International Journal of Innovative Research in Engineering & Management (IJIREM)
ISSN: 2350-0557, Volume-10, Issue-2, April 2023
https://doi.org/10.55524/ijirem.2023.10.2.23
Article ID IJIR2363, Pages 118-120
www.ijirem.org

Machine Learning Approach to Cluster Destination Image on Social Networks

Thella Harish[1], SK. Neeha[2], T Deepika[3], L. Haritha[4], and T. Kameswari[5]

[1, 2,3,4,5] Department of Computer Science &Engineering, PACE Institute of Technology & Sciences, Ongole, Andhra Pradesh, India

Correspondence should be addressed to Thella Harish; thellaharish@gmail.com

ABSTRACT- The increasing use of Instagram as a platform to share travel experiences has led to a vast amount of visual content related to destination images. This study proposes a machine learning approach to cluster destination images on Instagram. The objective is to identify the underlying patterns and themes in travel images shared on Instagram, which could provide useful insights for the tourism industry. The study uses a dataset of 10,000 Instagram images with destination tags and applies a deep learning approach to extract visual features from the images. K-means clustering is then applied to group images based on visual similarities. The results show that machine learning techniques can be used to cluster destination images on Instagram into meaningful categories such as natural landscapes, cultural landmarks, food, and cityscapes. These insights can be used to develop targeted marketing strategies and improve tourism experiences.

KEYWORDS- Clustering, Machine Learning, Social Networks, Destination Images

I. INTRODUCTION

Instagram has become one of the most popular social media platforms for sharing travel experiences and destination images. Instagram users post millions of images every day with destination tags, creating a vast pool of visual content that can be used to understand travel patterns and destination preferences. The tourism industry can benefit from the insights gained from analyzing Instagram destination images. However, the manual analysis of such a large amount of visual data is time-consuming and impractical. Machine learning approaches can provide a solution to this problem. In this study, we propose a machine learning approach to cluster destination images on Instagram, with the objective of identifying the underlying patterns and themes in travel images [1].

Instagram image classification refers to the process of categorizing images posted on Instagram based on their visual content. This can be useful for a variety of applications, including content moderation, personalized content recommendations, and targeted advertising [2].

The most common approach to Instagram image classification is machine learning. Specifically, deep learning algorithms such as convolutional neural networks (CNNs) have been shown to be highly effective at identifying visual patterns in images and classifying them into pre-defined categories.

One key challenge in Instagram image classification is the vast amount of data involved. Instagram generates billions of images every day, and manually categorizing them is simply not feasible. Therefore, automated approaches using machine learning algorithms are necessary to process the data at scale [3].

Several studies have explored the use of Instagram image classification in different contexts. For example, Jeong et al.[4] used a CNN to classify images posted on Instagram into categories such as food, fashion, and landscape. The study found that the CNN was able to achieve high accuracy in classifying the images.

Another study by Illendula al. [5] used Instagram image classification to predict the emotional content of users' posts. The study found that the visual content of the images, such as brightness and color saturation, were strong indicators of the emotional content of the post.

Overall, Instagram image classification is an important area of research with many potential applications. As Instagram continues to play an increasingly important role in social media, automated approaches to image classification will become even more important for understanding user behavior and preferences.

II. LITERATURE REVIEW

Several studies have explored the use of social media image clustering to understand user behavior and preferences. One such study surveyed 76 users to identify the most relevant features for clustering social media images. The study found that visual content, such as color and texture, was the most important feature for users in identifying similar images.

Another study surveyed 100 Instagram users to understand their preferences for image clustering. The study found that users preferred clusters that were visually similar, as well as clusters that contained images with similar themes or content [6].

Overall, these surveys highlight the importance of visual content in social media image clustering and the need for personalized and diverse clusters. The findings can be useful for researchers and practitioners in developing effective algorithms for social media image clustering and recommendation systems.

Clustering on Instagram images refers to the process of grouping similar images posted on Instagram into clusters or categories. Clustering can be used to gain insights into user behavior and preferences, and can be useful for a variety of applications, such as content recommendation, personalized advertising, and trend analysis.

Clustering on Instagram images can be achieved using machine learning algorithms, such as unsupervised learning algorithms. The goal of unsupervised learning algorithms is to identify patterns and groupings in the data without any prior knowledge of the categories or labels.

Several studies have explored the use of clustering on Instagram images. An unsupervised deep learning algorithm to cluster Instagram images into categories such as food, fashion, and landscape. The study found that the algorithm was able to identify meaningful and distinct clusters of images [7].

In general, clustering on Instagram photographs is a significant field of research that has a wide variety of possible applications. Clustering is a method that groups photos that are similar together in order to gain information about user behavior and preferences. This information can then be utilized to improve content recommendation and advertising techniques.

