

Sentiment analiz texnologiyaları və Azərbaycan dili

Günay İskəndərli

İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

gunayniftali@gmail.com

Xülasə — Sentiment analiz mətnin sentiment polyarlığının çıxarılması məqsədi daşıyır. Müxtəlif yanaşmalarda sözlərin sentiment polyarlığını müəyyən etmək üçün müxtəlif resurslardan istifadə olunur. İngilis dili bu cür resurslara sahib olmaq baxımından ən zəngin dil olsa da, Azərbaycan dili də daxil olmaqla, digər dillərin əksəriyyətində bu cür resurslar yoxdur. Bunu nəzərə alaraq, məqalədə sentiment analizin ingilis dilindən fərqli dillərdə mövcud problemləri araşdırılmış və Azərbaycan dilində mövcud vəziyyət nəzərdən keçirilərək bu istiqamətdə görüləcək işlər müəyyən olunmuşdur.

Açar sözlər — *sentiment analiz; dil; sözlər şəbəkəsi.*

I. GİRİŞ

Bu gün internet şəbəkələrindən və sosial mediadan istifadə insanlara digər insanlar və ya qurumlar haqqında fikirlərini asanlıqla və aydın şəkildə ifadə etmək və şərh etmək imkanı verir. Bu isə öz növbəsində internetdə məlumatların kəskin artmasına yol açır. Bunun nəticəsidir ki, internetdən əldə edilən böyük verilənlərin təhlili, onların faydalı informasiyaya çevrilməsi və istifadəsi son illərdə intensiv şəkildə araşdırılan sahələrdən biridir. Sentiment analiz isə böyük verilənlərdə subyektiv məlumatların sistemli təhlilinə imkan verən təbii dilin emalı üsuludur. Mətnin

öyrənilməsi və kompüter linqvistikası alətlərindən istifadə edərək, tədqiqatçılara mətndəki emosiyaları müəyyən etməyə, onları emosiya kateqoriyalarına görə təsnif etməyə və kateqoriyaların emosional qütblərini tapmağa kömək edir.

Sentiment analiz mətnin sistemlik şəkildə tədqiqi, onun emosiya kateqoriyasının və qütblüyünün müəyyən edilməsi prosesidir. O cümlədən sentiment analiz vətəndaşların müəyyən bir mövzuya münasibətinin müsbət, mənfi və ya neytral olduğunu müəyyən etmək üçün mətnə ifadə olunan fikirlərin hesablamaya yolu ilə müəyyən edilməsi və kateqoriyalara bölünməsi prosesidir [1].

Sentiment analiz dilçilik, maliyyə bazarları, marketing və sosial media təhlili kimi bir çox sahələrdə tez-tez istifadə olunur. Sentiment analiz üzrə aparılan tədqiqatlar populyar olsa da, hazırlanmış alqoritmlər və kitabxanalar nəzərə alındıqda, ingilis dili üzrə tədqiqatlar ön plandadır və müxtəlif dillərin təhlilinə imkan verən metod və kitabxanaların daha da inkişaf etdirilməsinə ehtiyac var. İngilis dili dünyada danışılan ümumi dil olduğundan, ədəbiyyatda ingiliscə mətnlər üzərində sentiment analiz üzrə çoxlu tədqiqatlar mövcuddur [2]. Bununla belə, digər dillərdə olan mətnlər üçün də sentiment analiz üzrə bir sıra işlər

görülmüşdür. Lakin Azərbaycan dilində olan mətnlərdə sentiment analiz hələ də açıq bir araşdırma sahəsi olaraq qalır.

Bu tədqiqatda deyilənləri nəzərə alaraq, sentiment analiz texnologiyaları nəzərdən keçirilmiş, digər dillərdə sentiment analiz üzrə mövcud problemlər araşdırılmış və Azərbaycan dili üzrə bu texnologiyaların tətbiqində qarşıya çıxan problemlər müəyyən edilmişdir.

