Recommendation System using Hybrid Classification Algorithm in E – Commerce Application

P.Mahalakshmi[1],

[1]AssistantProfessor, Department of Computer Science and Engineering

Nadar Saraswathi College of Engineering and Technology,Theni,Tamilnadu

R.Hemalatha[2]**S.Janani**[3]**V.Susmitha**[4]**P.Vaishnavi**[5]

UGStudents

[2,3,4,5]Department of Computer Science and Engineering

Nadar Saraswathi College of Engineering and Technology,Theni,Tamilnadu

Abstract— Purchase prediction can help e-commerce planners plan their stock and personalized offers.Recommender systems help the customers to find accurate product from a large database. Implement Hybridfilteringtechniqueinrecommendationsystemwithfeedbackanalysistoimprovetherecommendationsystem.Forrecommendation,admintrainadatabasewhichhassentimentbasedkeywordswithpositivityornegativityweight Analyses fake contextual information posted by online users Identifying the user details along withreview posting patterns**.**

Keywords—AgglomerativeHierarchicalClustering,BigDataAnalysis,CollaborativeFiltering,ImplicitRating

I.INTRODUCTION

Big data isa fieldthat treats waystoanalyze, systematicallyextract informationfrom, orotherwisedealwith data sets thatare toolargeorcomplextobedealswithbytraditional data-processing applicationsoftware.Datawithmanycases(rows)offergreaterstatisticalpower,whiledatawithhighercomplexity(moreattributesorcolumns)mayleadtoa higher falsediscoveryrate.Bigdatachallengesinclude capturingdata, datastorage, dataanalysis,search, sharing, transfer, visualization, querying,updating, informationprivacy anddatasource.Bigdatawasoriginallyassociatedwiththreekeyconcepts: volume, variety,and velocity. Big data analytics is the process of examining large data sets containing a variety of data typesto uncover hidden patterns, unknown correlations, market trends, customer preferences and other usefulbusinessinformation.Theanalyticalfindingscanleadtomoreeffectivemarketing,newrevenueopportunities, bettercustomerservice,improvedoperationalefficiency,competitiveadvantagesoverrivalorganizationsandother business benefits. The primary goal of big data analytics is to help companies make more informedbusinessdecisionsbyenablingdatascientists,predictivemodelersandotheranalyticsprofessionalstoanalyzelargevolumesoftransactiondata,aswellasotherformsofdatathatmaybeuntappedbyconventionalbusinessintelligenceprograms.BigDataapplicationswheredatacollectionhasgrowntremendouslyandisbeyondtheability of commonly used software tools to capture, manage, and process within a “tolerable elapsed time” ison the rise. The most fundamental challenge for the Big Data applications is to explore the large volumes ofdataandextractusefulinformationorknowledgeforfutureactions.Uponenteringthe21st century,theglobaleconomicstructureistransferringfromindustrialeconomytoserviceeconomy.Serviceusershavenowadaysencounterunprecedenteddifficulties infinding idealones fromtheoverwhelmingservices.

Arecommendersystemisanintermediaryprogramoranagentthatintelligentlycompilesalistofrequisiteinformationwhichsuitsauser’stastesandneeds.Manyrecommendersystemshavebeendesignedand implementedforvarioustypesofitemsincludingnewspapers,researchpapers,emails,books,movies,music,restaurants, Web pages and other e-commerce products. It proposes a new approach to develop a frameworkforanefficientrecommendersystemthatassistusersinadecisionmakingprocesswheretheywanttochoosesome items among a potentially overwhelming set of alternative products or services. The collaborativefilteringapproachhasbeenusedtoachievethedesiredframeworkforourrecommendersystem.Recommender System predicts new items of interest for a user on the basis of predictive relationshipsdiscovered between the user concerned and the other users sharing the same tastes and interests. The aim ofcollaborativefilteringinarecommendersystemisthereforetorecommenditemstoatargetuserbasedontheopinion of other users.

