

Recommending Diverse and Serendipitous Movies by employing user knowledge from Tag-genomes

Aniruddha Kalburgi

March 4, 2019

1 Deliverables

1. Data per user about their genre preference in decimal term vector form.
2. Implementation of an algorithm for aggregating genre preference in decimal term vector form using genre-based multi-sub-profiling approach.
3. Reduced tags data of Tag-genomes dataset using the stemming approach.
4. Implementation of an algorithm for recommendations focused on diversity and serendipity using clustering techniques.
5. Recommender system implementations to demonstrate the working of prototypes of both recommendation algorithms.
6. A file stating the movie recommendations list and their relevance scores per user for each approach.
7. A file, describing all approaches and their respective ID's.
8. Observations and evaluation results for all approaches, charts, and figures.

1.1 Evaluation methodology

There are two main algorithms; one uses users genre term vector in decimal form, while the other is uses tag-genomes as term vectors and produces diverse and serendipitous movie recommendations using clustering techniques.

To evaluate the benefits of both algorithms, we collect the movie recommendations list and its relevance scores for each user under test, from the baseline systems using traditional binary genre and tag-genome based term vectors as stated below.

1. Recommendation results using traditional genre-based term vectors.

2. Recommendations results using tag-genome based term vectors.

We then compare these results with the results from newly developed algorithms for the following combinations.

1. Recommendations results for an algorithm to recommend diverse and serendipitous movies through cluster techniques using:
 - (a) Tag-genome term vector without stemming.
 - (b) Tag-genome term vector with stemming.
2. Recommendation results for an algorithm that recommends diverse movies using decimal genre term vector formed by mean aggregating genre preferences using multi-sub-profiling approach per user.
3. Recommendation results obtained by hybridizing above two approaches using both with-stemming and without stemming tag-genome term vectors.

The comparison of the above combinational results shall be carried out based on the following list of criteria.

1. Compare in terms of the number of movies recommended by each approach.
2. Compare in terms of the number of diverse movies recommended per user.
3. Compare the number of serendipitous movies recommended per user.
4. Compare the mean relevance scores to the user profile for recommended movies:
 - (a) Per user.
 - (b) Across all users.
5. Compare the mean number of diverse and serendipitous movies recommended across all users.

The expected outcomes of the above evaluations are:

1. Prove that two newly proposed algorithms to produce better results than the traditional baseline system.
2. Answer the question of whether aggregating related tag-genomes using stemming approach has a positive impact on diverse and serendipitous movie recommendations or not.
3. Help understand the strengths and weaknesses of both algorithms in terms of diversity and serendipity.

2 Project Plan

The following images show the Work Breakdown Structure and Gantt Chart.

