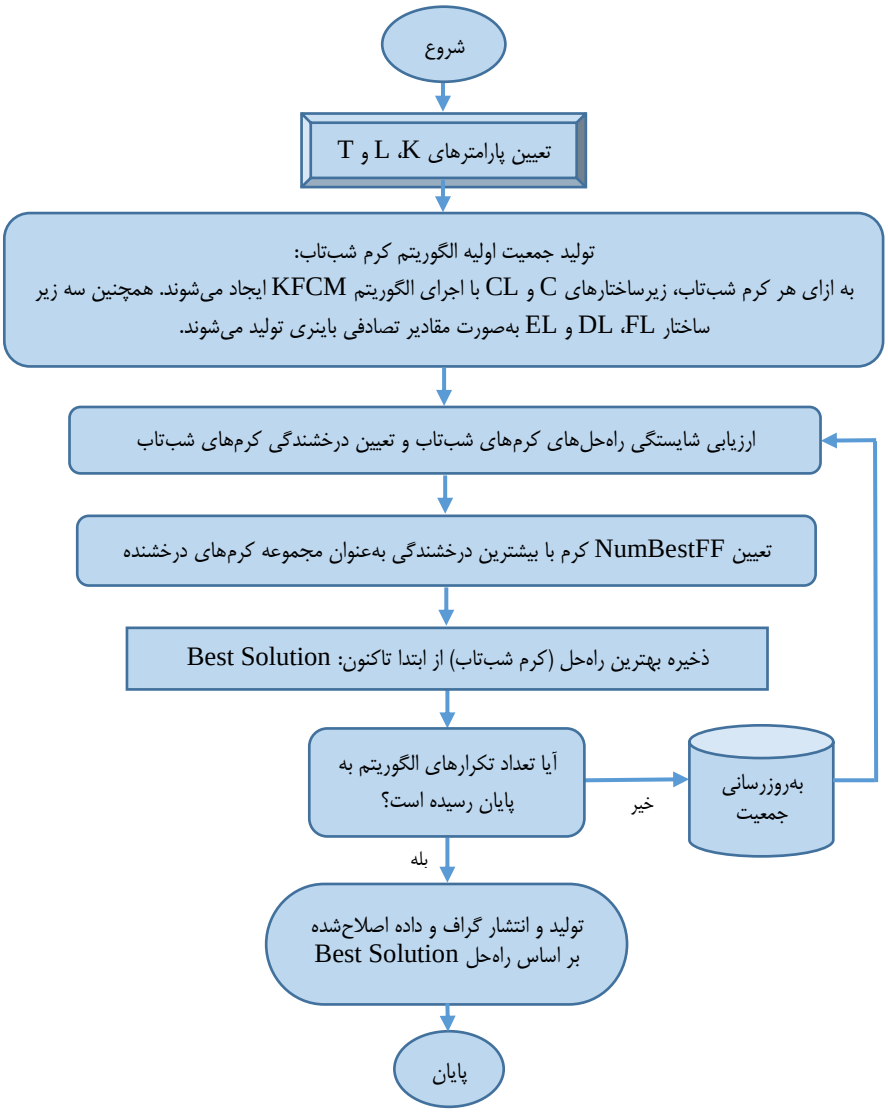
فرآیند جستجو در الگوریتم کرم شب‌تاب با تولید یک جمعیت اولیه تصادفی آغاز می‌شود. برای تولید یک راه‌حل تصادفی، زیرساختارهای FL، DL و EL به‌صورت تصادفی با عناصر صفر یا یک تولید می‌شوند. برای زیر ساختار C، یک عدد صحیح تصادفی بین C_{min} و C_{max} انتخاب می‌شود و زیر ساختار CL با اعداد پیوسته تصادفی در بازه [0,1) ایجاد می‌شود. بااین‌وجود، در الگوریتم ترکیبی FFA-KFCM دو زیر ساختار C و CL از اجرای الگوریتم‌های KFCM حاصل می‌شوند. در هر تکرار از الگوریتم، ابتدا ارزیابی شایستگی راه‌حل‌های تولیدشده و سپس به‌روزرسانی جمعیت صورت می‌گیرد. این دو مرحله پی‌درپی اجرا می‌شوند تا شرط خاتمه الگوریتم فرا برسد.

