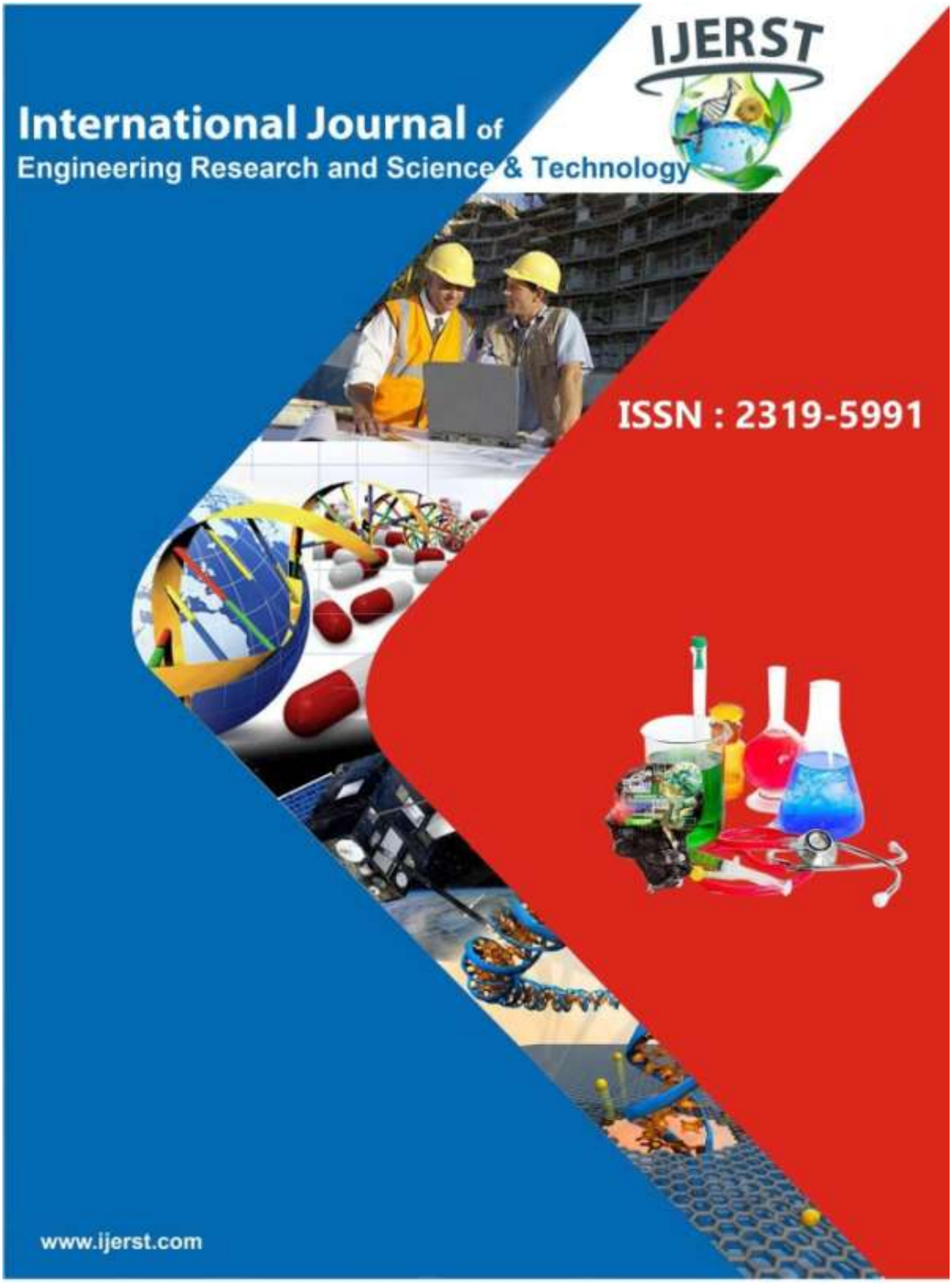


Email: editor@ijerst.com or editor.ijerst@gmail.com

ISSN 2319-5991 www.ijerst.com

Vol. 9, Issuse.2,May 2021

Recommendation System for Hotels in Big Data Based on user Preferences

1M.SathyaDevi,2Ch.Sarada

Abstract:TheRecommendationSystemforHotelsisavaluabletoolusedtorecommendhotelsforusers.Inthe existing system, hotel ratings and rankings are typically calculated based on reviews of previous users.Because of which the preferences of the current user are not considered anywhere. The current servicerecommender system presents the same ratings and services to different users without considering diverse users'preferences. Further, the present system will not work for a massive number of reviews. In this paper, a newcomputingapproachisproposedtopredictandrecommendhotelsbasedonuserpreferences,anditisscalable.In this paper, "User Preference based Comparison" algorithm is suggested to offer top hotels in services.MapReduceframeworkisusedtoimplementit.