II. SENTİMENT ANALİZ VƏ TƏTBİQ SAHƏLƏRİ

Sentiment analiz rəqəmsal məlumatların bol olduğu bu dövrdə biznes, hökumət və istehlakçı baxımından bir çox fayda və imkanlara malikdir. Bundan əlavə, Blokçeyn, Əşyaların İnterneti, Bulud Hesablama və Big Data kimi texnoloji yeniliklər sentiment analizin tətbiq dairəsini genişləndirərək, ondan praktiki olaraq istənilən sahədə istifadə etməyə imkan verir.

Sentiment analiz sosial mediada biznes, tibb, əyləncə və s. kimi sahələr üzrə yazılan rəylərin analizində uğurla tətbiq olunur. Aşağıda bu sahələrdə sentiment analizin tətbiqi imkanları nəzərdən keçirilmişdir [3]:

Biznes sahəsi: İstehlakçı firmalar məhsullarını təkmilləşdirmək, müştərilərin rəylərini araşdırmaq və innovativ marketing strategiyası hazırlamaq üçün sentiment analiz texnologiyalarından istifadə edə bilərlər.

E-ticarət inkişaf etdikcə satılan məhsulların sayı və müştərilərin rəyləri də artır. Sentiment analiz müştərilərə daha yaxşı məhsullar seçməyə kömək edə bilər. Sentiment analizin məqsədi hansı rəqiblərin

meydana çıxdığını və marketing kampaniyalarının necə müqayisə olunduğunu müəyyən etməkdir. Bir markanın və onun rəqiblərinin istehlakçı bazası haqqında tam təsəvvür əldə etmək üçün istifadə edilə bilər. Sentiment analiz bir neçə platformadan (*Twitter, Facebook, bloqlar*) məlumat toplayaraq nəzərə çarpacaq nəticələr verə və biznes analizindəki çətinlikləri aradan qaldıra bilər.

Səhiyyə və tibb sahəsi son dövrlərdə sentiment analizin istifadə olunduğu sahələrdən biridir. Məlumatlar sorğular, *Twitter*, bloqlar, xəbər məqalələri, rəylər və s. kimi müxtəlif mənbələrdən toplana bilər. Bu məlumatlar daha sonra müxtəlif məqsədlər üçün təhlil edilə bilər. Məsələn, sentiment analiz standartların qiymətləndirilməsi və tibb sahəsindəki yeniliklərin təhlili məqsədlə istifadə oluna bilər. Sentiment analiz səhiyyə xidmətlərini yaxşılaşdırmaq üçün xəstələrin əhval-ruhiyyəsini, epidemiyaları, dərmanların mənfi reaksiyalarını və xəstəlikləri toplamaqda və qiymətləndirməkdə kömək edə bilər.

Əyləncə sahəsi: Sentiment analiz əyləncə sahəsində geniş istifadə olunur. Tamaşaçının reaksiyasını müəyyən etmək üçün film, şou və qısa filmlərin rəyləri təhlil edilə bilər. Bu, nəinki izləyicilərin daha yaxşı seçim etməsinə, həm də yaxşı məzmunun populyarlaşmasına kömək edə bilər.

Otellər və restoranlara dair rəylərin sentiment analizi müştərilərə daha yaxşı seçim etməyə, həmçinin sahibkarlara işlərini təkmilləşdirməsinə kömək edə bilər. Sentiment analiz vasitəsilə xidmət

təminatçıları daha çox mənfəət əldə edə bilirlər. Belə ki, onlar ən çox mənfə rəy alan xidmətləri aşkarlaya və onları təkmilləşdirə bilər.

Fond bazarı: Sentiment analizinin tətbiqlərindən biri səhm qiymətlərinin proqnozlaşdırılmasıdır. Bu, birja ilə bağlı bütün xəbərləri təhlil edərək və birja qiymətlərinin trendlərini proqnozlaşdırmaqla edilə bilər. Məlumatlar *Twitter*, xəbər məqalələri, bloqlar və s. kimi müxtəlif mənbələrdən toplanır. [4]-də aparılan tədqiqatda sentiment analiz fond bazarında tendensiyanın yüksələcəyini və ya azalacağını müəyyən etmək üçün istifadə olunur. Müsbət xəbərlər yüksəliş trendinə, mənfə xəbərlər isə aşağıya doğru meyl kimi qiymətləndirilir.