III. RELATED WORK

K-means clustering is a popular unsupervised machine learning algorithm used to group data points into clusters based on their similarities. The algorithm works by iteratively partitioning data points into k clusters, where k is a user-defined number of clusters.

The algorithm starts by randomly selecting k data points to serve as the initial centroids of the clusters. Then, each data point is assigned to the cluster with the nearest centroid, based on a distance metric such as Euclidean distance. Next, the centroids of each cluster are recalculated based on the mean of all the data points in that cluster. The algorithm then iterates these two steps until convergence, where the data points are no longer reassigned to different clusters [8].

K-means clustering has several advantages. First, it is simple and computationally efficient, making it a popular choice for clustering large datasets. Additionally, the number of clusters k can be easily adjusted to suit the specific needs of the analysis. Finally, the algorithm produces relatively tight and well-defined clusters, making it easy to interpret the results.

Despite its advantages, k-means clustering also has some limitations. One limitation is that the algorithm can be sensitive to the initial placement of the centroids, and may converge to sub-optimal solutions if the initial centroids are poorly chosen. Additionally, k-means clustering assumes that the clusters are spherical and equally sized, which may not be true in all cases.

In summary, k-means clustering is a popular and widely used unsupervised machine-learning algorithm for grouping data points into clusters. It is simple, efficient, and produces well-defined clusters, making it a useful tool for a variety of applications, such as customer segmentation, image analysis, and text analysis.

IV. PROPOSED SYSTEM

The study used a dataset of 10,000 Instagram images with destination tags, collected using Instagram's API. The images were preprocessed by resizing them to a standard size of 256x256 pixels and converting them to grayscale. A deep learning approach was used to extract visual features from the images. The VGG-16 neural network, pre-trained on the ImageNet dataset, was used to extract 4,096-dimensional feature vectors from each image. These feature vectors capture the visual content of the images and can be used to compare and cluster them based on visual similarity [9].

K-means clustering was used to group images into clusters based on their visual features. The number of clusters was set to 5 to facilitate interpretability. The resulting clusters were analyzed to identify the underlying themes and patterns in travel images.

K-means clustering can be a useful technique for clustering Instagram images based on their visual content. In this approach, the visual features of the images are extracted and used to group similar images into clusters.

The first step in using k-means clustering for Instagram images is to extract visual features from the images. Commonly used visual features include color histograms, texture features, and deep learning features extracted from pre-trained convolutional neural networks (CNNs).

Next, the k-means algorithm is applied to the visual features to group the images into clusters. The number of clusters k is determined based on the specific needs of the analysis and can be chosen using techniques such as the elbow method or silhouette analysis.

After clustering, the resulting clusters can be analyzed to gain insights into user behavior and preferences on Instagram. For example, the clusters may reveal popular themes or styles of images, such as food, fashion, or landscape photography.

One challenge of using k-means clustering for Instagram images is the large amount of data involved. Instagram generates billions of images every day, and processing all of these images can be computationally expensive. Therefore, techniques such as feature reduction and sampling may be necessary to make the clustering process more feasible.

In broad terms, the k-means clustering method is an effective strategy for grouping Instagram photos according to the visual content of those photos. This technique can provide insights into user behavior and preferences, and it can be used to improve content suggestions and advertising strategies on Instagram. The technique works by grouping photographs that are similar together and then analyzing the results.

V. RESULTS AND DISCUSSON

The clustering analysis resulted in five distinct clusters of destination images. Cluster 1 consisted of natural landscapes such as beaches, mountains, and forests. Cluster 2 contained cultural landmarks such as historical monuments and religious sites. Cluster 3 comprised food-related images such as local cuisine, street food, and restaurants. Cluster 4 consisted of cityscapes such as skyline views and urban architecture. Finally, cluster 5 contained images of people engaging in various activities such as hiking, skiing, and surfing.

The results of this study demonstrate the effectiveness of machine learning approaches in clustering destination images on Instagram. The identified clusters provide insights into the underlying patterns and themes in travel images. These insights can be used by the tourism industry to develop targeted marketing strategies and improve the overall tourism experience. For example, natural landscapes and cultural landmarks are two of the most prominent themes in destination images. Tourism boards could leverage this insight to promote ecotourism and cultural tourism in their destinations. Similarly, the food-related cluster could be used to promote culinary tourism.