Keywords:UserPreferences,Keywords,Recommendationsystem,MapReduce

1. INTRODUCTION

Hotelrecommendationsystemsuggestsbesthotelstotheusers.

Itbecameanimportantresearchareabecauseof a growing number of different services. Service recommendation frameworks appear as essential tool to assistand provide suitable service recommendations for the users. Examples of such recommendation systems includeCDs, books, web pages, hotels and various other products [1]. The existing system does not consider userpreferenceswhilegivingservicerecommendations.Anovelapproachisproposedtoovercomethisprobleminhotelservice recommendation. The proposed system provides service ratings based on user likings. In the new scenario, hotels' reviews arepruned,andacomparativeanalysisisdonebetweenprunedreviewsandcurrentuserpreferencestodeterminetophotels.A"ServiceFilter"algorithmisproposedtoprunetheassessments.Thentoaccomplishacomparison, "User Preference based Comparison" algorithm is designed. In the real scenario, the proposedsystemneedstocomeacrossahugenumberofreviews.

Hence,itisimplementedinMapReduceframework.Theentireworkisaccomplishedintwophases.Inthefirstphase, previous user reviews are trimmed using "Service Filter" algorithm. Service Filter algorithm filters onlyservice/facility information and removes unnecessary information from reviews. In the Second phase ofMapReduce, a personalized rating of each preference of the current user would be calculated using thecomparison algorithm. Here the comparison algorithm compares calculated

1,2DepartmentofComputerScienceandEngineering,CVRCollegeofEngineeringHyderabad,India
satyamaranganti@gmail.com,sharada.ch@gmail.com

ratings with the current userpreference ratings. Finally, it lists out top hotels based on user preferences. Further the paper is organized asfollows: Section 2 provides a brief outline of the research work carried out in this area. Section 3 providesproposed system architecture. Section 4 gives details about the experimental evaluation. The conclusion andfutureworkarebrieflysummarizedinSection 5.

2. RELATEDWORK

Service Recommendation system is a subdivision of information filtering system that predicts service ratings.Recommender frameworks have turned out to be prominent, and are used in different zones including hotels,movies,music, news,books,investigatingarticles,lookinquiries, politicallabels,anditemsingeneral[1]. Recommenderframeworkcharacterizesamethodforhelpingandexpandingthetypicalprocedureofutilizingproposalsofotherstosettleondecisionswhenthereisnoadequateindividualinformationorexperienceoftheoptions [2]. Various approaches exist to develop recommendation systems.Well-known procedures areCollaborativeFiltering,Content-basedFilteringorHybridFiltering[3]. CollaborativeFiltering(CF)techniqueisthemostdevelopedandthemostusuallyimplementedone.Collaborative Filtering suggests things by recognizing different clients with comparable taste; it utilizes theirsentiment to prescribe things to the dynamic client. Collaborative Filtering has two variants like user-basedCollaborativeFilteringanditem-basedCollaborativeFiltering.Inuser-basedCollaborativeFiltering,theforecastoftheratingofathingforaclientreliesontheevaluationsofasimilaritemappraisedbyacomparableclient. VarioussimilaritymeasuresusedinCollaborativeFiltering[7]are Euclidean Distance, Pearson Correlation, Tanimoto Coefficient, Uncentered Cosine, City Block, LogLikelihoodandSpearmanCorrelation. Item-based Collaborative Filtering is a type of Collaborative Filtering for recommender systems filtering thelikenessbetweenthingscomputedutilizingindividuals'appraisalsofthosethings.

Cold-Start problem is the major challenge with Collaborative Filtering. The Cold-Start problem says that thesystemmaynotknowenoughaboutthenewusertodecidewhosimilar[9]is.

