UOT 004.94 DOI: 10.25045/jpit.v12.i1.07

Əhmədova X.V.
AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
ehmedova1xeyale@gmail.com

KLASTERLƏŞDİRMƏ METODLARININ TƏTBİQİ İLƏ SOSİAL ŞƏBƏKƏLƏRDƏ SAXTA PROFİLLƏRİN AŞKARLANMASI

Daxil olmuşdur: 12.11.2020 Düzəliş olunmuşdur: 27.11.2020 Qəbul olunmuşdur: 04.12.2020

Sosial şəbəkələr istifadəçiləri üçün yeni dostlar tapmaq, xəbər oxumaq, faydalı məlumatlar əldə etmək, əyləncə kimi bir sıra imkanlar təklif etdiyindən milyonlarla aktiv istifadəçisi vardır. Sosial şəbəkələrin milyonlarla aktiv istifadəçiyə malik olması insanların gizli şəkildə idarə edilməsi (manipulyasiya), müxtəlif növ çağırışlar, insan və ya təşkilatların nüfuzdan salınması kimi zərərli məqsədlərin icrası üçün şərait yaradır. Bu zaman troll profillər, sibil hesablar, kuklalar, bot hesablar və s. kimi qrup şəklində fəaliyyət göstərən saxta profillər geniş şəkildə istifadə edilir. Klassifikasiya alqoritmlərinin tətbiqi ilə saxta profillərin aşkarlanması zamanı verilənlərin sinif nişanlarına (ing. label) malik olmaları, çox sayda profilin tək-tək təsnif edilməsi zamanı sərf edilən vaxt və s. kimi problemlər ortaya çıxır. Bu məqalədə sosial şəbəkələrdə saxta profillərin qruplaşdırılması üçün k-means, Gaussian Mixture, aqlomerativ klasterləşdirmə, spektral klasterləşdirmə alqoritmləri istifadə edilmişdir. Klasterizasiya alqoritmləri saxta profillərin aşkarlanmasında klassifikasiya metodlarına nisbətən pis nəticə göstərdiyindən bu məqalədə saxta profillərin aşkarlanması üçün tətbiq edilmiş klasterizasiya metodlarının hansı verilənlər üzrə daha yaxşı nəticə verməsi məsələsinə baxılmışdır. Alqoritmlərin tətbiqi zamanı profiləsaslı verilənləri ehtiva edən əlyetər bazalardan istifadə edilmişdir. Klasterizasiya metodlarının nizamlanmış rand indeksi, homogenlik, dolğunluq və s. kimi qiymətləndirmə metrikalarının tətbiqi ilə performansının qiymətləndirilməsi zamanı əldə edilmiş nəticələrə əsasən, aqlomerativ klasterləşdirmə alqoritmi digər tətbiq edilmiş klasterizasiya alqoritmlərinə nisbətən daha yaxşı nəticə göstərmişdir.

Açar sözlər: *saxta profil, klasterləşdirmə, k-means, aqlomerativ klasterləşdirmə, əlamətlər.*

Giriş

Müəyyən şəxs və ya təşkilatlar haqqında məlumat əldə edilməsi, müxtəlif növ reklamlar, fərqli dünyagörüşü və maraq dairəsinə sahib olan insanların bir araya gəlməsi, yeni dostluq əlaqələrinin qurulması və s. kimi məqsədlərlə sosial şəbəkələr gün ərzində milyonlarla istifadəçi tərəfindən istifadə edilməkdədir. Bununla yanaşı sosial şəbəkələrdən istifadə zamanı bir sıra təhdidlərə də rast gəlinir: sosial mühəndislik – hədəf *(ing. target)* şirkət haqqında məlumatların əldə edilməsi üçün o şirkətin işçiləri ilə dostluq əlaqələri qurmaq; fişinq hücumları– tanışlıqəsaslı *(ing. dating-based)* sosial şəbəkələrdə əlaqə qurulmuş hədəf istifadəçilərdən şantaj yolu ilə pul tələb etmək; saxta hesablar – şəxsin saxta verilənlər istifadə edərək yaratdığı profillərdən istifadə etməklə insanları manipulyasiya etməsi, nüfuzdansalma, qanunsuz satışlar kimi zərərli fəaliyyətlərlə məşğul olması və s. [1, 2].

Sosial mediada bir sıra risk ssenariləri qurulmaqdadır. [3]-də sosial mediada olan risk ssenariləri təsnif edilmişdir: terrorçu, ekstremist, radikal qruplar tərəfindən sosial şəbəkələrin istifadə edilməsi, dinə aid həssas tvitlər, siyasətçilər və ictimai xadimlər tərəfindən sosial şəbəkə istifadəsi, sosial medianın kəşfiyyat qurumları üçün yaratdığı imkanlar və s. Sosial şəbəkələrdən istifadə zamanı qarşılaşılan bir sıra təhdidlərdən yan keçilməsi üçün tədbirlər görülməkdədir: sosial şəbəkə operatorlarının təhlükəsizlik tədbirləri; sosial şəbəkələrdə paylaşımların görünməsinin *(ing. visibility)* tənzimlənməsi; elmi tədqiqatçılar tərəfindən təklif olunmuş həllər və s. [4].

Sosial şəbəkələrdə sibil hesablar, klon profillər, troll profillər, ələ keçirilmiş profillər və s. kimi saxta profil növləri vardır. Bu saxta profilləri yaradılma və idarə edilmə üsullarına əsasən 3 qrupa cəmləmək olar: insan, bot və həm insan, həm də bot tərəfindən (*ing. cyborg*) yaradılıb, idarə edilən

saxta profillər [5]. Bu halda tədqiqatlar, əsasən, saxta profilin müəyyən növünün (məsələn, troll profillər) aşkarlanması istiqamətində olur. Bununla yanaşı, saxta profillərin ortaq əlamətləri olduğundan (məsələn, profil şəkli istifadə etməmək, yaşayış yeri, təhsil və s. haqqında məlumat qeyd etməmək kimi) saxta profillərin ümumi şəkildə aşkarlanması istiqamətində də tədqiqatlar aparılır.

