

ӘОК 004.85

МАШИНАЛЫҚ ОҚЫТУ НЕГІЗІНДЕ МЕЙРАМХАНАЛАР ТУРАЛЫ ПІКІРЛЕРДІ АСПЕКТІЛЕРГЕ ЖІКТЕУ

Оралбекова Іңкәр Тасқынқызы

iinkar9822@gmail.com

Л.Н. Гумилев атындағы Еуразия ұлттық университеті,
«Жасанды интеллект технологиялары» кафедрасының 2 курс магистранты
Ғылыми жетекшісі – PhD, доцент м.а. Ергеш Б.Ж.

Аңдатпа. Мақалада машиналық оқыту негізінде мейрамхана туралы пікірлерді аспектілерге бөлу мәселесін шешу қарастырылады. Аспект негізіндегі көңіл-күйді талдау (ABSA) - деректерді аспектілерге жіктейтін және олардың әрқайсысына қатысты көңіл-күйді анықтайтын мәтінді талдау әдісі. Интернеттегі мейрамханалар туралы пайдаланушылардың пікірлері мен мейрамханалады таңдау кезінде маңызды ақпарат көзі болып табылады. Сондықтан бұл пікірлер туралы ақпарат сапаны бақылау және мейрамхананы басқару үшін маңызды. Бұл зерттеу мейрамхана туралы пікірлерді талдайды және рейтингтер жіберіп алуы мүмкін ақпаратты береді. Пікірлер веб-сайттан сканерленеді және жалпы аспектілерге сәйкес алдын ала анықталған сыныптарға жіктеледі.

Кілттік сөздер: машиналық оқыту, Aspect Based Sentiment Analysis, сентиментті талдау классификаторлары.

Кіріспе

Мейрамхана индустриясы жаһандануға байланысты бәсекеге қабілетті болды. Жаңа технологиялар, соның ішінде үлкен деректер, жасанды интеллект және машиналық оқыту пайда болуда. Бұл жаңа технологиялар адамның күрделі мәселелерін сәтті шешуде. Кейбір

елдер осы жаңа технологияларды енгізген кезде, басқалары оларды өз салаларында сапалы қызмет көрсету және кірісті ұлғайту үшін пайдалануға тырысады.

Негізінен, сайттар мейрамханаларды алдыңғы қонақтардың пікірлері бойынша ұсынатын "рейтингтерге" сәйкес орналастырады. Кейбір веб-сайттар кез-келген аспектке негізделген рейтингтерді ұсына алады, бірақ зерттеулер клиенттердің қажеттіліктерін дұрыс қанағаттандыру үшін қызмет жеткізушілерімен және басқа клиенттермен өзара әрекеттесу қажет екенін көрсетеді[1]. Бұл рейтингтердің сенімділігі де күмәнді [2]. Ол үшін веб сайтта аталған қолданушылардың белгілі бір мейрамханаға барған кезде басқа тұтынушылардың не көргені туралы жақсы түсінік беретін пікірлері керек.

Тұтынушы пікірлерде көрсетілген жеке ақпаратқа сезімтал және көбінесе оны шешім қабылдау үшін негіз ретінде пайдаланады. Дәстүрлі шолулар жиынтығы пайдалы болғанымен, клиенттердің мүдделері шолуларда берілген ақпараттан тыс болған кезде сәтсіздікке ұшырайды[3]. Мұндай зерттеулерде көптеген қиындықтар бар. Деректерді талдауды қиындататын бірқатар факторлар бар [4]. Мысалы, бір пайдаланушының талаптары басқалардың талаптарынан айтарлықтай өзгеше болуы мүмкін; олардың бірі іскерлік сапар, екіншісі демалыста болуы мүмкін. Сонымен қатар, әр рецензенттің жазу стилі басқасынан өзгеше. Рецензенттің әртүрлі аспектілері олардың өмір сүру деңгейі, туған жері, өмір салты, бюджет және т.б. сияқты шешім қабылдауға әсер етеді. Клиенттердің көпшілігі алғашқы елу немесе ең танымал пікірлерді оқиды, олар басқалар сипаттаған кейбір мүмкіндіктерге ие болмауы мүмкін.