Content-Based (CB) methodologies match content resource to customer qualities. Content-Based dividingtechniquesordinarilybuildtheirestimatesinlightofcustomer'sdata[4,5].Aframeworkthatutilizescontent-basedseparationtoenableclientstodiscoverdataontheInternetincludeLetizia[6].InhybridFiltering,bothCFandCBarecombinedtogetthebenefitsofbothcollaborativeandcontent-basedfiltering.

Despite the presence of many service recommender frameworks, that exhibit similar evaluations and rankings ofservices to various clients, they do not consider differing clients' inclinations, and consequently neglectingclients' customized necessities. Moreover, the existing system is not scalable. Hence, a novel approach isproposed for the Hotel Service Recommendation System wherein recommendations are made based on userpreferences.Hence,thispaperproposestheservicesystemwhichisscalable.

Hadoop MapReduce is a framework for effectively composing distributed applications which can process a vastamountofdatainparallelonlargeclustersofcommodityhardwareinafaulttolerantway. HDFS(HadoopDistributedFileSystem)splitstheinputdata-setintosovereignpieceswhichareprocessedby mappers in a parallel manner. The system sorts the yields of mappers, which are then inputted to the reducer.Usually,boththedataandtheyieldofthejobareputawayinafile-system.Theframeworktakescareoftaskscheduling,monitoringthemandre-executesthefailedones.

3. PROPOSEDSYSTEMARCHITECTURE

The proposed framework prescribes hotels considering client inclinations. The architecture of the proposedsystem is designed and shown in figure 1. Two-phase MapReduce structure is utilized to develop the system.The First phase of the scheme determines pruned reviews from reviews of previous customers. The SecondphaseoftheMapReducepreparesahotel servicerecommendationlist.

In the first phase of MapReduce, reviews get split among mappers and each mapper prune the reviews of itssplit. Subsequently, pruned reviews of mapper yield will go as an input to the reducer. Reducer composes all theprunedauditsinto afile.

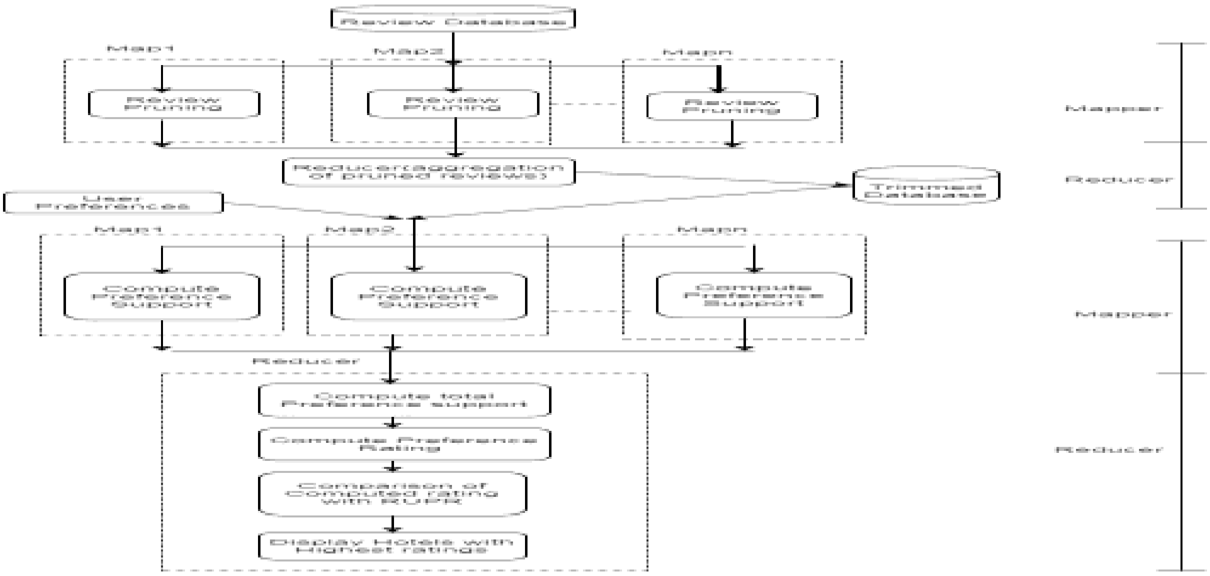

Review Database
Map1
Map2
Mapn
Review Pruning
Review Pruning
Review Pruning
Mapper
Reducer(aggregation of pruned reviews)
Trimmed Database
Reducer
User Preferences
Map1
Map2
Mapn
Compute Preference Support
Compute Preference Support
Compute Preference Support
Mapper
Reducer
Compute total Preference support
Compute Preference Rating
Comparison of Computed rating with RUPR
Display Hotels with Highest ratings
Reducer

