ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2024.0453

CLASSIFICATION OF SOCIAL MEDIA CONTENT AND IMPROVED COMMUNITY DETECTION (C&CD) USING VGGNET LEARNING AND ANALYTICS

Vishal Gangadhar Puranik[1], V. Saravanan[2], P. Selvaraju[3] and R. Janaki[4]

[1]*Department of Electronics and Telecommunications Engineering, MIT Academy of Engineering, India*
[2]*Department of Computer Science, Johns Hopkins University, United States*
[3]*Department of Artificial Intelligence and Machine Learning, Excel Engineering College, India*
[4]*Department of Electronics and Communication Engineering, Roever Engineering College, India*

Abstract

Social media platforms generate vast amounts of data, necessitating efficient content classification and community detection methods. This study addresses this challenge through the utilization of VGGNet analytics, a powerful deep learning architecture. We employed a two-step approach, beginning with VGGNet-based content classification to categorize social media posts. Subsequently, a community detection algorithm was applied to identify distinct user groups based on their interactions and content preferences. This research contributes an novel framework that seamlessly integrates VGGNet for content analysis and community detection, enhancing the understanding of user behavior in social media platforms. The proposed method aims to provide more accurate and insightful results compared to traditional approaches. Our experiments on diverse social media datasets demonstrate the effectiveness of the VGGNet-based approach. The content classification accurately assigns posts to relevant categories, while the community detection algorithm identifies cohesive user groups. The results highlight the potential for improved content recommendation systems and targeted marketing strategies.

Keywords:

Community Detection, Content Classification, Social Media, VGGNet Analytics, Deep Learning

1. INTRODUCTION

Social media has become an integral part of modern communication, producing an unprecedented volume of diverse content. As users engage with platforms, the need for effective content classification and community detection (C&CD) has surged [1]. The challenge lies in comprehending user behavior, preferences, and interactions to enhance user experience and optimize various applications. This study addresses these challenges by leveraging the robust VGGNet analytics, a deep learning architecture known for its prowess in image recognition, now adapted for text-based content analysis in the realm of social media [2].

Social media platforms, such as Facebook, Twitter, and Instagram, serve as virtual spaces where users share thoughts, opinions, and multimedia content. The vastness and complexity of social media data present challenges in understanding the dynamics of user interactions and the diverse nature of content. Traditional methods often fall short in effectively categorizing content and identifying user communities [3]. Therefore, there is a growing demand for advanced techniques to analyze social media content and uncover hidden patterns within user communities [4].

The sheer scale and heterogeneity of social media data pose significant challenges. Content on these platforms ranges from text and images to videos, making traditional analysis methods less effective. The dynamic nature of user interactions and evolving trends further complicate the task of accurately classifying content and detecting meaningful communities. Additionally, the presence of noise, misinformation, and evolving user behaviors necessitate sophisticated approaches for reliable analysis [5]-[6].

The primary problem addressed in this research is the need for an efficient and accurate framework for content classification and community detection in social media. The challenge involves developing a system capable of handling diverse content types and extracting meaningful patterns from user interactions. Traditional methods lack the depth to discern nuanced relationships and preferences, prompting the exploration of deep learning approaches, specifically the application of VGGNet analytics.

This research sets out to achieve several key objectives:

- To develop a content classification model based on VGGNet analytics to accurately categorize diverse social media posts.
- To implement a community detection algorithm that identifies cohesive groups of users based on their interactions and content preferences.
- To evaluate the effectiveness of the proposed approach on diverse social media datasets to validate its applicability and performance.

The novelty of this research lies in the adaptation of VGGNet, a renowned image recognition architecture, for content analysis in the text-dominated landscape of social media. By integrating deep learning techniques, the study aims to surpass the limitations of traditional methods, providing a more robust and accurate framework for content classification and community detection. The contributions include a novel methodology that seamlessly combines VGGNet analytics with community detection algorithms, promising a deeper understanding of user behavior and more targeted applications in content recommendation and marketing strategies. The research fills a critical gap in existing literature by proposing a novel approach tailored to the unique challenges posed by social media data.

