

*Hacırahimova M.Ş.<sup>1</sup>, İsmaylova M.İ.<sup>2</sup>*

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan  
<sup>1</sup>makrufa@science.az, <sup>2</sup>imarziya@gmail.com

## SENTİMENT ANALİZ: PROBLEMLƏRİ VƏ HƏLLİ YOLLARI

Daxil olmuşdur: 29.05.2020 Düzəliş olunmuşdur: 11.06.2020 Qəbul olunmuşdur: 30.06.2020

*Sentiment analiz və ya rəylərin intellektual analizi məhsullar, xidmətlər, təşkilatlar, fərdlər, problemlər, hadisələr, mövzular və onların xüsusiyyətləri kimi obyektləri araşdıran insanların fikirlərini, hisslərini, duyğularını, qiymətləndirmələrini və münasibətlərini təhlil edən tədqiqat sahəsidir. Sentiment analizin inkişafı sosial şəbəkələrdə, forumlarda, bloqlarda, mikrobloqlarda aparılan geniş müzakirələr və s. generasiya edilən böyük həcmdə rəqəmsal verilənlərin yaranması ilə bağlıdır. Son illər sentiment analiz təbii dilin emalı, verilənlərin intellektual analizi, mətnlərin analizi və informasiya axtarışı kimi sahələrdə geniş tədqiqat obyektinə çevrilmişdir. O, bütövlükdə cəmiyyətin bütün sahələrində, o cümlədən marketing, maliyyə, politologiya, sosial elmlər, tibb elmləri və s. geniş yayılmışdır. Bu populyarlıq rəylərin praktiki olaraq insan fəaliyyətinin bütün istiqamətlərində həlledici əhəmiyyətə malik olması və insanların davranışlarına təsir edən əsas faktor olması ilə bağlıdır. Sosial şəbəkələr, bloqlar, forumlar və digər rəy mənbələrində yayımlanan yüz minlərlə fikir mətnlərinin adi üsullarla analiz oluna bilməməsi kompüter linqvistikası, sentiment analiz və ya rəylərin, fikirlərin çıxarılması sistemlərinin istifadəsini zəruri etmişdir. Məqalədə sentiment analizin tarixi və terminoloji aspektlərinə baxılmış, informasiya mənbələri, tətbiq sahələri haqqında məlumat verilmişdir. İşdə, həmçinin sentiment analizin əsas məsələləri və səviyyələri tədqiq olunmuşdur. Mövcud problemlər və həll yolları təhlil edilmişdir.*

*Açar sözlər: sentiment analiz, opinion mining, sentiment klassifikasiya, sentiment analizin səviyyələri, maşın təlimi.*

### 1. Giriş

Bəşəriyyətin inkişafının indiki mərhələsi informasiyanın həcmının sürətlə artması ilə xarakterizə olunur. İnformasiyanın ən geniş yayılmış formalarından biri təbii dildə yazılmış mətnlərdir. İnformasiyanın mətn forması insanlar üçün təbiidir və onlar tərəfindən asanlıqla qəbul edilir. Sosial şəbəkələrdə, forumlarda, bloqlarda, mikrobloqlarda və digər rəy mənbələrində yayımlanan mətnlərdə ifadə edilən fikir və ya rəylərin sayı yüz minlərə çatır ki, bunun da əl ilə emalı mümkün deyil və avtomatlaşdırılmış emal tələb olunur. Problemi həll etmək üçün kompüter linqvistikası, sentiment analiz (*ing. sentiment analysis*) və ya “opinion mining” kimi analiz sistemləri geniş yayılmışdır [1, 2]. Bu yanaşmalar sayəsində müəllifin mətnə ifadə etdiyi rəyi avtomatik olaraq əldə etmək, onun pozitiv, neqativ, neytral olub-olmadığını qiymətləndirmək mümkün olmuşdur. Müəyyən bir obyekt haqqında müəllifin ifadə etdiyi rəyi qiymətləndirmək üçün maşın təlimi (MT) metodları, lüğətlərə və qaydalara əsaslanan vasitələrdən istifadə olunur.

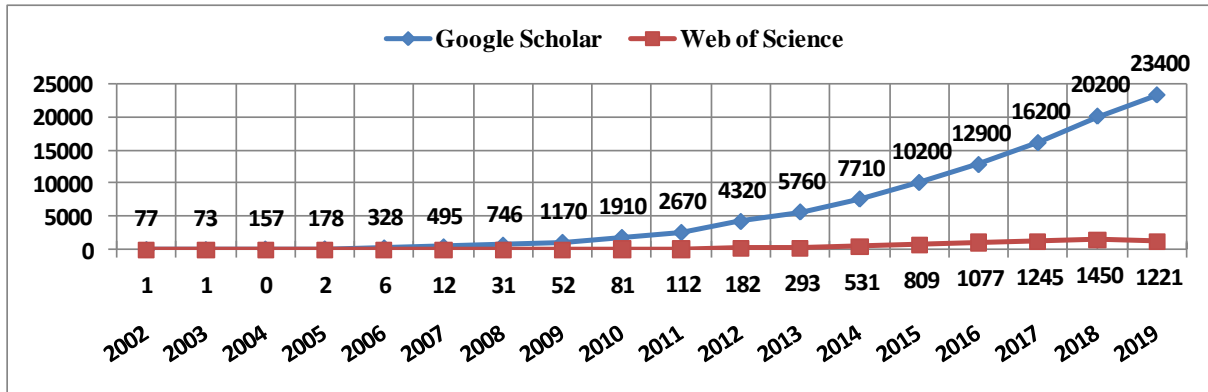
Sentiment analiz və ya rəylərin, fikirlərin çıxarılması rəylərə yönəlmiş təbii dilin emalı (*ing. Natural Language Processing, NLP*), statistika və ya MT metodlarından istifadə edərək mətn sənədlərindən onun emosional məzmununun çıxarılması, tanınmasını xarakterizə edir. Termin kimi sentiment analiz ilk dəfə 2003-cü ildə T.Nasukawa və J.Yinin məqaləsində [3], “opinion mining” termini isə K.Dave, S.Lawrence və D.M.Pennockun məqaləsində [4] istifadə olunmuşdur. Bu iki termin eyni mənalı olsa da, bəzi tədqiqatçılar onların fərqli anlayışlar olduğunu bildirirlər [5]. “Opinion mining” mətn sənədlərində subyektiv informasiyanın aşkar edilməsi, çıxarılması üçün bir üsuldur. Sentiment analiz isə mətnə ifadə edilmiş rəyləri müəyyən edir və sonra onları analiz edir. Yəni, sentiment analizin məqsədi rəyləri aşkarlamaq, onların ifadə etdiyi fikirləri müəyyən etmək və sonra onları qütblərinə (pozitiv və ya neqativ) görə klassifikasiya etməkdir [1, 5, 6].

Sentiment analizin kökləri 20-ci əsrin əvvəllərində ictimai rəyin analizi və 1990-cı illərdə kompüter linqvistikası sahəsində həyata keçirilmiş mətnin subyektivliyinin analizi üzrə

tədqiqatlara əsaslanır. İnternetdə böyük həcmdə mətn və rəylərin meydana çıxmasından əvvəl tədqiqatlar əsasən istifadəçilərin və ya müştərilərin rəylərinə deyil, ictimai və ya ekspert rəylərini nəzərə alan sorğu metodlarına əsaslanırdı. “İctimai rəyin analizində bir kəsişmə üsulu metod kimi” (ing. *The Cross-Out Technique as a Method in Public Opinion Analysis*) adlı ilk məqalə 1940-cı ildə nəşr olunmuşdur. İkinci dünya müharibəsindən sonra müharibə dövründə əziyyət çəkən bir neçə ölkədə (Yaponiya, İtaliya və Çexoslovakiya) ictimai rəyin ölçülməsinə həsr olunmuş üç məqalə 1945 və 1947-ci illərdə "İctimai rəy" (ing. *Public Opinion*) jurnalında nəşr edilmişdir. 1990-cı ilin ortalarında kompüter əsaslı sistemlər ortaya çıxmağa başladı. Bu sistemlərin inkişafı da tədqiqatlarda əks olunmağa başladı. 1995-ci ildə dərc olunmuş bir məqalədə sənaye təhlükəsizliyi sahəsində ekspert rəylərinin analizi üçün rəyləri birləşdirməyə imkan verən kompüter sistemindən istifadə olunmuşdur [7].