Sosial şəbəkələrdə saxta profillərin aşkarlanması zamanı klassifikasiya və klasterizasiya alqoritmləri geniş şəkildə istifadə edilir. Sosial şəbəkələr geniş istifadəçi kütləsinə malikdir. Buna görə də milyonlarla istifadəçidən real və saxta olanların fərqləndirilməsi məsələsi çətinləşir. Klassifikasiya alqoritmləri saxta profil aşkarlanmasında daha yaxşı nəticə verdiyindən bu istiqamətdə bir sıra tədqiqatlar aparılmışdır. Lakin günbəgün artan istifadəçilərin tək-tək klassifikasiya edilməsinə sərf edilən zaman, böyük həcmli sinif nişanına malik verilənlərin toplanması kimi problemlər ortaya çıxır. Bu məsələlər nəzərə alınaraq, sosial şəbəkələrdə saxta profillərin klasterizasiyası istiqamətində tədqiqatlar aparılır.

Bu məqalədə sosial şəbəkələrdə fəaliyyət göstərən saxta profillərin klasterizasiyası istiqamətində tədqiqat aparılmışdır. Məqalədə sosial şəbəkələrdə saxta profil qrupları, saxta profillərin klasterləşdirilməsi, klasterləşdirmə zamanı istifadə edilən əlamətlər və alqoritmlər kimi məsələlərə baxılmışdır.

Sosial şəbəkələrdə qrup şəklində fəaliyyət göstərən saxta profillər

Sosial şəbəkələrdə bir sıra saxta profil növləri mövcuddur. Amma bu saxta profillərin bir qismi qruplar şəklində zərərli əməliyyatlar icra edir. Bu tip saxta profil növlərinə sibil *(ing. sybil)* hesablar, troll hesablar, botlar, kukla hesablar aid edilir. Bu saxta profillər, əsasən, insanları manipulyasiya etmək məqsədilə istifadə edilir. Onlar yaradılma məqsədi və formalarına, icra etdikləri iş və icra etmək tərzlərinə görə bir-birindən fərqlənirlər.

Funksionallıq baxımından botlar 5 kateqoriyaya bölünür: spam botları, sosial botlar, bəyənmə botları, təsir botları və botnet [6]:

- Spam botlar yalnız zərərli fəaliyyətlər üçün hazırlanmış kompüter proqramıdır. Bunlar çox sayda istənilməyən əlaqələr yaratmaqla şəbəkəni çirkləndirmək, şəxsi bloq, ödənişli məzmun və reklamlara yönləndirmək və s. kimi məqsədlərlə hazırlanmışdır [7].
- Sosial bot avtomatik olaraq məzmun generasiya edən və sosial mediada insanlarla qarşılıqlı əlaqədə olan, davranışlarını təqlid etməyə və dəyişdirməyə çalışan bir kompüter alqoritmidir. Sosial botlar son bir neçə il ərzində sosial media platformalarında yer almışdır [8].
- Botnet onlayn sosial şəbəkələrdə avtomatlaşdırılmış kompüter proqramları şəbəkəsinə deyilir. Bu şəbəkədəki hər bir proqrama (bot) avtomatlaşdırılmış şəkildə yerinə yetiriləcək oxşar və ya fərqli tapşırıqlar verilir. Botnet "nəzarət kanalı" tərəfindən idarə olunan və qanunsuz fəaliyyət göstərmək üçün əmrlər verən kompüter proqramı toplusudur [9].

Sibil hesablar (ing. Sybil account). Zərərli istifadəçi məlumatların bütövlüyünü pozmaq, trollinq, populyarlığa təsir, dələduzluq və s. kimi əməliyyatları icra etmək məqsədilə çox saxta şəxsiyyət yarada bilər. Bu hücumlar sibil hücum növləridir [10, 11].

Kukla hesablar (ing. Sockpuppet). Wikipedia kontekstində birdən çox hesabın düzgün istifadə edilməməsinə kukla hesablar deyilir. Kukla birgə layihə saytlarında aldatma məqsədi ilə istifadə olunan saxta hesabdır [12].

Trol hesablar (ing. Troll profile). Troll hesablar fermalar şəklində fəaliyyət göstərirlər. Bu fermalardan ən məşhuru İnternet Araşdırma Agentliyidir (*ing. Internet Research Agency, IRA*). IRA 2016-cı ildə ABŞ-da prezident seçkilərinə təsir etmək məqsədilə istifadə edilmişdir [13].

Əlaqədar işlərin icmalı

Sosial şəbəkələrdə saxta profillərin aşkarlanması istiqamətində müxtəlif növ yanaşmalar təklif edilmişdir. Bu yanaşmalar klassifikasiya və klasterizasiya alqoritmlərini ehtiva edir. Bu bölmədə sosial şəbəkələrdə fəaliyyət göstərən saxta profillərin qruplaşdırılması istiqamətində aparılmış tədqiqatlar qeyd edilmişdir.

Facebook sosial şəbəkəsində saxta profillərin aşkarlanması üçün supervizorlu (ID3 qərar ağacı, k-NN və SVM) və supervizorsuz (k-means and k-medoids) maşın təlimi alqoritmləri RapidMiner Studio istifadə edilməklə tətbiq edilmişdir. 982 (781 real və 201 saxta) profil üzrə 12 davranış və qeyri-davranış əsaslı əlamətdən istifadə edilmişdir [14].

Eyni aktorla qeydiyyatdan keçmiş saxta profil qruplarının aşkarlanması üçün yanaşma təklif edilmişdir. Təklif edilmiş yanaşmanın əsası bütün klasterin zərərli və ya qanuni olaraq klassifikasiya edilməsi üçün supervizorlu maşın təlimi yanaşmalarının tətbiqidir. İstifadə edilmiş əsas əlamətlər istifadəçilər tərəfindən generasiya edilmiş mətnlərin (məsələn, ad, e-mail ünvanı, universitet və s.) statistikasıdır. Bu çərçivə *(ing. framework)* LinkedIn-də IP ünvan və qeydiyyatdan keçmə vaxtı əsasında qruplaşdırılmış profil verilənlərinin analizinə tətbiq edilmişdir [15].

