A REPORT ON

FASHION RECOMMENDATION SYSTEM

SUBMITTED BY

SAKSHI BUDHIA

SHREYA BHONGALE

AKHIL PANDEY

PRATYAKSHA PANDEY

Under the Guidance of

Prof. SASHIKALA MISHRA

SYMBIOSIS INSTITUTE OF TECHNOLOGY

A CONSTITUTENT OF SYMBIOSIS INTERNATIONAL (DEEMED UNIVERSITY)

Pune – 412115

YEAR 2019-23

ABSTRACT - Fast fashion has seen remarkable expansion in the textile and garment sectors in recent years. An effective recommendation system is required for e-commerce platforms with several choices to filter, organise, and quickly deliver relevant product content or information to consumers. FRSs (image-based fashion recommendation systems) have piqued the interest of fast fashion retailers since they provide a more personalised shopping experience for clients.

Thanks to technological advancements, this field of artificial intelligence has a lot of potential in image processing, parsing, classification, and segmentation. Despite its immense potential, few academic research has been conducted on this topic. The available study does not provide a comprehensive assessment of fashion recommendation systems and their accompanying filtering methods. [9]

The study focuses on assisting the user in finding the optimal matching pair of garments by considering complex characteristics such as style, patterns, colors, textures, and user variables such as age, skin tone, preferred color, and so on. Its goal is to assist the customer in selecting trendy clothing and organizing their wardrobe. It seeks to assist the user in selecting clothing that is appropriate for the occasion and in purchasing clothing that is appropriate for their style. In this article, an in-depth examination of numerous systems is conducted to determine the many elements that must be considered while developing a robust system that discovers and recommends matching garments for the user.

1. INTRODUCTION

Clothing is a form of symbol that displays an individual's inner thoughts via their appearance. It provides information on their personal preferences, religious beliefs, personalities, careers, social standing, and attitude on life. As a result, clothing is seen as a nonverbal means of communication as well as an important part of people's overall appearance. Demographics, region, personal tastes, interpersonal influences, age, gender, season, and culture are all factors that impact consumer fashion choices. Combining fashion preferences with the previously identified features associated with clothing choices may be able to transmit visual aspects for a better understanding of client preferences.

As a result, examining customer preferences and recommendations benefits fashion designers and merchants. Additionally, words, perspectives, and images or pictures describing customers' dress selections and product preferences are increasingly available on the Internet. Because these images include information about people all over the world, they are being used by both online and physical fashion retailers to reach billions of Online consumers. As a result, e-commerce has recently overtaken retail as the preferred option of purchase. The ability of recommendation systems to provide personalised recommendations and respond quickly to customer selections has greatly assisted the rise of e-commerce sales. Fashion recommendation algorithms were first used by e-commerce companies in the early 2000s. [15]

However, implementation was mostly in the planning stages until 2007–2008. Fashion items, like other items such as gadgets and books, were recommended based on the user's previous purchases. As computer vision algorithms have developed, personalised suggestions based on personal factors and user ratings have become more popular. [4]

2. RELATED WORK

a. Myntra - Matching Clothes Recommendation - When you select a certain item to purchase, Myntra will immediately suggest a matching pair of apparel. When a user chooses a t-shirt, the system generates a set of watches, shoes, pants, and other accessories that go with the t-shirt. This approach does not take into account the customer's personal attributes such as skin color or prior apparel. Only things that are already in its database will be recommended. [13]

b. Your Closet - This is a smartphone app for organizing closets. The user interface is shown in Figure 1. Customers must fill out the application with information about their clothing. Each fabric is then paired with other outfits by the algorithm. For example, when there are four shirts and four pants, the computer will pair each shirt with every pant, giving you 16 alternatives. Garments are not matched based on patterns, color, or texture in this software. It also doesn't have a way to make recommendations. [19]