Müasir sentiment analizin yaranmasına 1962-ci ildə əsas qoyulmuş Kompüter Linqvistikası Assosiasiyası böyük təsir göstərmişdir. Məhz bu cəmiyyətdə kompüter əsaslı sentiment analiz meydana çıxmış və subyektiv cümlələrin aşkarlanması üsulları təklif edilmiş və buna aid ilk məqalə 1990-cı ildə nəşr olunmuşdur. Buna baxmayaraq, müasir sentiment analizin inkişafı təxminən 10 il sonra 2000-ci illərin ortalarında baş verdi. O, internetdə mövcud olan məhsulların araşdırılmasına fokuslanmışdır. Həmin vaxtdan etibarən sentiment analizin istifadəsi maliyyə bazarlarının proqnozlaşdırılması, elektron kommersiya, tibb və terror hücumları və s. kimi bir çox digər sahələrdə də öz tətbiqini tapdı. Bundan başqa, sentiment analiz və təbii dilin emalına əsaslanan tədqiqatlar çoxdilliliyin dəstəklənməsi kimi bir çox problemlərin həllinə imkan verdi [7].

Son zamanlar nüfuzlu beynəlxalq təşkilatlar və elmi qurumlar tərəfindən sentiment analizə aid müxtəlif mövzular üzrə konfrans, simpozium, seminar və forumlar keçirilməkdədir. Hazırda dünyanın tanınmış elmi mərkəzlərində sentiment analiz istiqamətində fundamental elmi-tədqiqat işləri aparılmaqdadır. Aparığımız araşdırmaların nəticəsi olaraq, Google Scholar və Web of Science elmi bazalarında sentiment analizə aid dərc olunmuş məqalələrin 2002–2019-cu illər üzrə dinamikası aşağıdakı qrafikdə öz əksini tapmışdır (şəkil 1). Qrafikdən görüldüyü kimi, son illər bu sahəyə maraq artmışdır və bu da sentiment analiz probleminin aktuallığını göstərir.



Şəkil 1. Sentiment analiz üzrə Google Scholar və Web of Science bazalarında məqalələrin illər üzrə paylanması

Bu məqalədə sentiment analizin qısa icmalı verilmiş, onun terminoloji aspektləri, informasiya mənbələri, əsas məsələləri və səviyyələri şərh edilmişdir. Sentiment analiz üçün dərin təlim (ing. *deep learning*) yanaşmaları, tətbiq sahələri təhlil olunmuş, əsas problemləri müəyyənləşdirilmişdir.

## 2. Sentiment analizinin terminoloji aspektləri

Terminlər haqqında Sokratın belə bir aforizmi var: “Müdrikliyin başlanğıcı terminlərin təyin edilməsidir”. Formal olaraq sentiment analiz məndə ifadə edilmiş fikirlərin, rəylərin və hisslərin hesablanması işidir. Sentiment analizinin məqsədi müxtəlif mənbələrdə saxlanan subyektiv informasiyaları aşkar etmək və müəllifin hər hansı bir problem və ya sənədin ümumi vəziyyəti haqqında fikirlərini müəyyən etməkdir. Subyektivlik kiminsə rəylərinin, fikirlərinin, baxışlarının, qiymətləndirmələrinin, inamlarının, fərziyyələrinin linqvistik ifadəsi deməkdir [7]. [1–3]-də rəy, fikir, baxış və inam sözləri eynimənalı sözlər kimi istifadə olunsalar da, onların fərqli olduqları bildirilir.

[1]-də rəy riyazi olaraq  $(o, f, so, h, t)$  aşağıdakı kimi beşlik şəklində təsvir edilir.

$o$  -obyekti;  $f$  -  $o$  obyektinin xüsusiyyəti;  $so$  -  $o$  obyektinin  $f$  xüsusiyyəti haqqında rəyin orientasiyası və ya qütbü;  $h$  - rəy sahibi;  $t$  - zamanı göstərir.

- *Obyekt*: məhsul, insan, hadisə, təşkilat və mövzu ola bilər. Obyekt onunla əlaqəli atributlara, xüsusiyyətlərə və komponentlərə malik ola bilər;
- *Xüsusiyyət* (ing. *feature*): qiymətləndirilməsi aparılmış obyektin bir aspekti/xüsusiyyəti;
- *Rəyin orientasiyası və ya qütbü* (ing. *opinion orientation or polarity*):  $f$  xüsusiyyəti üzrə rəyin pozitiv, neqativ və ya neytral olmasını göstərir;
- *Rəy sahibi* (ing. *opinion holder*): rəy bildirən şəxs və ya təşkilatdır;
- *Zaman* (ing. *time*): rəyin bildirildiyi tarixdir.

## 3. Sentiment analiz və verilənlərin mənbəyi

İstifadəçi rəyləri xidmətlərin keyfiyyətinin yaxşılaşdırılması üçün mühüm bir hədəfdir. Bloqlar, sosial şəbəkə saytları, araşdırma saytları, məlumat və mikrobloqlar bu hədəf üçün yaxşı bir platformadır.

*Baxış saytları* (ing. *review sites*) istifadəçilərə insanlar, şirkətlər, məhsullar və ya xidmətlər haqqında tənqidi fikir verən rəyləri dərc etməyə imkan verən veb-saytdır. Sentiment analiz üzrə araşdırmaların çoxu film və məhsul üzrə baxış saytlarında aparılmışdır [4, 8, 9]. Baxış saytlarında sentiment analiz həm istehsalçılar, həm də istehlakçılar üçün faydalıdır. İstehsalçılar rəylərə əsasən məhsulun qəbulunu qiymətləndirə bilər, həm də rəyçilərin bəyəndikləri və bəyənmədikləri xüsusiyyətləri əldə edə bilərlər.

*Bloqlar* (ing. *blogs*) veb-jurnal və ya hadisələr gündəliyi onlayn jurnal üslubunda xronoloji olaraq təşkil edilmiş fikirləri, informasiyanı, şərhləri və ya müxtəlif tipli hiper əlaqələri özündə cəmləşdirən veb-səhifədir. İnternetdən istifadənin artması ilə bloq və bloq səhifələri sürətlə böyüyür. Bloq səhifələri şəxsi fikirləri ifadə etmək üçün ən populyar vasitəyə çevrilmişdir. Bloqlar gündəlik həyatda baş verən hadisələri qeyd edir, öz fikirlərini, hisslərini və emosiyalarını bir bloqda ifadə edirlər [1, 8, 10–13]. Bu bloqların çoxu bir çox məhsullar, problemlər və s. haqqında icmalları təşkil edir. Sentiment analiz ilə əlaqəli tədqiqatların çoxunda bloqlardan fikir mənbəyi kimi istifadə olunur. [11–13]-də filmlərin satışı, siyasi əhval-ruhiyyə və satışların analizini proqnozlaşdırmaq üçün bloq əsasında sentiment analizdən istifadə edilmişdir.

*Sosial şəbəkə saytları* (ing. *social networks*) bir-birini tanıyan və ya ümumi maraqları bölüşən insanlar arasında sosial münasibətləri təmin edən onlayn xidmətlər və ya saytlardır. Sosial şəbəkə saytları istifadəçilərə fikirlərini, fəaliyyətlərini, hadisələri və maraqlarını paylaşmalarına imkan verir. Bu paylaşımalar ən son satın alınan telefonlar, izlənən filmlər, siyasi məsələlər və müxtəlif məsələlər haqqında ola bilər [2, 14].