ارزیابی شایستگی راه‌حل‌های تولید شده

در هر تکرار از الگوریتم، پس از به‌روزرسانی جمعیت، مقدار خطای راه‌حل‌های تولیدشده با استفاده از تابع چندهدفه مقید پیشنهادی مطابق رابطه ۱ ارزیابی می‌شوند. سپس میزان درخشندگی (شایستگی) هر راه‌حل به‌صورت معکوس تابع هدف محاسبه می‌گردد؛ بنابراین میزان درخشندگی یک کرم شب‌تاب با استفاده از Fitness=1/ObjFun قابل محاسبه است.

به‌روزرسانی جمعیت

در هر تکرار، پس از ارزیابی شایستگی و تعیین میزان درخشندگی کرم‌های شب‌تاب، تعداد NumBestFF کرم شب‌تاب که بیشترین میزان درخشندگی را حاصل کرده‌اند، به‌عنوان مجموعه کرم‌های شب‌تاب برتر (درخشنده) انتخاب می‌شوند. سپس از بین کرم‌های شب‌تاب باقی‌مانده، هرکدام فقط به سمت کرم‌های شب‌تاب درخشنده حرکت می‌کند. فاصله کرم شب‌تاب i تا کرم شب‌تاب j به‌صورت فاصله اقلیدسی دو کرم به‌صورت r_{ij} بیان می‌شود که این فاصله می‌تواند در هرکدام از پنج زیر ساختار C، CL، FL، DL و EL محاسبه شود.



شکل ٤: فلوچارت کلی الگوریتم پیشنهادی

یافته‌های پژوهش

وضعیت عملکرد مدل پیشنهادی KFCM با چهار روش منتخب از پیشینه پژوهش؛ شامل KA، PKA، KFKA و TLK3L ازنظر معیارهای مختلف، مورد مقایسه و ارزیابی قرار

گرفته است.

معیارهای ارزیابی

معیارهای ارزیابی برای پیاده‌سازی و مقایسه روش پیشنهادی با روش‌های منتخب، به شرح ذیل است:

▪ **خطای خوشه‌بندی**[1] (CE)**:** این معیار کیفیت خوشه‌بندی را نشان می‌دهد و به‌صورت نسبت میانگین فواصل درون‌خوشه‌ای به میانگین فواصل بین‌خوشه‌ای محاسبه می‌شود. مقدار این معیار عددی بین صفر و یک است که هر چه کمتر باشد، حاکی از خوشه‌بندی مطلوب‌تر است (رحیمی، ۲۰۱۱).

▪ **میانگین گروه‌های متوازن** (CAVG)**:** این معیار کیفیت **K**-گمنامی را بر اساس توازن توزیع نمونه‌ها در خوشه‌های مختلف با محاسبه میانگین اعضای خوشه‌ها بر پارامتر **K** به‌صورت رابطه ۴ بیان می‌کند. هر چه مقدار **CAVG** برای یک داده گمنام سازی کمتر باشد، نمونه‌ها با توازن مطلوب‌تری بین خوشه‌ها توزیع شده‌اند (وال و ولینبرگ[2]، ۱۹۹۹).

▪ **میانگین نسبت انحراف**[3] (ADR)**:** میانگین نسبت انحراف جدول اصلاح‌شده نسبت به جدول اولیه است که به‌صورت میانگین تغییرات ایجادشده در ویژگی‌ها ($\mathbf{DR_F}$)، تغییرات ایجادشده در داده ($\mathbf{DR_D}$) و تغییرات ایجادشده در گراف ($\mathbf{DR_E}$) محاسبه می‌شود. مقدار این معیار نیز عددی بین صفر و یک است که هرچه کوچک‌تر باشد، مناسب‌تر است. برای این معیار، صفر به معنای عدم وجود هیچ اصلاحی در ویژگی‌ها، داده و گراف است و یک به معنای تغییر تمام عناصر ویژگی‌ها، داده و گراف است.

▪ **خطای کلی گمنام سازی** (Cost)**:** این معیار خطای کلی روش گمنام سازی را بر اساس میانگین وزنی خطاهای **CE**، **CAVG** و **ADR** بر اساس رابطه ۲ محاسبه می‌کند.

▪ **تابع جریمه** K (P_K)[4]**:** تعداد خوشه‌هایی که شرط **K**-گمنامی را برقرار نکرده‌اند.