c. Magic Closet - The purpose of this technique is to locate ensembles that match the clothes that have been entered from online merchants. These outfits must be suitable for the occasion. The user takes a photo of oneself using this approach, stating whether they want to wear the top or bottom clothing, as well as the occasion for which they want to wear them. The algorithm will hunt for clothing that matches the user's search and satisfies the requirements of looking good and fitting well.
d. Personalized Clothing Recommendation Based on Knowledge Graph - This system seeks to use the knowledge graph to give the user clothing ideas while considering the user's context. The recommendation is made by determining how similar the garments ontology and the user's collection are.
e. Clothes Recommendation based on Knowledge Graph - Three kinks of core ontology are presented for related things such as user, fabric, and context. The user entity description includes user characteristics such as height, weight, and skin color. Cloth characteristics such as color, texture, pattern, and fabric are utilized to distinguish the clothing. The context entity stores information about the current weather, events, and other factors. To uncover correlation among clothing and contextual attributes, the Apriori approach is applied, and the commonly used item set is produced. The knowledge rules are applied to the ontology and knowledge graphs. This approach is known as knowledge reasoning. After learning about knowledge reasoning, several association rules emerge. Relevant qualities are connected using edges to construct ontology and the knowledge graph for the clothing domain.

Merits and Demerits of Systems:

a. Myntra:
To maximise suggestions, the algorithm does not consider particular user characteristics such as skin colour or existing clothing.
b. Your Closet App:
Each piece of clothes is matched with all other types of clothing, for example, each shirt is matched with each pair of pants. Doesn't know what clothing go together based on colour, texture, and other factors. Users must organise their own closets because the method is not automated.
c. Magic Closet:
The system is built around the Kinect, which may or may not be available to all users. It prefers to propose clothing from online stores above clothing from the user's own closet.

3. OVERVIEW OF MACHINE LEARNING MODELS

a. K-Means Clustering:
K-means clustering is one of the easiest and most popular ways to learn unsupervised machine learning. K-means measurement combinations of equations use the normal distance of a straight line (Euclidean distance, in other words). Creates collections by placing a few points, called centroids, within the element space. Each point in the database is assigned a collection of any centroid closest to it. For the algorithm to work first we start with k points, called methods, at random. Then, we divide each item by its very closest scale, and we review the mean links, which are the average ratings of the items divided that way so far. The process is repeated with a given number of repetitions and finally, we have our collections.
To find the number of clusters in data, we need to use the K-Means integration algorithm for different K values and compare results. Thus, the performance of the K-Means algorithm depends on the value of K. We have to choose the total K value that gives us the best performance. There are various strategies available to find the total value of K. The most commonly used is the elbow method. The elbow method determines the amount of cost work produced by different K values.

If K increases, the median deviation will decrease. Then each collection will have a few shapes, and shapes will be closer to their centroids. However, the development of moderate distortion will decrease as K increases. The value of K when the distortion development decreases significantly is called the elbow, where we should stop dividing the data into additional clusters.

b. Gaussian Mixture Model:
Gaussian hybrid models can be used to combine non-labelled data in the same way

as k-means. The Gaussian hybrid model incorporates a hybrid (i.e., high-volume) multi-Gaussian distribution. Here instead of identifying collections with "near" centroids, we equate a set of gaussian k in data. We also measure gaussian distribution parameters as definition and variation for each collection and group weight. After reading the parameters of each data point, we can calculate the probability of belonging to each set.

All distributions are multiplied by weight π ($\pi 1 + \pi 2 + \pi 3 = 1\pi 1 + \pi 2 + \pi 3 = 1$) to calculate the fact that we do not have the same number of samples in each category. In other words, we can only include 1000 people from the red collection class and 100,000 people from the green collection class.

c. Hierarchal Agglomerative Clustering:

Hierarchical clustering algorithms can be top-down or bottom-up. The bottom-up algorithms treat each document as a singleton initially and then sequence (or agglomerate) pairs of collections until all collections are grouped together into one document containing all the documents. The lower hierarchical cluster is therefore called the hierarchical agglomerative clustering. Blending up and down requires a method of dividing the collection. It continues to divide collections over and over until each document is reached. The required space for the Hierarchical clustering Technique is very high if the data points are high as we need to keep the matrix in RAM. The complexity of space is the order of the square of n.