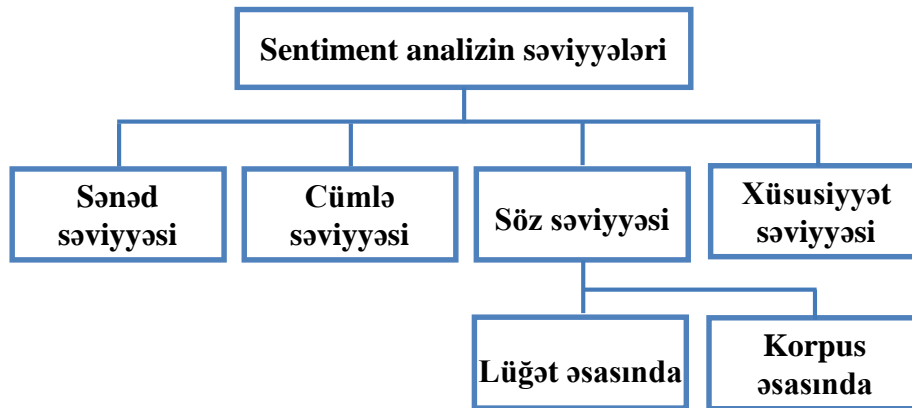
## 4. Sentiment analizinin məsələləri və səviyyələri

Sentiment analiz təbii dilin emalı, veb analiz, mətn analizi və MT metodlarını özündə cəmləşdirən mürəkkəb bir fənlərəmə məsələdir. Bu mürəkkəb məsələ subyektivlik klassifikasiyası və sentiment klassifikasiya məsələlərindən ibarətdir [2].

*Subyektivlik klassifikasiyası.* Ümumiyyətlə, sənəd faktlar ifadə edilmiş obyektiv və müəlliflərin fikirləri, baxışları və hissləri ifadə edilmiş subyektiv cümlələr toplusundan ibarətdir. Subyektivlik klassifikasiyası rəylər əsasında ifadə edilmiş cümlələrin klassifikasiyasını nəzərdə tutur [15, 16]. [10]-da subyektivlik klassifikasiyası aşağıdakı kimi ifadə edilmişdir:  $S = \{s_1, s_2, \dots, s_n\}$   $D$  sənədindəki cümlələr dəstidir. Subyektivlik klassifikasiyasının məqsədi subyektiv cümlələri ( $S_s$  - subyektiv cümlələr dəsti), obyektiv cümlələrdən ( $S_0$  - obyektiv cümlələr dəsti) ayırmaqdan ibarətdir, harada ki,  $S_s \cup S_0 = S$ . Bu məsələ xüsusilə xəbərlər və İnternet forumlar üçün aktualdır.

*Sentiment klassifikasiya.* Sentiment klassifikasiya rəylər əsasında fikir ifadə edən mətnlər müəyyən edildikdən sonra mətni qütbünə görə klasifikasiya edir, yəni onun pozitiv və ya neqativ fikir ifadə etməməsini müəyyənləşdirir. Sentiment klassifikasiya binar sentiment klassifikasiya (pozitiv, neqativ), çox sinifli sentiment klassifikasiya (pozitiv, ifrat pozitiv, neytral, neqativ, ifrat neqativ), reqressiya və ranqlaşdırılmış ola bilər [9, 10]. [10]-da binar sentiment klassifikasiya aşağıdakı kimi ifadə edilmişdir. Binar sentiment klassifikasiya  $D = \{d_1, d_2, \dots, d_n\}$  sənədlər dəstinin hər bir  $d_i$  sənədinin  $C = \{pozitiv, neqativ\}$  siniflər dəstində ifadə olunan bir etiket ilə klassifikasiyasını nəzərdə tutur.

Sentiment analizin yuxarıda qeyd olunan məsələləri bir neçə səviyyədə, yəni sənəd, cümlə, söz və xüsusiyyət səviyyəsində həyata keçirilir (şəkil 2).



Şəkil 2. Sentiment analizin səviyyələri

*Sənəd səviyyəsində sentiment analiz.* Sənəd səviyyəsində əsas məsələ rəy bildirilən sənədin pozitiv və ya neqativ fikir ifadə etməsini təyin etməkdən ibarətdir. Bu səviyyədə sentiment oriyentasiyası müəyyən olunan bütöv sənədə əsas vahid kimi baxılır. Məsələni asanlaşdırmaq üçün hər bir mətnin ümumi rəyinin tək bir rəy sahibi tərəfindən verildiyi və vahid bir obyekt haqqında olması ehtimal edilir [1, 2].

Bu səviyyədə MT-nin müxtəlif yanaşmalarından istifadə edilir. [17]-də müəlliflər rəylərin pozitiv və ya neqativ fikir ifadə etməsini klassifikasiya etmək üçün MT-nin ənənəvi metodlarından istifadə etmişlər. Müəlliflər *Naive Bayes* (NB), *Maximum Entropy* (ME), *Support Vektor Machines* (SVM) klassifikatorlarından və uniqramlar, biqramlar, terminlərin istifadə tezliyi, terminlərin mövqeyi və nitq hissələri kimi xüsusiyyətlərdən istifadə edərək eksperiment aparmışlar. Eksperiment nəticəsində məlum olmuşdur ki, SVM klassifikatoru digər klassifikatorlardan daha üstündür və uniqramlar (sözlər) iştirak edən informasiya daha effektivdir. [18]-də sənəd səviyyəsində sentiment analiz reqressiya problemi kimi xülasə edilmişdir. Reyting ballarını proqnozlaşdırmaq üçün nəzarət edilən (*ing. supervised*) təlim metodu istifadə edilmişdir.

*Cümlə səviyyəsində sentiment analiz.* Bu səviyyədə hər bir cümlənin pozitiv, neqativ və ya neytral fikir ifadə etməsi müəyyən olunur. “Neytral” adətən fikrin olmamasını bildirir. Analizin bu səviyyəsi subyektiv klassifikasiya ilə sıx bağlıdır [1, 2].

[19]-də cümlə səviyyəsində obyektiv və subyektiv cümlələrin qarışığından ibarət olan sənəddə subyektiv cümlələrin aşkarlanması üzrə araşdırmalar aparılmış və sonra bu subyektiv cümlələrin klassifikasiyası və eyni zamanda sentiment oriyentasiyaları (pozitiv, neqativ və neytral) müəyyən edilmişdir. Subyektiv və ya rəy cümlələrinin müəyyənləşdirilməsi üçün nəzarət edilən təlim metodundan istifadə olunmuşdur. Aşkarlanmış hər bir subyektiv cümlənin rəyini klassifikasiya etmək üçün [20]-də istifadə olunan nəzarətsiz (*ing. unsupervision*) təlim metoduna oxşar, lakin daha çox sayda ilkin sözlərdən və qiymətləndirmə funksiyası kimi loqarifmik metoddan istifadə edilmişdir. [21]-də müəlliflər tərəfindən istifadə olunan metod rəy cümlələrinin bütün qütblərini əldə etmək üçün cümlədəki sözlərin oriyentasiyalarını toplamaqdan ibarət idi.

*Söz səviyyəsində sentiment analiz.* Sentiment analizin əsas məsələlərindən biri də söz səviyyəsində semantik oriyentasiyanın aşkar edilməsi məsələsidir. Bir çox işlərdə cümlə və sənəd səviyyələrində rəylərin klassifikasiyası üçün söz və söz birləşmələrinin məlum qütbləri istifadə edilir [22]. Bu səbəbdən söz lüğətinin semantik oriyentasiyasının əl üsulu ilə və ya yarım avtomatik qurulması geniş yayılmışdır. Söz əhval-ruhiyyəsinin klassifikasiyası əsasən xüsusiyyətlər kimi sifətlərdən istifadə edir, lakin zərflər və bəzi fellər, isimlər də tədqiqatçılar tərəfindən istifadə olunur [23, 24]. Söz səviyyəsində rəyi avtomatik şərh edən iki: lüğət əsaslı və korpus əsaslı yanaşmadan istifadə olunur.