II. DEFININGTHEPROBLEM

The existing system relies on collaborative filtering recommendation algorithm that recommends a user theitems that are similar to what he/she has preferred before. Using the word2vec embedding method, Wordembedding is used to represent words as vectors that describe the word based on its context, such assurroundingwordsinthesentence.Andthetechniquesareusedintheexistingismatrixfactorizationforlearnthelatentfeaturesofusersanditemsandprobabilisticmatrixfactorizationmodelforpredictuserratings.But,abovetechniques aretraditionaland itencounters somechallenges in bigdataapplication.

Issuestobeaddressed

- Analyzedratingsfromuserreviews
- Inanexistingwork,Fakereviewscan’tbeanalyzed
- Thegenuinereviews can’tbe identifyingbytheusers
- Handleonlylimitednumberofproductreviews.

Collaborativefilteringisawidelyusedandprovenmethodofprovidingrecommendations.Mostcollaborativefiltering-based recommender systems rely on explicit feedback that is collected directly from users. Rating istypical examples of explicit feedback. Because it is easier to quantify ratings than reviews, in practice mostcollaborative filtering methods use rating data. Collaborative filtering algorithms focus on similarity amongusersorsimilarityamongitemsusingusers’ratings.Whenusersratehonestly,usingratinginformationisoneof thebestwaysto quantify user preferences.

Issuestobeaddressed

- Manyusers assignarbitraryratingsthatdonotreflecttheirtrueopinions.
- Itisnotpracticaltoexpectusers’activeparticipationinratings.
- Usersrateonlyasmallportionofallavailableproducts(sparsityProblem).

III. METHODOLOGY

HybridClassificationAlgorithm

Hybrid classification is a concept that employs basic classification algorithms for model induction and fordatapreprocessing.ThealgorithmsinvolvedintheproposedhybridclassificationalgorithmsaredecisiontreeinductionandnaïveBayesianclassifier.Basicclassificationalgorithmsinduceamodelfromdata,whichwillbe used to classify every new instance. Data preprocessing techniques, such as feature selection and instancefiltering, can enhance the performance of basic classification algorithms. The selective naive Bayesian hasbeenshowntobeasuccessfulwrapperforimprovingtheperformanceofnaïveBayesianclassifier.

In proposed system, a hybrid classification algorithm is used as a processing step. These consists of fiveimportant steps.The first step is collection of data in this section data will be collecting from variousresources.Andthenthesecondstepisdatapre-processing,itinvolvestransformingraw data intoanunderstandable format.Real-worlddataisoftenincomplete,inconsistent,and/orlackingincertainbehaviors or trends, and is likely to contain many errors.The next step is generation of categories; this will create thedifferentcategoriesof data.Finally, Classified thedatabasedon thecategory.

Basicclassificationalgorithmsinduceamodelfromdata,whichwillbeusedtoclassifyevery newinstance. Data preprocessing techniques, such as feature selection and instance filtering, can enhance theperformanceofbasicclassificationalgorithms.TheselectivenaiveBayesianhasbeenshowntobeasuccessfulwrapper for improving the performance of naïve Bayesian classifier. Instance filtering helps reduce data sizewithout sacrificing classification performance. A deep belief network is used to select features for supportvector machine in processing data sets with a large amount of class values. Genetic algorithm can also beemployed for feature selection in training classification algorithms such as support vector machine, naïveBayesian classifier, and decision tree induction. For predicting rear-end crashes, attributes are divided intodisjoints subsetswhentraining decisiontreeandnaiveBayesianclassifier.