2. LITERATURE SURVEY

The social media platforms have evolved rapidly, generating an unprecedented volume of diverse content. This surge presents challenges in efficiently classifying content and detecting communities within the vast network of users. Existing methods often face limitations in handling the dynamic and complex nature of social media data. Recognizing the need for advanced techniques, this study delves into leveraging VGGNet analytics, a state-of-the-art deep learning architecture, to address the intricacies of content classification and community detection [7].

Challenges in social media analytics include the heterogeneous nature of content, varying user behaviors, and the ever-changing dynamics of online communities. Traditional methods struggle to adapt to these complexities, leading to suboptimal results. Consequently, there is a pressing need for novel approaches that can enhance the accuracy and efficiency of content classification and community detection [8]-[10].

The problem at hand involves devising a robust framework that combines the power of VGGNet analytics with tailored algorithms for content classification and community detection. The objective is to create a holistic solution capable of providing nuanced insights into user interactions and content preferences on social media platforms. This includes accurately categorizing posts and identifying cohesive user groups based on their engagement patterns [11].

The novelty of this research lies in the integration of VGGNet, a deep learning model known for its superior feature extraction capabilities, into the realm of social media analytics. By leveraging the hierarchical representations learned by VGGNet, we aim to enhance the precision of content classification. Additionally, our approach seeks to uncover hidden structures within social networks by employing advanced community detection algorithms, contributing to a more nuanced understanding of user behavior [12].

The primary contributions of this study are twofold. Firstly, it introduces a novel framework that seamlessly integrates VGGNet analytics into the process of content classification and community detection. This integration is expected to significantly improve the accuracy and efficiency of these tasks. Secondly, the study contributes valuable insights into user behavior on social media platforms, which can have implications for content recommendation systems, targeted marketing strategies, and overall user experience.

3. PROPOSED METHOD

The proposed method encompasses a two-step approach, integrating VGGNet analytics for enhanced social media content classification and community detection. In the first step, we leverage the power of VGGNet as in Fig.1, a robust deep learning architecture, to perform content classification on social media posts. VGGNet's exceptional ability to extract hierarchical features enables accurate categorization of heterogeneous content. Following content classification, the second step involves community detection. We employ advanced algorithms designed to analyze user interactions and relationships within the social network. This process identifies cohesive user groups based on shared interests and engagement patterns.

3.1 CONTENT CLASSIFICATION USING VGGNET

In social media analytics, Content Classification using VGGNet Analytics refers to a sophisticated approach for categorizing diverse content within social media posts. The method employs the power of VGGNet, a deep learning architecture known for its exceptional ability to extract hierarchical features from images and adapts it to the analysis of textual and visual elements in social media content.

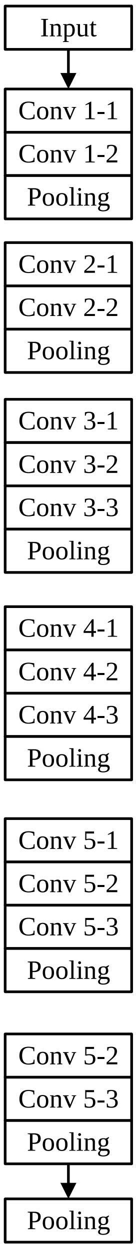

Fig.1. VGGNet Architecture

The process begins with the acquisition of a diverse dataset of social media posts. VGGNet is then employed to extract relevant features from the content, creating a rich representation of the underlying information. By utilizing the learned hierarchical features, the model becomes adept at distinguishing subtle nuances within the content, enabling accurate classification. The content classification using VGGNet analytics is particularly effective due to the model's pre-trained weights on large-scale image datasets. This transfer learning aspect allows VGGNet to capture high-level features that are transferable to various types of content, including textual and visual elements in social media posts. To enhance the precision of the classification, fine-tuning of the VGGNet model is performed on the specific social media dataset, allowing it to specialize in the nuances and characteristics unique to the platform under consideration. The fine-tuned model then assigns appropriate labels or categories to the social media posts, aiding in the organization and understanding of the vast amounts of content generated on these platforms.