Fig1:SystemArchitecture

InthesecondphaseofMapReducealsod atagetssplitamongmappers.However,trimmed reviewsareusedasdata. Support for each current user preference is determined at each mapper. Subsequently, the reduceraggregates the total support, calculates preferred service ratings, then compare them with the user requiredpreferenceratingsandconsequentlypr oduceafinalrecommendationlist.

Intheproposedsystemtwoalgorithmsar eusedlike“Servicefilter”forreviewpruningand“ UserPreferencebased comparison” for determining preference support from previous users. Table 1 illustrates various symbolsusedinthispaperandtheirmeanings.

TABLE1:SymbolsandtheirMeanings

SYMBOL	MEANING
PR	PrunedReview
R	Review
PURS	PreviousUser Reviews
SERVICE	AnyServiceorFacilityofhotel
PUR	PreviousUserReview
PRS	PrunedReviews
CUPS	CurrentuserPreferences
PRGPS	PrunedReviewGroupsbasedonHotel Name
PRGP	PrunedReviewGroup
P	Preference
RPR	RequiredPreferenceRating

HN	HotelName
APR	Aggregate PreferenceRating
TNR	TotalnumberofReviews

Table2:SampleService_Facility_Listof

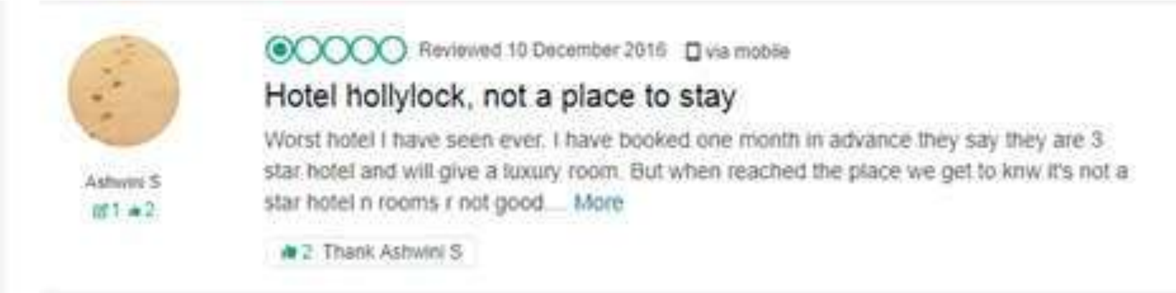
S.No	Service_Facility	S.No.	Service_Facility
1	Room	6	Wi-Fi
2	Shopping	7	Beach
3	Cleanness	8	Transportation
4	Airport	9	Gym
5	Environment	10	Family

3.1 ReviewPruning

Table2:SampleService_Facility_Listof
S.No Service_Facility S.No. Service_Facility
1 Room 6 Wi-Fi
2 Shopping 7 Beach
3 Cleanness 8 Transportation
4 Airport 9 Gym
5 Environment 10 Family

ThecrucialworkofthefirstphaseofMapReduceis reviewpruning.Inthisphase,unnecessaryinformationiswipedoutfromreviews.Servicefilteralgorithmisusedtodopruning&itispresentedinalgorithm1.