Space complexity = $O(n^2)$ where n is the number of data points -------(1)

The time complexity is also very high. The time complexity has the order of cube of n.

Time complexity = $O(n^3)$ where, the number of data points are n

4. EVALUATION METRICS

a. Silhouette Score:

The silhouette coefficient, or silhouette score, is a metric used to calculate the quality of a clustering method. The range of its value is from -1 to 1.

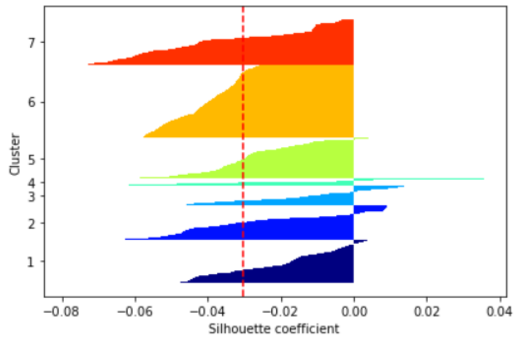
1: Means clusters are distinguishable and apart from each other.
0: Distance between means cluster is not significant and they are indifferent.
-1: Means clusters are not correctly assigned.

Silhouette Score = (b-a)/max(a,b) ------(2)

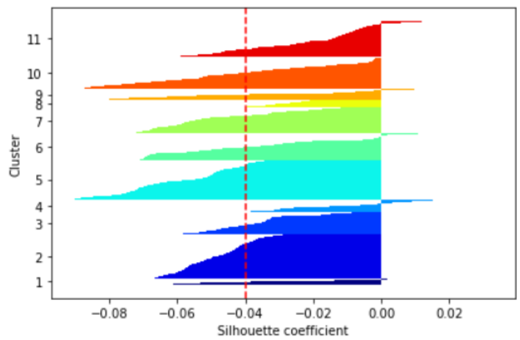
Where,

a= average intra-cluster distance i.e. the average distance between every point inside a cluster.
b= average inter-cluster distance i.e. the average distance between all clusters.

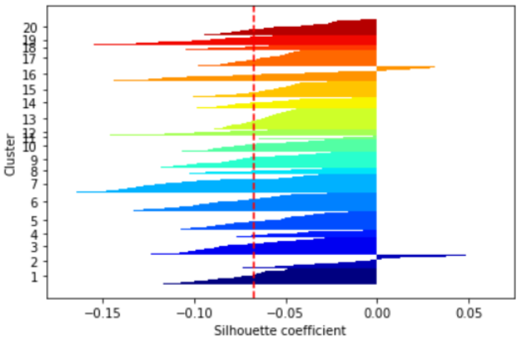
Silhouette coefficient when k=7:



Silhouette coefficient when k=11:



Silhouette coefficient when k=20:



b. Elbow Method:
The idea behind the elbow method is to take a plus k over a range of k values of k (k is 1 to 10 in the example above) and each value of k computes the sum of the squared error (e.g., 1 In k up to 10 in the example above), each value of k computes the sum of the square root error (SSE). Then edit the SSE line chart for each value of k. If the line graph looks like a hand, the "elbow" of the hand is the best k value. The idea is that we want a smaller SSE, but the SSE tends to decrease to zero as k increases (since k is equal to the number of data points in the database, the SSE is zero, since each data point does not belong to each other and there are no errors between them. and the center of his collection). So, our goal is to choose the smallest k with the smallest SSE, and kink shows where we usually start to get reduced returns by increasing k.