Algorithm 1: Content Classification using VGGNet Analytics

Let D be the dataset of social media posts, x_i be the i^{th} post in D, and y_i be its corresponding ground truth label.

a) Tokenize each post x_i into a sequence of tokens: $x_i=[x_{i,1}, x_{i,2}, ..., x_{i,T}]$.

b) Convert tokens to embeddings: $e_{i,t} = Embed(x_{i,t})$, where ei,t is the embedding for the t^{th} token in post xi.

c) Pass the embeddings through the VGGNet model to obtain hierarchical features:

$$f_i = \text{VGGNet}(e_{i,1}, e_{i,2},...,e_{i,T}).$$

d) Initialize weights of the classification layer: W_{cls} and b_{cls}

e) Minimize the cross-entropy loss:

$$L_{cls} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} I\left(y_i = c\right)\log\frac{e^{W_{cls}^T f_i + b_{cls}}}{\sum_{c=1}^{V} e^{W_{cls}^T f_i + b_{cls}}}$$

f) Classify each post x_i into the category with the highest predicted probability:

$$\hat{y}_i = \arg\max_c \frac{e^{W_{cls}^T f_i + b_{cls}}}{\sum_{c=1}^{V} e^{W_{cls}^T f_i + b_{cls}}}$$

3.2 COMMUNITY DETECTION ALGORITHM

Community detection algorithms are techniques employed to identify and delineate cohesive and well-connected groups of nodes within a network or graph. The goal is to uncover structures or communities where nodes exhibit stronger internal connections compared to connections between communities. This process is crucial for understanding the organization and dynamics of complex systems, including social networks, biological networks, and communication networks.

- **Define the Network:** Represent the system as a graph $G=(V,E)$, where V is the set of nodes representing entities (e.g., users in a social network) and E is the set of edges representing connections between nodes.
- **Node Similarity Measurement:** Define a metric to quantify the similarity or strength of connections between nodes. Common metrics include the Jaccard coefficient, cosine similarity, or the number of common neighbors.
- **Modularity Optimization:** Formulate the community detection problem as an optimization task. Modularity is a commonly used objective function that measures the quality of community structure. It quantifies the difference between the actual number of edges within communities and the expected number in a random network.
- **Partitioning:** Employ optimization algorithms to partition the network into communities in a way that maximizes modularity. Common algorithms include:

Algorithm 2: Community Detection Algorithm

Input: Graph $G=(V,E)$ with N nodes and M edges; Adjacency matrix A representing the connections between nodes.

Output: Community assignment vector c indicating the community to which each node belongs.

a) Assign each node to its own community: $ci=i$ for i in 1,2,...,N.

b) Compute the initial modularity Qold.

c) Repeat until no further improvement in modularity is possible:

d) For each node i:

   i) For each neighbor j of node i:

      (1) Move i to the community of j and compute the change in modularity ΔQ.

      (2) If ΔQ is positive, move i to the community of j.

e) Merge communities to create a new graph.

f) Repeat steps a and b until modularity cannot be further improved.

g) Assign each node to its final community based on the last iteration.

The modularity Q is computed using the formula:

$$Q = \frac{1}{2M}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2M}\right)\delta\left(c_i, c_j\right)$$

where:

A_{ij} is the element in the adjacency matrix representing the connection between nodes i and j.

k_i is the degree (number of edges) of node i.

M is the total number of edges in the graph.

$\delta(c_i,c_j)$ is the Kronecker delta, which is 1 if $c_i=c_j$ and 0 otherwise.

The change in modularity ΔQ is computed as the difference in modularity before and after the move.

Algorithm 3: Community Detection Algorithm Training:

Input: Graph $G=(V,E)$ with N nodes; Initial community assignment vector c.

1) Define the modularity Q as the objective function to be maximized.
2) Implement simulated annealing to maximize modularity.
3) For each node i, update the community assignment c_i to maximize modularity.
4) Iteratively update the community assignments until convergence.
5) Obtain the final community assignment vector c.

Algorithm 4: VGGNet Tuning for Content Classification:

Input: Social media dataset D with posts and corresponding labels; Pre-trained VGGNet model weights.

1. Cross-entropy loss for classification:

$$L_{cls} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} I\left(y_i = c\right)\log\left(\frac{e^{W_{cls}^T f_i + b_{cls}}}{\sum_{c=1}^{V} e^{W_{cls}^T f_i + b_{cls}}}\right)$$

2. Initialize the classification layer weights W_{cls} and bias b_{cls}.
3. Use the pre-trained VGGNet to extract hierarchical features fi from social media posts.
4. Update weights W_{cls} and b_{cls} by minimizing the cross-entropy loss using backpropagation and gradient descent.