Algorithm1:Servicefilter



Reviewed 10 December 2016 via mobile

Hotel hollylock, not a place to stay

Worst hotel I have seen ever. I have booked one month in advance they say they are 3 star hotel and will give a luxury room. But when reached the place we get to know it's not a star hotel n rooms r not good.... More

Input:Previoususerreviewsfromwebsiteslikesitewww.tripadvisor.in.orfromothersources(PURS)Mapper:

```
1.  IntializePR□ empty
2.  For∀R∈PURS
3.      For∀SERVICE∈SERVICE_AND_FACILTY_LIST
4.          IfSERVICEisfoundinthereviewR
5.              PUR□ FiltertheuserreviewonSERVICE
6.              AppendPURtothePR
7.      EndFor
8.      IfPRnotempty thenWritePRtothemapperoutputfile
9.  EndFor10.EndofMapper
```

Mapper_output:PrunedReviewsofanassociatedsplitSorting_and_Shuffling:

Prunedreviewsarearrangedinascendingorderconcerningthehotelname Reducer:

Reducer_Input:PrunedReview(PRS)

```
1.  For∀PR∈PRS
2.      WritePRtothereduceroutputfile
3.  EndFor4.EndofReducer
```

Thisalgorithmextractsstatusinformationabouteachservice/facilityfromtheprevioususerreviews.Table2showsSampleService_Facility_Listofhotels.It'sinputisprevioususerreviews(PURS)andtheoutputisPrunedReviews(PRS).

MapReduce approach is used here due to the need of processing millions of reviews. Every Mapper processprune reviews of previous users. It is an iterative process. In each iteration, one review is processed. Each reviewmay contain feedback about multiple hotel services. The above algorithm finds information about every servicewritten in the preview. The resultant review can be called as pruned view. Every Mapper prune a set of

Reviews.Output of all Mappers is given as input to the Reducer. Sample review and the concerned pruned review isshownhere.

Fig.2Samplereview

Outputofabovereview:

ReducersimplytakeinputfromallMappersandstorethemintheprunedreviews(PRS)file.

3.2 CalculationOfServiceRatings

In the second phase of MapReduce, for each current user preference, rating is calculated. Then the calculatedratingsofpreferredservicesofthecurrentuserarecomparedwithhis/herrequiredpreferenceratings.Atlast,the hotel recommendation list will be out. To accomplish this, “User Preference based comparison” algorithm ispresentedinalgorithm 2. UserPreferencebasedcomparisonprocedurepreparestopHotellistbasedonratingsofhotelservicesagainstUserPreferenceratings.ItusesMapReduceapproachtodothiscomparisonprocess.EveryMapperdeterminessupport metric of every User Preference of every hotel. At every Mapper, iterative procedure is applied on everyreviewtodeterminesupportcountofevery UserPreferenceofeveryhotel.

ReducertakesservicecountofeveryUserPreferenceofeveryhotelasinputfromMappers.ItdeterminestotalsupportofeveryUserpreferredserviceofeveryhotel.Thenitwillcalculateserviceratingsofeveryhotelforevery User Preference specified. Followed by it arranges calculated ratings in descending order and segregatethem with respect to Hotel. Finally, it lists out top hotels that satisfy service ratings against User Preferenceratings.