5. DATASET DESCRIPTION

The dataset used in the project is the clothing dataset that can be found in Kaggle.
It has Over 5,000 images of 20 different classes.

The dataset consists of a folder with the following attributes:

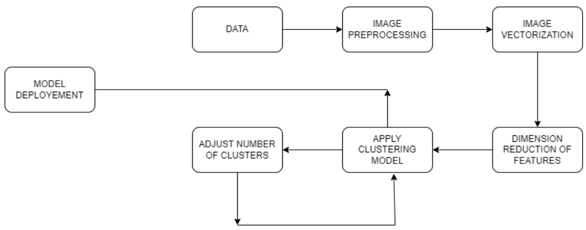
- Images of Clothes: This attribute consists of 5000 images of clothes.
- ID: Each image of clothes has an ID associated with it. This attribute stores the IDs of the images.

It also consists of a csv file which has 4 attributes

- Image id
- Sender_id-the ID of a person who contributed the image
- Label-the type of clothes on the image
- Boolean value-if the clothes are for kids or not.

The dataset consists of images of T-Shirts, Long Sleeve clothes, Pants, Shoes, Shirts, Dresses, Outwear, Shorts, Hats, Skirt, Polo, Undershirt, Blazers, Hoodie, Tops, Blouses, kids wear etc. [3]

6. PROPOSED MODEL



The figure above shows the proposed model and the steps involved in the process image preprocessing to vectorization ,applying algorithm and model deployment.

7. DATA PREPROCESSING

The Fashion MNIST dataset contained 5756 images, belonging to 20 different classes. In order to apply clustering algorithms to the data (images), the data was:

i. *Reshaped*, i.e., all the images were reshaped into numpy arrays with size 224x224. [11]
In order to get more accurate results, it is a good practise to resize the images into the same format.

ii. VGG16, a CNN model was applied to all the reshaped images, to extract features from the images. The model extracted 4096 features.
The VGG16 is a convolution neural network (CNN) architecture that is regarded as one of the best vision model architectures to date. Instead of having many hyper-parameters, VGG16 has convolution layers of 3x3 filter with stride 1 and always uses same padding and maxpool layers of 2x2 filter with stride 2.[21]

iii. PCA (Principal Component Analysis) was applied to reduce the number of features from 4096 to 100. The PCA method examines the interrelationships between a set of variables in an unsupervised manner.
Thus, the images in the dataset were successfully converted into vectors with 100 features. [7]

8. MODEL EVALUATION AND RESULTS:

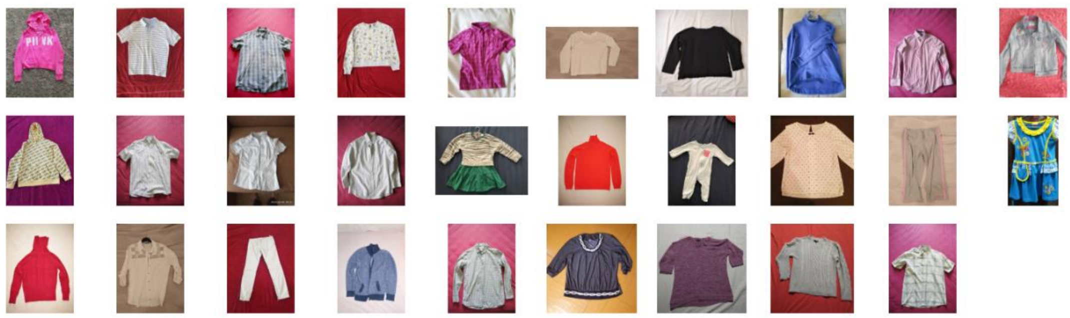
The models used for training and evaluation purposes are:

- K-means Clustering
- Gaussian Mixture Model
- Agglomerative hierarchical clustering