Классификаторлар

Белгілі болғандай, мәтіннің реңкін талдау үшін көбінесе Naive Bayes классификаторы және SVM әдісі қолданылады. Бұл ретте мәселенің типтік тұжырымы – фильмдер немесе тауарлар туралы, саяси тұлғалар немесе оқиға туралы (блогта, твиттерде және т.б.) пікірлерді оң және теріс болып жіктеу қажет. Naive Bayes моделі әдетте негізгі, ең қарапайым үлгі ретінде пайдаланылады, ал SVM әдісі күрделірек, өйткені ол тиімдірек деп саналады. Дегенмен, Бермингем мен Смитон [5] және Ванг пен Мэннинг [6] Naive Bayes классификаторы твиттер сияқты қысқа мәтіндерде SVM-ге қарағанда жақсырақ жұмыс істейтінін көрсетті. Сонымен қатар, кейбір жұмыстарда сөздердің векторлық көрінісі бар нейрондық желілерді белгілер ретінде қолдануға негізделген тәсіл ұсынылады. Бұл тәсіл тілге тәуелді емес және семантикалық тезаурусты қажет етпейді (тек мәтіндерді оқыту корпусы қолданылады). Біздің жұмысымыздың бір бөлігі ретінде Naive Bayes классификаторы, логистикалық регрессия сияқты модельдермен эксперименттер жүргізілді.

Пікірлер корпусы

Пікірлер (www.TripAdvisor.com) және (astana.restoran.kz) сайттарынан алынды.

Пікірлердің ұзақтығы 1-ден 20 сөйлемге дейін өзгереді, орташа есеппен 10 сөйлемді құрайды. Корпустың шағын бөлігі кейінгі машиналық оқыту үшін белгіленеді (Сурет 2). Бұл корпуста Нұр-Сұлтан қаласындағы 5 мейрамхананың 250 пікірі кіреді.

D12				negative
	A	B	C	D
1	Comment	Aspect category	Aspect term	Polarity
2	Барлығы супер	General	барлығы супер	positive
3	тамаша сервис	Service	тамаша сервис	positive
4	алуан түрлі ас мәзірі	Food	алуан түрлі	positive
5	Өткен күні мен досыммен бірге болдым, пицца жедім, өте дәмді қытырлақ камыр	Food	өте дәмді	positive
6	жартылай тәтті ақ шарап іштім, даяшы бізге кеңес берді, біз өте риза болдық, рахмет	Food	өте риза	positive
7	Дәмді асхана	Food	дәмді	positive
8	Тамаша қызмет!	Service	тамаша	positive
9	орналасуы жақсы!	Location	жақсы	positive
10	Олар сондай-ақ abt + жүйесін/ смартфондағы қосымшаны пайдаланатыны ыңғайлы	General	ыңғайлы	positive
11	Қызмет өте жақсы.	Service	өте жақсы	positive
12	Жалғыз кемшілік егер жылытқыштың жанына отырсаң / пісіп кетесің	General	кемшілік	negative
13	Ыңғайлы және тыныш жер	Ambiance	ыңғайлы	positive
14	Орналасқан жері керемет!	Location	керемет	positive
15	Жақсы баға, тамаша тағам!	Food	тамаша	positive
16	Орташа баға санатынан сәл жоғары жақсы мекеме	General	жақсы	positive
17	Жақсы/ сыпайы қызмет. Олар өте тырысалды)))	Service	сыпайы	positive
18	Жайлы шағын кафе	Ambiance	жайлы шағын	positive
19	Жайлы шағын кафе.Балалармен демалу үшін қолайлы-балалар алаңы бар.	Ambiance	қолайлы	positive
20	Ас үй жақсы / мен макаронға тапсырыс беруге кеңес беремін. Жақсы лимонадтар.	Food	жақсы	positive
21	Жақсы мейрамхана, интерьер әдемі	Ambiance	әдемі	positive
22	Біз мұнда пиццаға тапсырыс бердік, жалпы ұнады	Food	ұнады	positive

Сурет 2. Пікірлер корпусы

Мейрамхананың аспектілері

Бұл жұмыста объектілердің аспектілерінің тізімі (біздің жағдайда мейрамханалар) пікірден автоматты түрде алынбайды, бірақ көңіл-күйді талдау модулі болатын жүйені пайдаланушылардың қажеттіліктері негізінде қолмен құрастырылатын тәсіл қолданылады. Пікірлерден алынған аспектілер тізімінде мекеме түрі мен асханасы, тағам мен қызмет көрсету сапасы, жайлылықтың болуы, романтикалық атмосфера, бардың, би алаңының, балалар бөлмесінің болуы, жақын жерде қонақ үйдің болуы, автотұрақ және т.б кіреді. Осы мақалада қарастырылатын мейрамхананың аспектілері Кесте 1-де көрсетілген. Әр аспекті үшін кестеде осы аспект қабылдай алатын мәндер жиынтығы көрсетілген. Мейрамхананың объективті аспектілері де (мысалы, би алаңының, бардың, балалар бөлмесінің және т.б. болуы) және субъективтік аспектілері бар (Кесте 1). Объективті аспектілер үшін аспектілерді жіктеу тапсырмасы ақпаратты алу тапсырмасы болып табылады, ал субъективті аспектілер үшін бұл сезімді талдау тапсырмасы.