*Lüğət əsaslı yanaşmalar.* Bu yanaşmada əvvəlcədən qütbləri məlum olan sözlərin kiçik bir ilkin siyahısı yaradılır. Bu siyahı daha sonra WordNet kimi onlayn-lüğət mənbələrindən sinonimlərin və ya antonimlərin iterativ çıxarılması yolu ilə genişləndirilir. Yeni tapılmış sözlər ilkin siyahıya əlavə edilir və bundan sonra növbəti iterasiya başlayır. Bu proses yeni söz tapılmayana qədər davam edir. Proses başa çatdıqdan sonra isə səhvləri aradan qaldırmaq və ya düzəltmək üçün əl üsulu ilə yoxlama yerinə yetirilə bilər [5]. [25]-də müəlliflər əl üsulu ilə pozitiv və neqativ fel və sifətlərin iki ilkin siyahısını yaratmışlar. Sonra bu siyahıları WordNet-dən ilkin siyahıdakı sözlərin sinonim və antonimlərini çıxararaq müvafiq siyahılara (sinonimləri bir siyahıya, antonimləri əks siyahıya) mənimsəməklə genişləndirmişlər. Sözlərin sentiment gücü (*ing. strength*) müəyyən edilmiş, hər bir söz üçün həm pozitiv və həm də neqativ rəyin gücü hesablanmış və onların nisbi əmsalları müqayisə edilmişdir. [26]-də müəlliflər WordNet-in leksik əlaqəsinə əsaslanaraq sözlərin semantik oriyentasiyasını müəyyən etmiş, eyni dəstdən olan sözləri və onların bütün sinonimlərini WordNet-ə yığaraq sinonim sözləri tillərlə birləşdirərək qraf yaratmışlar. Sözlərin semantik oriyentasiyası onların nisbi məsafəsinə görə iki ifadə (“yaxşı” və “pis”) üçün hesablanmışdır. Bu məsafə iki söz arasındakı ən qısa yolun uzunluğunu bildirir. Oriyentasiyanın gücünü göstərən qiymətlər mütləq qiymətlə [-1, 1] intervalında dəyişir.

[27]-də müəlliflər reklam kontentində sentiment cümlələri müəyyən etmək üçün lüğətə əsaslanan yanaşmadan istifadə edərək istifadəçilər üçün reklamın relevantlığını və istifadəçi təcrübəsini yaxşılaşdırmaq məqsədilə reklam strategiyası təklif etmişlər. Onlar sintaktik təhlil və sentiment lüğətdən istifadə edərək reklam açar sözləri kimi qəbul edilən neqativ hisslərlə əlaqəli rəy cümlələrinin tematik sözlərini çıxarmaq üçün qayda əsaslı bir yanaşma təklif etmişlər. Eksperimental nəticə təklif olunan yanaşmanın səmərəliliyini nümayiş etdirmişdir.

Lüğət əsasında yanaşmanın çatışmazlığı universal olmamasıdır, yəni hər bir sahənin öz lüğəti yaradılmalıdır. Parametrlərin dəqiq qiymətləndirilməsi üçün tələb olunan ilkin verilənlərin sayının az olması isə onun üstün cəhətidir.

*Korpus əsaslı yanaşmalar.* Bu yanaşma kontekstə uyğun oriyentasiyalara malik rəy sözlərinin tapılması problemini həll etməyə kömək edir. Korpus əsaslı yanaşma sintaktik və ya statistik metodlara əsaslanır. [28]-də bu metodlardan biri təqdim edilmişdir. Müəlliflər cüt-cüt birləşmiş sifətlərin eyni oriyentasiyaya (əgər “və” bağlayıcısı ilə birləşibsə) və əks oriyentasiyaya (əgər “ancaq” bağlayıcısı ilə birləşibsə) malik olduğunu fərz edərək sifətlərin oriyentasiyasının

proqnozunu vermişlər. Onlar loqarifmik xətti regressiya modelindən istifadə edərək oxşar və əks oriyentasiyalı sözlərin klasterini yaratmaq üçün “pozulmuş və çox pis” və ya “bəsit, lakin yaxşı qəbul edilmiş” kimi birləşmələrdən istifadə etmişlər. Daha yüksək orta tezlik ifadələri ehtiva edən klasteri intuitiv olaraq pozitiv siyahı adlandırmışlar. Bu metod nəzarətsiz klassifikasiya metodu olduğu üçün tələb olunan korpus çox böyük idi. [20]-də birləşmə vasitəsilə semantik oriyentasiya müəyyən edilmişdir. Əgər onlar yaxşı birləşməyə malikdirsə, onda pozitiv oriyentasiyaya malik olduqları bildirilir. Naməlum söz və əl ilə seçilmiş ilkin siyahılar (əla və pis kimi) arasında birləşmə əlaqəsi onu pozitiv və ya neqativ kimi klassifikasiya etmək üçün istifadə edilmişdir.

*Xüsusiyyət əsasında sentiment analiz.* Qeyd etmək lazımdır ki, sənəd, cümlə və söz səviyyəsinin analizi ilə rəy sahibinin nə barəsində rəy bildirdiyi məlum deyil. Bundan əlavə bu səviyyədə hansı xüsusiyyətlərdən bəhs edildiyi də bilinmir. Ona görə də xüsusiyyət əsasında sentiment analizdən istifadə olunur. Rəyçi bəzi xüsusiyyətləri bəyənilib və bəzilərini bəyənməməsinə baxmayaraq məhsulun ümumi rəyi müsbət və ya mənfi ola bilər. Bu sənəd səviyyəsində və ya cümlə səviyyəsində sentiment klassifikasiya ilə müəyyən edilə bilmir. Bu halda, xüsusiyyətə əsaslanan rəylərin sentiment analizi tələb olunur [29–31]. Buraya məhsulun xüsusiyyəti və bununla bağlı uyğun rəyin çıxarılması daxildir. İnstinkt olaraq, məhsulun xüsusiyyətlərinin isim və isim ifadələri ilə ifadə edildiyini hesab etmək olar, lakin bütün isim və isim ifadələri məhsulun xüsusiyyətləri olmaya bilər. [23]-də müəlliflər yalnız baza isim ifadələrini (*ing. Base Noun Phrases, BNP*), müəyyən baza isim ifadələrini (*ing. Definite Base Noun Phrases, dBNP*) və başlanğıc müəyyən isim ifadələrini (*ing. Beginning Definite Base Noun Phrases, bBNP*) çıxarmaqla namizəd sözləri daha da məhdudlaşdırırlar. Aşkar edilmiş hər bir rəy ifadəsi üçün hədəf və son ziddiyyət rəy nümunəsi verilənlər bazası (VB) əsasında müəyyən edilir.

[30]-da tədqiqatçılar insanları ən çox maraqlandıran xüsusiyyəti, yəni daha tez-tez rast gəlinən isim və ya isim ifadələrini çıxarırlar. Onlar sentiment oriyentasiyanı müəyyənləşdirmək məqsədilə bir xüsusiyyətə ən yaxın rəy sözünü təyin etmək üçün sadə evristik metoddan istifadə etmişlər. [31]-də müəlliflər WordNet resursundan və morfoloji təhlildən istifadə edərək məhsulun xüsusiyyətlərinə aid olmayan isim ifadələrini kənarlaşdırmaqla xüsusiyyətlərin çıxarılması məsələsini əhəmiyyətli dərəcədə yaxşılaşdırmışlar.

Xüsusiyyət əsasında sentiment analiz bir neçə alt məsələni əhatə edir:

- Rəy sahibinin şərh etdiyi obyektin xüsusiyyətlərinin aşkarlanması və çıxarılması;
- Xüsusiyyətlər haqqında rəylərin pozitiv, neqativ və ya neytral olmasının müəyyənləşdirilməsi;
- Xüsusiyyətlərin sinonimlərinin çıxarılması.

