

Table of Contents

Abstract	xvi
Acknowledgement	xvii
List of Abbreviations	xxv
List of Symbols	xxvii
List of Figures	xxxi
List of Tables	xxxii
CHAPTER – 1: INTRODUCTION	1
1.1 Introduction to Recommender Systems	1
1.2 Types of Accuracy-oriented Recommender Systems	2
1.2.1 Collaborative Recommender Systems	3
1.2.2 Content-based Recommender Systems	4
1.2.3 Knowledge-based Recommender Systems	7
1.2.4 Context-aware Recommender Systems	9
1.2.5 Demographic Recommender Systems	12
1.2.6 Hybrid Recommender Systems	13
1.3 Characteristics of Recommender Systems	15
1.3.1 Recommendation Accuracy	15
1.3.2 Trustworthiness	16
1.3.3 Diversity of Recommendations	16
1.3.4 Recommendation Value	16
1.3.5 Robustness of the Recommender System	17
1.3.6 Privacy Issue	18
1.4 Applications of Recommender Systems	20

1.5 Evaluation Strategies of Recommender Systems	21
1.6 Evaluation Metrics of Accuracy-Oriented Recommender Systems	23
1.6.1 Prediction Accuracy	23
1.6.2 Rank Accuracy	23
1.7 Challenges of Recommender Systems	25
1.8 Ethical issues of Recommender Systems	28
1.8.1 User Privacy	28
1.8.2 Avoiding Filter Bubbles	30
1.8.3 Bias and Fairness	31
1.8.4 Manipulation and Influence	33
1.8.5 Transparency and Explainability	34
1.8.6 Informed Consent	36
1.8.7 Algorithmic Accountability	37
1.8.8 Serendipity at the Expense of Accuracy	39
1.8.9 Long-Term Impacts	40
1.8.10 User Control and Autonomy	42
1.9 Motivation	43
1.10 What is Serendipity?	44
1.11 Components of Serendipity	45
1.12 Thesis Organization	48
1.13 Summary	49
CHAPTER – 2: LITERATURE SURVEY	51
2.1 Overview	51
2.2 Types of Serendipitous Recommender Systems	53
2.3 Existing Serendipitous Recommender Systems	54
2.3.1 Reranking algorithms (Reranking)	54

2.3.1.1 Greedy Algorithm	54
2.3.1.2 Personalized Re-ranking Model	55
2.3.1.3 Utility Model	56
2.3.1.4 Auralist	57
2.3.1.5 PSR (Probabilistic Serendipitous Reranking model)	58
2.3.1.6 News Reco Model	59
2.3.2 Serendipity-oriented modification (Modification)	59
2.3.2.1 SPR (Serendipitous Personalized Ranking)	59
2.3.2.2 Unexpectedness-augmented Utility Model	60
2.3.2.3 kFN (k Furthest Neighbor)	61
2.3.2.4 kNN with item taxonomy	62
2.3.2.5 Search-time-based-RS	62
2.3.3 Novel algorithms (New)	63
2.3.3.1 CURUMIM	63
2.3.3.2 Serendipity based on folksonomy	64
2.3.3.3 RWR-KI	64
2.3.3.4 TANGENT	65
2.3.3.5 Distance-based Model	66
2.3.4 Examples and Analysis of Real-world Serendipitous Recommender Systems	66
2.3.4.1 Examples of Real-world Serendipitous Recommender Systems	66
2.3.4.2 Evaluative summary of Real-world Serendipitous Recommender Systems	67
2.4 Comparative Analysis of Existing Serendipitous Recommender Systems	68
2.5 Evaluation Metrics of Serendipity-Oriented Recommender Systems	70
2.5.1 Component Metrics	71
2.5.1.1 Novelty Metrics	71
2.5.1.2 Unexpectedness Metrics	71