Instagram sosial şəbəkəsində təqlidçilərin *(ing. impersonator)* aşkarlanması istiqamətində tədqiqat aparılmışdır. Bu zaman k-means, GMM (Gaussian Mixture Model) və spektral klasterləşdirmə alqoritmlərinin tətbiqi ilə profillər 3 klasterdə cəmlənmişdir. Müəlliflər bu klasterləri C0 – Fan-Pages, C1 – Ordinary-Users və C2 – BotLike kimi adlandırırlar [16].

Twitter sosial şəbəkəsində cins və məkan (ing. *location*) əlamətləri daxil olmaqla profil əlamətlərinə Bayes klassifikatoru və k-means klasterləşdirmə metodları aldatmanın aşkarlanmasına tətbiq edilmişdir [17]. Şilinq hücumlarının aşkarlanması üçün hücum profillərinin ikili qərar ağacının yarpaq təpəsində toplanmasını təmin edən Bisecting k-means alqoritmindən istifadə edən üsul təklif edilmişdir [18].

Tvitlərin axın təbiətini effektiv şəkildə idarə etmək üçün 2 axın klasterləşdirmə alqoritmi: StreamKM++ və DenStream spam aşkarlanmasını asanlaşdırmaq üçün dəyişdirilmişdir. Hər iki alqoritm normal Twitter istifadəçilərini bir araya gətirərək, digər profilləri spammer olaraq dəyərləndirmişdir [19].

[20]-də Twitter məlumatından turist hərəkəti modellərini müəyyənləşdirmək üçün qrafəsaslı üsul təqdim edilmişdir. İstifadəçilərin fəaliyyət mərkəzlərini müəyyənləşdirmək üçün tanınmış məkan qruplaşma alqoritmi olan DBSCAN alqoritmi istifadə edilmişdir [21].

[22]-də müəlliflər sosial mediada kuklaların avtomatik aşkarlanması və qruplaşdırılması üçün SocksCatch adlanan, verilənlərin toplanması və seçilməsi, ML alqoritmlərinin (SVM, RF, NB, k-NN, Bayesian network, Adaptive boosting) tətbiqilə kukla hesablarının aşkarlanması, qraflar nəzəriyyəsindən istifadə edilməklə kukla hesabların qruplaşdırılması kimi 3 mərhələdən ibarət model təklif edirlər.

[23]-də botnetlər tərəfindən idarə edilən hesabların aşkarlanması üçün onlayn hesab və IP ünvanı arasındakı xəritədən istifadə edərək fəaliyyət göstərən EvilCohort adlı sistem tətbiq edilmişdir. Sistem verilənlərin toplanması, proyeksiyalanmış qraf təsvirinin qurulması və icmaların tapılması kimi 3 addımdan ibarətdir.

[24]-də veb-dən çıxarılmış sosial şəbəkələrdə icma aşkarlanması üçün yeni bir metod təklif edilmişdir. Təklif edilmiş yeni metodda məqsədlər bütün seçilmiş klasterlərin ümumi çəkisininin maksimallaşdırılması və seçilmiş klasterlər arasındakı oxşarlığın minimallaşdırılması, hər bir verilənlər obyektinin yalnız bir klasterə təyin olunması, hər bir klasterin ən azı bir obyekti ehtiva etməsi, sadəcə *k* klasterin seçilməsinin təmini, ona təyin edilmiş hər hansı bir verilənlər obyekti varsa, klasterin seçilməsinə zəmanət verilməsi şəklindədir.

[25]-də çəkili sosial şəbəkələrdə icma strukturlarının təyini üçün bölgü klasterləşdirməsi əsaslı metod işlənmişdir. Bu tədqiqat işində *n* verilənlər obyektinin sabit *k* sayda klaster sayına vahid bölgüsünün tapılması problemi həll edilmişdir. Yəni, hər bir verilənlər obyekti həmin qruplardan dəqiqliklə birinə təyin edilir və hər bir klasterdə ən azı bir obyekt vardır.

[26]-da zərərli hesabların aşkarlanması üçün SynchroTrap adlı sistem təklif edilmişdir. Bu zaman istifadəçilərin qruplaşdırılması üçün 2 istifadəçi bənzərliyinə əsaslanan *single-linkage hierarchical clustering* alqoritmi istifadə edilmişdir.

Sosial şəbəkələrdə saxta profillərin aşkarlanması üçün istifadə edilən əlamətlər

Sosial şəbəkə və saxta profillərin növlərindən asılı olaraq saxta profillərin fəaliyyəti fərqlənir. Buna görə də saxta profillərin aşkarlanması üzrə aparılmış tədqiqatlarda müxtəlif növ əlamətlərdən istifadə edilmişdir. Tədqiqat işinin bu bölməsində saxta profillərin aşkarlanması istiqamətində aparılmış tədqiqatlar üzrə istifadə edilmiş əlamətlər qeyd edilmişdir

Sosial media platformlarında (SMP) təklif edilmiş əlamətlər 3 qrupa bölünür: hesabın kimliyini *(ing. identity),* hesabın başqaları ilə əlaqələrini və son olaraq hesabın davranışını və ya mesajlarını təsvir edən məlumatlar. SMP-lərdə saxta bot hesablarının aşkarlanması üçün qeyd edilmiş əlamətlərin müxtəlif kombinasiyalarından istifadə edilmişdir [5].

Sətirlərin qeyri-səlis məntiq əsasında müqayisəsi üsulu istifadə edilməklə 2 dostun oxşarlığının ölçülməsi üçün 4 əlamətdən istifadə edilmişdir: iş, təhsil, doğma şəhər və indiki yaşayış şəhəri [27].

Twitter-də spammerlərin aşkarlanması üçün qrafəsaslı (dostların sayı, izləyicilər sayı və istifadəçinin nüfuzu hesablanmaqla), məzmunəsaslı (təkrarlanan tvitlər, HTTP linklər, cavablar və bəhslər *(ing. mentions)*, trend mövzular və s.) əlamətlər istifadə edilmişdir [28].