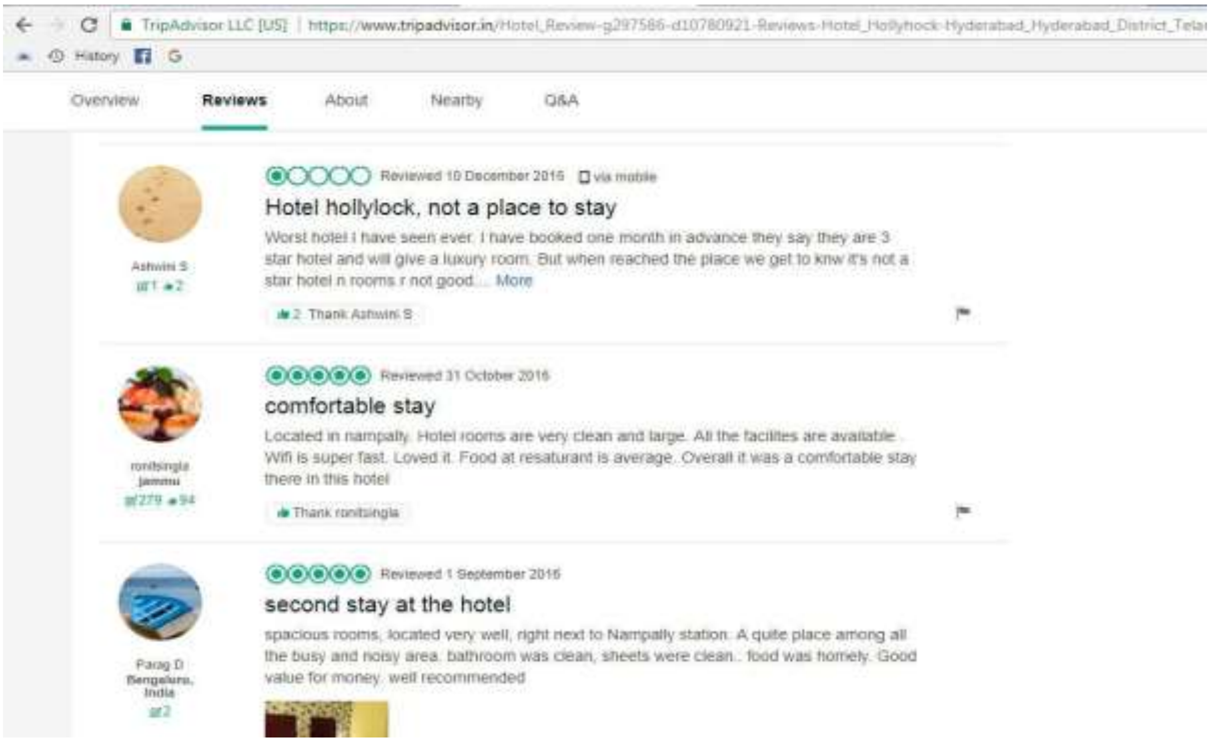
```
Algorithm2:UserPreferencebasedcomparison
Input:PrunedReviews-PRS,CurrentUserpreferences-CUPSMapper:
3.      PRGPS DividePrunedReviewsintogroupsconcerningHotelnames
4.      For∀PRGP∈PRGPS
5.      For∀P∈CUPS
6.      HN StoreHotelNameofPRG
6.      For∀PR∈PRGP
7.      IfPfoundinPRThenSupport Support+1
8.      EndForPGRP
9.
        KEY ConcatenateHN,P,andRPRwithspaceasseparator
11.     Write(KEY,Support)intomapperoutput
12.     End For CUPS13.EndForPRGPS14.EndofMapper
Mapper_Output:SupportforallCUPSbasedonhotelname.

Sorting_and_Shuffling:
1.
        HNS,P,RPRSandthecalculatedSupportareshuffledandsortedsothatthesupportforPfromallmapperswill form as one record. In this record HNS, P, and RPRforms the key and Support for the associatedpreferencefrommapperswillbecomethevalue.
Reducer:
Reducer_Input:KEYSthatcontainsHNS,P,RPRS
:VALUESthatholdsCalculatedSupportforPfrom allmappers1.For∀KEY ∈KEYS
2.      Sum 0
For∀VAL∈VALUES
Sum Sum+VALEndFor
ASR Sum/(TNR)IfASR>=RPR
SR_TABLEHN StoreASRPintothenextcolumnof
ST_TABLEattherowcorrespondingtoHNEndFor
3.
        SR_TABLEissortedinthedescendingorderofAPSRusingMulti-ColumnsortingtechniquePrinttopthreeHotelNamesalongwithAPRofeachCUP
4.      EndofReducer

Output:Top3hotelsrecommendationlist
```

4. EXPERIMENTALEVALUATION

The proposed system is portrayed under the MapReduce framework. In this section, the points of interest ofsample hotel reviews and pruned reviews appear. At last, this section exhibits the result of the experimentalstudy. ProposedsystemisportrayedunderMapReduce framework.Inthissection,thepointsofinterestof samplehotelreviewsandprunedreviewsareappeared.Atlast,theresultofexperimentalstudyisexhibited.

4.1 Reviews

Hotelreviewsaretakenfromthesitewww.tripadvisor.in.Thisanalysisconsidersabout100000surveysforthisstudy.Figure3showssomeexamplereviews.


