Asian Journal of Engineering and Applied Technology
ISSN: 2249-068X (P) Vol.6 No.2, 2017, pp.18-31
© The Research Publication, www.trp.org.in
DOI: https://doi.org/10.51983/ajeat-2017.6.2.821

Semantic Approach for Multifarious Ranked Novelty and Diversity in News Recommendations

Shikha Agarwal
Assistant Professor, Department of Computer Science,
IP College for Women, University of Delhi, Delhi, India
E-Mail: shikha_8june@rediffmail.com

(Received 3 July 2017; Revised 18 July 2017; Accepted 9 August 2017; Available online 16 August 2017)

***Abstract* - Information overload on web creates lots of inconvenience for end users to unearth requisite information which instigate the demand of personalized recommendations. Recommender systems strive to attain just accuracy in recommendations based on the history of user preferences, resulting in over specialized recommendations. This leads to gradual loss of user's interest in the system. These approaches also fail to recommend other products in long list which could be of user interest but user is not aware of. This in-turn leads to sparse user profiles.**

In dynamic domains, novelty and diversity in coverage can handle over specialization problem. To bring useful novelty and diversity, it is proposed to semantically expand the user preferences. For this purpose our designed multi level ontology is used with linkages among concepts, entities and properties, carrying different weights. Ontology is also enriched with different relations from online semantic lexicon and annotated with additional information about entities from open and linked external knowledge source.

For relevant accurate recommendations in highly dynamic domains, we propose semantic approach to implicitly capture individual user preferences in separate profiles. One profile is to capture short term dynamic interests with temporal effects and another is to capture long term static interests. Semantic profiling helps to handle preferences in manageable number of news categories which are arranged hierarchically. It also helps to handle synonymy and polysemy problems, reducing ambiguity in profiling. An additional important step of outlier analysis and rectification is proposed to handle the effects of sudden unexpected temporary drifts in dynamic user preferences. This focused analysis helps to maintain correct ranking of user preferences which is generally overlooked. The rectified ranking of user interests is fetched for semantic expansion of user preferences, to bring correct ranking in diversity and novelty.

Proposed approach semantically brings reasonable multifarious novelty and diversity by analyzing different features of item. Testing is based on live data gathered from RSS feeds of popular news providers for trustworthiness. Transparency is achieved by presenting new recommendations in separate sections along with explicit trace path options, based on difference in approaches. Transparency assists end user in making decision. User's interests in diverse and novel recommendations are updated in their individual profiles, resulting in reduction of sparsity.

***Keywords*: User profile, OWL ontology, recommender system, Semantic Expansion, Novelty, Diversity**

I. INTRODUCTION

Information overload (popularized by Alvin Toffler 1970 and Bertman Gross 1964) is an immense problem in today's online world because the vastness of on hand information causing anxiety in user. It also hampers the making of correct decision with quality. Therefore relevant information is required not just for decision making but also to save valued time. Approaches for information filtering can assist the user by providing required information. Different approaches (Goldberg et al., 1992, Melville et al., 2010, Ricci et al., 2011) are used depending upon the required services in different domains to analyze product to be recommended and end user of the product. In present web scenario, recommendation systems play a vital role in delivering the required information to the required users (Shawn et al., 2010). In past few years different recommender systems are developed for different domains like News Dude, Pandora Radio, Movie Lens, Netflix and so on.

User profiling is the basis of achieving personalization in recommendations (Chen et al., 2008). It usually captures user preferences based on past history to bring accuracy in recommendations. This leads to over-specialized recommendations. It means user preferences are stored for few of the categories only in which user has shown interest in the past. It also leads to the problem of sparsity. In some domains with high dynamism in available data some diversity and novelty in recommendations is desirable. Therefore profiling approach must be modified to handle the issues of over-specialized recommendations. Novelty and Diversity in recommendation coverage (MouZhi et al., 2010) can solve the problem of over specialization.

In dynamic domains like news, we observed and concluded that a single user profiling will not be able to serve the purpose of capturing user preferences. Therefore it becomes a need to maintain two profiles of each user. One profile is to maintain user interests likely to remain static for longer durations. Another profile is to maintain dynamic interests which changes after every short span of time. In proposed approach two profiles are created for each of the registered user. In user profiles current preferences must be given more importance to adapt user likings with changing time therefore current preferences are given more weight age as compared to previous ones in the history of user's click behavior. Moreover we are considering the actual active duration of the sessions in calculations by removing idle time from total session duration. Therefore superfluous session durations will not be given undue importance. User interests are captured in news concepts as well as news entities. Deeply analyzing dynamic and static profiles it was observed that two profiles are not independent. There are sudden notable with a peak in dynamic profiles which drifts away the actual static preferences of users. These unexceptional changes (outlier values) in dynamic preferences can't be ignored while analyzing static profile of user for recommendations. Therefore an algorithm has been designed for relative outlier detection and rectification of ranking in user preferences (Agarwal et al., 2014a). It also improves the ranking of novel and diverse recommendations because correct ranking of preferences are fetched to the algorithm proposed for semantic diffusion of user preferences.

Semantic diffusion is required to handle the problem of over specialized recommendations by inferring other related concepts and entities of user interest which could not have been identified directly through past user preferences. It makes use of concepts, sub concepts, property, annotations, individuals and semantic linkages in our designed Semantically Enriched News Domain Ontology (SENDO). Linkages among concepts, sub concepts, entities and properties in ontology are given different weights based on their part in expansion. The ontology is enriched with semantic relations from online semantic lexicon and also annotated with additional information from open and linked external knowledge source. Semantics based profiling helps to handle ambiguity due to synonymy and polysemy. Ontology is implemented using Web Ontology Language OWL and is designed based on news industry standards given by International Press Telecommunication Council (IPTC) (Agarwal et al., 2012a). Semantic diffusion of user interests making use of hierarchical ontology structure brings reasonable diversity and novelty. Semantic diffusion of user interest in different entities makes use of semantic linkages in the ontology as well as information captured in ontology, about the individual entities from DBpedia (Agarwal et al., 2013). It results in recommending new unseen news items related to the entity of user interest as well as new unseen news items related to different entities belonging to same occupation like news about all Australian cricketers or all singers. Additional information from Dbpedia also helps in bridging the gap of missing information about the entities (like occupation, birth date or birth place of an individual of entity type person) in recommended news items.

A portal has been designed for the purpose of testing different proposed algorithms. News items from different trusted RSS feeds are gathered from popular trusted sources and semantically classified into different categories of ontology (Agarwal et al., 2012b). Pre-classified items help to curb the issue of cold start for new item in recommendations. Experimental study shows that proposed approach is capable of handling the problem of over specialized recommendations and sparse user profiles. Coverage of recommended concepts as well as entity curbs the limitation of over specialized recommendations. Local diversity and local novelty in recommendations have been shown based on semantic diffusion of individual user preferences, both in entity as well as concept. Global novelty and diversity has been achieved considering preferences of all the users in news source, news category and news entity. Novelty and diversity is desirable to retain user's interests in news domain (Lv et al., 2011). These multifarious novel and diverse recommendations due to different proposed approaches are presented to user with proper explanations. It brings transparency for users which facilitate them to understand reasons behind new recommendations and also to decide whether it is worth to accept new recommendations or not. When user clicks on the new recommendations, it reflects acceptance by user, and accordingly user preferences are updated for the new categories leading to the reduction of sparsity in user profiles.

Rest of the paper is organized as follows. Next section gives idea about the background information. Section III focuses on proposed approach for the problem. Section IV gives experimental results and evaluations. Section V gives insight of the related work done in the area. Section VI is about conclusions and directions for future work. Last Section contains list of References.

II. BACKGROUND

In this age of information overload, people uses a variety of strategies to make choices about what to buy, how to spend their leisure time, and making various decisions. In news domain user wants to read latest news of their interest area. Our approach has the goal of providing personal, and high-quality recommendations of RSS feed news items to end user. We have tried to curb the limitations found in approaches used by different recommender systems developed earlier. The limitations are over specialization, sparsity, limited coverage, and ranking of diverse and novel items, missing data, usefulness of diverse and novel recommendations, transparency and trustworthiness in recommendations.