[29, 30]-da istifadəçilərin qoşulduğu qruplar, qəbul edilən dostluq təkliflərinin sayı (dərəcə ilə), göndərilən dostluq istəklərinin sayı (dərəcə şəklində) və s. kimi bir neçə qrafəsaslı əlamətdən istifadə edilmişdir.

Twitter spamerlərinin aşkarlanması üçün 6 əsas kateqoriyaya: profiləsaslı, məzmunəsaslı, qrafəsaslı, qonşuəsaslı, vaxtəsaslı və avtomatlaşdırmaəsaslı kateqoriyaya bölünən 25 əlamətdən istifadə edilmişdir [31, 32].

[15]-də onlayn sosial şəbəkələrdə saxta profil klasterlərinin aşkarlanması üçün yanaşma təklif edilmişdir. Bu yanaşmada əsas paylanma əlamətləri, obraz əlamətləri və tezlik əlamətləri kimi əlamət siniflərindən istifadə edilmişdir.

[16]-də Instagram sosial şəbəkəsində təqlidçilərin aşkarlanması üçün istifadəçi adı oxşarlığı, ad oxşarlığı, bioloji oxşarlıq, şəkil oxşarlığı və s. kimi 10 əlamətdən istifadə edilmişdir. Nəticədə 3 klaster yaranmışdır: C0–Fan-Pages, C1–Ordinary– Users, C2 – Botlike.

[26]-da böyük zərərli profil qruplarının aşkarlanması üçün SynchroTrap adlı sistem işlənmişdir. Bu zaman əsas əlamətlər kimi istifadəçi ID-si, istifadəçi əməliyyatının zaman damğası, tətbiq identifikatoru (məsələn, paylaşma, mesajlaşma), istifadəçi əməliyyatı ilə əlaqəli obyekt ID-si, istifadəçinin IP ünvanı istifadə edilmişdir.

Bu məqalədə eksperimentlərin aparılması üçün profiləsaslı verilənləri ehtiva edən açıq bazalardan istifadə edilmişdir. Bu bazalar istifadəçi adı, izləyici sayı, izlənənlər sayı və s. kimi profiləsaslı verilənləri ehtiva edir. Klasterizasiya metodlarının tətbiqi zamanı lazımi əlamətlərin (komponentlərin) seçilməsi zamanı əsas komponent analizi (*ing. principal component analysis* (*PCA*)) üsulundan istifadə edimişdir.

Aparılmış eksperimentlər

Maşın təlimi alqoritmləri verilənlərə tətbiq edilməzdən əvvəl verilənlər bir sıra proseslərdən keçirlər: verilənlərin toplanması, ötürülmüş verilənlərin təmizlənməsi, verilənlərin normallaşdırılması və ya standartlaşdırılması, əsas komponentlərin çıxarılması və s. Bütün bu əməliyyatların icrasından əvvəl verilənlərin tipinin ədədi olmasına diqqət yetirmək lazımdır. Əgər ədədi deyilsə, verilənlərin müəyyən qaydalarla ədədiləşdirilməsi yerinə yetirilir.

Verilənlərin təsviri. Aparılmış eksperimentlərdə əlyetər olan 4 bazadan istifadə edilmişdir. Bu bazalar daha sonra D1, D2, D3, D4 olaraq adlandırılmışdır. Bazaların təsviri bu şəkildədir:

- *D1 bazası.* Github-da Harshit Kumar Gupta tərəfindən paylaşılmış, ölçüləri uyğun olaraq 34×1337, 34×1481 olan "users" və "fusers" adlı fayllardan D1 bazası kimi istifadə edilmişdir. Qeyd edilmiş verilənlər Twitter sosial şəbəkəsi üzrə saxta və real profilləri ehtiva etməkdədir [33].
- *D2 bazası.* Kaggle-də Bardiya Bakhshandeh tərəfindən paylaşılmış 12×696 ölçülü "train" və 12×120 ölçülü "test" adlı fayllardan D2 bazası kimi istifadə edilmişdir. Qeyd edilmiş verilənlər Instagram sosial şəbəkəsində spam və spam olmayan profilləri ehtiva etməkdədir [34].
- *D3 bazası.* Github-da Fatih Çağatay Akyon tərəfindən paylaşılmış, ölçüləri uyğun olaraq 9×200, 9×994 olan FakeAccountData və RealAccountData adlı 2 dataset D3 bazası olaraq istifadə edilmişdir. D3 bazası Instagram sosial şəbəkəsində olan saxta və real profil verilənlərini ehtiva edir [35].
- *D4 bazası.* Github-da Radheshyam Saharan tərəfindən Twitter-də olan spamerlər və spamer olmayan profilləri ehtiva edən, ölçüsü 17×552 olan "Total_data" adlı data D4 bazası imi istifadə edilmişdir [36].

Verilənlərin işlənməsi. Qeyd edilmiş bazalara alqoritmlərin tətbiq edilməsi üçün ədədi olmayan verilənlər üzərində işlənilmişdir. Məsələn, D1 bazası 34 əlaməti ehtiva edirdi, amma bu əlamətlərdən bəziləri çox sayda ötürülmüş verilən ehtiva etdiyindən uyğun əlamətlər silinmişdir. Eyni zamanda profilin gizlilyi, profil şəkli, url və s. kimi əlamətlərə aid verilənlərin olub-olmamasına əsasən, 0 və ya 1 (verilən vardırsa 1, əks halda 0) qiyməti mənimsədilmişdir. Məkan, dil, saat qurşağı kimi əlamətləri ehtiva etdiyi unikal verilənlərin nömrəsi mənimsədilmişdir. Bazalara bu kimi əməliyyatların tətbiqi nəticəsində D1 bazasında hədəf "fake" sütunu da daxil olmaqla 19 əlamət, D2 bazasında hədəf "fake" sütunu da daxil olmaqla 13 əlamət, D3 bazasında hədəf "isfake" sütunu da daxil olmaqla 10 əlamət, D4 bazasında hədəf "SpammerOrNot" sütunu da daxil olmaqla 15 əlamət olmuşdur. Daha sonra ötürülmüş verilənlərin təmizlənməsi üçün Python-da olan KNNImputer paketindən istifadə edilmişdir [37].