1. Clustering Error(CE)
2. Waal & Willenborg
3. Average Distortion Rate(ADR)
4. Penalty Function for K-Anonymity

- **تابع جریمه** L (P_L)[1]: تعداد خوشه‌هایی که شرط **L**-تنوع را برقرار نکرده‌اند.
- **تابع جریمه** T (P_T)[2]: تعداد خوشه‌هایی که شرط **T**- همسایگی را برقرار نکرده‌اند.
- **تعداد خوشه‌ها** (C): تعداد خوشه‌های نهایی ایجادشده پس از اجرای الگوریتم است. هر چه تعداد خوشه‌های ایجادشده بیشتر باشد، میانگین تعداد نمونه‌های موجود در خوشه‌ها به **K** نزدیک‌تر شده و معیار **CAVG** کمتر می‌شود.
- **زمان اجرای الگوریتم** (CPU Time): زمان اجرای الگوریتم گمنام سازی برحسب ثانیه است.

تجزیه‌و تحلیل داده‌ها

نتایج به‌دست‌آمده از معیارهای مختلف نظیر تابع خطا (Cost)، تابع جریمه (Penalty)، تابع هدف (ObjFun) و زمان اجرا (برحسب ثانیه) برای مجموعه داده فیس‌بوک، گوگل پلاس، یوتیوب و توییتر در جداول ۱ تا ۴ نشان داده شده است.

جدول ۱: مقایسه معیارهای مختلف برای مجموعه داده فیس‌بوک

Parameter	KA	PKA	KFKA	TLK3L	FFA	FFA-KFCM
C	۲۱/۲	۱۷/۹	۸۶	۱۴	۴۳/۸	۴۷/۱
CE	۰/۰۸۱	۰/۱۱۸	۰/۰۱۷	۰/۱۹	۰/۰۸۸	۰/۰۷۱
CAVG	۴/۰۸	۴/۸۳	۱/۰۰۱	۶/۱۹	۱/۹۸	۱/۸۴
ADR	۰	۰/۰۰۱	۰	۰/۰۷	۰/۰۰۴	۰/۰۲۴
P_K	۰	۰	۰	۰	۰	۰
P_L	۲/۳	۰	۲۳/۶	۰	۰	۰
P_T	۲۲/۷	۱۷/۴	۷۱/۵	۰	۰	۰
Cost	۰/۴۵۷	۰/۵۵۴	۰/۱۱۱	۰/۷۵۸	۰/۲۵۷	۰/۲۲۵
Penalty	۲۶	۱۷/۴	۹۵/۱	۰	۰	۰
ObjFun	۱۲/۳۶	۱۰/۱۷	۱۰/۷	۰/۷۵۸	۰/۲۵۷	۰/۲۲۵
CPU Time	۰/۷	۳/۱	۳/۴	۲/۳	۴۳	۴۵

1. Penalty Function for L-Diversity

2. Penalty Function for T-Closeness

جدول ۲: مقایسه معیارهای مختلف برای مجموعه داده گوگل پلاس

Parameter	KA	PKA	KFKA	TLK3L	FFA	FFA-KFCM
CE	۰/۱۲۲	۰/۱۷۲	۰/۰۱۹	۰/۲۲۵	۰/۱۴۳	۰/۱۳۳
CAVG	۸	۸/۸	۱/۰۱۱	۴	۲/۷۵	۲/۸۳
ADR	۰	۰/۰۰۱	۰	۰/۲۱۸	۰/۱۵۵	۰/۰۱۴
ObjFun	۷/۸۵	۴/۹۱	۶/۶۴	۰/۶	۰/۴۰۷	۰/۳۶۸