RSS (Really Simple Syndication) is XML based format to syndicate web information or metadata of a web site to multiple other sites. News feeds allow you to see when websites have added new content at your place, without having to visit the websites from where you have taken the feed. In proposed approach News items from RSS feeds are aggregated, classified and mapped to the domain ontology. Ontology is an explicit formal specification of shared conceptualization of a domain (Gruber, 1993). It is used for describing, connecting and reusing knowledge (Neches et al., 1991, Uschold et al., 1996). Ontology is a tree like structure consisting of concepts, sub concepts, properties and relationships of a domain. Ontology together with a set of individual instances of classes constitutes a knowledge base. Our ontology design is based on IPTC standards. The International Press Telecommunications Council, based in London, United Kingdom, is a consortium of the world's major news agencies and news industry vendors. It develops and maintains technical standards for improved news exchange that are used by virtually every major news organization in the world. IPTC has given 28 news codes, out of which two are subject code and subject qualifier, used for categorization and can be seen as a three level hierarchy. Categories in hierarchy are represented by sequence of fixed 8 decimal digit strings in the order of more general to more specific. First two digits (01-17) represent subject code which describes the content of a set of terms to provide a description of the editorial content of news for example 09 for Labor, 12 for Religion and Belief, 13 for Science and Technology, next three digits (000) represent subject matter (optional) which provides a description at a more precise level and last three digits (000 means no value) represent subject detail at a more specific level.

RSS feed news item are also classified based on different entity types in the news items. Named entities (NE) of different types are identified in RSS feed news items. NE is a phrase in the text which uniquely refers to an entity of the world. In general examples of named entity types are first and last names, geographic locations, ages, addresses, phone numbers, organization, monetary values etc. NE Recognition (NER) is a subtask of information extraction that identifies atomic elements in text and maps into above given appropriate predefined categories. Many tools are available for NER which can be trained for a domain. Information about entity has been extracted from open and linked external knowledge source DBpedia. It provides structured information about four million things, from Wikipedia. Entity individuals in ontology are populated with additional information extracted from DBpedia. Usage of this information has been given in proposed approach in next section.

III. PROPOSED APPROACH

We propose semantics based approach for user profiling to bring accuracy in recommendations. Dynamic domain may catch user attention to a particular news category for a short span of time but it may deviates user’s actual preferences captured in profile for longer duration. It will change the actual ranking of user preferences. To rectify the incorrect ranking of preferences, a crucial step has been taken in proposed approach to detect these outlier categories and re-rank the preferences.

Accuracy in recommendations leads to the issue of over-specialization and sparse user profiles. To curb the issue of over-specialization, an approach for semantic expansion of user preferences is proposed. It brings ranked and reasonable diversity and novelty in recommendations. Improvement in recommendation coverage helps to bring down sparsity. Dynamic nature of user profile in terms of preferred news source motivated us to give choice of news source also (fig. 18).

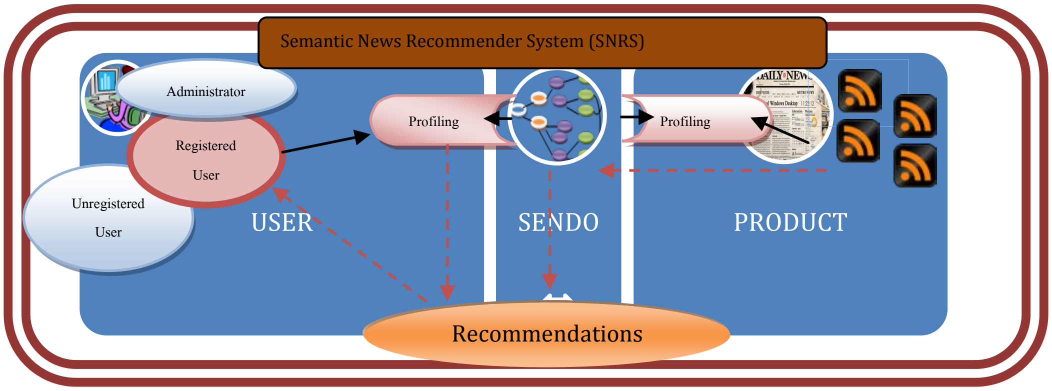


Fig.1 Overview of SNRS

To test the algorithms designed for proposed approach a portal: ‘SNRS-Semantic News Recommender System’ has been designed (fig. 1). Different types of users (Aadministrator, Unregistered users, registered users: first user, every new user and existing experienced user) have different privileges, needs and expectations. Algorithms are capable of fulfilling the need of different users and also curb the problem of zero recommendations for first user and news users.

Proposed approach of user profiling makes use of ontology knowledge to capture user preferences without ambiguity. Therefore a Semantically Enriched News Domain Ontology (SENDO) has been constructed (fig. 2). The Structure of SENDO is based on IPTC standards (Agarwal et al., 2012a), having multi level categorization of news categories. The categories are distributed into 17 major concepts, 390 sub concepts and about 986 leaf concepts. In designed Ontology all concepts have been given IPTC newscodes (Agarwal et al., 2012a) which makes it easier to traverse and modify large hierarchy, without remembering all concept names. Ontology is also enriched with semantic relations (Synonym, Hyponym and Meronym) from online semantic lexicon WordNet (Agarwal et al., 2012b) to capture user preferences without ambiguity. Ambiguity (synonymy and polysemy) in keyword based profiles motivated us to perform semantic user profiling; adopting semantically enriched ontology. Data type and Object properties have been added based on news domain analysis, which gives semantic linkages among the ontology concepts, properties and entities. Besides this, individuals of identified named entities are automatically created in the ontology. These individuals get populated with information (Agarwal et al., 2013, Ivo Lasek 2011) from external, open and linked knowledge source, DBpedia.

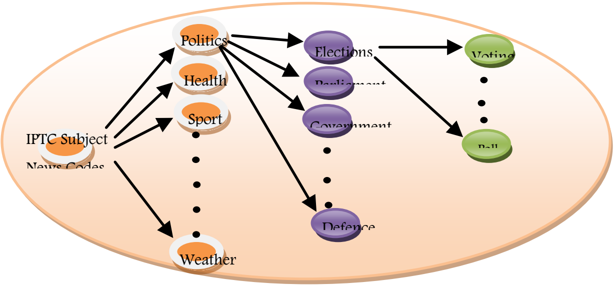


Fig.2 A section of hierarchical structure of SENDO

News items to be recommended are already classified and mapped into ontology categories. News items from trusted RSS feeds are automatically gathered daily and stored in active database for processing and indexing. Recommending news from popular RSS feeds enhances trustworthiness of users. Aggregated News items are semantically classified into major and leaf categories of SENDO. Apart from category based classification news items are also classified on the basis of named entities. For this purpose a layer has been added in the SNRS to identify named entities in the news items. News items are mapped to the news concepts as well as individuals of identified entities in the ontology. This classification of news items apart from concept based classification helps to better understand likings of users. For example let's take a case of news headline "Flood hits Assam" which user has read because of his/her interest in entity individual: "Assam", belonging to entity type: "Location". The same news has been mapped to sub-concept "Flood" within in major concept "Disaster and Accident" in news domain ontology. Ignoring the user interest in entities will recommend all the news related to concept "Flood". Thus in news profiling we have considered all the aspects which can affect user likings.

Major steps of News Profiling are as follows:-

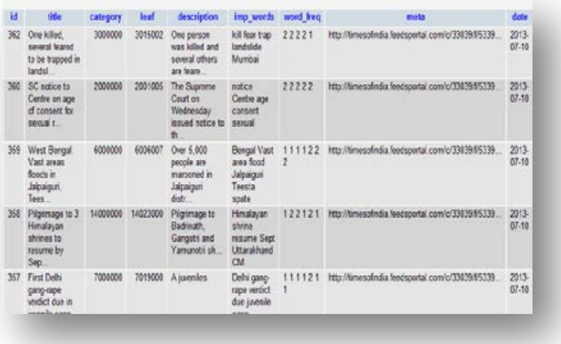
1. Pre-processing and Indexing of news items
2. Concept based semantic Classification of news items
3. Entity Based Semantic Classification of news items

A.Pre-processing and Indexing of News Items

News items from RSS feeds originally comes in XML format. News items are first transformed into a form which can be understood by system. Title, description, date, source, concept and sub concept of news items are stored into database (fig. 3) for further processing. Stop words in news title and description are removed for feature reduction. Words are stemmed to get their root form to reduce ambiguity. Pre-processed news items are semantically indexed based on ontology concepts and entities.

B.Concept based semantic classification of news items

News items are classified into different major and leaf categories of SENDO. Classified news is mapped to leaf categories of SENDO. Semantic linkages (synonyms, meronyms, and hyponyms) extracted from online dictionary WordNet has also been used to disambiguate the classification process (Agarwal et al., 2012b). This classification assists in ranking the order of the preferred categories in recommendations.