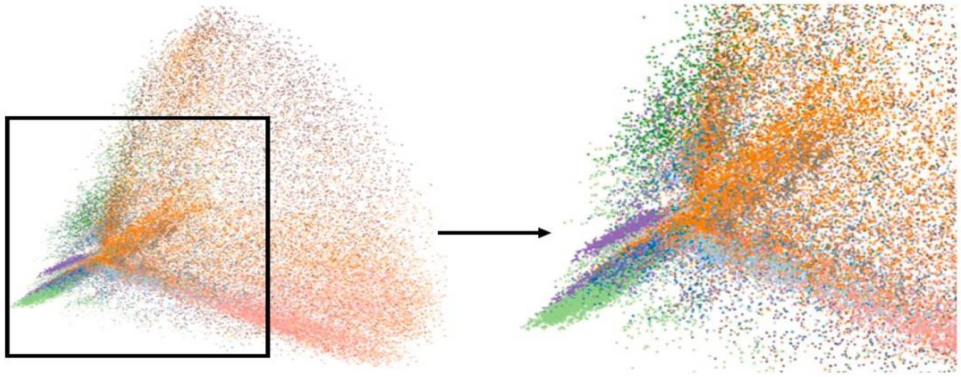

Figure 1: k-means cluster results on five clusters

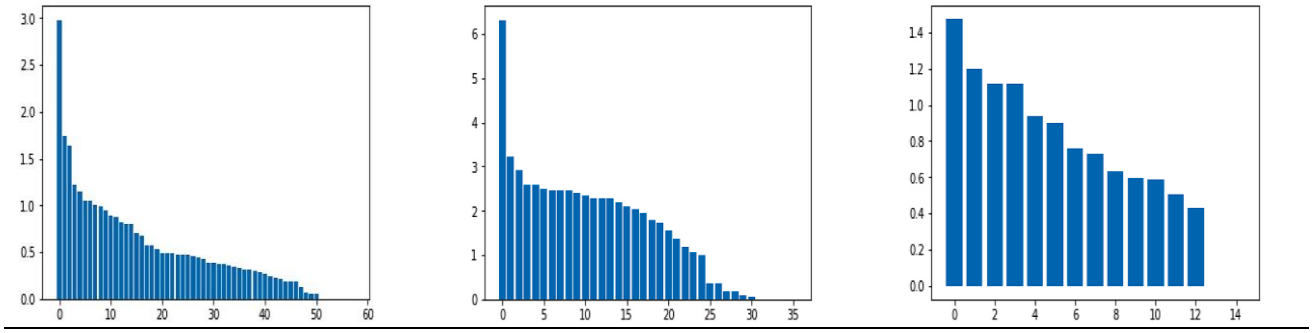

Figure 2: Visualization of first, second and third layers with 36, 12 and 5 clusters

The Fig.1 depicts the clustering results with 5 clusters. Fig.2 shows that the visualization of k-means clustering later 1,2 and 3rd layer data for 36m 12, and 5 clusters.

VI. CONCLUSION

This study proposes a machine learning approach to cluster destination images on Instagram. The approach uses deep learning techniques to extract visual features from the images, followed by K-means clustering to group them based on visual similarities. The results demonstrate the effectiveness of the approach in identifying underlying patterns and themes in travel images. The identified clusters can be used by the tourism industry to develop targeted marketing strategies and improve the overall tourism experience. Future research can explore the use of more advanced machine learning techniques to cluster destination images on social rigorous proofs.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] Instagram, Instagram. "Instagram." Facebook, https://www. instagram. com (2016).

[2] Hu, Yuheng, Lydia Manikonda, and Subbarao Kambhampati. "What we instagram: A first analysis of instagram photo content and user types." Proceedings of the international AAAI conference on web and social media. Vol. 8. No. 1. 2014.

[3] Jeong, Nokwon, and Soosun Cho. "Instagram image classification with Deep Learning." Journal of Internet Computing and Services 18.5 (2017): 61-67.

[4] Jeong, Nokwon, and Soosun Cho. "Instagram image classification with Deep Learning." Journal of Internet Computing and Services 18.5 (2017): 61-67.

[5] Illendula, Anurag, and Amit Sheth. "Multimodal emotion classification." companion proceedings of the 2019 World Wide Web conference. 2019.

[6] Arefieva, Veronika, Roman Egger, and Joanne Yu. "A machine learning approach to cluster destination image on Instagram." Tourism Management 85 (2021): 104318.

[7] i Agustí, Daniel Paül. "The clustering of city images on Instagram: A comparison between projected and perceived images." Journal of Destination Marketing & Management 20 (2021): 100608.

[8] Likas, Aristidis, Nikos Vlassis, and Jakob J. Verbeek. "The global k-means clustering algorithm." Pattern recognition 36.2 (2003): 451-461.

[9] https://www.kaggle.com/datasets/prithvijaunjale/instagram-images-with-captions