Əlamətlərin seçilməsi. Bazalar üzrə verilənlərin hədəf verilənlər ilə korrelyasiyasına əsasən hesablanması üçün Orange-də olan "Rank" adlı moduldan istifadə edilmişdir [38]. Bu zaman Information Gain Ratio və Gini Decrease hesablama metodlarına əsasən, yuxarı qiymət almış ilk 10 əlamət seçilmişdir. Bazalar üzrə seçilmiş əlamətlər cədvəl 1-də qeyd edilmişdir. Eksperimentlərin icra edilməsi zamanı PCA istifadə edilməklə dispersiya əsasında komponentlər seçilmişdir. Dispersiya faizinin kəskin düşməsi 2-ci komponentdə müşahidə edildiyindən, alqoritmlər 1 komponent əsasında icra edilmişdir [39].

Cədvəl 1

Ranqlar əsasında seçilmiş əlamətlər

D1	D2	D3	D4
Followers to Friends ratio	Posts	Follower to Following ratio	Friends to Followers ratio
Statuses count	Followers	Usermediacount	Avgfavcount
Lang	Description length	Userfollowercount	Userfollowerscount
Favourites count	Profile pic	Userhasprofilpic	Userlocation
Followers count	Nums/length username	Userfollowingcount	Userdescriptionlength
İd	Follows	Usernamedigitcount	Userfriendscount
Default profile	Followers to follows ratio	Userısprivate	Avgretweet

Time zone	Fullname words	Userbiographylength	Userıd
Profile banner url	External url	Usernamelength	Usercreatedat
Listed count	Nums/length fullname	İsfake	Avghashtag
Fake	Fake		Spammerornot

Klasterləşdirmə metodlarının tətbiq edilməsi. Eksperimentlərin icrası zamanı Python proqramlaşdırma dilində olan Scikit-learn adlı bağlamanın təklif etdiyi klasterləşdirmə metodlarından aşağıda qeyd edilənlər tətbiq edilmişdir. İstifadə edilmiş verilənlər 2 sinif nişanına malik olduğundan alqoritmlərin icrası zamanı klaster sayı 2 təyin edilmişdir [40]:

- *K-Means alqoritmində parametrlərin seçilməsi*. Bu alqoritmin istifadə etdiyi metrika nöqtələrarası məsafədir. Alqoritmin tətbiqi zamanı seçilmiş parametrlər n_clusters = 2, init='k-means++' olmuş, digər parametrlər defolt saxlanılmışdır.
- *GMM alqoritmində parametrlərin seçilməsi.* Bu alqoritmin istifadə etdiyi metrika mərkəzlərə olan Mahalanobis məsafələridir. Alqoritmin tətbiqi zamanı n_components=2 olaraq seçilmiş, digər parametrlər defolt olaraq qalmışdır.
- *Aqlomerativ klasterləşdirmə (AC) alqoritmində parametrlərin seçilməsi.* Bu alqoritmin istifadə etdiyi metrika istənilən cüt məsafəsidir. Alqoritmin tətbiqi zamanı n_clusters=2 seçilmiş, digər parametrlər defolt saxlanılmışdır.
- *Spektral klasterləşdirmə (SCl) alqoritmində parametrlərin seçilməsi.* Bu alqoritmin istifadə etdiyi metrika qraf məsafəsidir (məsələn, ən yaxın qonşu qrafı). Alqoritmin tətbiqi zamanı parametrlər olaraq n_clusters=2, assign_labels="discretize", random_state=0 seçilmiş, digər parametrlər defolt olaraq saxlanılmışdır.

Klasterləşdirmə metodlarının qiymətləndirilməsi metrikaları. Python proqramlaşdırma dilində olan scikit-learn adlı alət ML alqoritmlərinin qiymətləndirilməsi üçün bir sıra metrikalar təklif edir. Klasterləşdirmə alqoritmlərinin qiymətləndirilməsi üçün bir sıra metrikalar təklif edilir [41]:

- *Nizamlanmış rand indeksi (ing. Adjusted Rand Index, ARI).* Əgər *C* tam həqiqi sinif təyinatı və *K* klaster çoxluğudursa, bu halda a və b aşağıdakı kimi təyin edilir:
  - a *C* və *K*-da eyni çoxluqda olan elementlər cütlərinin sayıdır.
  - b C və *K*-da fərqli çoxluqlarda olan elementlər cütlərinin sayıdır.

$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]}$$

$$RI = \frac{a + b}{C_2^{n_{samples}}}$$

Burada ARI nizamlanmış rand indeksi, RI nizamlanmamış rand indeksidir.

- *Qarşılıqlı informasiyaəsaslı indeks (ing. Mutual Information, MI).* U və V kimi iki nişan tapşırığını (eyni *N* obyektin) fərz edək. Onların entropiyası, bölgü çoxluğu üçün aşağıdakı kimi təyin olunan qeyri-müəyyənlik qiymətidir:

$$H(U) = -\sum_{i=1}^{|U|} P(i)\log(P(i)).$$

Burada $P(i) = |U_i|/N$ U -dan təsadüfi olaraq seçilmiş bir obyektin U_i sinfinə düşməsi ehtimalıdır. Bu, V üçün də keçərlidir:

$$H(V) = -\sum_{j=1}^{|V|} P'(j)\log(P'(j)).$$

$$P'(j) = |V_j|/N.$$

U və V arasındakı *MI* qarşılıqlı informasiyası aşağıdakı kimi hesablanır:

$$MI(U,V) = \sum_{i=1}^{|U|}\sum_{j=1}^{|V|} P(i,j)\log\left(\frac{P(i,j)}{P(i)P'(j)}\right).$$

Burada, $P(i,j) = \left|U_i \cap V_j\right|/N$ təsadüfi olaraq seçilmiş bir obyektin həm U, həm də V sinfinə düşməsi ehtimalıdır.