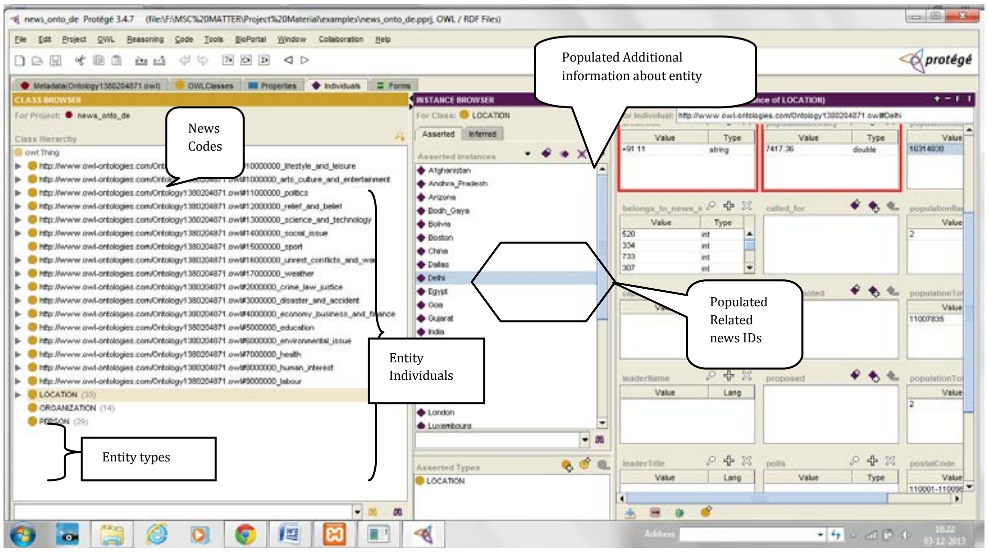
id	title	category	leaf	description	imp_words	word_freq	meta	date
362	One killed, several feared to be trapped in landsl...	3000000	3015002	One person was killed and several others are feare...	kill fear trap landslide Mumbai	2 2 2 2 1	http://timesofindia.feedsportal.com/c/33039/f/5339...	2013-07-10
360	SC notice to Centre on age of consent for sexual r...	2000000	2001005	The Supreme Court on Wednesday issued notice to th...	notice Centre age consent sexual	2 2 2 2 2	http://timesofindia.feedsportal.com/c/33039/f/5339...	2013-07-10
359	West Bengal: Vast areas floods in Jalpaiguri, Tees...	6000000	6006007	Over 5,000 people are marooned in Jalpaiguri distr...	Bengal Vast area flood Jalpaiguri Teesta spate	1 1 1 1 2 2 2	http://timesofindia.feedsportal.com/c/33039/f/5339...	2013-07-10
358	Pilgrimage to 3 Himalayan shrines to resume by Sep...	14000000	14023000	Pilgrimage to Badrinath, Gangotri and Yamunotri sh...	Himalayan shrine resume Sept Uttarakhand CM	1 2 2 1 2 1	http://timesofindia.feedsportal.com/c/33039/f/5339...	2013-07-10
357	First Delhi gang-rape verdict due in	7000000	7019000	A juveniles	Delhi gang-rape verdict due juvenile	1 1 1 1 2 1 1	http://timesofindia.feedsportal.com/c/33039/f/5339...	2013-07-10

Fig.3 A section of aggregated news items: Preprocessed and Labeled

C.Entity Based semantic classification of news items

In title and description of RSS feed news items, three entity types: Person, Place and Organization are recognized using Named Entity Recognition tool (Agarwal et al., 2013). Ontology is populated with information about the identified entities (Agarwal et al., 2013) from external knowledge source DBpedia. While recommending news items to users, this additional information can bridge the gap of missing information in news. All the news items are classified based on the entities occurring in them. If a news items contains more than one entity, then proposed approach is able to classify news into all those belonging entities. Classified news items are mapped with concerned entities in the ontology. In ontology only unique entity individuals are created. In Ontology each entity is populated with news ids of other related news having same entity. This results in automation of multi label entity based semantic classification of news items. Ontology is populated with relations among entity class and their properties. Figure 4 shows entity ontology as part of main news ontology.

Middle part of the figure shows the individuals created for the identified entities. Right part of figure shows the entity properties populated autonomously with knowledge from DBpedia. This entity based ontological news classification and annotations, helps in diversifying the coverage of recommendations. News of person entity is also linked with other news item about the persons of same profile. This information helps in semantic expansion of preferences based on related entity and similar entity profile.


news_onto_de Protégé 3.4.7 (file:\F:\MSC%20MATTER\Project%20Material\examples\news_onto_de.pprj, OWL / RDF Files)
File Edit Project OWL Reasoning Code Tools BioPortal Window Collaboration Help
Metadata(Ontology1380204871.owl) OWLClasses Properties Individuals Forms
CLASS BROWSER
For Project: news_onto_de
Class Hierarchy
INSTANCE BROWSER
For Class: LOCATION
Asserted Inferred
Asserted Instances
Asserted Types
protégé
News Codes
Entity Individuals
Entity types
Populated Additional information about entity
Populated Related news IDs

owl:Thing
http://www.owl-ontologies.com/Ont...#10000000_lifestyle_and_leisure
http://www.owl-ontologies.com/Onto...owl#1000000_arts_culture_and_entertainment
http://www.owl-ontologies.com/Ontology1380204871.owl#11000000_politics
http://www.owl-ontologies.com/Ontology1380204871.owl#12000000_relief_and_belief
http://www.owl-ontologies.com/Ontology1380204871.owl#13000000_science_and_technology
http://www.owl-ontologies.com/Ontology1380204871.owl#14000000_social_issue
http://www.owl-ontologies.com/Ontology1380204871.owl#15000000_sport
http://www.owl-ontologies.com/Ontology1380204871.owl#16000000_unrest_conflicts_and_war
http://www.owl-ontologies.com/Ontology1380204871.owl#17000000_weather
http://www.owl-ontologies.com/Ontology1380204871.owl#2000000_crime_law_justice
http://www.owl-ontologies.com/Ontology1380204871.owl#3000000_disaster_and_accident
http://www.owl-ontologies.com/Ontology1380204871.owl#4000000_economy_business_and_finance
http://www.owl-ontologies.com/Ontology1380204871.owl#5000000_education
http://www.owl-ontologies.com/Ontology1380204871.owl#6000000_environmental_issue
http://www.owl-ontologies.com/Ontology1380204871.owl#7000000_health
http://www.owl-ontologies.com/Ontology1380204871.owl#8000000_human_interest
http://www.owl-ontologies.com/Ontology1380204871.owl#9000000_labour
LOCATION (33)
ORGANIZATION (14)
PERSON (29)

Afghanistan
Andhra_Pradesh
Arizona
Bodh_Gaya
Bolivia
Boston
China
Dallas
Delhi
Egypt
Goa
Gujarat
India
London
Luxembourg

LOCATION

For Individual: http://www.owl-ontologies.com/Ontology1380204871.owl#Delhi

Value	Type
+91 11	string

Value	Type
7417.36	double

Value
16314838

belongs_to_news_i

Value	Type
520	int
334	int
733	int
307	int

Fig.4 Entity individuals in SENDO

To know the preferences of end user using proposed approach of user profiling, algorithms are designed for

1. Capturing User preferences implicitly as well as explicitly
2. Outlier detection and rectification of anomalies in preferences
3. Dual Semantic Diffusion of User Preferences bringing Novelty and Diversity in Recommendations

D.Capturing User preferences Implicitly

For the end user of news domain certain important factors are necessary to consider. As the news domain is highly dynamic in nature, any news item, news category or news source may gain users' attention for a short period, due to natural calamity, political rally or sports event. This will result in occurrence of sudden peak in user preferences captured on the basis of past history. This is not the actual user interest in long run. This problem generates the demand to create two profiles of each user. One profile is for relatively stable interests for longer duration. Another profile is for dynamic interests for short duration. Both the interests are captured implicitly. As the user interests in news are not stable, a time based decay factor has been added to capture the interests of user in ranked order of recency. Users static and dynamic ranked interests are captured both in news concept and news entity (fig. 5). Algorithms used for profiling of dynamic and static user preferences have been given below:

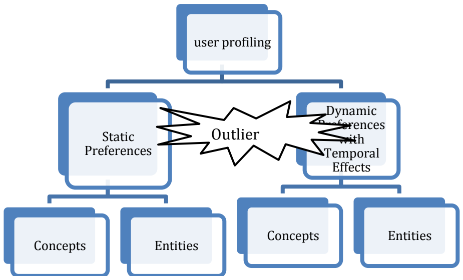


Fig.5 User profiling on different features

Algorithm for profiling of dynamic user preferences:

```
U = Set of registered users
C = Set of all categories.
Short = Set of sessions in most recent DS days.
//Dynamic user profiling
For each $u_i$ Є U
$clicks_{ci} = 0 \quad \forall\, ci$ Є C
Total_sessions = 0
For each $s_i$Є Short
If(username($s_i$) == $u_i$) // counts total session of a user
Then Total_sessions += 1 End if
 End for
Recency_wt = $\frac{1}{Total\_sessions}$   // temporal drift in session wise user preferences
Session_recency = 0
For each $s_i$ Є Short AND username($s_i$) == $u_i$
Total_clicks($s_i$) = 0
DampningFactor($s_i$) = Session_recency + Receny_wt
Session_recency = DampningFactor($s_i$)
```

```
Total_clicks($s_i$) += $clicks_{ci}$(si)                ∀ $ci$ Є C
Active_time($s_i$) = NC * Total_clicks($s_i$)//NC is genuine time gap between 2 clicks
Dur_secs($s_i$) = hours *3600 + minutes * 60 + seconds // calculates actual duration of session
If (dur_secs($s_i$) <Active_time($s_i$))
Then $Wt_{dur}(s_i) = 0.5 *( 1 - \frac{dur\_secs(si)}{\sqrt{(dur\_secs(si))^2+100}} )$ Else
                    $Wt_{dur}(s_i) = 0.5 *( 1 - \frac{Active\_time(si)}{\sqrt{(Active\_time(si))^2+100}} )$
End if
$clicks_{ci}$ += (0.5 *$clicks_{ci}$(si)) + (0.3 * $Wt_{dur}(s_i)$)+ (0.2 * DampningFactor($s_i$))  ∀ $ci$ Є C
End for //Make an entry for user $u_i$ in table for short term profile.
End for
```

Algorithm for profiling of static user preferences:

```
U = Set of registered users
C = Set of all categories.
ses = Set of all Sessions
S = Set of sessions in past N months.
// Static user profile
For each $ses_i$Є ses //session wise calculations
$clicks_{ci}$ = 0       ∀ $ci$ Є C
Total_clicks(sesi) = Total_clicks(sesi) + $clicks_{ci}$(sesi)
                    ∀ $ci$ Є C
                    $clicks_{ci} = \frac{clicks_{ci}(sesi)}{Total\_clicks(sesi)}$ ∀ $ci$ Є C
Update database for session's frequency.
End for //all sessions in past N months
For each $u_i$ Є U
$clicks_{ci}$ = 0       ∀ $ci$ Є C
        For each $s_i$Є S
If(username($s_i$) == $u_i$) Then $clicks_{ci}$ = $clicks_{ci}$ + $clicks_{ci}$(si)          ∀ $ci$ Є C
End if // updates category wise total clicks of each user
End for Update database of static user preferences.
End for
```

These algorithms are proposed to capture static and dynamic user preferences separately, in different news concepts. Same is the process of capturing static and dynamic user preferences in different news entities. These proposed algorithms makes four profiles for each user as shown in figure 4, which are analyzed intelligently for outlier detection and rectification in second part of profiling.

E.Capturing user preferences explicitly

New user will be explicitly given the list of major categories and entities, to select preferences (fig 19). Pre-classified items of preferred category will be recommended to the user. User preferences are fine tuned as the user interacts with the system. We propose this step to handle cold start problem for new users and first user.

F.Outlier Detection and Rectification

Separate profiles are proposed in our approach for each user due to high dynamism in their interests: two profiles for static interests in news category and entity and another two profiles for dynamic interests in news category and entity. Now the question arises "Are these dynamic and static profiles of users disjoint with each other?" Answer is big NO. Deep analysis of user profile shows that it is not just enough to capture the static and dynamic interests of user separately for correct recommendations. Certain situations drags user's short term interest towards a specific class (or entity) of news, which deviates long term user interests remarkably. This analysis of the profiles helped to detect the loophole because of which the results were deviating and did not match the user's expectations. In proposed approach we have added few steps to identify relative outlier values in user interests. Computation takes place daily for all the sessions for all the users. These calculations are the basis of actual relative outlier values in user's short term profile (duration for analysis of dynamic profiles is sessions in past 15 days); affecting user's long term static interests (duration for analysis of static profiles is sessions in past 3 months). Algorithm has been designed and implemented to handle this problem. Results show improvement in recommendation accuracy after outlier analysis. Algorithm can handle the situations where there are more than one outlier values. The skewed results causes wrong ranking of recommendations. Proposed designed algorithm is as follows:

In the algorithm, U = Set of registered users, C = Set of all categories, D = Set of most recent DS days for short term analysis, S = set of sessions in DS days. D_long = set of most recent N months for long term analysis and S_long = set of sessions in N months.

```
For each ui Є U // Algorithm for outlier detection
$clicks_{ci}$ = 0       ∀ ci Є C
clck_lngci = 0    ∀ ci Є C
 For each di Є D
$clicks_{ci}$+= $clicks_{ci}$(si)    ∀siЄ S : si(logindate) Є di,  ∀ciЄ C
avg = $\sum_{ci \in C} \frac{clicks_{ci}}{|C|}$ // average of category wise clicks

Threshold = avg *x%  // sets new threshold each day based on avg clicks
cum_sum = 0
For each ci Є C
cum_sum = cum_sum + $clicks_{ci}$ – avg
 If (cum_sum> Threshold)  Then
ci = outlier in di   // day wise outlier category
                                        Threshold = cum_sum
                                Endif
                        End for
        End for
avg = 0, tot = 0
For each di Є D
For each ci Є C
If (ci == outlier in di)  Then tot += z% * clicksci // considers only z% clicks of outlier category
                        Else  tot += clicksci
                        End if
                        End for
End for
For each di Є D_long
clck_lngci += clck_lngci (si)      ∀ci Є C, si Є S_long
//Calculates category wise clicks in long term

End for
            avg = tot/|C|
Threshold = avg, cum_sum = 0
For each ci Є C
            cum_sum = cum_sum + $clicks_{ci}$ – Threshold
            If (cum_sum> Threshold)   Then
            Identify if clicksci>= y% * clck_lngci, then
            clicksci in long_term = Threshold
            Threshold = cum_sum
            End if
            End for
End for
```

Proposed approach does not remove an outlier category but only re-rank it in static user preferences. Experimental analysis shows improvement in ranking of recommendations after outlier analysis. Semantic News profiling and focused deep analysis of user profiling is followed by recommendations capable of handling issue of diversity in different perspectives, which is ignored by most of the recommender systems.

Correct ranking of actual user preferences becomes the basis of semantic diffusion to move in appropriate direction.

G.Dual Semantic Diffusion of User Preferences

We propose dual semantic diffusion of user preferences. Diffusion improves recommendation coverage because number of recommended news items increases. Novelty and Diversity is also achieved in coverage of recommendations.

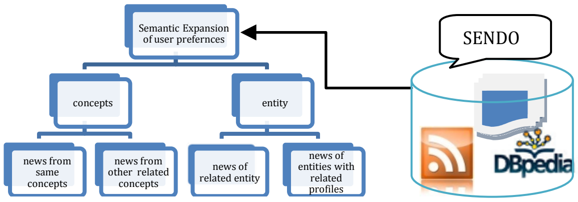

Semantic Expansion of user prefernces
SENDO
concepts
entity
news from same concepts
news from other related concepts
news of related entity
news of entities with related profiles
DBpedia

Fig.6 Semantic Expansion using SENDO

Proposed approach is ontology based diffusion of user interest in news concept as well as news entities (fig. 6). Proposed algorithm makes use of generalization and specialization scores given to ontology links. Generalization is moving from lower level towards upper level. Specialization is moving from upper level towards lower level. Apart from these links it also makes use of property link connecting two or more leaf nodes semantically via an object property in SENDO. Basically in content based recommender systems, recommended items are those items which are too similar to those previously rated by the user leading to over-specialization problem. Proposed approach successfully curbs the problem of over-specialization. It not just recommends the entities of user's direct interest but also the news of other semantically related entities. It also recommends news items of new categories which are inferred and found related to the preferred categories. Designed algorithms for concept based semantic expansion and entity based semantic expansion are as follows:

Required Input for concept based semantic expansion algorithm is (1) ranked long term and short term interests of users after outlier analysis and (2) SENDO. In output algorithm will give news items of predicted new concepts bringing diversity and novelty

In algorithm, [C]=Set of concepts in O, [S]=Set of sub concepts of a concept C', [L]=Set of leaf nodes of a sub concept S', [U]=Set of users.