TripAdvisor LLC [US]
https://www.tripadvisor.in/Hotel_Review-g297586-d10780921-Reviews-Hotel_Hollyhock-Hyderabad_Hyderabad_District_Tela
History
Overview
Reviews
About
Nearby
Q&A

Ashwini S

Reviewed 10 December 2016 via mobile

Hotel hollylock, not a place to stay

Worst hotel i have seen ever. I have booked one month in advance they say they are 3 star hotel and will give a luxury room. But when reached the place we get to knw it's not a star hotel n rooms r not good.... More

2 Thank Ashwini S

ronitsingla jammu

Reviewed 31 October 2016

comfortable stay

Located in nampally. Hotel rooms are very clean and large. All the facilites are available . Wifi is super fast. Loved it. Food at resaturant is average. Overall it was a comfortable stay there in this hotel

Thank ronitsingla

Parag D Bengaluru, India

Reviewed 1 September 2016

second stay at the hotel

spacious rooms, located very well, right next to Nampally station. A quite place among all the busy and noisy area. bathroom was clean, sheets were clean.. food was homely. Good value for money. well recommended

Fig3.Samplereviews

4.2 ExperimentalAnalysis

Intheexploratoryexamination,reviewsareprunedfirstthenratingsforpreferredservices/facilitiesarecalculatedandcomparedwithuserrequiredpreferenceratings.

4.2.1 PrunedReviews

Reviewpruningprocessextractsservice/facilityinformationfromtheoriginalreviews.Prunedreviewsareessentialtodetermineserviceratings.Table3showsprunedreviewsofsurveyslistedinfigure3.

Table3:Prunedreviews

Review1	Hotelhollylock,Hyderabad,worsthotel
Review2	Hotelhollylock,Hotelroomsareveryclean,wifiissuperfast,foodisaverage
Review3	Hotelhollylock,spaciousrooms,locatedverywell,nexttoNampallystation,foodwas homely

4.2.2.FinalRecommendations

In this experiment, ratings of preferences of the current user are calculated based on pruned reviews andcomparedthemwiththeirrequiredpreferenceratings.Atlast,therecommendationlistisdisplayedalongwithratingsofeachpreferredservice/facilityofthecurrentuser.Table4,showsthehotelservicerecommendationlistalongwithratingsforthefollowinguserpreferences:

Table4:Userinputwithpreferencesalongwithrequiredratings

ServicePreferred	Preferredrating
GoodService	35%
Friendlystaff	25%
Cleanliness	20%

Table5:Finallistofhotelservicerecommendations

HotelName	Review	Rating%
CentralCourtHotel	Clean lobby Friendlystaff Serviceisgood	45% 28% 25%
TajDeccan	Clean lobby Friendlystaff Serviceisgood	42% 26% 23%
TajKrishna	Clean lobby Friendlystaff Serviceisgood	39% 25% 22%

5. CONCLUSIONANDFUTUREWORK

Theproposedsystemrecommendshotelsbased onuserpreferences.Thewholesystemisimplementedinthe MapReduce environment, to make it scalable. In the proposed solution, “Service Filter” algorithm is the pivotalone,anditisusedtoprunethesurveys,yetitisn'tperformingpruningflawlessly.Asapartoffuturework,wewouldlike toimprovethisalgorithm.

REFERENCES

[1] https://en.wikipedia.org/wiki/Recommender_system

[2] Resnick P, Varian HR. Recommender systems. Commun ACM 1997;40(3):56–8.http://dx.doi.org/10.1145/245108.24512.

[3] AcilarAM,ArslanA.Acollaborativefilteringmethodbasedon ArtificialImmuneNetwork.ExpSystAppl2009;36(4):8324–32.

[4] MinSH,HanI.Detectionofthecustomertime-variantpatternforimprovingrecommendersystem.ExpSystApplicant2010;37(4):2911–22.

[5] CelmaO,SerraX.FOAFingtheMusic:bridgingthesemanticgapinmusicrecommendation. WebSemant:SciServAgentsWorldWideWeb2008;16(4):250–6.

[6] LiebermanH.Letizia:anagentthatassistswebbrowsing.In:Proceedingsofthe1995internationaljointconferenceonartificialintelligence. Montreal,Canada;1995.p.924–9.

[7] SaikatBagchi,PerformanceandQualityAssessmentofSimilarityMeasuresinCollaborativeFilteringUsingMahout.ProcediaComputerScience50 (2015)229– 234.

[8] https://www.Tripadvsior.com.

[9] Cold-startProbleminCollaborativeRecommenderSystems:EfficientMethodsBasedonAsk-to-rateTechnique.