## 5. Sentiment analiz üçün dərin təlim yanaşmaları

Yuxarıda qeyd etdiyimiz kimi, sentiment analiz üçün MT metodları geniş şəkildə istifadə olunur. Son zamanlar mətnlərin sentiment analizində MT metodlarından biri olan dərin təlim metodları geniş istifadə olunmağa başlanmışdır. Onlardan bir neçəsi aşağıda şərh olunmuşdur.

Qısa mətnlərin sentiment analizi, ehtiva etdikləri məhdud kontekstli informasiyalara görə çətindir. Qısa mətnlərin sentiment analizinə dair mövcud tədqiqatlar əsasən emosional məlumatlara əsaslanan metodlar və xüsusiyyətlərə əsaslanan klassifikasiya metodlarını əhatə edir. [32]-də konvolsional neyron şəbəkə (*ing. Convolutional Neural Network, CNN*) və rekurrent neyron şəbəkənin (*ing. Recurrent Neural Network, RNN*) üstünlüklərindən faydalanan və onları qısa mətnlərin sentiment analizi üçün birlikdə istifadə edən dərin neyron şəbəkə modeli təqdim edilmişdir. Təqdim olunmuş modeldə CNN vasitəsilə çıxarılan lokal xüsusiyyətlər RNN üçün giriş veriləni kimi istifadə edilmişdir. Eksperimentlər MR, SST1, SST2 film rəyləri VB-də həyata keçirilmişdir. Tədqiqatlar CNN və RNN hibrid modelinin qısa mətnlərin sentiment klassifikasiyasında göstərdiyi nəticələrin ayrı-ayrılıqda CNN və RNN-nin göstərdiyi nəticələrdən daha üstün olduğunu göstərmişdir.

[33]-də müəlliflər qısa mətnlərin sentiment analizini həyata keçirmək üçün xüsusiyyət, söz və cümlə səviyyəsindən birgə istifadə edərək yeni dərin CNN modeli təqdim etmişlər. Təklif olunan CNN cümləsinin xüsusiyyəti (*ing. Character to Sentence Convolutional Neural Network, CharSCNN*) adlı model sözlərdən və istənilən ölçülü cümlələrdən uyğun xüsusiyyətlərin çıxarılması üçün iki laydan istifadə edir. Eksperimentlər film rəyləri (*ing. Stanford Sentiment Treebank, SSTb*) və tvit mesajları (*ing. Stanford Twitter Sentiment, STS*) VB-də həyata keçirilmişdir. Eksperimentlər göstərdi ki, nümayiş olunan bu model irəli istiqamətlənmiş neyron şəbəkə modeli (*ing. Recursive Neural Tensor Network, RNTN*) kimi effektiv ola bilər.

Sənəd səviyyəsində sentiment klassifikasiya, sənəddəki məhsula dair istifadəçinin ümumi əhval-ruhiyyəsini proqnozlaşdırmağa yönəlmiş əsas problemdir. Problemin həlli üçün müxtəlif üsullar təklif edilmişdir, lakin onların çoxu rəy bildirən istifadəçilərin və qiymətləndirilən məhsulların təsirini nəzərə almırlar. [34]-də problemin həlli üçün sənəd səviyyəsində sentiment klassifikasiya üçün eyni zamanda istifadəçi və məhsul haqqında məlumatları toplaya bilən dərin yaddaş şəbəkəsi təklif edilmişdir. Nəticədə alqoritmin bir neçə mövcud metoddan daha məhsuldar olmasına nail olunmuşdur.

[35]-də müəlliflər sənəd səviyyəsində sentiment klassifikasiya üçün SVM və süni neyron şəbəkələri (*ing. Artificial Neural Network, ANN*) arasında empirik müqayisə aparmışlar. Nəticədə ANN-in əksər hallarda SVM-dən daha üstün olduğu nümayiş olunmuşdur.

[36]-da sənəd səviyyəsində sentiment klassifikasiya üçün neyron şəbəkə modelləri (*ing. Convolutional Gated Recurrent Neural Network (Conv-GRNN) və Long Short-Term Memory-Gated Recurrent Neural Network (LSTM-GRNN)*) təqdim edilmişdir. Bu yanaşma sənəd nümunəsindəki cümlələrin semantikasını və onların əlaqələrini kodlaşdırır və sentiment klassifikasiya nəzarət edilən təlim metodu ilə effektiv şəkildə öyrədilir. İki qiymətləndirmə metrikası üçün IMDB film rəyləri və Yelp (2013–2015) restoran rəyləri VB-də eksperimentlər aparılmışdır. Empirik nəticələr yanaşmanın ən yaxşı nəticələrə nail olduğunu göstərmişdir. RNN-nin sənədin məzmununun modelləşdirilməsində son dərəcə zəif olduğu, eyni zamanda ənənəvi neyron keçidlər (*ing. gates*) əlavə etməklə məhsuldarlığı əhəmiyyətli dərəcədə artırdığı, cümlə nümunəsinin modelləşdirilməsində uzun qısamüddətli yaddaşın (*ing. Long Short-Term Memory, LSTM*) multi filtirlənmiş CNN-dən daha yaxşı olduğu göstərilir.

Son zamanlarda neyron şəbəkələr texniki xüsusiyyətlərini asanlaşdırmaq qabiliyyətinə görə sentiment klassifikasiyada böyük uğur qazanmışdır. Ancaq, problemlərdən biri yaddaş blokunun çatışmazlığı səbəbindən təkrarlanan (*ing. recurrent*) arxitektura çərçivəsində sənəd səviyyəsində sentiment klassifikasiya üçün müxtəlif formatlı uzun mətnlərin modelləşdirilməsidir. [37]-də problemin həlli üçün, uzun mətnlərdə ümumi semantik məlumatları əldə etmək məqsədilə keşləmiş uzun qısamüddətli yaddaş (*ing. Cached Long Short-Term Memory, CLSTM*) neyron şəbəkə təqdim edilmişdir. CLSTM yaddaşı müxtəlif unutmə dərəcələrinə malik bir neçə qrupa bölən keş (*ing. cache*) mexanizm təqdim edir və bununla da şəbəkəyə təkrarlanan blok daxilində sentiment məlumatları daha yaxşı saxlamağa imkan verir. Təqdim olunan CLSTM-in, sənəd səviyyəsində sentiment analizi üç Yelp 2013, Yelp 2014 və IMDB VB-də həyata keçirilmişdir və onun ən müasir modellərdən daha yaxşı nəticə göstərdiyi qeyd edilmişdir.

## 6. Sentiment analizin tətbiqləri

Müxtəlif baxış saytlarından, bloqlardan, forumlardan və sosial şəbəkələrdən rəylərə əsaslanan və emosional həyəcan yaradan məlumatların əlçatanlığı həm elmi dairələr, həm də iş adamları tərəfindən sentiment analizə maraq yaratmışdır. Sentiment analiz təşkilatlara və servis provayderlərinə müştərilərinin və istifadəçilərinin fikirlərini öyrənməyə, məhsul və xidmətləri onların ehtiyaclarına uyğunlaşdırmağa kömək edir. Bu tətbiqlərin bir neçəsi aşağıda şərh edilmişdir:

*Biznes tətbiqləri (ing. Business applications)*. Sentiment analiz “bazar əhval-ruhiyyəsindən” yararlanmaq istəyən bir çox şirkətlər tərəfindən qəbul edilmişdir [38]. Sentiment analizin tətbiqi

ilə məhsulların araşdırılması, brendlərin izlənməsi, marketing strategiyalarının dəyişdirilməsi və maliyyə yeniliklərinin əldə edilməsinə nail olmaq olar. Sentiment analizinin dəstəklədiyi əsas fəaliyyətlər aşağıdakılardır:

- Birləşdirilmiş istifadəçi rəylərinin və icmal saytlarından markaların, məhsulların və xidmətlərin reytinglərinin avtomatik izlənməsi [39];
- Alıcıların, rəqiblərin və bazar tendensiyalarının meyllərinin analizi;
- Şirkətlə əlaqəli hadisələrə və insidentlərə verilən reaksiyaların qiymətləndirilməsi. Yeni bir məhsul satışa çıxarılan zaman məhsulun qəbulu barədə anında əks əlaqə yarada bilər. Marka imicinin bəyənilib-bəyənilmədiyini qiymətləndirə bilər.
- Zərərli virus təsirlərinin qarşısını almaq üçün həlledici məsələlərin monitorinqi.