T1, T2 and T3 are thresholds such that T3<T1<T2

```
For each user Ui∈ [U]
For each Ci∈ [C] and Cw>T1 where Cw is weight calculated for a concept based on user's long term interest.
For each max{L'}:L'∈Ci and L'w>T2
V=Gs*Ss*max (Ln) //Gs=generalization score in O, Ss=Specialization score in O,
Ln=normalized interest score of leaf node with maximum value in O
For each L'∈S'i
Lnew=L'w+V
If Lnew>T3
Recommend news at Lnew
If Lnew∈domain of any property in ontology
            Recommend all those news
If Lnew∈ range of any property in Ontology
Recommend all those news
Endif
Endfor
Endfor
 For each Si∈Ci: Si has no further leaf nodes
If Si∈domain of any property in ontology
        Recommend all those news
If Si∈ range of any property in Ontology
Recommend all those news
Endfor
Endfor
Endfor
```

H.Similarly user preferences in entities are semantically expanded using proposed given algorithm

For given Set of logged in users 'UL' and Set of Named Entities 'E', Login_date (ui) = last date of most recent N months for user ui and first_date (ui) = first date of most recent N months for user ui.

```
//ENTITY EXPANSION
For each ui Є UL
```

```
Clicksei = 0   ∀ei Є E
If (current session date <login_date (ui) AND current session date >= first_date (ui)) Then
For each news item NW viewed by ui
        If (NW contains ei)   ∀ei Є E
        Then
        Clicksei += 1
Update table for user ui with entity ei and clicksi
                End if
        End for
        Else
setfirst_date (ui) = current session date
setlogin_date(ui) = current session date + N months(N*30 days)
        delete all entries for user ui from the table
        For each news item NW viewed by ui
        If (NW contains ei)   ∀ei Є E
        Then
                Clicksei += 1
Update table for user ui with entity ei and clicksi
End if
End for
End if
max_entity (ui) = $\max_{ei \in E}$ clicks(ei)   from the table for user ui
Store news ids for max_entity (ui) from system ontology in new table
Recommends news for news id from new table
End for
```

Achieved diffusion in user preferences assists in recommending news of new concepts and entities semantically linked with user basic preferences.

I.Bringing Multifarious Ranked Novelty and Diversity

Using the proposed approach of semantic expansion, recommendation coverage improves. New recommendations are analyzed for novelty and diversity in coverage. Novelty and diversity are different though related notions. Novelty is user specific and diversity is item set based. Both are related as novel recommendations bring diversity. Both are different as it cannot be true that diverse recommendations are novel also.

Proposed approach brings global and local novelty and diversity, based upon following factors:

1. News Source
2. News Entity
3. News Concept

The ontology which is being considered for semantic diffusion of user preferences is a large structure. Numbers of subcategories in a major category ranges from minimum 8 to maximum 103. Similarly total number of leaf nodes per sub category ranges from minimum 1 to maximum 475. News items are mapped at leaf nodes of the hierarchy. While bringing diversity and novelty in coverage, if a new major category is added in user interest then it will subsequently add large number of new leaf categories as new recommendations. It brings so much diversity in coverage that it may confuse and irritate the user, resulting in user's total distrust in the system. Therefore in bringing diversity proposed approach puts some checks. A new major category will not be added in user's basic interests because all major categories like Politics, Education and Sports are disjoint. All semantically linked sub-categories and leaf categories are added in recommendations. This will bring limited reasonable logical diversity and novelty in recommendations (Vargas et al., 2011), acceptable by end users.

Diversity generally applies to a set of items, and is related to how different the items are with respect to each other. This is related to novelty in that when a set is diverse, each item is "novel" with respect to the rest of the set. As mentioned above, we have considered the change in user preferences with passing time in both dynamic and static interests. Similarly diversity in news recommendations must also show temporal effects. Results of diversity will also be affected by ranking of user preferences and will also be affected by outlier values in user preferences. We consider fine tuned preferences of users to bring ranked diversity (Vargas et al., 2011) in recommendations based on the temporal changes in preferences of individual user or user community as a whole. Diversity can be Local Diversity or Global Diversity based on the preferences of specific user or aggregate preferences of all the users.

For Global Diversity preferences of all the registered users are analyzed and:

1. Entity of maximum interests in each entity type among all the users is recommend to all
2. Major Concept and its leaf concept of maximum interest among all the users is recommend to all
3. Most preferred news source is recommended to all

For local diversity preferences of individual user are analyzed:

1. New news of new related sub concept from major concepts of user preference
2. Entity (person profile) based diversity (as shown in figure 15). For example, User interested in news of entity X say a singer/player will be recommended latest news about all the singers/players respectively.

An example is given (as shown in figure 7) to explain local diversity due to new sub concepts of a major concept (of user liking). In SENDO, there is a major category "Politics" with news code '11,000,000'. It has 28 sub concepts and 54 leaf concepts as mentioned earlier also. A user is found interested in news of sub category "Elections". Few of the leaf concepts of user interest are "Political Candidates", "National Elections" and "Regional Elections". Leaf concept “Political Candidate” is related via SENDO object property “member_of” with other two leaf concepts “Upper house” and “lower house”. These two leaf concepts belong to sub concept “Parliament”. Recommending news of this new sub concept “parliament” (having news mapped at leaf concepts “upper house” and “lower house”) will bring diversity in coverage of recommendations.

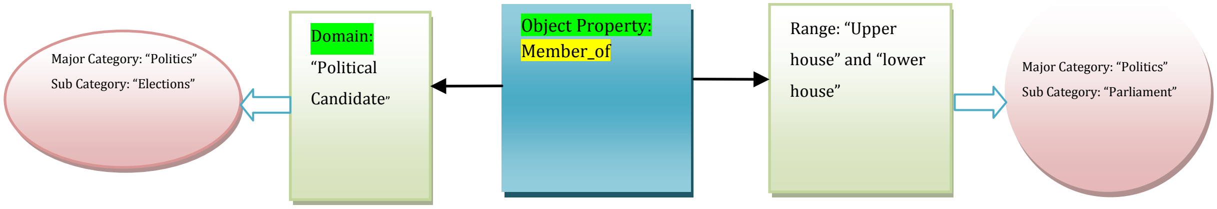

Major Category: “Politics”
Sub Category: “Elections”
Domain:
“Political Candidate”
Object Property:
Member_of
Range: “Upper house” and “lower house”
Major Category: “Politics”
Sub Category: “Parliament”

Fig.7 Example to explain domain and range of an object property having same major category but different sub category

The novelty of a piece of information generally refers to how different it is with respect to “what has been previously seen”, by a specific user, or by a community as a whole. Novelty is based on popularity and distance. Recommending popular items brings low novelty. Similarly recommending items of the similar nature what user has picked earlier also, brings low novelty. In case of news recommender which is highly dynamic domain, we have to consider the change in user preferences with passing time in both dynamic and static interests. Novelty is based primarily on user preferences. So while capturing user preferences neutralizing the effect of outlier values shows positive improvement in ranking of short term and long term recommendations. This will fine tune ranking of user preferences. Novelty can be Global Novelty or Local Novelty based on the preferences of specific user and preferences of all the users. We are considering temporal effects in both local and global novelty in news recommendations. We will consider fine tuned preferences of users to bring novelty in recommendations based on the temporal changes in preferences of user or user community as a whole.

For global novelty preferences of all the registered users are analyzed:

1. News items most preferred and least preferred among all the categories
2. News items most preferred and least preferred among each category (only in those categories in which user has interest)
3. Most preferred and least preferred concept/ category
4. Most preferred and least preferred individual (only having same profile of entity of user interests)

For local novelty preferences of individual user has been analysed

1. New news of same entity of user preference (From any concept) (as shown in figure 15)
2. New news of same sub concept because of new leaf concept
3. Next section experimentally proves the achieved results.

Global Diversity and global novelty also helps to curb the zero recommendation problems for new users.