جدول ۳: مقایسه معیارهای مختلف برای مجموعه داده توییتر

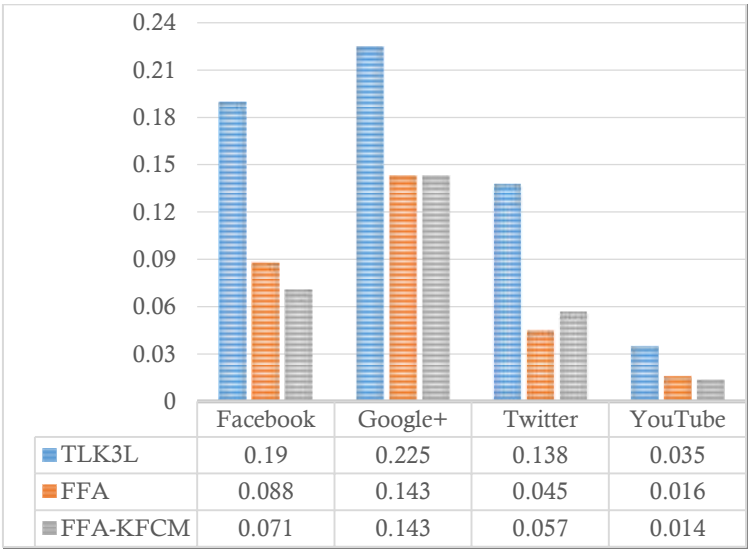
Parameter	KA	PKA	KFKA	TLK3L	FFA	FFA-KFCM
CE	۰/۶۱	۰/۵۷۹	۰/۰۴۳	۰/۱۳۸	۰/۰۴۵	۰/۰۵۷
CAVG	۱۰/۱۶	۸/۷۱	۱/۰۱۶	۲/۶۵	۱/۰۸	۱/۲۷
ADR	۰	۰/۰۰۱	۰	۰/۰۴۲	۰/۱۵	۰/۰۱۲
ObjFun	۲/۷۶	۲/۴۳	۳/۶۶	۰/۳۶	۰/۱۸۱	۰/۱۶۳

جدول ٤: مقایسه معیارهای مختلف برای مجموعه داده یوتیوب

Parameter	KA	PKA	KFKA	TLK3L	FFA	FFA-KFCM
CE	۰/۰۳۱	۰/۰۳	۰/۰۱۳	۰/۰۳۵	۰/۰۱۶	۰/۰۱۴
CAVG	۱/۸۴	۱/۸۷	۱/۰۰۴	۲/۰۰۱	۱/۰۱۶	۱/۰۱
ADR	۰	۰/۰۰۱	۰	۰/۰۵۱	۰/۰۲	۰/۰۱۶
ObjFun	۰/۲۱	۰/۲۰۷	۰/۱۰۸	۰/۲۲۴	۰/۱۱۶	۰/۱۱۳

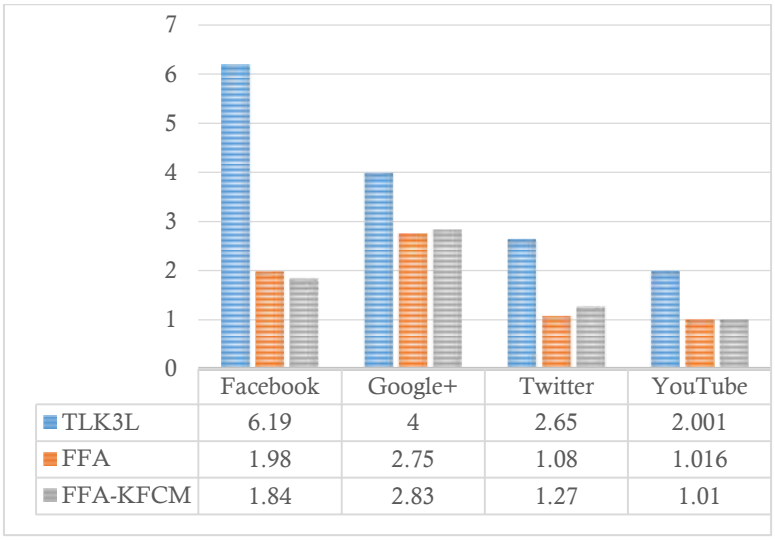
مطابق با جداول ۱ تا ۴ روش‌های KA و KFKA فقط معیار K-گمنامی را برقرار کرده‌اند و روش PKA دو معیار K-گمنامی و L-تنوع را ارضا می‌کند. ولی هر سه روش پیشنهادی و نیز الگوریتم TLK3L به‌طور هم‌زمان معیارهای K-گمنامی، L-تنوع و T-همسایگی را تضمین می‌کنند. هرچند الگوریتم KFKA، همان‌طور که انتظار می‌رفت بهترین مقدار را برای معیارهای خطای CE، CAVG و ADR حاصل کرده است؛ اما با توجه به اینکه معیارهای L-تنوع و T-همسایگی را برقرار نمی‌کند، نمی‌تواند به‌عنوان روش برگزیده انتخاب شود؛ بنابراین از بین چهار روش موجود در مقالات قبلی، فقط الگوریتم TLK3L با