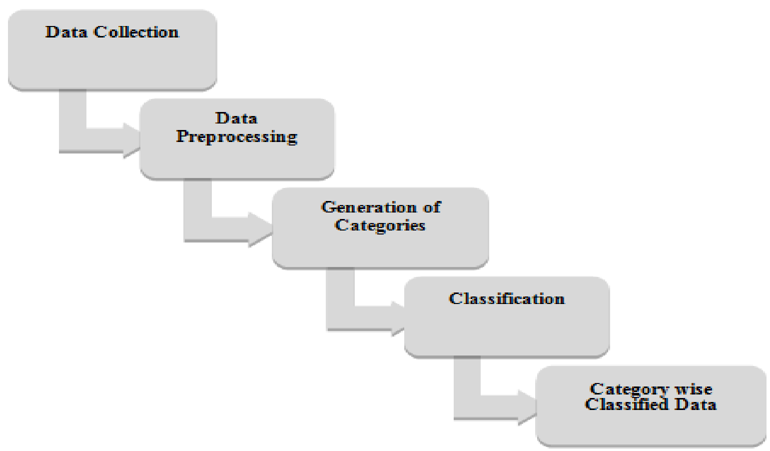

Figure1:ProcessingstepsofhybridclassificationalgorithmHy

brid RecommendationSystem

Hybridrecommendersystemscombinetwoormorerecommendationstrategiesindifferentwaystobenefitfrom their complementary advantages.We address the most relevant problems considered and present theassociateddataminingandrecommendationtechniquesusedtoovercomethem.
Fourmajorrecommendationtechniques constructinghybrids arecollaborativefiltering(CF),content-based(CN),demographic,
andknowledge-
based(KB).UnlikethefirstthreewhichmakeuseoflearningalgorithmsKBexploitsdomainknowledgeand makesinferences aboutusers'needsandpreferences.

Most recommender systems now use a hybrid approach, combining collaborative filtering, content-basedfiltering,and different procedures.Thereis no reason why anumber of different techniques ofthe same kind should nolonger behybridized.Hybrid strategies canbe carriedout in numerous ways: bymeans of making content-based and collaborative-based algorithm one after the other and then combiningthem.Collaborativefilteringisoftenusedalongwithotherfilteringtechniqueslikecontent-based,knowledge-basedtechniques.
ArecommendersystemisCollaborativefiltering methodsareestablishedongatheringandexaminingalargeamount of information which based on user’s demeanor, activities or preferences andwaiting for style ofthatuniqueperson by way of theuseof their similaritywith other user
Thetechniquesusedincollaborativefilteringkn
earestneighbors

1.UserbasedNearestNeighbor2
.ItembasedNearestNeighbor

Dimensionalityreductiontechniques

Arecommendersystemis Content-basedfilteringtriestorecommenditemstothecurrentuserbasedonsimilaritycountwhichisratedbythatuserpositivelyin thepast.Content-basedrecommendersystem

provides user independence through exclusive ratings which are used by the active user to build their ownprofile.
A recommender system is knowledge-based when it makes recommendations based not on a user's ratinghistory, but on specific queries made by the user. It might prompt the user to give a series of rules orguidelinesonwhattheresultsshouldlooklike,oranexampleofanitem.Thesystemthensearchesthroughits databaseofitemsandreturnssimilar results.
Therefore,thecollaborativefiltering,contentbasedfilteringandknowledgebasedfilteringareusedtofiltertheinformation.itprovidesarecommendation accuracy
Collaborative:"Tellmewhat'spopularamongmypeers"
Itrecommendsmewhatotherusersmostlysearchedandbuyed.itshowthepopularproductbasdonratingContentbased:"ShowmemoreofthesamewhatI'veliked"
ItrecommendsmetheproductwhatImostlylikedandsearched.forexampleifisearchformobileitalsorecommendmearealtedthingsofmoblielikeheadset,pendrive,RAM,mouse,etc.
Knowledgebased:"Tellmewhatfitsbasedonmyneeds"
ItrecommendsmeamysearchbasedthingforexampleifisearchformobileitshowsmeamobilepouchIn method 1, the CF and CBF estimate recommendations individually and subsequently combine them toyield better recommendations. In method 2, the characteristics of CBF is integrate into the CF approach. Inmethod 3, By combining some features of CBF and CF one unified model is constructed that can improveeffectiveness of recommendation process. In method 4, that incorporate CF characteristics into a CBFapproachto overcometheproblemwhilecombiningand toyield therecommendation