Bu tətbiqlərdə tədqiqatçılar tərəfindən aşkar edilən əsas problemlər aşağıdakılardır:

- məhsulun aspektlərini müəyyənləşdirmək;
- rəyləri məhsulun aspektləri ilə əlaqələndirmək;
- saxta resenziyaları müəyyənləşdirmək və qaydaya uyğun olmayan rəyləri emal etmək.

*Siyasət sahəsində tətbiqlər.* Sentiment analiz siyasi bloqlarda bloqerlərin problemləri və subyektiv fikirlərini izləməyə imkan verir. Sentiment analiz siyasi təşkilata seçicilərə hansı mövzuların yaxın olduğunu anlamalarına kömək edə bilər [40]. Müəlliflər ABŞ Konqresindəki debatların stenoqramında təklif olunan qanunvericiliyin dəstəklənib-dəstəklənməməsini müəyyənləşdirmişlər. Seçicilərə verilən məlumatların əhəmiyyətini artırmaq üçün ictimai xadimlərin mövqeyini, yəni onların dəstəklədikləri və ya dəstəkləmədikləri səbəbləri də müəyyən etmək olar [41]. [42]-də orijinal posta müxalif siyasi baxışların olub-olmadığını mühakimə etmək üçün siyasi müzakirə qrupu yazılarında statistik sentiment analiz metodu təklif edilmişdir.

*Tövsiyə sistemi (ing. Recommender system).* Bu sistemlər məndən istifadəçi reytingini çıxarmağa kömək edə bilər. [8]-də müəlliflər film rəylərini “tövsiyə edilən” və “tövsiyə edilməyən” kimi klassifikasiya etmişlər.

*Ekspert tapma (ing. Expert finding).* [43]-də müəlliflər ədəbi nüfuzu izləmək üçün sentiment analiz üsullarından istifadə etmişlər. [44]-də hər bir qrup üzvünün bloqunu və əldə edilən şərhlərin sentiment analizindən istifadə edərək virtual cəmiyyətdən bilik əldə etməyi təklif edirlər. Birləşdirilmiş oriyentasiya gücü, bloqların sıralanmasına imkan verən bloq hesabını müəyyən edir və ekspertin ən yüksək bloq dərəcəsinə sahib olduğunu müəyyənləşdirir.

*Yekunlaşdırma (ing. Summarization).* Rəyin yekunlaşdırılması bir məhsula onlayn baxış sayı çox olduqda tətbiq olunur. Bu həm müştəri, həm də istehsal olunan məhsul üçün çətinləşə bilər. İstehlakçı bütün rəyləri oxuya və əsaslandırılmış qərar qəbul etmək imkanına, istehsalçı isə istehlakçının rəylərini izləmək imkanına malik olmaya bilər. [27]-də müəlliflər müəyyən məhsul üzrə araşdırmalar aparmışlar:

- 1) məhsulun aşkarlanmış xüsusiyyətləri şərh edilmişdir;
- 2) hər xüsusiyyət üçün fikir bildirilən rəy cümlələri müəyyən edilmişdir;
- 3) aşkarlanmış informasiyadan istifadə edərək xülasə hazırlanmışdır.

Tək və ya birdən çox sənədlərin yekunlaşdırılması sentiment analizi yaxşılaşdırıla bilər bir tətbiqdır [45].

*Dövlət kəşfiyyatı (ing. Government intelligence).* Monitorinq mənbələrinin təklif etdiyi bu tətbiq sahəsi ilə antaqonist və ya düşmən əlaqələrin artması izlənilə bilər [46].

## 7. Sentiment analizinin problemləri

Sentiment analizinin bir sıra problemləri mövcuddur. Bunlardan bəziləri aşağıda sadalanmışdır [3, 4].

*Açar sözlərin seçilməsi (ing. keyword selection).* Adətən mətnləri klassifikasiya etmək üçün açar sözlər dəstindən istifadə edilir. Sentiment analizdə mətni bir-birindən çox fərqli olan iki sinfə (pozitiv və neqativ) görə klassifikasiya edirlər. Ancaq düzgün açar sözlər dəstini tapmaq çox da asan məsələ deyildir. Əhval-ruhiyyələr çox vaxt həssaslıqla ifadə edilə bilər, bu da cümlə və ya



sənəddəki bir ifadənin ayrılıqda nəzərdən keçirilməsini çətinləşdirə bilər. Məsələn, “Əgər siz hesab edirsiniz ki, bu sizin sevimli ətirinizdir, zəhmət olmasa ondan evdə istifadə edin və pəncərələri bağlayın” cümləsində heç bir neqativ sözə rast gəlinmir.

*Rəyin predmet sahəsindən asılı olması (ing. sentiment is domain specific).* Rəy predmet sahəsindən asılıdır və sözlərin mənası istifadə olunan məzmunun asılı olaraq dəyişir. “Get kitab oxu” ifadəsi bir kitaba yazılmış resenziya üçün pozitivdir, ancaq filmə yazılmış reseziyada neqativdir, yəni əks nəticəyə malikdir.

*Cümlədəki çoxsaylı fikirlər (ing. multiple opinions in a sentence).* Bir cümlədə subyektiv və faktiki hissələrlə yanaşı çoxlu sayda fikirlər ola bilər. Belə hissələri ayırmaq faydalıdır. Məsələn, “Bu kameranın şəkil keyfiyyəti heyvətəmizdir və batareyanın ömrü də çoxdur, ancaq belə əla kameranın görüntü axtararı çox kiçikdir” cümləsində həm pozitiv və həm də neqativ fikirlər ifadə edilir.

*İnkarnın idarə edilməsi (ing. negation handling).* Sentiment analizdə inkarnın idarə edilməsi çətin ola bilər. Məsələn, “Bu kitab mənim xoşuma gəlir” və “Bu kitab mənim xoşuma gəlmir” bir-birindən yalnız bir əlamətə görə fərqlənir, buna görə fərqli və əks siniflərə aid edilməlidir. Əsas çətin problem digər dillərin istifadəsində, inkar ifadələri, məhsul xüsusiyyətləri və ya atributları, cümlə və ya sənədin mürəkkəbliyi, gizli məhsul xüsusiyyətləri və s. ilə əlaqəli rəylərin xülasəsinin hazırlanmasıdır.

*İstehza (ing. sarcasm).* Siyasi və onlayn müzakirələrdə tez-tez istehzalı və kinayəli cümlələrə rast gəlinir. Bu sözlər dildən asılı olaraq dəyişdiyi üçün onları müəyyən etməkdə çətinlik yaradır. Bu mövzuda çox az sayda araşdırmalar aparılmışdır.

*Gizli rəylər (ing. implicit opinion).* Mətnə olan rəyləri aşkar və gizli şəkildə xarakterizə etmək olar: “Biz vaxtımızı çox gözəl keçirdik” cümləsində fikir aşkar, “Batareyaya üç saat davam etdi” cümləsində isə gizli şəkildə ifadə edilmişdir. Mövcud sentiment analiz modelləri gizli şəkildə ifadə edilmiş rəyi neqativ rəy kimi aşkar edə bilmir.