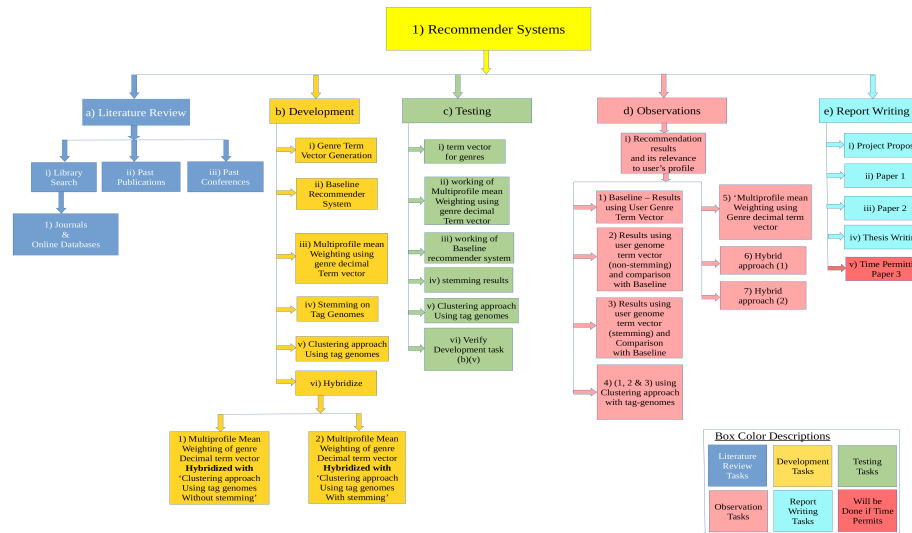


Figure 1: Project Work Breakdown Structure

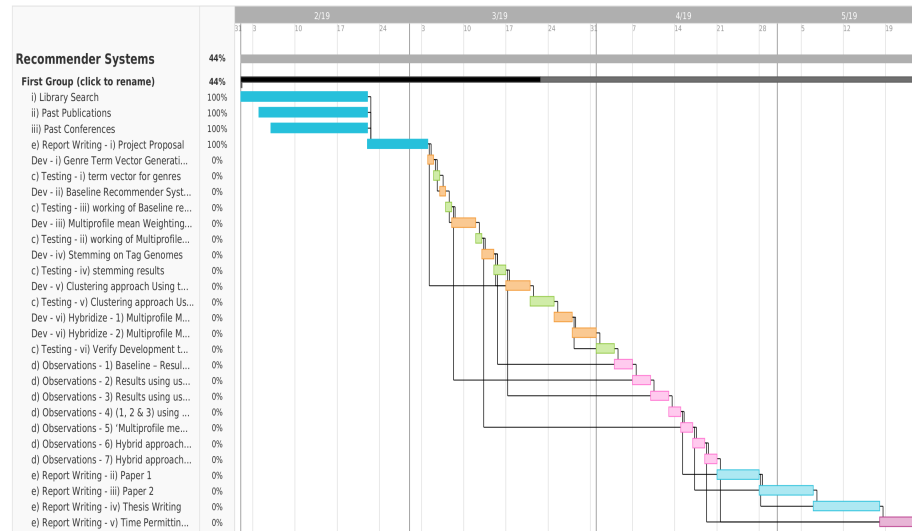


Figure 2: Project Gantt Chart

3 Literature Review

In today’s world, an online user has a massive number of items to choose from, when it comes to making choices regarding purchasing stuff online, reading the new book, listening to music, watching movies, and loads of other things available as options. The Recommender Systems helps a user by suggesting them with options of several relevant items to their profile. Often these systems are critiqued to keep the user in the bubble by recommending them the obvious items that only relate to the user’s existing preferences, also failing to surprise the user.

As stated by [4], diversity can help broaden the user’s preferences and boost chances to convince the user. A study by [7, 4, 6, 8] helps us understand that serendipity plays a vital role in diversifying users preferences. Hence in this paper, we aim to broaden user preferences and by offering serendipitous and diverse movie recommendations to the user.

We target to achieve our goals by concentrating on two important feature aspects:

1. The genre information available for each movie in the movielens dataset [2].
2. The tag-genome - almost 1100 tags to describe the movie by stating tag relevance. This dataset is published to the movielens by [10].

The following sources inspire our approaches in this paper.

Subprofiling is one of the crucial aspect used in this paper to gain broader knowledge about the user’s taste. The results achieved by [4] influence the decision to use the profiling approach. They create multiple sub-profiles for each user such that each sub-profile represents the distinct taste of a user. They further rerank the recommendations in such a manner that promotes the diversity factor. [4] achieve up to 14% improvement in precision and diversity over the baseline systems.

To learn more about serendipity definitions, we referred the work of [6], who state that crucial components of serendipity are ”relevance, novelty, and unexpectedness” with many variations for each. They consider eight different definitions of serendipity and find most of them to broaden the user’s preferences by surveying 475 real users using eight types of questions to assess each definition of serendipity. [6] claims to achieve 69% serendipity score for 437 movies serendipitous recommendations they offered to the real users. They also suggest that the entire Movielens dataset has 8.5% of serendipitous movies based on eight considered definitions. Although, this assumption relies on the serendipitous movies have an average rating equal or above 3.5 out of 5. As this work makes user rating an essential factor, although we don’t restrict the minimum rating

for serendipity, but consider it as a weight while ranking the serendipitous and diverse recommendations. As per information is shown in below table referred from [6], all definitions of serendipity seem to broaden user preferences, and six definitions seem to improve user satisfaction, complementing further exploration about serendipity techniques.