SystemArchitecture

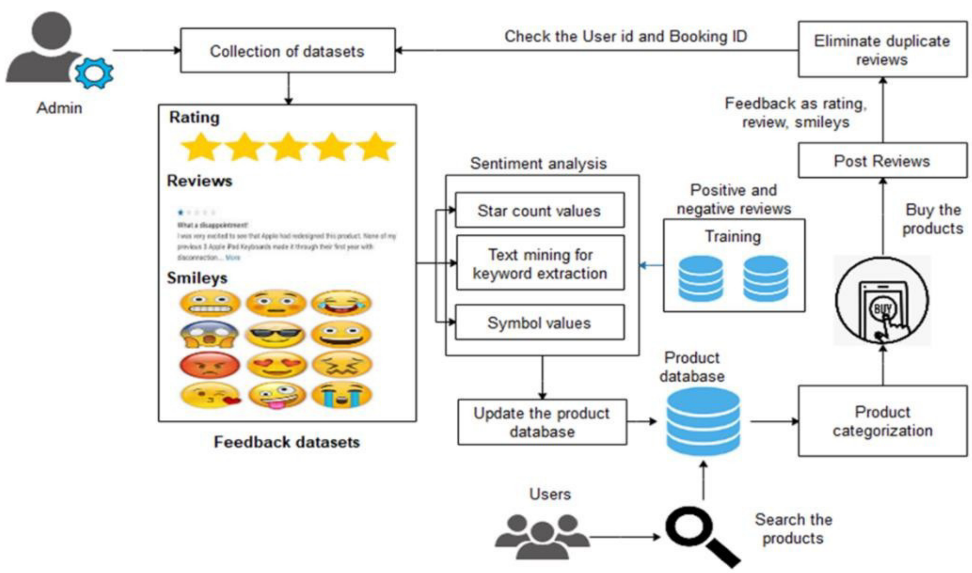

Figure2:ArchitecturalDesignofoverallProcess

MODULES:
In our project contains five modules online e-commerce framework, Review collection, Sentimentanalysis,Recommendationsystem,Fakereviewsmonitoring.Thismoduleshelpsustofindthegenuine productsbyeliminatingfakereviews andoftenrecommendthegoodproducttothecurrentusers.

i) ONLINEE-COMMERCEFRAMEWORK:

Thismoduleis usedtocreatewebsiteandbuyorpost products for users. There are two accountssuchasadminanduseraccount.Admincanlogintothesystemandpostproductswithfeatures.Usercan

logintothesystemtoviewproductdetails.Admincancategorizetheproducts basedontype,genderandsoon.

ii) REVIEWSCOLLECTION:

Admin collect reviews and have various types of reviews. Reviews may be rating reviews, text reviewsandsmileys reviews.Allreviews arestoredindatabaseforfutureevaluation.Ratingsareintheformofstarvalues.Reviews maybeunigrams,bigrams orn-grams.Smileyspecifythesymbols ofhappyandsad.

iii) SENTIMENTANALYSIS:

Admincananalyzewhethertheproductis positiveornegative.Instarrating,wecancalculatestarcountvalues. In text reviews, extract keywords and matched with database. Then smileys reviews are calculatedbasedhappy and sad symbols

iv) RECOMMENDATIONSYSTEM:

User can search the product in search bar. And view the list of products based on features and reviews.Using support vector machine, recommend the products based on product categorization. If the product hasnegativereview means, automaticallythepositiveproductsinrecommendation panel.

v) FAKEREVIEWSMONITORING:

Inthis module,fakereviewsareanalyzedbyadmin.Admincangetuseraccountdetails,MACaddressandOrder id details. So oneuser can postonereviewsthatwillbegenuinereviews.