این ادعا که مقدار تابع جریمه برابر با صفر را تضمین می‌کند، می‌تواند در مقام مقایسه با روش‌های پیشنهادی قرار گیرد. الگوریتم TLK3L برای دستیابی به تابع جریمه صفر مجبور به ادغام خوشه‌های زیادی شده است که درنتیجه تعداد خوشه‌ها برابر با ۱۴ شده است. این مسأله منجر به افزایش چشم‌گیر معیار CAVG به مقدار ۶/۱۹ شده است که حدود ۳ تا ۴ برابر روش‌های پیشنهادی است. مقدار خطای خوشه‌بندی الگوریتم TLK3L نیز به‌صورت تقریبی دو برابر روش‌های پیشنهادی است. همچنین، خطای نسبت انحراف میانگین ADR نیز در الگوریتم TLK3L نسبت به روش‌های پیشنهادی بسیار بالاتر است. روش پیشنهادی FFA و FFA-KFCM قادر به تشکیل خوشه‌های بسیار متوازن‌تر و کاهش خطای خوشه‌بندی، معیار CAVG و ADR به‌مراتب بهتر از روش TLK3L هستند.



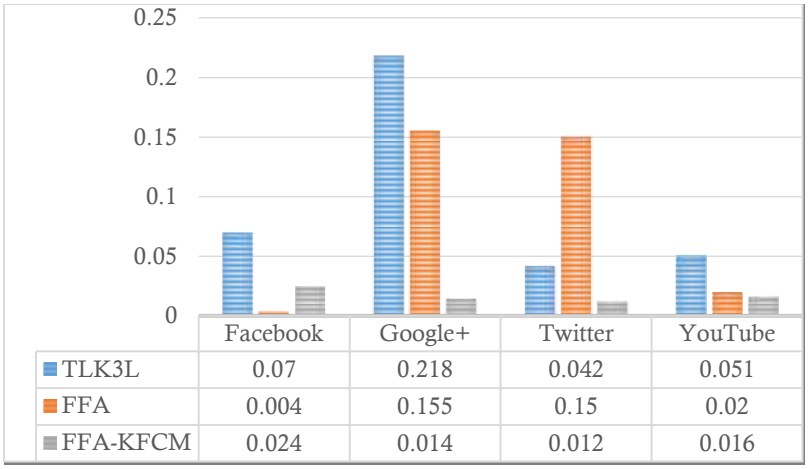
	Facebook	Google+	Twitter	YouTube
TLK3L	0.19	0.225	0.138	0.035
FFA	0.088	0.143	0.045	0.016
FFA-KFCM	0.071	0.143	0.057	0.014

شکل ۵: مقایسه خطای CE الگوریتم‌های مختلف به ازای چهار مجموعه داده



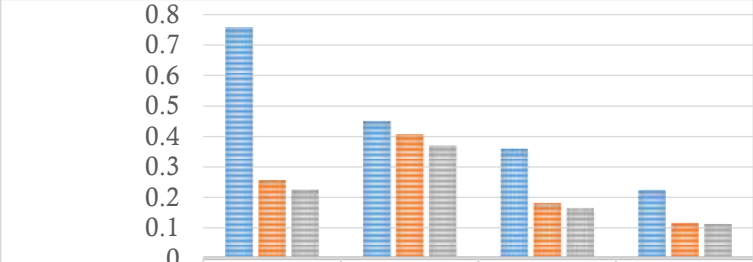
	Facebook	Google+	Twitter	YouTube
TLK3L	6.19	4	2.65	2.001
FFA	1.98	2.75	1.08	1.016
FFA-KFCM	1.84	2.83	1.27	1.01

شکل ٦: مقایسه خطای CAVG الگوریتم‌های مختلف به ازای چهار مجموعه داده



	Facebook	Google+	Twitter	YouTube
TLK3L	0.07	0.218	0.042	0.051
FFA	0.004	0.155	0.15	0.02
FFA-KFCM	0.024	0.014	0.012	0.016

شکل ۷: مقایسه خطای ADR الگوریتم‌های مختلف به ازای چهار مجموعه داده



	Facebook	Google+	Twitter	YouTube
TLK3L	0.758	0.45	0.36	0.224
FFA	0.257	0.407	0.181	0.116
FFA-KFCM	0.225	0.368	0.163	0.113