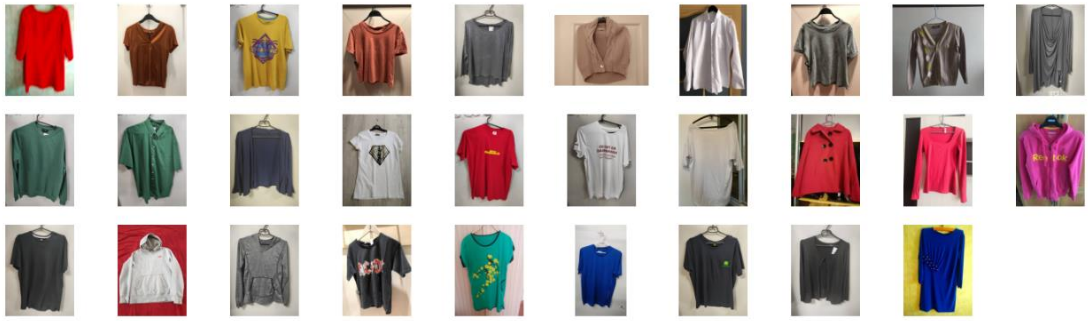
Following are the results of clusters formed by each model:

- K-means Clustering

- Gaussian Mixture Model



- Agglomerative hierarchical clustering



Note: As we have used clustering models, the accuracy of the model cannot be determined, as there are no labels to the images. So there is no way to find the LABEL of each image using the machine

We can still find if the number of clusters taken are correct or not, with the help of the Silhouette Coefficient and Elbow Method.

After testing and trying out different number of clusters, the best number of clusters were found to be 20.

Agglomerative Hierarchical Clustering is the best out of the other algorithms, as the clusters formed by the model were better and more accurate than the other algorithms.

9. CONCLUSION

- After examining three different clustering models on our we then compared their efficiency and accuracy of methodologies utilized, as well as their benefits and drawbacks.
- K-means is the quickest and easiest algorithm but suffers to noise in the data and outliers can't be identified
- Agglomerative algorithm is flexible easy to visualize and can handle any form of similarity but is computationally expensive and can't handle outliers
- GMM is robust to outliers but is highly complex and very slow.
- In our dataset we noted that agglomerative clustering has given us the best clusters
- After a thorough examination, we concluded that agglomerative clustering works best for our dataset

REFERENCES

1. Guan, C., Qin, S., Ling, W. and Ding, G. (2016), "Apparel recommendation system evolution: an empirical review", *International Journal of Clothing Science and Technology*, Vol. 28 No. 6, pp. 854-879.
2. Qianqian Deng, Ruomei Wang, Zixiao Gong, Guifeng Zheng, Zhuo Su, "Research and Implementation of Personalized Clothing Recommendation Algorithm", *Digital Home (ICDH) 2018 7th International Conference on*, pp. 219-223, 2018.
3. Deepti Goel, Santanu Chaudhury, Hiranmay Ghosh, "Multimedia ontology based complementary garment recommendation", *Multimedia & Expo Workshops (ICMEW) 2017 IEEE International Conference on*, pp. 208-213, 2017.
4. Shinya Miura, Toshihiko Yamasaki, Kiyoharu Aizawa, "SNAPPER: Fashion Coordinate Image Retrieval System", *Signal-Image Technology & Internet-Based Systems (SITIS)*