DataSet

In order to evaluate the performance, Dress attributes sales dataset is used which contains Attributes ofdresses. Sales are monitor on the basis of alternate days. Data has to be classified based on the hybridclassificationtechnique.

Productdataset

uniq_id	crawl_tim	product_u	product_n	product_c	pid	retail_pric	discounte	image	is_FK_Adv	descriptio	product_r	overall_ra	brand
c2d766ca9	2016-03-2	http://ww	Alisha Sol	["Clothing	SRTEH2FF	999	379	["http://ir	FALSE	Key Featu	5	4	good
7f7036a6d	2016-03-2	http://ww	FabHomel	["Furnitur	SBEEH3QG	32157	22646	["http://ir	FALSE	FabHomel	2	4	ok
f449ec65d	2016-03-2	http://ww	AW Bellie	["Footwe	SHOEH4GI	999	499	["http://ir	FALSE	Key Featu	1	4	better
0973b37ac	2016-03-2	http://ww	Alisha Sol	["Clothing	SRTEH2F6I	699	267	["http://ir	FALSE	Key Featu	5	5	excellent
bc940ea42	2016-03-2	http://ww	Sicons All	["Pet Sup	PSOEH3ZY	220	210	["http://ir	FALSE	Specificat	4	No rating	ok
c2a173139	2016-03-2	http://ww	Eternal Ga	["Eternal (	PWTEB7H:	430	430	["http://ir	FALSE	Key Featu	2	4	super
ce5a6818f	2016-03-2	http://ww	Alisha Sol	["Clothing	SRTEH2FV	1199	479	["http://ir	FALSE	Key Featu	3	No rating	bad
8542703ca	2016-03-2	http://ww	FabHomel	["Furnitur	SBEEH3QG	32157	22646	["http://ir	FALSE	FabHomel	5	4	wow
29c8d290c	2016-03-2	http://ww	dilli bazaa	["Footwe	SHOEH3D	699	349	["http://ir	FALSE	Key Featu	2	1	worst

Userdataset

Email	Address	Avatar	Avg. Sessi	Time on A	Time on V	Length of	Yearly Amount Spen
mstepher	835 Frank	Violet	34.49727	12.65565	39.57767	4.082621	587.9511
hduke@h	4547	DarkGree	31.92627	11.10946	37.26896	2.664034	392.2049
pallen@y	24645	Bisque	33.00091	11.33028	37.1106	4.104543	487.5475
riverarebe	1414	SaddleBrc	34.30556	13.71751	36.72128	3.120179	581.8523
mstepher	14023	MediumA	33.33067	12.79519	37.53665	4.446308	599.4061
alvarezna	645	FloralWhi	33.87104	12.02693	34.47688	5.493507	637.1024
katherine	68388	DarkSlate	32.0216	11.36635	36.68378	4.685017	521.5722
awatkins@	Unit 6538	Aqua	32.73914	12.35196	37.37336	4.434273	549.9041
vchurch@	860 Lee	Salmon	33.98777	13.38624	37.5345	3.273434	570.2004

EmpiricalResults

In this experiment, accuracy can be defined as the ratio of the number of items recommended andpurchasedtothenumberofitemsrecommendedbythesystem.TheresultofRecommendationsystem

usinghybridclassificationiscomparedwiththepure collaborativefiltering-basedrecommendersystemusingtheword2vecalgorithm.

Table1:ResultwithPureCF

Item	PureCF	ProposedAlgorithm
ItemsrecommendedandPurchased	234	310

ThefollowingfigdepictstheaccuracyestimationrelatedtoCFmodelswheneithernoclusteringisperformeor clusteringisappliedonaratingmatrix usingimplicitrating.