IV. EXPERIMENTAL STUDY AND EVALUATIONS

A.Setup

Code of the proposed approach has been written in PHP. Aggregated and classified news items are stored in MySql on XAMPP server. The knowledge base consisting of ontology has been designed using Protégé. To interact with Ontologies as well as for reasoning, Protégé OWL API (a Java open source library) has been used. NetBeans IDE is used to write java codes. For Identification of named entities Stanford and Alchemy Named Entity Recognition tools have been used.

B.Experimental Study

Evaluations in table 1 shows that outlier analysis results in correct ranking of preferences. Same is applicable for all those users, having outlier existence in their profiles (figure 8 and 9). Encircled values in table 1 represent the detected outlier in user profiles using proposed algorithm. Shaded part shows correct ranking of user preference after making corrections. In the table News Category P=”Politics”, E=”Education”, ACE=”Art, Culture and Entertainment”, DA=”Disaster and Accident”, HI=”Human Interest”, SI=”Social Issues”, W=”Weather”, UCW=”Unrest Conflict and War”, ST=”Science and Technology”, CLJ=”Crime Law and Justice”.

TABLE 1 OUTLIER DETECTION AND CORRECTION

Rectified Ranking of Preferences

Users	News Category(User Preference score)

tripti@gmail.com	DA(2.91781), HI(2.53530), SI(0.64124), W(0.53778), E (0.36785)
	HI(2.53530), SI(0.64124), W(0.53778), D(0.37407), E(0.36785)

asheesh@gmail.com	HI (2.30556), E(2.00000), SI(1.00000) W(0.47222) DA(0.22222) HI(2.30556) SI(1.00000) W(0.47222) E(0.36890) DA(0.22222)

soham@gmail.com	DA(2.33334), UCW(2.00000), ST(1.16666), P(0.50000), CLJ(0.50000)
	UCW(2.00000), DA(1.57480), ST(1.16666), CLJ(0.50000), P(0.04412)

shikha@rediffmail.com	E(2.25000), P(1.70692), CLJ(0.61538), W(0.15384), ST(0.15384)
	P(1.70692), E(1.5407), CLJ(0.61538), W(0.15384), ST(0.15384)

satvika@gmail.com	P(1.51164), ACE(0.64286), E(0.21429), DA(0.08403) P(1.51164), ACE(0.64286), E(0.21429), DA(0.08403)

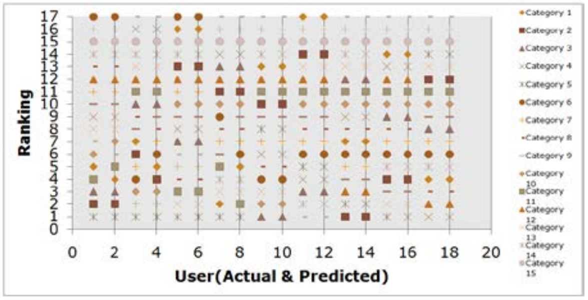


Fig.8 User's actual and predicted ranking

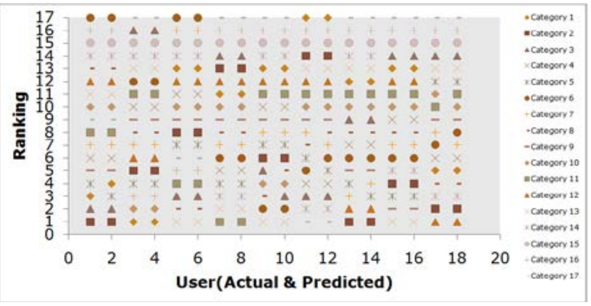


Fig.9 User's actual and predicted ranking

In Figure 8 each pair of bar depicts user's actual and predicted ranking of items for captured short term interest in news categories. Similarly each pair of bar in figure 9 shows user's actual and predicted ranking of items for captured long term interest in news categories using proposed approach of user profiling for accurate recommendations in correct ranking.

C.Evaluations to prove Diversity and Novelty in Recommendation coverage

For all the users, in each major category having user interests if number of new recommendations are larger than number of previous recommendations then it depicts improvement in recommendation coverage.

Coverage = Novelty + Diversity (as shown in figure 14)
If a new leaf node category within preferred sub concept is found in recommendations: It means Novelty in recommendations is achieved and if new sub category is found in preferred category: It means diversity in recommendations is achieved.

Evaluations are based on the formula NR=Y-X, where, X=items recommended before proposed approach of expansion for all leaf nodes of sub category i, Y=items recommended after proposed approach of concept expansion for all leaf nodes of sub cat category I and number of new recommendations NR=new leaf nodes recommended in sub category i. In calculations,

If X=0 and Y>0, It means this new sub category has been added (as shown in figure 10 and 11) in the recommendation list of a user

If Y=X and X>=0, It means that no new leaf of this new sub category has been recommended to the user. (Zero novelty in category i)

If X>0 and Y>X, it means news leaf nodes of the same subcategory of user interests has been added in the recommendation list (figure 10 and 11)
It is never possible that X=num and Y<num where num =+ive integer

D.Results showing improvement in Recommendations Diversity and Recommendations Novelty in ranked order

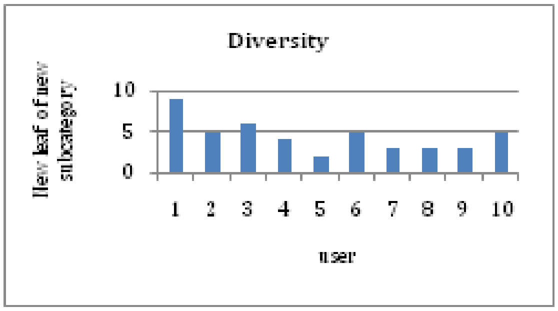


Fig.10 Diversity in recommendations in all categories

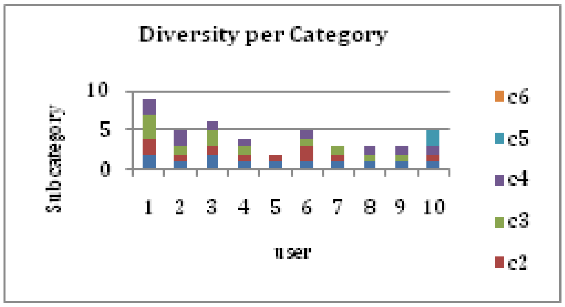


Fig.11 Diversity in recommendations due to different sub categories

Figure 10 shows Diversity in recommendations in all categories of user preference and figure 11 shows Diversity in recommendations due to different sub categories per categories of user preference.

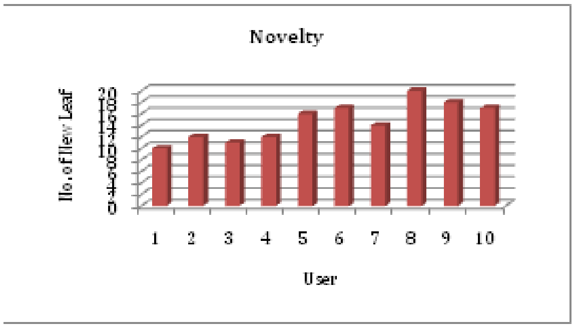


Fig.12 Novelty in recommendation

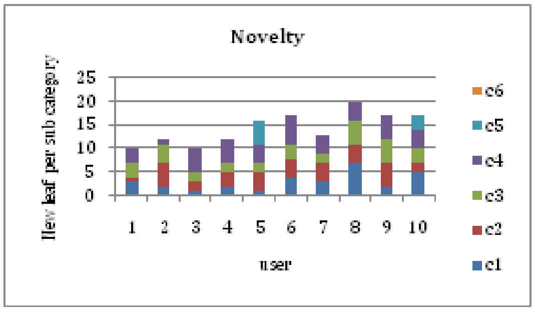


Fig.13 recommendations due to new leaf category

Figure 12 shows Novelty in recommendations in total leaf category of all sub categories of user preference and Figure 13 shows Novelty in recommendations due to new leaf category per sub categories of user preference.