*Müqayisəli cümlələr (ing. comparative sentences).* Müqayisəli cümlə birdən çox obyektin oxşarlıqlarına və ya fərqliliklərinə əsaslanan əlaqəni ifadə edir. Müqayisəli cümlələrin rəy bildirilən və ya bildirilməyən kimi klassifikasiya edilməsi ilə bağlı tədqiqatlar məhduddur. Eyni zamanda müqayisəli cümlələrdəki sözlərin sırası rəyin oriyentasiyasının müəyyənləşdirilməsində fərqliliyi açıq-aydın göstərir. Məsələn “X avtomobili Y avtomobilindən yaxşıdır” cümləsi “Y avtomobili X avtomobilindən daha yaxşıdır” cümləsinin tamamilə əksinə bir fikir bildirir.

*Çoxdilli sentiment analiz (ing. multilingual sentiment analysis).* Sentiment analiz üzrə tədqiqatların çoxu, əsasən lüğətlər və əl ilə işarələnmiş korpuslar kimi resursların mövcudluğuna görə ingilis dilindəki məlumatlara yönəldilmişdir. İnternet istifadəçilərinin böyük bir qismi ingilis dilində danışdığından, digər dillərdə sentiment analiz üçün resursların və alətlərin yaradılmasına ehtiyac vardır. [47]-də dillərarası layihələrdən istifadə edərək ingilis dilində mövcud olan resurs və alətlərdən yararlanmaq üçün bir neçə metod təklif edilmişdir.

*Küylü mətnlər (ing. noisy texts).* Orfoqrafik və qrammatik səhvlər, buraxılmış və ya problemlili durğu işarələri və jarqonlar sentiment analiz sistemlərinin çoxunda problem olaraq qalır.

*Rəy spamı (ing. opinion spam).* Rəy spamı, obyektləri təşviq etmək və / və ya nüfuzuna xələl gətirmək məqsədilə digər obyektlərə mənfi rəylər vermək üçün bəzi hədəf obyektlərə müsbət rəylər verməklə oxucuları və ya avtomatlaşdırılmış sistemləri bilərəkdən çaşdırmağa çalışan saxtalaşdırılmış və ya saxta rəylərə aiddir. İcmalların aqreqasiyasının bir çox saytları, rəyçiyə hər bir baxış üçün rəyin faydalılığını qiymətləndirməyə imkan yaratmaqla, həmin rəylərin əsasında spamları müəyyən etməyə çalışırlar.

## 8. Nəticə

Çoxsaylı tədqiqatlar göstərir ki, sentiment analiz sosial şəbəkələrdə, forumlarda, bloqlarda, mikrobloqlarda və s. mənbələrdə yayımlanan rəylərin analizinin ən perspektivli istiqamətlərindəndir. Aparılan araşdırmalardan görüldüyü kimi, rəylərin analizində sentiment

analizin bir-biri ilə sıx əlaqəli subyektivlik klassifikasiyası və sentiment klassifikasiya məsələlərindən istifadə olunur. Həmçinin, tədqiqatlar sentiment analizin sənəd, cümlə, söz, xüsusiyyət səviyyələrində həyata keçirilən bu məsələlərin həllində MT metodları, onlayn lüğətlər, sintaktik, statistik metodlar və dərin təlim yanaşmalarının geniş tətbiq olunduğunu göstərir. Sentiment analizin tətbiqi biznes, siyasət, dövlət kəşfiyyatı və s. sahələrdə mühüm rol oynaya bilər. Sonda onu da qeyd etmək olar ki, sentiment analizdə istifadə olunan metod və alqoritmlərin inkişaf etməsinə baxmayaraq bu sahədə açar sözlərin seçilməsi, rəyin predmet sahəsindən asılı olması, cümlədə bir neçə fikrin olması, müqayisəli cümlələr, gizli rəylər, coxdilli sentiment analiz, küylü mətnlər və rəy spamı kimi bir çox problemlər öz həllini tapmamışdır və tədqiqatçıların gələcək fəaliyyəti üçün açıq istiqamətlərdəndir.

### Ədəbiyyat

1. Liu. B. Sentiment analysis and opinion mining. Morgan & Claypool Publishers, 2012, p.168.
2. Kumar A., Sebastian T.M. Sentiment Analysis: A Perspective on its Past, Present and Future // International Journal of Intelligent Systems and Applications, 2012, vol.4, no.10, pp.1–14.
3. Nasukawa T., Yi J. Sentiment analysis: Capturing favorability using natural language processing / Proc. of the 2nd International Conference on Knowledge Capture, 2003, pp.70–77.
4. Dave K., Lawrence S., Pennock D.M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews / Proc. of the 12th international conference on World Wide Web, 2003, pp.519–528.
5. Medhat W., Hassan A., Korashy H. Sentiment analysis algorithms and applications: A survey // Ain Shams Engineering Journal, 2014, vol.5, no.4, pp.1093–1113.
6. Padmaja S., Fatima S. Opinion Mining and Sentiment Analysis –An Assessment of Peoples’ Belief: A Survey // International Journal of Ad hoc, Sensor & Ubiquitous Computing, 2013, vol.4, no.1, pp.21–33.
7. Mäntylä V., Graziotin D., Kuutila M. The Evolution of Sentiment Analysis - A Review of Research Topics, and Top Cited Papers // Computer Science Review, 2018, vol.27, pp.16–32.
8. Vinodhini G., Chandrasekaran RM. Sentiment Analysis and Opinion Mining: A Survey // International Journal of Advanced Research in Computer Science and Software Engineering, 2012, vol.2, no.6, pp.282–293.
9. Tang H., Tan S., Cheng X. A survey on sentiment detection of reviews // International Journal Expert Systems with Applications, 2009, vol.36, no.7, pp.10760–10773.
10. Chau M., Xu C. Mining communities and their relationships in blogs: A study of online hate groups // International Journal Human-Computer Studies, 2007, vol.65, no.1, pp.57–70.
11. Mishne G., Glance N. Predicting movie sales from blogger sentiment / AAAI Symposium on Computational Approaches to Analyzing Weblogs (AAAI-CAAW), 2006, pp.155–158.
12. Liu Y., Huang J., An A., Yu X. ARSA: A sentiment-aware model for predicting sales performance using blogs / Proc. of the 30th Annual International ACM Conference on Research and Development in Information Retrieval, 2007, pp.607–614.
13. Melville P., Gryc W., Lawrence R.D. Sentiment analysis of blogs by combining lexical knowledge with text classification / Proc. of the 15th ACM SIGKDD International conference on Knowledge discovery and data mining, 2009, pp.1275–1284.
14. Arya P., Bhagat A. Deep Survey on Sentiment Analysis and Opinion Mining on Social Networking Sites and E-Commerce Website // International Journal of Engineering Science and Computing, 2017, vol.7, no.3, pp.4796–4810.
15. Hatzivassiloglou V., Wiebe J. Effects of adjective orientation and gradability on sentence subjectivity / Proc. of the International Conference on Computational Linguistics (COLING), 2000, vol.1 pp.299–305.
16. Riloff E., Wiebe J. Learning extraction patterns for subjective expressions / Proc. of the Conference on Empirical Methods in Natural Language Processing, 2003, pp.105–112.