Siyasət: Dövlət qurumları sentiment analizdən istifadə edərək vətəndaşların ən çox hansı problemlərdən əziyyət çəkdiyini, hansı xidmətlərdən razı və narazı olduğunu müəyyən edə bilirlər. Sentiment analiz tendensiyaları, təkrarlanan problemləri və narahatlıqları aşkar etmək üçün məlumatları təsnif və təşkil etməyə imkan verə bilər.

Göründüyü kimi, sentiment analizinin tətbiq sahəsi olduqca genişdir və bu sahələrin hər birində həllini gözləyən bir sıra məsələlər mövcuddur.

III. SENTİMENT ANALİZİN ÜSTÜNLÜKLƏRİ VƏ ÇATIŞMAZLIQLARI

Opinion mining kimi də tanınan sentiment analiz, bir sıra sözlərin arxasındakı emosional çaların müəyyən edilməsi prosesidir və son illərdə

tədqiqatçıların daha çox diqqət mərkəzindədir. Bu sahədə həll edilməli olan bir-biri ilə əlaqəli bir çox altproblemlər var ki, bu da bu sahəni daha da çətinləşdirir. Bununla bağlı sentiment analizinin bir sıra üstünlükləri və çatışmazlıqları mövcuddur. Sentiment analizinin əsas üstünlükləri bunlardır:

Miqyaslılıq (genişlənmə qabiliyyəti) – sentiment analiz məlumatları səmərəli və sərfəli şəkildə emal etməyə imkan verir.

Real vaxt analizi – sentiment analiz vəziyyətləri dərhal müəyyən etməyə və hərəkətə keçməyə kömək edir.

Ardıcıl kriteriyalar – mərkəzləşdirilmiş sentiment analiz sistemindən istifadə edərək, şirkətlər bütün məlumatlarına eyni meyarları tətbiq edə bilirlər. Bu, səhvləri azaltmağa və məlumatların ardıcılığını yaxşılaşdırmağa kömək edə bilər.

Sentiment analizinin əsas çatışmazlıqları bunlardır:

- Mətnə istehzanın müəyyən olunması;
- İnkarnın idarə edilməsi;
- Cümlədə çoxsaylı fikirlərin olması;
- Çoxdilli məlumatların mövcudluğu;
- Audio-video məlumatların olması və s.

Bu problemlərin aradan qaldırılması üçün tədqiqatçılar tərəfindən yeni metod və alqoritmlər təklif olunmasına baxmayaraq, hələ də həllini gözləyən problemlər qalmaqdadır.

IV. SENTİMENT ANALİZ VƏ DİLLƏR

Yuxarıda qeyd etdiyimiz kimi, sentiment analizin problemlərindən biri də dildir. İngilis dili, şübhəsiz ki, sentiment analizin əsas tədqiqat obyektidir. Bununla belə, ingilis dili fikirlərin onlayn ifadə olduğu yeganə dil deyil. Belə ki, internet məlumatlarının əksəriyyəti ingilis dilindən fərqli dillərdədir. Lakin zaman keçdikcə internetdə daha çox mənbə əldə olunmasına baxmayaraq, mənbələrin əksəriyyəti ilk növbədə ingilis dili üçün mövcuddur və bu isə müxtəlif dillərdə sentiment analizin tətbiqini çətinləşdirir. Bu isə tədqiqatçılar qarşısında belə bir mühüm sual çıxarır: niyə fərqli dillərdə sentiment analizin tətbiqi çətinləşir? Bunun bir neçə səbəbi var. Bunlardan bir neçəsi aşağıda nəzərdən keçirilmişdir [5]:

A. Fərqli kodlaşdırma ilə ilkin emal

Mətnin ilkin emalı istənilən təbii dilin emalı prosesinin tətbiqi üçün vacib bir addımdır. Kompüterlərin mətni emal etməsi üçün bugünkü sənaye standartı *Unicode*-dir. Qeyri-latin əlifbalarından və ya köhnə mətn məlumat mənbələrindən istifadə edən dillər kiril əlifbası üçün istifadə edilən KOI8 kimi fərqli kodlaşdırma üslubuna malik ola bilər. Bu isə həmin dillər üçün təbii dilin emalı prosesini ən başlanğıcda dayandırır.