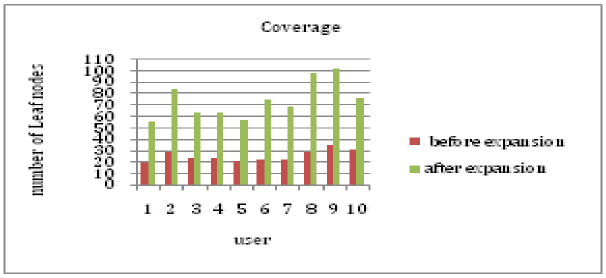


Fig.14 Comparison of coverage

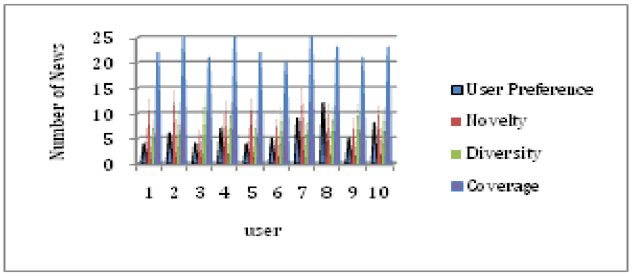


Fig.15 coverage, novelty (due to same entity)

Figure 14 shows Comparison of coverage in recommendations before expansion and after expansion of user preferences using proposed approach. Figure 15 shows comparison of coverage, novelty (due to same entity) and diversity (due to same profile of more than one person entity) after proposed expansion approach, based on user's maximum preference in Named Entity.

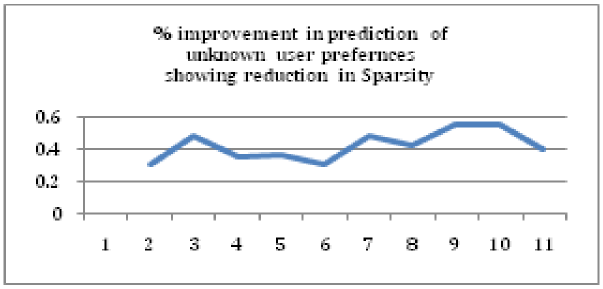


Fig.16 new predictions after expansion

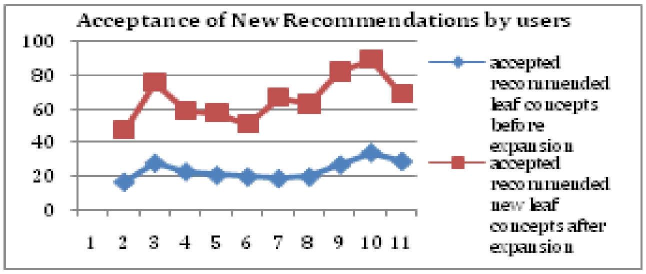
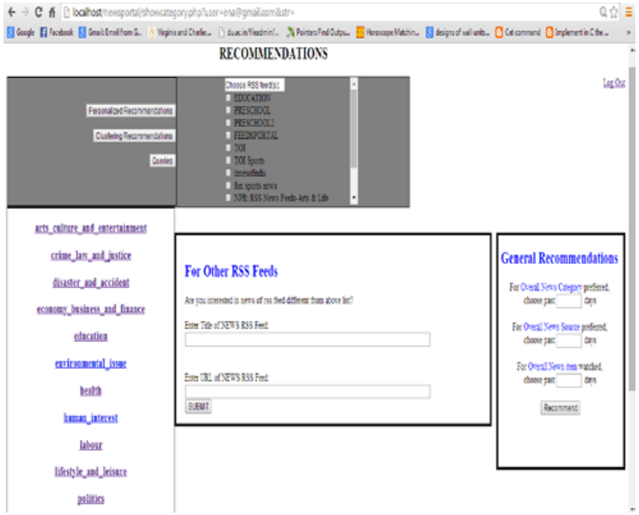


Fig.17 comparison of acceptance of recommendations



RECOMMENDATIONS

Personalized Recommendations

Clustering Recommendations

Queries

Choose RSS feeds:

- EDUCATION
- PRESCHOOL
- PRESCHOOL1
- FEEDSPORTAL
- TOI
- TOI Sports
- timesofindia
- fox sports news
- NPR: RSS News Feeds-Arts & Life

Log Out

arts_culture_and_entertainment

crime_law_and_justice

disaster_and_accident

economy_business_and_finance

education

environmental_issue

health

human_interest

labour

lifestyle_and_leisure

politics

For Other RSS Feeds

Are you interested in news of rss feed different from above list?

Enter Title of NEWS RSS Feed:

Enter URL of NEWS RSS Feed:

SUBMIT

General Recommendations

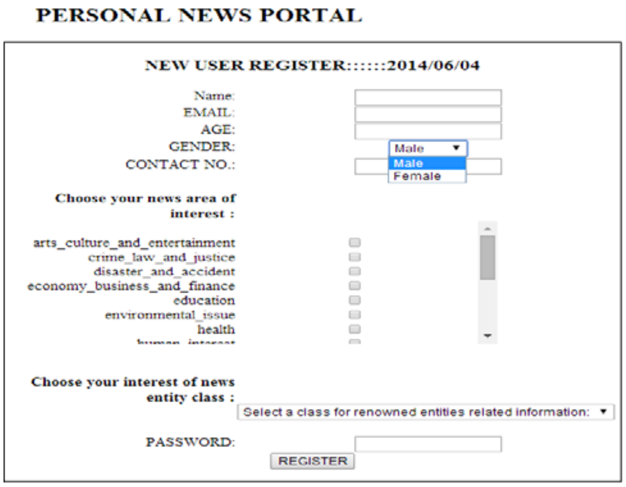
For Overall News Category preferred, choose past ____ days

For Overall News Source preferred, choose past ____ days

For Overall News item watched, choose past ____ days

Recommend

Fig.18 A snapshot of the recommendations made on the portal


PERSONAL NEWS PORTAL

NEW USER REGISTER:::::2014/06/04

Name:

EMAIL:

AGE:

GENDER: Male

Male

Female

CONTACT NO.:

Choose your news area of interest :

arts_culture_and_entertainment □

crime_law_and_justice □

disaster_and_accident □

economy_business_and_finance □

education □

environmental_issue □

health □

human_interest □

Choose your interest of news entity class : Select a class for renowned entities related information:

PASSWORD:

REGISTER

Fig.19 A snapshot of explicit user profiling

Figure 16 shows percentage of new predictions after expansion, resulting in reduction of sparsity in user profiles and figure 17 shows comparison of acceptance of recommendations before expansion and acceptance after expansion. This shows user's interest in recommendations

In fig 18 snapshot shows the options for personalized and general recommendations given on the designed portal. Hyperlinked list of news categories contains news mapped into these categories after classification. Option to choose news source from a list has also been given considering the dynamic preference of user. New RSS feed can also be added in the list. General recommendations are open for all the users. In Fig 19 it is shown that user can explicitly specify preferred news category or entity at the time of registration. Pre-classified news will be offered to the user based on specified preferences. It helps to curb the cold start problem for new user.

V. RELATED WORK

We have scanned the work done in the area in past few years and compared with our approach as given here. Alajendro et al. (2013) have considered just user preference history to give just accurate recommendations in news domain. This domain is highly dynamic and authors have made a single profile for each user which is unable to capture both the static and the dynamic user preferences. Moreover outlier analysis is not done in user preferences. They have not even considered user preference in news entity apart from news concept, which we have considered. Authors have handled the issue of novelty in recommendations which according to authors means recommending latest news. This is not so, because novelty means recommending the new items of category or entity of user preference which was there in the data but user was not aware of and therefore was not watching. Using expansion of user preferences based on ontology knowledge we have handled the issue of over specialization, diversity and novelty in recommendations. It helps to reduce sparsity also.

Zhanglian *et al.* (2012) have applied semantic expansion. Improvement over cold start, diversity and accuracy has been shown. Usefulness of this diversity for end user has not been shown. Ting Peng et al. (2008) have worked on explicit user profiling which is not based on ontology. Network used to capture user preferences is not able to consider different linkages. Recommendations made are not evaluated by authors for bringing novelty and diversity which are important issues in recommendations.

Hao et al. (2012) have captured user preferences reflecting changes with time but one very crucial issue of unexpected change in user preference has not been given any focus. Without this user profile will surely deviate from actual preferences. Recommender system is facing the problem of over specialization because recommendations are just based on basic user preferences.

Ivo Lasek (2011) has used information from Linked Open Data to enrich content based RS. We have used it to capture user preferences both in news category as well as entity. It also helps to bring entity based diversity in user preferences along with content based diversity in preferences. Entity based diversity is dependent on two factors: same entity individual and same entity profile. The information is extracted from DBpedia. Major limitation of approach of

Ahu *et al.* (2007) is that no temporal analysis and outlier analysis is done for user profiling, failing which system is surely unable to capture actual user preferences. They have then re-ranked search results based on ontological user profiles and spreading activation method in the domain of web search. Without outlier detection and correction of wrong ranking of user preferences, spreading activation will give misleading results which becomes the cause of user's distrust in the system.