2.5.2 Full Metrics	72
2.6 Summary	73
CHAPTER – 3: RESEARCH GAP AND PROBLEM DEFINITION	75
3.1 Research Gap	75
3.2 Problem Definition	76
3.3 Summary	76
CHAPTER – 4: A NOVEL SERENDIPITOUS RECOMMENDER SYSTEM FOR LONG TAIL ITEMS INCULCATING RELEVANCE SCORES	78
4.1 Introduction	78
4.2 Proposed Algorithm [244]	79
4.2.1 Differentiating Between Regular (R) and Occasional (O) Users	79
4.2.2 Finding Unpopular Products for the intended User	80
4.2.3 Create a list of the target user's closest neighbors	81
4.2.4 Determine an item's relevance for the intended user by applying the Bhattacharyya coefficient	81
4.2.4.1 How is the Bhattacharyya coefficient, or BC, calculated?	82
4.2.4.2 Example: Bhattacharyya' coefficient determination	82
4.2.5 Determine the Novelty Score	83
4.2.6 Suggest Top N Items	83
4.3 Experiments	83
4.3.1 Dataset	83
4.3.2 Evaluation Metric	84
4.4 Results and Analysis	84
4.4.1 Analysis of User Recommendation Count	84
4.4.2 Novelty Score Evaluation	86
4.4.2.1 Novelty Score Analysis with the System's default settings	86
4.4.2.2 System parameters - Variant 1: Novelty Score Analysis	87

4.4.2.3 System parameters - Variant 2: Novelty Score Analysis	88
4.4.2.4 System parameters - Variant 3: Novelty Score Analysis	89
4.4.2.5 System parameters - Variant 4: Novelty Score Analysis	90
4.4.2.6 System parameters - Variant 5: Novelty Score Analysis	91
4.4.2.7 System parameters - Variant 6: Novelty Score Analysis	92
4.5 Summary	93
CHAPTER – 5: AN EVALUATION AND COMPARISION OF DIFFERENT RECOMMENDER SYSTEM ALGORITHMS FOR SERENDIPITY USING NOVELTY SCORES	95
5.1 Introduction	95
5.2 Methodology	96
5.2.1 Dataset	96
5.2.2 Algorithms	96
5.2.2.1 K Furthest Neighbor Algorithm	96
5.2.2.2 An Innovative Serendipitous Recommender System	97
5.2.2.3 Tangent Algorithm	97
5.3 System Flow	98
5.3.1 Novelty Score Metric	99
5.4 Results and Analysis	99
5.4.1 Novelty Score Evaluation	99
5.4.1.1 Score Distribution for Novelty	99
5.4.1.1.1 K Furthest Neighbor Algorithm	99
5.4.1.1.2 Innovative Serendipitous Recommender System	100
5.4.1.1.3 Tangent Algorithm	101
5.4.1.2 User IDs versus Novelty Scores	101
5.4.1.3 Score Plots with Novelty	103
5.4.2 An analysis comparing the mean scores for novelty	104

5.5 Summary	104
CHAPTER – 6: CONCLUSION	106
CHAPTER – 7: FUTURE SCOPE	108
LIST OF REFERENCES	109
LIST OF PUBLICATIONS	131

List of Abbreviations

Sr No	Abbreviations	Full Forms
1	RS	Recommender Syetsm
2	IoT	Internet of Things
3	CARS	Context-Aware Recommender Systems
4	GPS	Global Positioning System
5	MAE	Mean Absolute Error
6	RMSE	Root Mean Squared Error
7	P	Precision
8	R	Recall
9	F1	F1 Measure
10	MAP	Mean Average Precision
11	LOD	Linked Open Data
12	SVD	Single Value Decomposition
13	LSI	Latent Semantic Indexing
14	CF	Collaborative Filtering
15	CB	Content-Based Filterin
16	NLP	Natural language processing
17	SOG	Serendipity-Oriented Greedy
18	TD	Topic Diversification
19	xQuAD	EXplicit Query Aspect Diversification
20	LDA	Latent Dirichlet Allocation
21	PSR	Probabilistic Serendipitous Reranking

22	SPR	Serendipitous Personalized Ranking
23	kFN	k Furthest Neighbor
24	RWR	Random Walk with Restarts
25	RWR-KI	Random Walk with Restarts enhanced with Knowledge Infusion
26	PMI	Point-wise Mutual information
27	CAF	Context Aware Filtering
28	BC	Bhattacharyya coefficient
29	TF-IDF	Term Frequency - Inverse Document Frequency
30	IF-IUF	Item Frequency -Inverse User Frequency
31	NSRS	Novel Serendipitous Recommender System