- *Homogenlik, dolğunluq (ing. completeness) və V-ölçü.* Homogenlik və dolğunluq balları riyazi olaraq aşağıdakı kimi verilir:

$$h = 1 - \frac{H(C|K)}{H(C)}$$

$$c = 1 - \frac{H(K|C)}{H(K)}$$

Burada $H(C|K)$ klaster tapşırıqları verilən siniflərin şərti entropiyasıdır və aşağıdakı kimi ifadə olunur:

$$H(C|K) = -\sum_{c=1}^{|C|}\sum_{k=1}^{|K|}\frac{n_{c,k}}{n}\log\left(\frac{n_{c,k}}{n_k}\right)$$

$H(C)$ siniflərin entropiyasıdır, aşağıdakı şəkildə ifadə olunur:

$$H(C) = -\sum_{c=1}^{|C|}\frac{n_c}{n}\log\left(\frac{n_c}{n}\right)$$

n nümunələrin ümumi sayı, n_c və n_k c və k klasterlərinə aid nümunələrin sayı və $n_{c,k}$ k klasterinə təyin edilmiş c sinfindən olan nümunələrin sayıdır. Verilmiş $H(K|C)$ sinfi klasterlərinin şərti entropiyası və $H(K)$ klasterlərinin entropiyası simmetrik şəkildə müəyyən edilmişdir.

Rosenberg və Hirschberg daha sonra v-ölçünü homogenlik və bütövlüyün harmonik mənası olaraq təyin edirlər:

$$v = 2 \cdot \frac{h \cdot c}{h + c}$$

- *Fowlkes-Mallows indeksi (ing. Fowlkes-Mallows index, FMI).* Fowlkes-Mallows indeksindən nümunələrin həqiqi sinif nişanı təyinatları məlum olduqda istifadə edilə bilər. Fowlkes-Mallows indeksi (FMI) pairwise precision və recall-ın həndəsi ortası olaraq təyin olunur:

$$FMI = \frac{TP}{\sqrt{(TP + FP)(TP + FN)}}$$

- *Siluet əmsalı (ing. Silhouette Coefficient, SC).* Əgər verilənlərin həqiqi sinif nişanları məlum deyilsə, qiymətləndirmə modelin özündən istifadə edilməklə icra edilməlidir. Siluet əmsalı bu tip qiymətləndirməyə misaldır, harada ki, yüksək *SC* indeksi modelin daha yaxşı təyin edilmiş klasterləri olduğunu göstərir. *SC* hər nümunə üçün təyin edilir və 2 qiymətdən ibarətdir:
  - a: eyni klasterdə olan bir nümunə və bütün digər nöqtələr arasındakı orta məsafə;
  - b: növbəti ən yaxın klasterdə olan bir nümunə və bütün digər nöqtələr arasındakı orta məsafə.

  Tək bir nümunə üçün s siluet əmsalı aşağıdakı kimi təyin edilir:

$$s = \frac{b - a}{\max(a, b)}$$

Nümunələr çoxluğu üçün *SC* hər nümunə üçün olan *SC*-nin ortası kimi verilir.

Alqoritmlərin performansının qiymətləndirilməsi nəticələri. Alqoritmlərin performansı qiymətləndirilərkən ARI, MI, Homogeneity, Completeness, V-measure, FM, SC kimi klasterləşdirmə alqoritmlərinin qiymətləndirmə metrikalarından istifadə edilmişdir. PCA əsasında seçilmiş əlamətlər üzrə alqoritmlərin performansının qiymətləndirilməsi nəticələri cədvəl 2-də qeyd edilmişdir.

Cədvəl 2

PCA əsasında seçilmiş əlamətlərə alqoritmlərin tətbiqinin nəticələri

	Alqoritmlər	ARI	MI	Homogenity	Completeness	V-measure	FM	SC
D1	**k-Means**	0.6821	0.6429	0.6404	0.6456	0.6429	0.8423	0.7260
	GMM	0.7705	0.7038	0.6982	0.7096	0.7039	0.8868	0.5611
	AC	0.9288	0.8735	0.8741	0.8730	0.8735	0.9644	0.6993
	SCl	0.8020	0.7431	0.7433	0.7430	0.7431	0.9012	0.7277
D2	**k-Means**	0.3449	0.3229	0.3131	0.3346	0.3235	0.6863	0.5274
	GMM	0.0300	0.0724	0.0575	0.1017	0.0735	0.6319	0.5842
	AC	0.5869	0.5134	0.5085	0.5193	0.5138	0.7962	0.5165
	SCl	-0.0001	0.0096	0.0061	0.1127	0.0116	0.7022	0.7933
D3	**k-Means**	0.6372	0.4543	0.4752	0.4360	0.4548	0.8944	0.6665
	GMM	0.4927	0.2973	0.2804	0.3180	0.2980	0.8716	0.7778
	AC	0.6446	0.4599	0.4315	0.4935	0.4604	0.9106	0.6834
	SCl	-0.0026	-0.0005	0.0006	0.0242	0.0013	0.8499	0.8336
D4	**k-Means**	0.0047	0.0014	0.0030	0.0027	0.0028	0.5339	0.6288
	GMM	0.0178	0.0022	0.0037	0.0037	0.0037	0.5713	0.4617
	AC	0.0056	0.0017	0.0032	0.0029	0.0030	0.5350	0.6303
	SCl	-0.0056	0.0029	0.0033	0.0624	0.0063	0.7449	0.8470

Nəticələrdən aydın olduğu kimi, D1 bazası üzrə alqoritmlər daha yaxşı, D4 bazası üzrə isə daha pis nəticə göstərmişdir. Tətbiq edilmiş alqoritmlərdən aqlomerativ klasterləşdirmə daha yaxşı, k-means və spektral klasterləşdirmə alqoritmləri isə nisbətən pis nəticə göstərmişdir. Cədvəl 2-də yaşıl xanalar bazalar əsasında alqoritmlərin metrikalar üzrə aldığı ən yaxşı, qırmızı xanalar isə ən pis nəticələri ifadə edir. Bütün kodlar və verilənlər [37]-də əlyetərdir.