	s_ser_rel	s_ser_find	s_ser_imp	s_ser_rec	m_ser_rel	m_ser_find	m_ser_imp	m_ser_rec
Broadening	0.979*	1.471*	1.581*	1.605*	0.663	1.667*	1.354*	1.307*
Satisfaction	-0.347	0.322	0.284	0.276	-0.166	0.486	0.164	0.265

Figure 3: Referred 'Table 5' from [6]

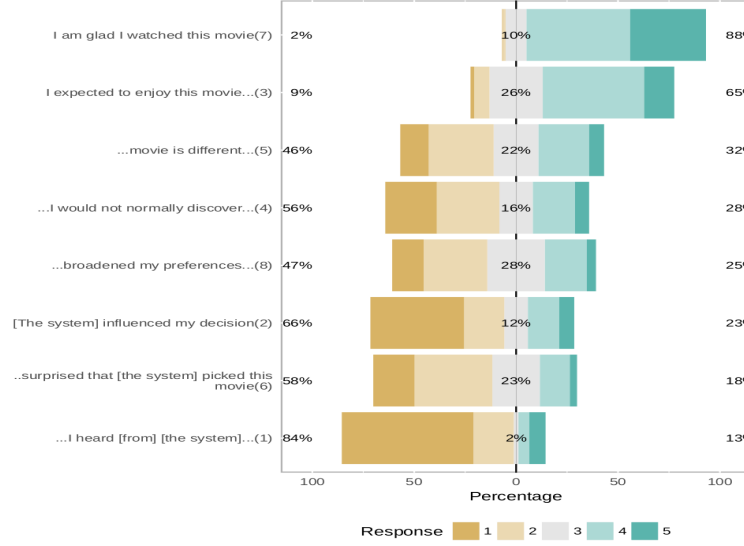


Figure 2: Distributions of answers (“1” - strongly disagree, “2” - disagree, “3” - neither agree nor disagree, “4” - agree, “5” - strongly agree)

Figure 4: Referred image 'Figure 2', along with the figure description from [6]

A further study from [6] reveals the graphical response as shown in 'Figure 4' of ratings from users for eight serendipity related questions. From novelty and unexpectedness perspective, it shows 88% users were happy to watch those movies while 9% users 'unexpectedly ended enjoying it,' and 18% were 'surprised to have those suggestions.' From the diversity perspective, 32% users "thought that the movie was different" where 28% said, "they wouldn't normally discover

it on their own.” and 25% users ‘agreed to have their preferences broadened.’ The evaluation metrics used by [6] focuses on ‘preference broadening and user feedback.’ Although due to time constraints, real user feedback isn’t possible in our case and we will consider that in our future work, yet we mainly focus on ‘preference broadening’ as one of our primitive evaluation metrics.

Movie Tuner (?)
Find similar movies with less or more of particular qualities. The movie list below will update as you indicate your preferences. Use the lock button to select multiple qualities.

This movie	What I want			
	less	ok	more	
action <input type="range" value="50"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	lock
violent <input type="range" value="80"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	
dark <input type="range" value="50"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	
crime <input type="range" value="80"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	
funny <input type="range" value="20"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	
<input type="text" value="Enter selection..."/> <input type="range" value="50"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

reset all

Similar to *Reservoir Dogs*, but **less violent**

- House of Games (1987)
- Out of Sight (1998)
- Killing, The (1956)
- Usual Suspects, The (1995)
- Four Rooms (1995)
- Blood Simple (1984)
- Get Shorty (1995)

Figure 5: Movie Tuner Interface - referred from [10]

Movielens released the tag-genome dataset [10] in 2012. The tag-genomes are approx 1100 distinct tags like 'fight, adventure, airplane, alien, and so on,' that describe the relevance scores to each movie using a decimal number between 0 and 1, where 0 indicates 'absolute no relevance,' and 1 indicates the 'strongest relevance' [10].