Jiahui *et al.* (2010) have built two profiles incorporating time based variations due to dynamic nature of the domain. No outlier analysis is done in user preferences. Handling the issue of over specialization in recommendations is not given focus.

Jose *et. al.* (2011) have put in efforts to tune the parameters in spreading activation algorithm to recommend new concepts in medical domain. We have worked on spreading activation using not only generalization, specialization semantic linkages but property linkage also. It helps to bring more logical diversity and novelty to the coverage. Authors have not shown the results in terms of improvement in recommendation accuracy and diversity. Our approach has shown improvement in recommendation novelty and diversity. We have also worked on bridging the gap of missing information in recommendations, transparency and trustworthiness in recommendations.

Qi Liu *et al.* (2012) have used expansion approach for user preferences. Their approach is not based on semantics leading to ambiguous results. LeiLi et al. (2011) have not worked on novelty, transparency in recommendations. They have not worked on providing missing data in news recommendations. Usage of semantically enriched ontology populated with related information from external knowledge source, provides extra information about entities.

VI. CONCLUSION AND FUTURE WORK

In this paper we have proposed an approach for semantic user profiling with focused analysis for outlier detection and rectification. Semantic profiling can handle issues of synonymy and polysemy. Outlier analysis helps to rectify the errors in ranking of user preferences due to sudden notable changes in preferences for short span of time. User interests are captured implicitly both for the news category as well as entity, in separate profiles for dynamic and static interests. Recommendation based on these accurate user preferences leads to over specialized recommendations and also sparse user profiles. Captured accurate user preferences are then fetched to the proposed algorithm for semantic expansion of preferences. Semantic expansion of these profiles making use of semantically enriched news domain ontology brings multifarious reasonable novelty and diversity in recommendation coverage, with accurate ranking. Recommendations due to different approaches are presented with proper reasons to bring transparency. This helped end user to decide whether to accept or reject new recommendations. Accepted recommendations are also updated in user profiles bringing down sparsity. We are also working on bringing reasonable diversity and positive serendipity using semantics based collaborative filtering approach.

REFERENCES

[1] Agarwal, Shikha, Singhal Archana and Bedi Punam, (2012a). IPTC based Ontological Representation of Educational News RSS Feeds, *In proceedings ITC 2012 – 3rd International Conference on Recent Trends in Information, Telecommunication and Computing*, Aug 03-04, 2012, Bangalore, India, Lecture Notes in Electrical Engineering, Vol .150, pp. 353-359, 2013.

[2] Agarwal, Shikha, Singhal Archana and Bedi Punam, (2012b). Classification of RSS Feed News Items using Ontology, *In proceedings ISDA 2012 - 12th International Conference on Intelligent System Design and Applications*, November 27-29, 2012, Kochi, India: 491- 496. USA: IEEE Xplore.

[3] Agarwal Shikha, Singhal Archana. (2013). “Autonomous Ontology Population From DBpedia based On Context Sensitive Entity Recognition” *in Fourth international joint conference on advances in engineering and technology, AET 2013*. Elsevier 2013. pp 580-589. ACEEE conference proceeding series 02.

[4] Agarwal Shikha, Singhal Archana. (2014a). “Handling Skewed Results In News Recommendations by Focused Analysis of Semantic User Profiles” *in International Conference on Reliability Optimization & Information Technology ICROIT 2014 IEEE Delhi Section.* 01/01/2014, pp. 74-79, 6p.

[5] Agarwal Shikha, Singhal Archana, Bedi Punam, Jain Ena, Gupta Gunjan. (2014b). “Proactive Predictions To Handle Issues In Recommendations” *in 4th IEEE International Advanced Computing Conference - IACC 2014*. pp.555-560, 6p.

[6] Ahu sieg, Bamshd Mobasher, Robin Burke. (2007). “Web search personalization with ontological user profiles” in CIKM’07 Lisboa, Portugal ‘07, ACM 978-1-59593-803-9/07/0011.

[7] Alejandro Montes-Garcia, Jose Maria, Jose Emilio, Marco. (2013). “Towards a journalist based news recommender system: The Wesomender approach*” in International Journal: Expert Systems with Applications 40* (2013) 6735-6741.

[8] C.C Chen, M.C Chen, Y. Sun, (2008). “ PVA: A self-adaptive personal view agent system” *in Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining.*

[9] D. Goldberg, D. Nichols, B. Oki, and D. Terry, (1992). Using collaborative filtering to weave an information tapestry. *Communications of the Association of Computing Machinery,* Vol. 35, No.12, pp.61-70, 1992.

[10] T.R. Gruber, (1993). “A translation approach to portable ontology specifications” in Knowledge Acquisition Vol. 5, No.2 , pp.199-220, 1993.

[11] Hao Wen, Liping Fan, Ling Guan. (2012). “A *hybrid Approach for personalized recommendation of news on the web” in International Journal: Exper systems with Applications 39*, 5806-5814.

[12] Ivo Lasek. (2011). “DC Proposal: Model for News Filtering with Named Entities” *in International Semantic Web Conference (2) 2011*: pp.309-316.

[13] Jiahui Liu, Peer Dolan, Elin Ronby Pedersen. (2010). “Personalized News Recommendations based on click behavior” *in Proceedings of the 15th international conference on Intelligent user interfaces ACM*, pp. 31-40.

[14] Jose Maria Alvarez, Luis Polo, Weena et.al. (2011). “Application of the spreading activation Technique for recommending concepts of well known ontologies in medical systems” in ACM-BCB Chicago, USA 978-1-4503-0796-3/11/08.

[15] Lei Li, Dingding Wang, Tao Li et al. (2011). “SCENE : A Scalable Two-Stage Personalized News Recommendation System” in SIGIR’11, July 24–28, 2011, Beijing, China Copyright 2011 ACM 978-1-4503-0757-4/11/07.

[16] Liang Ting-Peng, Yung-Fang Yang, Deng-Neng Chen, Yi-Cheng Ku. (2008). “A semantic expansion approach to personalized knowledge recommendation” in ScienceDirect Decision Support Systems 45, pp. 401-412.

[17] Lv, Yuanhua., T. Moon, P. Kolari, Z. Zheng, X. Wang, Y. Chang, (2011). “Learning to model relatedness for news recommendation”. *In: Proceedings of the 20th international conference on World wide web*. pp. 57–66. WWW ’11, ACM, New York, NY, USA (2011).

[18] P. Melville, and V. Sindhwani, (2010) Recommender Systems. In C. Sammut and G. Webb (Eds.) *Encyclopedia of Machine Learning*, pp.829-837, Springer, 2010.

[19] MouZhi Ge, Carla, Dietmar, (2010). “Beyond Accuracy: Evalauting Recommender Systems by Coverage and Serendipity” *in RecSys Barcelona, Spain ACM 978-1-60558-906-0/10/09*.

[20] Neches Robert, Richard Fikes, Tim Finin, Thomas Gruber, Ramesh Patil, Ted Senator, and William R. Swartout., (1991). “Enabling

Technology for Knowledge Sharing" *in AI Magazine*, Vol. 12, No.3 ,pp.36-56, 1991.

[21] Qi Liu, Enhong Chen, Chris H. Q. Ding, (2012). "Enhancing Collaborative Filtering by User Interest Expansion via Personalized Ranking" in *IEEE Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics*, Vol. 42, No. 1, February 2012, 1083-4419/$26.00 © 2011 IEEE.

[22] F. Ricci, L. Rokach, B. Shapira and P.B. Kantor (2011). (Eds.) Recommender Systems Handbook, 842 pages. Springer, 2011.

[23] Shawn R. Wolfe and Yi Zhang, (2010). "Interaction and Personalization of Criteria in recommender Systems" *in LNCS 6075, Springer-Verlag Berlin Heidelberg*, pp. 183–194.

[24] Uschold, Mike, and Michael Gruninger, (1996). "Ontologies: Principles, methods and applications" in Knowledge engineering review Vol.11, No.2: pp.93-136.

[25] S. Vargas, and P. Castells, (2011). "Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems" in 8th ACM *International Conference on Recommender Systems (RecSys 2011). Chicago, IL*, 109-116.

[26] Zhenglian SU_, Jun YAN, Haifeng LING, Haisong CHEN. (2012). "Research on Personalized Recommendation Algorithm Based on Ontological User Interest Model" *in Journal of Computational Information Systems 8,* pp.1 169–181.