Nəticə

Sosial şəbəkələrin istifadəçi kütləsi genişlədikcə, ondan istifadə edən bədniyyətlilər və bədniyyətlərin məzmunu da böyüməkdədir. Məsələn, əvvəl saxta profil kimisə gizli şəkildə izləmək üçün istifadə edilirdisə, indi insanları manipulyasiya etmək üçün istifadə edilir. Sosial şəbəkələrdə əsas təhlükə yaradan profillər qruplar şəklində fəaliyyət göstərən saxta profillərdir. Buna görə də saxta profil qruplarının aşkarlanması məsələsi günümüzün aktual problemlərindəndir.

Bu məqalədə sosial şəbəkələrdə fəaliyyət göstərən saxta profillərin qruplaşdırılması istiqamətində görülmüş işlərə baxılmış, qeyd edilmiş saxta və real profilləri ehtiva edən 4 bazaya k-means, Gaussian Mixture, aqlomerativ klasterləşdirmə, spektral klasterləşdirmə alqoritmləri tətbiq edilmişdir. Twitter sosial şəbəkəsində olan saxta və real profilləri ehtiva edən 20 əlamətdən ibarət bazaya (D1) PCA tətbiqilə dispersiyaya əsaslanaraq seçilmiş 1 əlamət üzrə tətbiq edilmiş bütün klasterləşdirmə alqoritmlərinin yaxşı nəticə əldə etməsinə baxmayaraq digər bazalar üzrə nəticələr aşağı olmuşdur. Ümumi şəkildə qeyd etmək olar ki, aqlomerativ klasterləşdirmə alqoritmi saxta profillərin klasterizasiyası məsələsində digər alqoritmlərdən üstün nəticə əldə etmişdir.

Gələcək tədqiqat işində saxta profillərin qruplaşdırılması üzrə aqlomerativ klasterləşdirmə alqoritminin tətbiqilə seçilmiş əlamətlərə görə daha yaxşı nəticələrin əldə edilməsi üçün çalışmaq nəzərdə tutulur.

Ədəbiyyat

1. Security threats we face while using social media. https://novalisit.com/security-threats-we-face-while-using-social-media/
2. İmamverdiyev Y.N., Əhmədova X. Sosial şəbəkələrdə saxta profillərin aşkarlanması metodları haqqında / İnformasiya təhlükəsizliyinin aktual multidissiplinar elmi-praktiki problemləri V respublika konfransının əsərləri, 2019, s.45–48.
3. İmamverdiyev Y.N. Sosial media və təhlükəsizlik problemləri / İnformasiya təhlükəsizliyinin multidissiplinar problemləri üzrə II respublika elmi-praktiki konfransı, 2015, s.189–192.
4. Şıxəliyev R.H. Sosial şəbəkələrdə təhlükəsizlik problemləri // İnformasiya cəmiyyəti problemləri, 2016, №2, s.80–88.
5. Van Der Walt E., Eloff J. Using machine learning to detect fake identities: Bots vs humans // IEEE Access, 2018, vol.6, pp.6540–6549.
6. Wani M.A., Jabin S. A sneak into the devil's colony - fake profiles in online social networks // arXiv preprint arXiv:1705.09929, 2017.
7. Boshmaf Y., Muslukhov I., Beznosov K., & Ripeanu M. Design and analysis of a social botnet // Computer Networks, 2013, vol.57, no.2, pp.556–578.
8. Ferrara E., Varol O., Davis C., Menczer F., & Flammini A. The rise of social bots // Communications of the ACM, 2016, vol. 59, no.7, pp.96–104.
9. Silva S.S., Silva R.M., Pinto R.C., & Salles R.M. Botnets: A survey // Computer Networks, 2013, vol. 57, no. 2, pp.378–403.
10. Al-Qurishi M., Al-Rakhami M., Alamri A., Alrubaian M., Rahman S.M.M., & Hossain M.S. Sybil defense techniques in online social networks: a survey // IEEE Access, 2017, vol.5, pp.1200–1219.
11. Douceur J.R. The sybil attack // International Workshop on Peer-to-Peer Systems, 2002, pp.251–260.
12. Yamak Z., Saunier J., Vercouter L. Detection of multiple identity manipulation in collaborative projects / Proceedings of the 25th International Conference Companion on World Wide Web, 2016, pp.955–960.
13. Im J., Chandrasekharan E., Sargent J., Lighthammer P., et al. Still out there: Modeling and identifying russian troll accounts on twitter / 12th ACM Conference on Web Science, 2020, pp.1–10.
14. Albayati M.B., Altamimi A.M. Identifying Fake Facebook Profiles Using Data Mining Techniques // Journal of ICT Research and Applications, 2019, vol.13, no.2, pp.107–117.
15. Xiao C., Freeman D.M., Hwa T. Detecting clusters of fake accounts in online social networks // Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security, 2015, pp.91–101.