List of Symbols

Sr No	Symbols	Description of the Symbols
1	R_{test}	Set of ratings in the test part
2	r_{ui}	Rating given by user u to item i
3	$\hat{r}_{ui}$	Prediction of the rating that user u will give to item i
4	U	The set of all the users of the system
5	$RS_u(k)$	The set of top - k recommendations provided by the recommender system to user u
6	REL_u	The set of items relevant for user u (rated highly by user u)
7	R_u	The set of items recommended to user u
8	i_{uk}	The item at position k in the recommendation list generated for user u
9	$rel(i_{uk})$	Relevance of item i_{uk} for user u
10	U_i	The number of recommended items that are unknown to the user
11	T_r	The total number of recommendations
12	R_s	The suggestions provided by content-based filtering
13	R_d	The recommendations made by a dissimilarity-based recommender system
14	R_{si}	An element of the list R_s
15	$U_{nex}(R_{si})$	Unexpectedness of element R_{si} belonging to list R_s
16	$P(v u)$	Probability that user u ($u \in U$) is interested in item $v \in V$
17	λ	Parameter regulating the popularity bias
18	E_u	Set of expected items

19	I_u	Items previously consumed by user u
20	I	Set of all items in the catalog
21	I_u^+	Set of items liked by user u
22	S	The set of user-item pairs whose ratings are observed
23	$\bar{S}$	The set of user-item pairs whose ratings are unobserved
24	r^+	A high rating value depending on the rating span of the dataset
25	D_s	The training data
26	(u, i, j)	Triplet suggesting that user u prefers item i to item j
27	σ	Denotes a 0-1 loss function
28	$Pop(j)$	Measure of popularity of item j
29	$ U $	The count of all users
30	$S(u)$	The set of items selected by user u
31	$diff(i, j)$	Measures how different item i is from item j
32	$sim(i, j)$	The similarity between items i and j
33	$nov_{va}^{dist1}(i, u)$	Distance-based novelty metric - 1 (measure of how novel item i is for user u based on distance)
34	$dist(i, j)$	Distance between item i and item j
35	$nov_{va}^{dist2}(i, u)$	Distance-based novelty metric - 2 (measure of how novel item i is for user u based on distance)
36	I_u	The set of all the items user u has interacted with
37	$ I_u $	Count of the set of items user u has interacted with
38	$nov_{va}^{pop1}(i, u)$	Popularity-based novelty metric - 1 (measure of how novel item i is for user u based on popularity)
39	$nov_{va}^{pop2}(i, u)$	Popularity-based novelty metric - 2 (measure of how novel

		item i is for user u based on popularity)
40	$ U_i $	Count of all the users who have interacted with item i
41	$ U $	Count of all the users in the system
42	$nov_{na}(i, u)$	Novelty metric (how novel item i is for user u)
43	cls_k	A class of item k in an item taxonomy
44	$dis(cls_j, cls_i)$	Distance between class cls_j and class cls_i
45	$PMI(i, j)$	Point-wise Mutual Information measuring the similarity between item i and item j
46	$p(i)$	Probability that any user has rated item i
47	$p(i, j)$	Probability that items i and j are rated together
48	$unexp_{kam}^{co-occ1}(i, u)$	Unexpectedness metric - 1 for user u when items i and j are co-rated
49	$unexp_{kam}^{co-occ2}(i, u)$	Unexpectedness metric - 2 for user u when items i and j are co-rated
50	$ser_{mur}(u)$	Serendipity metric for user u
51	$Pr_u(i)$	Confidence of recommending item i to user u by the examined model
52	$Prim_u(i)$	Confidence of recommending item i to user u by the primitive model
53	$rel_u(i)$	Relevance of an item i for user u
54	$ser_r_{mur}(u)$	Rank-sensitive Serendipity metric for user u
55	$count_u(j)$	Number of applicable items for user u with a rank lower or equal to j
56	PM	Set of items generated by a primitive recommender model
57	E_u	Set of items that matches the interests of user u