B. Ümumişlək sözlər və mətn klassifikatoru

Təbii dilin emalı paketləri mənası olmayan sözləri (ümumişlək sözlər) təmizləmək üçün daxili söz siyahılarından kömək alır. Məsələn, ingilis dilindəki mətni emal üçün mənalı sözlərlə məhdudlaşdırmaq üçün mətndən “the, a, an və s.” kimi sözləri çıxarmaq lazımdır.

Ümumişlək sözlər müxtəlif dillərdə mövcuddur, lakin onların kompüterdə emal olunması üçün əl ilə işlənməlidir.

C. Dillər arasında struktur fərqləri

Sentiment analizin əsas addımlarından biri mətnin kiçik mənalı vahidlərə ayrılması prosesidir. Stanford Universitetində bir araşdırma qrupu tərəfindən müəyyən olunmuşdur ki, bu proses bir çox dillər üçün fərqlidir. Çünki müxtəlif dillərin qrammatik quruluşları fərqlidir.

Müxtəlif dillərdə sentiment analizin tətbiqi üçün əsas məsələlərdən biri sözlər şəbəkəsinin (*WordNet*) olmasıdır.

Hazırda 200-dən çox dil üçün nəzərdə tutulmuş “*WordNet*” şəbəkəsi mövcuddur. Qlobal *WordNet* Assosiasiyası dünyadakı bütün dillərə görə söz şəbəkələrini müzakirə etmək, paylaşmaq və birləşdirmək üçün platforma təmin edən açıq, ictimai və qeyri-kommersiya təşkilatıdır [6]. Bu platformadan istifadə edərək mövcud *WordNet* şəbəkələri ilə tanış olmaq olar.

Bir çox dillər üçün ingilis dilində yaradılmış *WordNet*-ə analoji olaraq belə söz şəbəkələrinin yaradılmasına baxmayaraq, bu şəbəkələr hələ də istənilən səviyyədə deyil. Belə ki, bu yaradılmış şəbəkələrin bir qismi açıq sistem olsa da, bir qismi açıq deyil. O cümlədən bu şəbəkələrdə semantik şəbəkələr istənilən səviyyədə deyil. Məsələn, ingilis dilində *WordNet*-də təxminən 117.000 semantik şəbəkə (synset) olmasına baxmayaraq, türkcə yaradılmış *WordNet*-də çox az – təxminən 15000 semantik şəbəkə (synset) var [7].

Bütün bunlar göstərir ki, müxtəlif dillərdə sentiment analizin tətbiq olunması üçün tədqiqatların daha da genişləndirilməsinə ehtiyac var.

V. AZƏRBAYCAN DİLİ VƏ ONUN SENTİMENT ANALİZ PROBLEMLƏRİ

Azərbaycan dili Altay dillərinin türk ailəsinin üzvüdür. Azərbaycan dilinin özünəməxsus xüsusiyyətləri bu dil üçün təbii dil emalını və sentiment analizin tətbiqini çətinləşdirir. Morfoloji cəhətdən Azərbaycan dili aqlütinativ bir dildir. Sözlər kökə çoxlu şəkilçilərin birləşməsi yolu ilə əmələ gəlir. Azərbaycan dili üçün sentiment analiz texnologiyalarının tətbiqinin əsas çətinlikləri bunlardır:

A. Azərbaycan dilinin morfoloqiyası

Azərbaycan dili aqlütinativ bir dildir. Bu törəmə və fleksiya şəkilçiləri sözün nitq hissəsini və semantik oriyentasiyasını dəyişə bilər (məsələn, “bəyəndim” (I liked it), “bəyənmədim” (I didn’t like it)).