2013 International Conference on, pp. 784-789, 2013.

5. Sang-Young Jo, Sun-Hye Jang, Hee-Eun Cho, Jin-Woo Jeong, "Scenery-Based Fashion Recommendation with Cross-Domain Geneartive Adverserial Networks", *Big Data and Smart Computing (BigComp) 2019 IEEE International Conference on*, pp. 1-4, 2019.
6. Yasutomo Kawanishi, Hiroshi Murase, Satoshi Komorita, Sei Naito, "Aggregating Everyday Outfits by Incremental Clustering With Interactive User Adaptation", *Access IEEE*, vol. 9, pp. 121467-121475, 2021.
7. Evandro de Barros Costa, Hemilis Joyse Barbosa Rocha, Emanuele Tuane Silva, Natalia Caroline Lima, Juliana Cavalcanti, *Recent Advances in Information Systems and Technologies*, vol. 569, pp. 841, 2017.
8. Hyunwoo Hwangbo, Yang Sok Kim, Kyung Jin Cha, "Recommendation system development for fashion retail e-commerce", *Electronic Commerce Research and Applications*, vol. 28, pp. 94, 2018.
9. Congying Guan, Shengfeng Qin, Wessie Ling, Guofu Ding, "Apparel recommendation system evolution: an empirical review", *International Journal of Clothing Science and Technology*, vol. 28, pp. 854, 2016.
10. Yan Hu, Weisong Shi, Hong Li, Xiaohui Hu, "Mitigating Data Sparsity Using Similarity Reinforcement-Enhanced Collaborative Filtering", *ACM Transactions on Internet Technology*, vol. 17, pp. 1, 2017.
11. Samit Chakraborty, Md. Saiful Hoque, Naimur Rahman Jeem, Manik Chandra Biswas, Deepayan Bardhan, Edgar Lobaton, "Fashion Recommendation Systems Models and Methods: A Review", *Informatics*, vol. 8, pp. 49, 2021.
12. Roland Graef, Mathias Klier, Andreas Obermeier, Jan Felix Zolitschka, "What to buy Pepper? – Bridging the Physical and the Digital World with Recommendations from Humanoid Robots", *Journal of Decision Systems*, pp. 1, 2022.
13. K. Kavitha, S. Laxman Kumar, Pudu Pravalika, Kotte Sruthi, R.V.S. Lalitha, N.V. Krishna Rao, "Fashion compatibility using convolutional neural networks", *Materials Today: Proceedings*, 2020.
14. Samit Chakraborty, Md. Saiful Hoque, Naimur Rahman Jeem, Manik Chandra Biswas, Deepayan Bardhan, Edgar Lobaton, "Fashion Recommendation Systems Models and Methods: A Review", *Informatics*, vol. 8, pp. 49, 2021
15. Guan, C., Qin, S., Ling, W. and Ding, G. (2016), "Apparel recommendation system evolution: an empirical review", *International Journal of Clothing Science and Technology*, Vol. 28 No. 6, pp. 854-879.
16. L. A. PI, E. Brannon, P. Ulrich, A. Presley, D. Woronka, M. Grasso, S. Gray, and C. Fit, "Understanding fitting preferences of female consumers: Development an expert system to enhance accurate sizing selection."
17. C.-H. Hsu, "Data mining to improve industrial standards and enhance production and marketing: An empirical study in apparel industry," Expert Systems with Applications, vol. 36, no. 3, pp. 4185–4191, 2009.
18. G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the stateof-the-art and possible extensions," IEEE transactions on knowledge and data engineering, vol. 17, no. 6, pp. 734–749, 2005.
19. Jyotikrishna Dass, Monika Sharma, Ehtesham Hassan, Hiranmay Ghosh, "A density based method for automatic hairstyle discovery and recognition", *Computer Vision Pattern Recognition Image Processing and Graphics (NCVPRIPG) 2013 Fourth National Conference on*, pp. 1-4, 2013.
20. Jose Luis Jorro-Aragoneses, Gineth M. Ceron-Rios, Ma Belén Díaz-Agudo, Juan A. Recio-García, Diego M. López-Gutierrez, "RecOnto: An Ontology to Model Recommender Systems and its Components", *Tools with Artificial Intelligence (ICTAI) 2017 IEEE 29th International Conference on*, pp. 815-821, 2017.
21. Bellini, P., Palesi, L.A.I., Nesi, P. *et al.* Multi Clustering Recommendation System for Fashion Retail. *Multimed Tools Appl* (2022). https://doi.org/10.1007/s11042-021-11837-5.