16. Zarei K., Farahbakhsh R., Crespi N. How impersonators exploit Instagram to generate fake engagement? // arXiv preprint arXiv:2002.07173, 2020.
17. Alowibdi J. S., Buy U. A., Philip S. Y., Ghani S., & Mokbel M.Deception detection in Twitter // Social network analysis and mining, 2015, vol.5, no.1, Article number 32, 13 p.
18. Bilge A., Ozdemir Z., Polat H. A novel shilling attack detection method // Procedia Computer Science, 2014, vol.31, pp.165–174.
19. Miller Z., Dickinson B., Deitrick W., Hu W., & Wang A.H. Twitter spammer detection using data stream clustering // Information Sciences, 2014, vol.260, pp.64–73.
20. Hu F., Li Z., Yang C., & Jiang Y. A graph-based approach to detecting tourist movement patterns using social media data // Cartography and Geographic Information Science, 2019, vol.46, no.4, pp.368–382.
21. Luo F., Cao G., Mulligan K., & Li X. Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago // Applied Geography, 2016, vol.70, pp.11–25.
22. Yamak Z., Saunier J., Vercouter L. SocksCatch: Automatic detection and grouping of sockpuppets in social media // Knowledge-Based Systems, 2018, vol.149, pp.124–142.
23. Stringhini G., Mourlanne P., Jacob G., Egele M., Kruegel C., & Vigna G. EvilCohort: Detecting communities of malicious accounts on online services / 24th USENIX Security Symposium, 2015, pp.563–578.
24. Ganjaliyev F. New method for community detection in social networks extracted from the Web / Proc. of the 4th International Conference "Problems of Cybernetics and Informatics" (PCI), 2012, pp.1–2.
25. Alguliev R.M., Aliguliyev R.M., Ganjaliyev F.S. Partition clustering-based method for detecting community structures in weighted social networks // International Journal of Information Processing and Management, 2013, vol.4, no.2, pp.60–72.
26. Cao Q., Yang X., Yu J., & Palow C. Uncovering large groups of active malicious accounts in online social networks / Proc. of the 2014 ACM SIGSAC Conference on Computer and Communications Security, 2014, pp.477–488.
27. Wani M.A., Jabin S. Mutual clustering coefficient-based suspicious-link detection approach for online social networks // Journal of King Saud University – Computer and Information Sciences, 2018.
28. Wang A.H. Don't follow me: Spam detection in twitter / International Conference on Security and Cryptography (SECRYPT), IEEE, 2010, pp.1–10.
29. Kwak H., Lee C., Park H., & Moon S. What is Twitter, a social network or a news media? / Proceedings of the 19th International Conference on World Wide Web, 2010, pp.591–600.
30. Zheng X., Zeng Z., Chen Z., Yu Y., & Rong C. Detecting spammers on social networks // Neurocomputing, 2015, vol.159, pp.27–34.
31. Yang C., Harkreader R., Gu G. Empirical evaluation and new design for fighting evolving twitter spammers // IEEE Transactions on Information Forensics and Security, 2013, vol.8, no.8, pp.1280–1293.
32. Wei F., Nguyen U. T. Twitter bot detection using bidirectional long short-term memory neural networks and word embeddings / Proc. of the 1st IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA), 2019, pp.101–109.
33. Harshitkgupta / Fake-Profile-Detection-using-ML. https://github.com/harshitkgupta/Fake-Profile-Detection-using-ML/tree/master/data
34. Instagram fake spammer genuine account. https://www.kaggle.com/free4ever1/instagram-fake-spammer-genuine-accounts
35. Fcakyon / instafake-dataset. https://github.com/fcakyon/instafake-dataset/tree/master/data/fake-v1.0

36. Radheysm / Fake-Profile-Detection. https://github.com/radheysm/Fake-Profile-Detection
37. KhayalaAhmadova / Fake_profile_clustering. https://github.com/KhayalaAhmadova/Fake_profile_clustering
38. Rank. https://orange3.readthedocs.io/projects/orange-visual-programming/en/latest/widgets/data/rank.html
39. Principal Component Analysis and k-means clustering to visualize a high dimensional dataset. https://medium.com/@dmitriy.kavyazin/principal-component-analysis-and-k-means-clustering-to-visualize-a-high-dimensional-dataset-577b2a7a5fe2
40. Clustering. https://scikit-learn.org/stable/modules/clustering.html#clustering
41. Clustering performance evaluation. https://scikit-learn.org/stable/modules/clustering.html#clustering-performance-evaluation

УДК 004.056:351

Ахмедова Хаяла В.
Институт Информационных Технологий НАНА, Баку, Азербайджан
ehmedova1xeyale@gmail.com

Обнаружение поддельных профилей в социальных сетях с применением методов кластеризации

В Сети существуют миллионы активных пользователей, так как она предлагает ряд возможностей для пользователей социальных сетей, чтобы завести новых друзей, прочитать новости, получить полезную информацию и повеселиться. Наличие миллионов активных пользователей социальных сетей создает условия для реализации злонамеренных целей, таких как манипулирование людьми, различные виды вызовов, дискредитация людей или организаций. В этом случае поддельные профили, работающие как группы, такие как профили троллей, учетные записи сивиллов, sockpuppets, учетные записи ботов и т.д., широко используются. Когда для обнаружения поддельных профилей используются алгоритмы классификации, данные должны иметь метки, время, затраченное на классификацию многих профилей, и т.д. Такие проблемы возникают. В этой статье используются k-means, Gaussian Mixture, агломеративная кластеризация, алгоритмы спектральной кластеризации для группировки поддельных профилей в социальных сетях. Поскольку алгоритмы кластеризации работают хуже, чем методы классификации при обнаружении поддельных профилей, в этой статье обсуждается, в каких данных методы кластеризации, используемые для обнаружения поддельных профилей, дают лучшие результаты. При применении алгоритмов использовались наборы данных открытого доступа, содержащие профильные данные. На основе результатов, полученных во время оценки эффективности методов кластеризации с использованием таких показателей оценки, как скорректированный рандомный индекс, однородность, полнота и т.д., алгоритм агломеративной кластеризации показал лучшие результаты, чем другие применяемые алгоритмы кластеризации.

Ключевые слова: *поддельный профиль, кластеризация, k-средних, агломеративная кластеризация.*

Khayala V. Ahmadova
Institute of Information Technology of ANAS, Baku, Azerbaijan
ehmedova1xeyale@gmail.com

Detection of fake profiles in social networks with the application of clustering methods

Social network has millions of active users, as it offers a number of opportunities for social network users to make new friends, read the news, get useful information, and have fun. Having millions of active users of social networks creates conditions for the implementation of malicious purposes, such as manipulation of people, various types of challenges, discrediting people or organizations. In this case, fake profiles operating as groups such as troll profiles, sibyl accounts, sockpuppets, bot accounts, etc. are widely used. When classifying the algorithms used to detect fake profiles, the problems such as, data must have labels, the time spent classifying many profiles, and so on. arise. This article uses k-means, Gaussian Mixture, agglomerative clustering, spectral clustering algorithms to group fake profiles on social networks. Since clustering algorithms perform worse than classification methods in detecting fake profiles, this article discusses in which data the clustering methods used to detect fake profiles give better results. During the application of the algorithms, open access datasets containing profile-based data are used. Based on the results obtained during the performance evaluation of clustering methods using evaluation metrics such as, adjusted rand index, homogeneity, completeness, etc. the agglomerative clustering algorithm shows better results than other applied clustering algorithms.

Keywords: *Fake profile, clustering, k-means, agglomerative clustering, feature.*