17. Pang B., Lee L., Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques / Proc. of the Conference on Empirical Methods in Natural Language Processing, 2002, pp.79–86.
18. Pang B., Lee L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales / Proc. of the Association for Computational Linguistics, 2005, pp. 115–124.
19. Yu H., Hatzivassiloglou V. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences / Proc. of the Conference on Empirical Methods in Natural Language Processing, 2003, pp.129–136.
20. Turney P. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews / Proc. of the Association for Computational Linguistics, 2005, pp. 417–424.
21. Liu B., Hu M., Cheng J. Opinion observer: Analyzing and comparing opinions on the web / Proc. of the 14th International world wide web conference, 2005. ACM Press, pp.10–14.
22. Wilson T., Wiebe J., Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis / Proc. of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, 2005, pp.347–354.
23. Esuli A., Sebastiani F. Determining the semantic orientation of terms through gloss classification / Proc. of the 14th ACM International Conference on Information and Knowledge Management, 2005, pp.617–624.
24. Aue A., Gamon M. Customizing sentiment classifiers to new domains: A case study / Proc. of the International Conference on Recent Advances in Natural Language Processing, 2005.
25. Kim S., Hovy E. Determining the sentiment of opinions / Proc. of the 20th International Conference on Computational Linguistics, 2004, pp.1367–1373.
26. Kamps J., Marx M., Mokken R.J., de Rijke M. Using WordNet to measure semantic orientation of adjectives / Proc. of the 4th International Conference on Language Resources and Evaluation, 2004, pp.1115–1118.
27. Guang Q., Xiaofei H., Feng Z., Yuan Sh., Jiajun B., ChunC. DASA: dissatisfaction-oriented advertising based on sentiment analysis // Expert System with Applications, 2010, vol.37, pp. 6182–6191.
28. Hatzivassiloglou V., McKeown K. Predicting the semantic orientation of adjectives / Proc. of the Joint ACL/EACL Conference, 2004, pp.174–181.
29. Yi J., Nasukawa T., Niblack W., Bunescu R. Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques / Proc. of the 3rd IEEE international conference on data mining (ICDM 2003), pp.427–434.
30. Hu M., Liu B. Mining opinion features in customer reviews / Proc. of AAAI, 2004, pp.755–760.
31. Popescu A-M., Etzioni O. Extracting product features and opinions from reviews / Proc. of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, 2005, pp.339–346.
32. Wang X, Jiang W, Luo Z. Combination of convolutional and recurrent neural network for sentiment analysis of short texts / Proc. of the International Conference on Computational Linguistics, 2016, pp.2428–2437.
33. Santos C. N., Gatti M. Deep convolutional neural networks for sentiment analysis for short texts / Proc. of the International Conference on Computational Linguistics, 2014, pp.69–78.
34. Dou Z.Y. Capturing user and product information for document level sentiment analysis with deep memory network / Proc. of the Conference on Empirical Methods on Natural Language Processing, 2017, pp.521–526.
35. Moraes R., Valiati J.F., Neto W.P. Document-level sentiment classification: an empirical comparison between SVM and ANN // Expert Systems with Applications, 2013, vol.40, no.2, pp.621–633.

36. Tang D., Qin B., Liu T. Document modelling with gated recurrent neural network for sentiment classification / Proc. of the Conference on Empirical Methods in Natural Language Processing, 2015, pp.1422–1432.
37. Xu J., Chen D., Qiu X., and Huang X. Cached long short-term memory neural networks for document-level sentiment classification / Proc. of the Conference on Empirical Methods in Natural Language Processing, 2016, pp.1660–1669.
38. Kaji N., Kitsuregawa M. Building lexicon for sentiment analysis from massive collection of html documents / Proc. of the Conference on Empirical Methods in Natural Language Processing, 2007.
39. Wilson T., Wiebe J., Hoffmann P. Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis // Computational Linguistics, 2009, vol.35 no.3, pp.399–433.
40. O'Connor B., Balasubramanyan R., Routledge B.R., Smith N.A. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series / Proc. of the 4th International Conference on Weblogs and Social Media, 2010, pp.122–129.
41. Thomas M., Pang B., Lee L. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts / Proc. of the Conference on Empirical Methods in Natural Language Processing, 2006, pp.327–335.
42. Mullen T., Malouf R. Taking sides: User classification for informal online political discourse // Internet Research, 2008, vol.18, no.2, pp.177–190.
43. Taboada M., Gillies M.A., McFetridge P. Sentiment classification techniques for tracking literary reputation / In LREC Workshop: Towards Computational Models of Literary Analysis, 2006, pp.36–43.
44. Kumar A., Ahmad N. ComEx Miner: Expert Mining in Virtual Communities // International Journal of Advanced Computer Science and Applications, 2012, vol.3, no.6.
45. Seki Y., Eguchi K., Kando N., Aono M. Multi-document summarization with subjectivity analysis / Proc. of the Document Understanding Conference, 2005.
46. Spertus E. Smokey: Automatic recognition of hostile message / Proceedings of Innovative Applications of Artificial Intelligence Conference (IAAI), 1997, pp.1058–1065.
47. Denecke K. Using SentiWordNet for Multilingual Sentiment Analysis / Proc. of the IEEE 24th International Conference on Data Engineering Workshop, 2008, IEEE Press, pp.507–512.

#### УДК 004.04

**Гаджирогимова Макруфа Ш.<sup>1</sup>, Исмайылова Марзия И.<sup>2</sup>**

Институт Информационных Технологий НАНА, Баку, Азербайджан

<sup>1</sup>[makrufa@science.az](mailto:makrufa@science.az), <sup>2</sup>[imarziya@gmail.com](mailto:imarziya@gmail.com)

#### **Сентимент анализ: проблемы и решения**

Сентимент анализ или интеллектуальный анализ мнений – это область исследований, которая анализирует мысли, чувства, эмоции, оценки и отношения людей, изучает такие объекты, как продукты, услуги, организации, люди, проблемы, события, темы и их характеристики. Развитие анализа настроений связано с появлением большого количества цифровых данных, генерируемых в социальных сетях, на форумах, в блогах, микроблогах и т.д. В последние годы сентимент анализ стал предметом обширных исследований в таких областях, как обработка естественного языка, интеллектуальный анализ данных, анализ текста и поиск информации. Сентимент анализ широко распространен в областях маркетинга, финансов, политологии, общественных наук, медицинских наук и т.д. Такая популярность объясняется тем, что мнения имеют решающее значение практически во всех областях человеческой деятельности и являются основным фактором, влияющим на поведение людей. Неспособность легко анализировать сотни тысяч мнений, опубликованных в социальных сетях, блогах, форумах и других источниках мнений, обусловила необходимость использования компьютерной лингвистики, сентимент анализа

или анализа мнений. В статье рассматриваются исторические и терминологические аспекты сентимент анализа, предоставляется информация об источниках информации и областях применения. Основные задачи и уровни сентимент анализа также изучаются в данной работе. Существующие проблемы и решения были проанализированы.

**Ключевые слова:** сентимент анализ, анализ мнений, классификация настроений, уровни сентимент анализа, machine learning.

**Makrufa Sh. Hajirahimova<sup>1</sup>, Marziya I. Ismayilova<sup>2</sup>**

Institute of Information Technology of ANAS, Baku, Azerbaijan

<sup>1</sup>[makrufa@science.az](mailto:makrufa@science.az), <sup>2</sup>[imarziya@gmail.com](mailto:imarziya@gmail.com)

### **Sentiment analysis: problems and solutions**

Sentiment analysis or opinion mining is a research area that analyses the thoughts, feelings, emotions, assessments and attitudes of people who study the objects such as products, services, organizations, people, problems, events, topics and their characteristics. The development of sentiment analysis is associated with the emergence of a large amount of digital data generated in social networks, forums, blogs, microblogs, etc. In recent years, sentiment analysis has become the subject of extensive research in the areas, such as natural language processing, data mining, analysis of texts and information retrieval. An analysis of sentiments is widely applied in the fields of marketing, finance, political science, social sciences, medical sciences, etc. This popularity is explained by the fact that opinions are crucial in almost all areas of human activity and are the main factor influencing people's behavior. The inability to easily analyze hundreds of thousands of opinions published on social networks, blogs, forums, and other sources of opinions has necessitated the use of computer linguistics, sentiment analysis, or opinion analysis systems. The article discusses the historical and terminological aspects of sentiment analysis, provides information on information sources and areas of application. The main tasks and levels of sentiment analysis are also studied in this paper. Existing problems and solutions are analyzed.

**Keywords:** sentiment analysis, opinion mining, sentiment analysis problems, sentiment classification, sentiment analysis levels, machine learning.