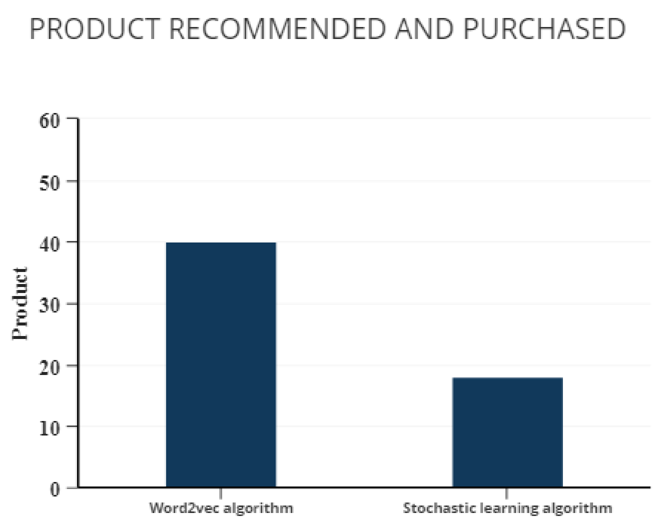

Figure3:AccuracyImprovement

Computation Time

Thefollowingfig.shows theresultoftotaltimetakentomakepredictionscomparepurecollaborativefiltering process.

Table2:PredictionEfficiency

Items	PureCF	Proposed Algorithm
PredictionandrecommendationTimeinSec	280	210

AHC algorithm based CF spends less computation time. Since the number of services in a cluster isfewer than the total number of services, the time of rating similarity computation between every pair ofservices willbegreatly reduced.

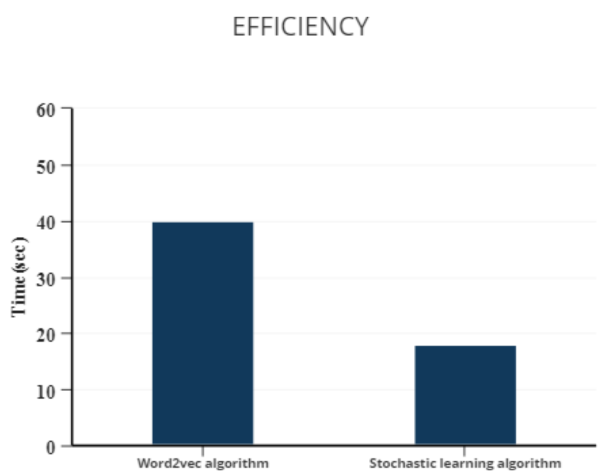

Figure4:PredictionEfficiencywithPureCF

IV. CONCLUSION

Purchase prediction is an important factor for e-commerce decision-makers to give offers andrecommendations to the customers. In this paper, we integrated product similarities as a feature ofclassificationmodelstoimprovepredictionaccuracy.WeclassifiedtheproductusingHybridClassificationalgorithm,andincludingthesentimentanalysishelps theusertofind outcorrectreview oftheproduct. Recommend the positive products to trust users. Automatic decisions making system in productrecommendation.Helps toeliminatefakereview postingusinguser identification.

V. REFERENCES

[1] Katarya, R., Verma, O.P.: An effective collaborative movie recommender system with cuckoo search.EgyptianInformaticsJournal18(2), 105– 112 (2017)

[2] Krismayer, T., Schedl, M., Knees, P., Rabiser, R.: Predicting user demographics from music listeninginformation.MultimediaToolsand Applications78(3), 2897–2920 (2019)

[3] Liu,Q.,Zeng,Y.,Mokhosi,R.,Zhang,H.:Stamp:short-termattention/memoryprioritymodelforsession-based recommendation. In: Proceedings of the 24th ACM SIGKDD International Conference on KnowledgeDiscovery&DataMining. pp. 1831–1839 (2018)