58	$\eta_{I_u, f(i)}$	Number of items matching attribute f from user profile I_u
59	$\eta_{I_u, max}$	Maximum number of items with the same attribute from the user profile
60	$f(i)$	Attribute of item i
61	$RS(u, i)$	Relevance of item i for user u
62	$\bar{r}(u)$	Average rating by target user u
63	Z	Normalized constant
64	$BC(u, n)$	Bhattacharyya coefficient between users u and n
65	N_k	Set of k neighbouring users for target user u
66	$r(n, i)$	Rating of item i by user n
67	$NS(i)$	Novelty score of item i
68	$IF(i)$	Item frequency for item i
69	$IUF(i)$	Inverse User Frequency for item i

List of Figures

Figure 1: Recommender system as a 'black box'	14
Figure 2: What is Serendipity?	45
Figure 3: Shortcomings of Accuracy-oriented Recommender Systems	51
Figure 4: Long Tail Problem	53
Figure 5: Popularity Bias Explanation	79
Figure 6: Proposed System Flow	79
Figure 7: Differentiating Between Regular (R) and Occasional (O) Users	80
Figure 8: Finding Unpopular Products for the intended User	81
Figure 9: User Vs Number of Recommendations	85
Figure 10: Regular Users' Novelty Scores (Default Settings)	87
Figure 11: Regular Users' Novelty Scores (Variant - 1)	88
Figure 12: Regular Users' Novelty Scores (Variant - 2)	89
Figure 13: Regular Users' Novelty Scores (Variant - 3)	90
Figure 14: Regular Users' Novelty Scores (Variant - 4)	91
Figure 15: Regular Users' Novelty Scores (Variant - 5)	92
Figure 16: Regular Users' Novelty Scores (Variant - 6)	93
Figure 17: The K-Furthest Neighbor Algorithm in Operation	97
Figure 18: Working of Innovative Serendipitous Recommender System	97
Figure 19: Working of Tangent Algorithm	98
Figure 20: Working of the overall proposed system	98
Figure 21: Distribution of Novelty Scores in the K Furthest Neighbor Algorithm	100
Figure 22: Distribution of Novelty Scores for the Innovative	100
Figure 23: Distribution of Novelty Scores for the Tangent Algorithm	101
Figure 24: K Furthest Neighbor Algorithm's Novelty Scores	102
Figure 25: Innovative Serendipitous Recommender System's Novelty Scores	102
Figure 26: Tangent Algorithm's Novelty Scores	102
Figure 27: K Furthest Neighbor Algorithm points for novelty score	103
Figure 28: Innovative Serendipitous Recommender System Points for Novelty Score	103
Figure 29: Tangent Algorithm points for novelty score	104
Figure 30: An overall comparison of the average novelty scores for all three algorithms	104

List of Tables

Table 1: Database of ratings for collaborative suggestion	4
Table 2: Book knowledge base	5
Table 3: Jyoti's Preference profile	6
Table 4: Example product assortment: digital cameras	8
Table 5: An illustration of user demographic information	13
Table 6: Evaluative summary of Real-world Serendipitous Recommender Systems.....	67
Table 7: Comparative Analysis of Existing Serendipitous Recommender Systems	68
Table 8: Statistics on user recommendations	85
Table 9: Example Suggestions	86
Table 10: System parameter default values	86
Table 11: Novelty Score Statistics (System Parameters by Default).....	87
Table 12: System parameters with altered values - Variant 1	87
Table 13: Variant - 1 Novelty Score Statistics.....	88
Table 14: System parameters with altered values - Variant 2.....	88
Table 15: Variant - 2 Novelty Score Statistics.....	89
Table 16: System parameters with altered values - Variant 3.....	89
Table 17: Variant - 3 Novelty Score Statistics.....	90
Table 18: System parameters with altered values - Variant 4.....	90
Table 19: Variant - 4 Novelty Score Statistics.....	91
Table 20: System parameters with altered values - Variant 5.....	91
Table 21: Variant - 5 Novelty Score Statistics.....	92
Table 22: System parameters with altered values - Variant 6.....	92
Table 23: Variant - 6 Novelty Score Statistics.....	93
Table 24: Statistics of Novelty Scores for the K-furthest Neighbor Algorithm	99
Table 25: Statistics on Novelty Scores for the Innovative.....	100
Table 26: Statistics on Novelty Scores for the Tangent Algorithm	101