B. Azərbaycan əlifbası

Azərbaycan dilində ingilis əlifbasında olmayan bir neçə simvol var: “ç”, “ğ”, “ı”, “ö”, “s”, “ü”, “ə”. Qeyri-rəsmi yazıda insanlar bu hərflərin bəziləri üçün ən yaxın ASCII simvollarını əvəz etməyə meyllidirlər (məsələn, “ç” “c” kimi yazılır). Bu sözlər kontekstdə insanlar tərəfindən oxuna bilsələr də, cümlədəki sözlərin dəqiq müəyyən edilməsində çətinliklər yaranır.

C. İnkarnın mürəkkəbliyi

Azərbaycan dilində bir sözü inkar etməyin bir neçə yolu var ki, bu da öz

növbaşində ifadə olunan hissi dəyişir. Bunlar aşağıdakılardır:

- ma/mə şəkilçiləri feli inkar edir (məsələn, “bəyənmədim” (I did not like it)).
- siz/sız/suz/süz şəkilçiləri, məsələn, “uğursuz” (without success – unsuccessful).
- “deyil” sözü (is/are not) , məsələn, “gözəl deyil” (is not beautiful).
- “Yox” sözü yoxluğu və ya əlçatmazlığı göstərir. (məsələn, “mövzu yoxdur” (does not have a topic))

NƏTİCƏ

Sentiment analiz mətnin verilmiş seqmentinin sentiment polyarlığını aşkarlamaq məqsədi daşıyır. Sentiment analizi ingilis dilindən fərqli dillərdə tətbiq etməyin əsas çətinliklərindən biri məhdud polyarlıq resurslarının mövcudluğudur. Belə ki, ingilis dili bu cür resurslara sahib olmaq baxımından ən zəngin dil olsa da, Azərbaycan dili də daxil olmaqla, digər dillərin əksəriyyətində bu cür resurslar yoxdur.

Yuxarıda qeyd etdiyimiz kimi, sentiment analizi Azərbaycan dilinə tətbiq etməyin əsas şərtlərindən biri sentiment analiz üzrə resursların mövcud olmasıdır. Bunun üçün isə ingilis dilində yaradılmış *WordNet* şəbəkəsinə analoji olaraq sözlər şəbəkəsi və SentiAzerNet şəbəkəsi yaradılmalıdır. Bu məqsədlə informasiya texnologiyaları və dilçilik üzrə tədqiqatçılar birgə araşdırma aparmalı və əməkdaşlıq etməlidirlər.

İSTİNADLAR

- [1] R.Dehkharghani, Y.Saygin, B.Yanikoglu and K.Oflazer, “SentiTurkNet: a Turkish polarity lexicon for sentiment analysis,” Language Resources & Evaluation, Vol. 50, No.3, pp.667-685, 2016.
- [2] S.Deshmukh, N.Patil, S.Rotiwari and J.Nunes, “SENTIMENT ANALYSIS OF MARATHI LANGUAGE”, International Journal of Research Publications in Engineering and Technology, Vol.3, No.6, pp.93-97, 2017.
- [3] M.Wankhade, A.Ch.S.Rao and Ch.Kulkarni, “A survey on sentiment analysis methods, applications, and challenges”, Artificial Intelligence Review, Vol.55, No.7, pp.5731-5780, 2022.
- [4] F.Z. Xing, E. Cambria, R.Welsch, “Natural Language based financial forecasting: a survey”, Artificial Intelligence Review, Vol. 50, No.1, pp. 49–73, 2018
- [5] C.Dilmegani, Challenges and methods for Multilingual sentiment analysis in 2023. <https://research.aimultiple.com/multilingual-sentiment-analysis/>
- [6] <http://globalwordnet.org/>
- [7] G.Geziç, B.A.Yanikoglu, “Sentiment Analysis in Turkish”, Turkish Natural Language Processing, pp.1-19, 2018.