Various sources of these tags include user reviews, comments, blogs, meta-data such as actor names, descriptions, release date, digital book copies and many other digital forms of information available online about the movies. They then apply Machine Learning techniques to predict the tag-relevance based on the information gathered from all sources. Because these tags include a lot of contextual data, there's a high potential to learn more about user's interests using 1100 tags. They then introduce the application called 'Movie Tuner,' shown in 'Figure 5' above, which allows users to explore movies which offer less, equal or more quantity of tags, such as violence, action, funny, etc. in correspondence to the currently selected movie.

[9] used tag-genomes to predict user ratings for movies using collaborative filtering technique. They attempted to learn users preferences across all the movies watched by a user using tags which had relevances only above the threshold called ' α ' ranging from 10% to 50%. Their objective was to "find a subset of quality tags that can explain the most about users' preferences" while giving less importance to the tag-relevance. However, to avoid the well-known Sparsity problem associated with Collaborative Filtering approach, they only considered users who had rated for minimum 99 movies. Most of their models to predict user ratings shown underperformance or no improvement in terms of MAE and Precision over the traditional collaborative filtering techniques which don't use tag-genomes.

Here are a few insights on why their approach might have failed. Firstly, they ignored tag-relevances below the ' α ' threshold. Because each movie has 1100 tag-genomes, the relevance scores are spread very thinly across them. Additionally, few distinct tags logically represent the same tags, but instead, their relevance scores have are scattered across the multiple. e.g. 'Fight, Fights, and Fighting' are logically the same tags, but just because they are considered distinct, the overall good relevance score of Fight is spread thinly across three tags, hence, increasing chances of them getting ignored when only considered above threshold tags is very high. This is supposed to get only worse as they increase the ' α threshold' which is true as they have stated in conclusion. They could have instead used stemming approach, as we propose in our paper, to aggregate relevance scores for logically similar tags under the single tag and eliminate the redundant ones.

Secondly, as they are using collaborative filtering, they suffer from rating sparsity issue, hence, significantly limiting the choice of k-Nearest Users to predict ratings. The system, therefore, will also greatly suffer from the cold start problem. Content-based recommender system can address such issues.

The sparsity problem in Collaborative Filtering(CF) seems to have tackled well by [5] using "A collective matrix factorization method using tag-genomes." They begin by stating the common sparsity problem and study the techniques to reduce sparsity level such as collective matrix factorization techniques. As these methods suffer when there's a 'less relevant target domain information available.' To address this issue, [5] proposes a Gradient Descent based technique primarily using tag-genomes, as they infer that "the tag information is strongly related to the user-item matrix that is the main information of CF." They use tag-genomes along with the sparse user-item matrix to learn and make user-item matrix dense. They also achieve very good prediction results over the baseline systems using Gradient Descent to reduce the error rate in determining the missing tag values. As a result, the predicted user ratings for items are more accurate. The 'figure 6' shows their improved results over the baseline model for different sizes of sparsity.

TABLE 4. PERFORMANCE IN DIFFERENT SPARSITY

Sparsity	1%	2%	3%	4%
<i>Baseline</i> <i>($\alpha = 0, \beta = 1$)</i>	0.9262	0.8368	0.7862	0.7522
<i>Proposed method</i> <i>($\alpha = 1, \beta = 1$)</i>	0.8566	0.7998	0.7802	0.7483
<i>Improvement (%)</i>	7.5146	4.4216	0.7632	0.5185

Figure 6: Referred 'Table 4' from [5]

[8], in their work, proposes a methodology to recommend serendipitous research papers by exploiting hidden relationships with documents from another domain. Their goal is to "help researchers access relevant, useful and novel research papers," and they plan to achieve it by "offering serendipitous research papers from domains that are normally regarded as unrelated."