Бұл жұмыста аспектілер 3 баллдық шкаламен бағаланады {-1; 0; 1;}. Сәйкесінше пікірге теріс, бейтарап және оң деген реңк беріледі.

Кесте 1

Аспект	Мәндер жиыны
Атмосфера	{-1; 0; 1;}
Баға/Сапа	{-1; 0; 1;}
Қызмет көрсету	{-1; 0; 1;}
Тағам	{-1; 0; 1;}
Жалпы	{-1; 0; 1;}

Аспектіге бағытталған сентиментті талдау

Aspect Based Sentiment Analysis (ABSA) [7] — мәтіндік деректердегі нысанның әртүрлі аспектілерінің сезімдерін табуға әрекеттенетін жоғары деңгейлі SA мәселесі. Мәтіндік деректерде нысанды әртүрлі сөздер жиыны немесе объектілерді нақты көрсететін мүмкіндіктер арқылы сипаттауға болады; бұл белгілер сәйкес субъектінің «аспектісі» деп аталады. Мәні бірнеше аспектілермен және осы аспектілерге қатысты кейбір көңіл-күйді білдіретін бірнеше көмекші сөздермен немесе тіпті сөйлемдермен нақты сипатталуы мүмкін. ABSA субъектіні нақты көрсететін сәйкес аспектілерді олардың қолдаушы сөздерімен бірге табуға тырысады және әрі қарай осы аспектілердің полярлығын немесе көңіл-күйін анықтайды. Келесі сөйлемді қарастырайық: «Таңғы ас дәмді болды, бірақ қызметкерлер жағымсыз болды». «Тағам» және «Қызмет көрсету» бұл жерде қарастырылатын екі аспект.

Ұсыныстың осы түріне арналған ABSA Сурет 2-де көрсетілгендей, «Тағам» аспектісіне оң көзқарасты және «Қызмет көрсету» аспектісіне теріс қатынасты белгілейді.

Таңғы ас дәмді болды, бірақ **қызметкерлер жағымсыз** болды

Aspect category	Aspect term	Sentiment
Food	дәмді	Positive
Service	жағымсыз	Negative

Сурет 2. Аспектiлiк сентименттi анықтау.

Қорытынды

Машиналық оқыту – бұл жасанды интеллект әдістерінің классы. Машиналық оқыту күнделікті біздің өмірімізде көптеген қолданыстарға ие болуда. Оның қолданылуының кеңдігіне байланысты ол IT технологияда маңызды орын алады. Қазіргі кезде машиналық оқыту әдістеріне негізделген көптеген қосымшалар жұмыс істейді. Сентимент талдау кез келген мәтіндік дереккөздерде айтылған пікірлерді ауқымды өндеудің қуатты құралына айналды. Аспектіге бағытталған сентимент талдау бұл мәтінді талдаудың маңызды бөлігі болып табылады. Машиналық оқыту эксперименттерінің нәтижесінде мейрамхана туралы пікірлерді сентиментті талдау үшін ең тиімді болып табылатын классификаторлар анықталды. Зерттелетін аспектіге қатысты пікірлерді санаттарға жіктеу үшін мейрамханалардың көптеген аспектілері үшін мультиномды Naive Bayes классификаторы ең тиімді болып табылады.

Қолданылған әдебиеттер тізімі

- [1] Potgieter, Marius, Johan W. de Jager, and Neels H. van Heerden. An innovative marketing information system: A management tool for South African tour operators. *Procedia-Social and Behavioral Sciences* 2013; 99: 733-741.
- [2] Wan, Yun, and Makoto Nakayama. The reliability of online review helpfulness. *Journal of Electronic Commerce Research* 2014; 15.3: 179.
- [3] Raut VB, Londhe DD. Survey on opinion mining and summarization of user reviews on web. *International Journal of Computer Science and Information Technologies*. 2014; 5(2): 1026-30.
- [4] Bansal P, Ahmad T. Methods and Techniques of Intrusion Detection: A Review. *In International Conference on Smart Trends for Information Technology and Computer Communications, Springer, Singapore* 2016; 518-529
- [5] Bermingham A., Smeaton A. Classifying Sentiment in Microblogs: Is Brevity an Advantage? // *Proceedings of the International Conference on Information and Knowledge Management (CIKM)*, 2010.
- [6] Wang S., Manning Ch. D. Baselines and Bigrams: Simple, Good Sentiment and Topic Classification // *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL): Short Papers – vol. 2*, pp. 90– 94, 2012.
- [7] Gupta D., Ekbal A. (2014), IITP: Supervised Machine Learning for Aspect based Sentiment Analysis, *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Dublin, pp. 319–323