شکل ۸: مقایسه مقدار تابع هدف (ObjFun) الگوریتم‌های مختلف به ازای چهار مجموعه داده

نتیجه‌گیری و پیشنهادها

نتایج شبیه‌سازی و معیارهای ارزیابی مختلف بر روی مجموعه داده‌هایی از فیس‌بوک، گوگل پلاس، یوتیوب و توییتر در سناریوهای مختلف، به‌وضوح بیانگر عملکرد بهتر روش پیشنهادی (FFA و FFA-KFCM) نسبت به سایر روش‌های موجود است. ایجاد خوشه‌های متوازن با در نظر گرفتن قیدهای مسئله در قالب یک تابع جریمه، معیارهای K-گمنامی، L-تنوع و T-همسایگی را برای تمام خوشه‌های مدل پیشنهادی تضمین کرده و حداقل نسبت انحراف در سطح ویژگی‌ها، گراف و داده در حین فرآیند گمنام سازی برقرار است. علاوه بر روش پیشنهادی، الگوریتم TLK3L نیز قادر به برقراری شرایط K-گمنامی، L-تنوع و T-همسایگی است. بااین‌وجود، خطاهای جزئی و نیز تابع هدف آن نسبت به الگوریتم‌های پیشنهادی حدوداً ۳ برابر است. دلیل این امر این است که الگوریتم TLK3L هر سه معیار K-گمنامی، L-تنوع و T- همسایگی را بکار گرفته است اما خطاهای مختلف ایجادشده در اثر فرآیند گمنام سازی را مدنظر قرار نداده است؛ بنابراین، می‌توان این‌گونه نتیجه گرفت که روش پیشنهادی FFA و KFCM قادر به تشکیل خوشه‌های متوازن‌تر و کاهش خطای خوشه‌بندی، معیار CAVG و ADR به‌مراتب بهتر از چهار الگوریتم منتخب پژوهش هستند.

دلیل اصلی این امر، در فرمول‌بندی مسئله به‌صورت یک مسئله بهینه‌سازی در روش پیشنهادی و استفاده از ساختار ترکیبی خوشه‌بندی فازی اولیه و بهینه‌سازی با استفاده از الگوریتم کرم شب‌تاب است. بااین‌وجود، روش پیشنهادی به دلیل وابستگی به اجرای الگوریتم کرم شب‌تاب که یک الگوریتم مبتنی بر جمعیت است که باید در یک فرآیند تکراری به دنبال پاسخ بهینه باشد، به زمان اجرای به‌مراتب بالاتری نیاز دارد. همچنین با تغییر هر یک از پارامترهای K، L و T الگوریتم پیشنهادی توانایی بهتری در پایین نگه‌داشتن میزان خطاهای مختلف (خطای CE، CAVG، ADR و ObjFun) نسبت به سایر الگوریتم‌ها از خود نشان داده است.

پیشنهاد برای پژوهش‌های آتی

- استفاده از روش هوشمند برای تعیین نحوه اعمال تغییرات بر روی ماتریس داده و گراف برای تولید جمعیت اولیه الگوریتم کرم شب‌تاب به‌منظور کاهش زمان اجرای الگوریتم.
- استفاده از وزن دهی برای بهبود سرعت همگرایی و به‌روزرسانی جمعیت در کرم شب‌تاب.

منابع

سرگلزایی، احسان؛ و عبدالهی ازگمی، محمد. (آذرماه ۱۳۹۲). ارائه الگوریتمی حریصانه برای حفظ حریم خصوصی داده‌های منتشرشده شبکه‌های اجتماعی. *اولین همایش ملی کاربرد سیستم‌های هوشمند (محاسبات نرم) در علوم و صنایع، دانشگاه آزاد اسلامی واحد قوچان.*

Afshari, M. H., Dehkordi, M. N., & Akbari, M. (2016). Association rule hiding using cuckoo optimization algorithm, *Expert Systems with Applications*, 64: 340-351.

Akshaya, T., Amrit, P. (2016). Data mining with big data and Privacy preservation, *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 5(4), 1121-1124.

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges, 28(3), 45-59.

Bingchun, L., & Guohua, L. (2011). The classification of k-anonymity data, *seventh international conference on computational intelligence and security*, pp.1374-1378.

Chen, R., Fung, B. C., Philip, S. Y., & Desai, B. C. (2014). Correlated network data publication via differential privacy, *The VLDB Journal*, 23(4), 653-676.