They start by trying to understand researchers profile by exploiting the semantic relationships based on the relevant information found about the researcher such as "email communications, social media interactions, project cooperation, domain knowledge, user background, etc." They then perform the visual and contextual analysis of the research papers from similar and different domains. Using all relational information, they form graphs with terms as nodes and related associations as connections between them, allowing to discover back and forth association with the nodes across other domains - quoted

as "BisoNets." Upon finding the 'latent relationships between different domains, they recommend such papers,' so presumed to be serendipitous.

[8], to evaluate their approach, queried the system with two keywords from the different domains - "magnesium" and "migraine." Despite the further restrictions on the search to 'return titles published only before the year 1988, the system was able to find correlations between two domains.

In the study of Recommender Systems done by [1], they calculate the diversity among the movies using the cosine similarity measure on the tag-genome vector and compare its performance against 'Item mean centered ratings' and 'Latent Feature Vectors.' From the results shown in the figure 7 taken from [1], they advise this - "our data suggest that the computing item similarities using the tag genome maps better to user-perceived list diversity than the rating or latent feature similarity." - taken from the [1], page 197, paragraph 3.

Sim. Metric	Correlation			SEM Fit	
	Genome	Rating	Lat. Feat.	Div Coef.	RMSEA
Genome	1.0	0.698	0.791	-0.345	0.0335
Rating		1.0	0.821	-0.329	0.0376
Lat. Feat.			1.0	-0.276	0.0303

Table 7.4: Summary of similarity metrics for measuring diversity

Figure 7: Referred 'Table 7.4' from [1] on page 195

From the 'Table 7.1' from [1] on page 185, we intend to consider the questions they have asked to evaluate 'Diversity, Satisfaction, and Novelty.' Some of those questions are: "Which list has more movies that are similar to each other? Which list has a more varied selection of movies?" and so on. We think these questions give us the right direction to evaluate the diversity and serendipity achieved by our models.

In paper [3], they propose a 'Genre-based Hybrid Collaborative Filtering (CGCF)' approach to predict the values for missing user ratings which yield them with up to 2.5% improvement in MAE. Instead of using the traditional binary genre term vector, they use TF-IDF technique to calculate the number of movies under each genre to form an integer value based genre term vector. Although, it suffers from the sparsity of ratings. But as using Integer Genre Term Vector helps improve MAE in CF, and also it helps understand user's taste across multiple genres, we aim using the similar term vector in content-based recommender system and explore possible trends of diversity and serendipity.

References

- [1] Michael Ekstrand. Towards recommender engineering: tools and experiments for identifying recommender differences. 2014.
- [2] F Maxwell Harper and Joseph A Konstan. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4):19, 2016.
- [3] Yajie Hu, Yi Yang, Caihong Li, Yachen Wang, and Lian Li. A hybrid genre-based personalized recommendation algorithm. In *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, pages 1369–1373. IEEE, 2016.
- [4] Mesut Kaya and Derek Bridge. Accurate and diverse recommendations using item-based subprofiles. In *The Thirty-First International Flairs Conference*, 2018.
- [5] Bu Sung Kim, Heera Kim, Jaedong Lee, and Jee-Hyong Lee. Improving a recommender system by collective matrix factorization with tag information. In *2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS)*, pages 980–984. IEEE, 2014.
- [6] Denis Kotkov, Joseph A Konstan, Qian Zhao, and Jari Veijalainen. Investigating serendipity in recommender systems based on real user feedback. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, pages 1341–1350. ACM, 2018.
- [7] Denis Kotkov, Jari Veijalainen, and Shuaiqiang Wang. How does serendipity affect diversity in recommender systems? a serendipity-oriented greedy algorithm. *Computing*, pages 1–19, 2018.
- [8] Benard Magara Maake, Sunday O Ojo, and Tranos Zuva. A serendipitous research paper recommender system.
- [9] Tien T Nguyen and John Riedl. Predicting users preference from tag relevance. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 274–280. Springer, 2013.
- [10] Jesse Vig, Shilad Sen, and John Riedl. The tag genome: Encoding community knowledge to support novel interaction. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(3):13, 2012.