[4] Lo, C., Frankowski, D., Leskovec, J.: Understanding behaviors that lead to purchasing: A case study ofpinterest. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery anddatamining. pp. 531–540. ACM(2016)

[5] Ludewig, M., Mauro, N., Latifi, S., Jannach, D.: Performance comparison of neural and non-neuralapproachestosession-basedrecommendation.In:Proceedingsofthe13thACMConferenceonRecommenderSystems.pp. 462–466 (2019)

[6] Mart´ınez,A.,Schmuck,C.,PereverzyevJr,S.,Pirker,C.,Haltmeier,M.:Amachinelearningframeworkfor customer purchase prediction in the non-contractual setting. European Journal of Operational Research281(3), 588–596 (2020)

[7] Zhang, D., Xu, H., Su, Z., Xu, Y.: Chinese comments sentiment classification based on word2vec andsvmperf.ExpertSystemswith Applications42(4), 1857–1863 (2015)

[8] Mohasseb, A., Bader-El-Den, M., Liu, H., Cocea, M.: Domain specific syntax based approach for textclassificationinmachinelearningcontext.In:2017internationalconferenceonmachinelearningandcybernetics (ICMLC). vol. 2, pp. 658–663. IEEE(2017)

[9] Mokryn,O.,Bogina,V.,Kuflik,T.:Willthissessionendwithapurchase?inferringcurrentpurchaseintentofanonymous visitors. ElectronicCommerceResearchandApplications34,100836 (2019)

[10] Musto, C., Semeraro, G., de Gemmis, M., Lops, P.: Learning word embeddings from wikipedia forcontent-basedrecommendersystems.In: EuropeanConference onInformationRetrieval.pp.729–734.Springer(2016)
[11] Ozsoy,M.G.:Fromwordembeddingstoitemrecommendation.arXivpreprintarXiv:1601.01356(2016)
[12] Parkhimenka, U., Tatur, M., Zhvakina, A.: Heuristic approach to online purchase prediction based oninternet store visitor'sclassificationusingdata miningmethods.In: 2017International ConferenceonInformationand DigitalTechnologies(IDT).pp. 304–307. IEEE(2017)
[13] Pennington,J.,Socher,R.,Manning,C.D.:Glove:Globalvectorsforwordrepresentation.In:Proceedingsofthe 2014conferenceonempiricalmethodsinnaturallanguageprocessing(EMNLP).pp.1532–1543 (2014)
[14] Qiu, J., Lin, Z., Li, Y.: Predicting customer purchase behavior in the ecommerce context. Electroniccommerce research 15(4), 427–452 (2015) [36] Rong, X.: word2vec parameter learning explained. arXivpreprintarXiv:1411.2738 (2014)
[15] Sheil, H., Rana, O., Reilly, R.: Predicting purchasing intent: automatic feature learning using recurrentneuralnetworks. arXiv preprintarXiv:1807.08207 (2018)
[16] Teran, L., Mensah, A.O., Estorelli, A.: A literature review for recom- ´ mender systems techniques usedinmicroblogs. ExpertSystemswith Applications103, 63–73 (2018)
[17] Yadav, S., Nagpal, S., et al.: An improved collaborative filtering based recommender system using batalgorithm.Procediacomputer science132, 1795–1803 (2018)
[18] Zhang, D., Xu, H., Su, Z., Xu, Y.: Chinese comments sentiment classification based on word2vec andsvmperf. Expert Systems with Applications 42(4), 1857–1863 (2015) [19] Zheng, B., Liu, B.: A scalablepurchase intention prediction systemusing extreme gradient boosting machines with browsingcontententropy.In:2018IEEEInternationalConferenceonConsumer Electronics (ICCE).pp.1–4. IEEE(2018)
[20] RamazanEsmeli, Mohamed Bader-El-Den, HassanaAbdull,: Using Word2Vec Recommendation forImprovedPurchasePrediction.In:2020 IEEEInternationalconference. (2020)