ElBarawy, Y., Mohamed, M., & Ghali, N. I. (2014). Improving social network community detection using DBSCAN algorithm, *World Symposium Computer Application & Research (WSCAR)*, pp. 1-6.

Ferrag, M. A., Maglaras, L., & Ahmim, A. (2017). Privacy-preserving schemes for ad hoc social networks: *A survey. IEEE Communications Surveys & Tutorials*, 19(4), pp. 3015-3045.

Hartung, S., Hoffmann, C., & Nichterlein, A. (2014). Improved upper and lower bound heuristics for degree anonymization in social networks, *International Symposium on Experimental Algorithm*, pp. 376-387.

Honda, K., Kawano, A., Notsu, A., & Ichihashi, H. (2012). A fuzzy variant of k-member clustering for collaborative filtering with data anonymization, *In IEEE Conference on Fuzzy Systems*, pp. 1-6.

Li, M., Liu, Z., & Dong, K. (2016). Privacy preservation in social network against Public neighborhood attacks, *In Trustcom/BigDataSE/I SPA, 2016 IEEE*, pp. 1575-1580.

Li, N., Li, T., & Venkatasubramanian, S. (2007). T-Closeness: Privacy Beyond k-anonymity and ℓ-Diversity, in: Proceedings of ICDE, pp.106-115.

Liu, P., Cui, L., & Li, X. (2014). A hybrid algorithm for privacy preserving social network publication, *In Advanced Data Mining and Applications*, pp. 267-278.

Machanavajjhala, A., Gehrke, J. & Kifer, D. (2006). L-diversity: Privacy beyond k-anonymity, *in: Proceedings of ICDE*, pp. 24.

Mandapati, S., Bhogapathi, R. B., Rao, M. C. S., & Vjiet, V. (2013). Swarm optimization algorithm for privacy preserving in data mining, *Int. J. Comput. Sci. Issues*, 10(2).

Rahimi, M., Bateni, M., & Mohammadinejad, H. (2015). Extended k-anonymity model for privacy preserving on micro data, *IJ Comput Netw INF Secur*, 7(12), 42-51.

Sandelowski, M., and Barroso, J, (2007). Handbook for synthesizing qualitative research, New York, NY: Springer.

Schlegel, R., Chow, C. Y., Huang, Q., & Wong, D. S. (2017). Privacy-preserving location sharing services for social networks, *IEEE Transactions on Services Computing*, Vol. 10(3), 811-825.

Sihag, V. K. (2012). A clustering approach for structural k-anonymity in social networks using genetic algorithm, *Proceedings of the CUBE International Information Technology Conference*, pp. 701-706.

Susan, V. S., & Christopher, T. (2016). Privacy preserving data mining using multiple objective optimization, *ICTACT Journal on Soft Computing*, 7(4), 1366-1371.

Sweeney, L. (2002). Achieving k-anonymity privacy protection using generalization & suppression, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5), 571-588.

Tan, Y., Takagi, H., & Shi, Y. (Eds.). (2017). Data mining and Big data: Second International Conference", *DMBD 2017, Fukuoka, Japan, Proceedings*, 10387: 9-21

Tai, C., Yu, P., & Chen, M. (2010). K-support anonymity based on pseudo taxonomy for outsourcings of frequent item set mining, *in: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 16(2), 473-482.

Tiwari, A., & Choudhary, M. (2017). A review on k-Anonymization techniques, *Scholars Journal of Engineering and Technology (SJET)*, 5(6), 238-245.

Torra, V., & Navarro-Arribas, G. (2015). Data privacy: A survey of results, *in advanced research in data privacy, springer international publishing*, 567(2), pp. 27-37.

Torra, G., & Erola, A., & Roca, J. (2012). User k-anonymity for privacy preserving data mining of query logs, *Information Processing and Management*,48(3), 476–487.

Truta, T. M., & Vinay, B. (2006). Privacy protection: p-sensitive k-anonymity property. *In 2nd International Workshop on Privacy Data Management PDM 2006, p. 94, Berlin Heidelberg*, 2006. IEEE Computer Society.

Waal, T., & Willenborg, L. (1999). Information loss through global recoding and local suppression, 14: 17-20.

Yang, C. C. (2011). Preserving privacy in social network integration with τ-tolerance, *In Intelligence and Security Informatics (ISI), IEEE International Conference o*n pp. 198-200.