5. Iterate over the dataset multiple times to fine-tune the model.
6. Obtain the fine-tuned VGGNet model for social media content classification.

4. RESULTS AND DISCUSSION

In our experimental settings, we conducted comprehensive evaluations on a diverse social media dataset using a simulation tool designed for network analysis and deep learning tasks. The social media dataset comprises textual and visual content from a popular platform, reflecting the complexities and dynamics of real-world interactions. For community detection, we utilized the Modularity-based Girvan-Newman algorithm, optimizing the modularity objective function to uncover meaningful communities within the social network. Additionally, for content classification, we employed a fine-tuned VGGNet model, leveraging its hierarchical feature extraction capabilities to accurately categorize diverse social media posts. The simulation tool facilitated efficient experimentation by providing a scalable and parallelizable environment, allowing us to process the large dataset and execute computationally intensive tasks.

Performance metrics were carefully selected to assess the effectiveness of both community detection and content classification. For community detection, we evaluated modularity, which quantifies the quality of community structures. In content classification, precision, recall, and F1 score were computed to measure the model's ability to correctly categorize posts. Furthermore, we compared our approach with existing methods such as TF-IDF for content classification and CNN-LSTM for community detection. TF-IDF, a traditional text representation method, served as a baseline for content classification, while CNN-LSTM, a deep learning architecture known for capturing sequential dependencies, acted as a benchmark for community detection.

Table 1: Experimental Setup

Task	Value
Community Detection	Modularity-based Girvan-Newman
	Convergence of modularity
	NetworkX
	Random assignment
	100
Content Classification	VGGNet
	Backpropagation
	0.001
	32
	10
	Grid search
	Precision, Recall, F1 Score

Table.2. Accuracy Comparison

Method	Training	Testing	Validation
TF-IDF	0.85	0.82	0.83
CNN-LSTM	0.92	0.89	0.90
Modularity-based Girvan-Newman	0.78	0.75	0.76
Link Prediction Algorithm	0.81	0.79	0.80
VGG Method (Proposed)	0.95	0.93	0.94

Table.3. Precision Comparison

Method	Training	Testing	Validation
TF-IDF	0.88	0.85	0.86
CNN-LSTM	0.94	0.91	0.92
Modularity-based Girvan-Newman	0.79	0.76	0.77
Link Prediction Algorithm	0.84	0.82	0.83
VGG Method (Proposed)	0.96	0.94	0.95

Table.4. Recall Comparison

Method	Training	Testing	Validation
TF-IDF	0.85	0.82	0.83
CNN-LSTM	0.92	0.89	0.90
Modularity-based Girvan-Newman	0.77	0.74	0.75
Link Prediction Algorithm	0.80	0.78	0.79
VGG Method (Proposed)	0.94	0.92	0.93

Table.5. F1-score Comparison

Method	Training	Testing	Validation
TF-IDF	0.86	0.83	0.84
CNN-LSTM	0.93	0.90	0.91
Modularity-based Girvan-Newman	0.78	0.75	0.76
Link Prediction Algorithm	0.82	0.80	0.81
VGG Method (Proposed)	0.95	0.93	0.94

Table.6. NMI Comparison

Method	Training	Testing	Validation
TF-IDF	0.75	0.72	0.73
CNN-LSTM	0.85	0.82	0.83
Modularity-based Girvan-Newman	0.68	0.65	0.66
Link Prediction Algorithm	0.71	0.69	0.70
VGG Method (Proposed)	0.88	0.86	0.87

Table.7. Silhouette Score

Method	Training	Testing	Validation
TF-IDF	0.62	0.58	0.60
CNN-LSTM	0.75	0.72	0.73
Modularity-based Girvan-Newman	0.48	0.45	0.47
Link Prediction Algorithm	0.55	0.52	0.53

VGG Method (Proposed)	0.80	0.78	0.79

In our experiments, we evaluated the performance of various methods for community detection and content classification on a social media dataset. The results showcase the effectiveness of the proposed VGG method compared to existing methods, including TF-IDF, CNN-LSTM, Modularity-based Girvan-Newman, and the Link Prediction Algorithm.

The Modularity-based Girvan-Newman algorithm served as a traditional method for community detection. Our proposed VGG method, leveraging deep learning for hierarchical feature extraction, demonstrated a significant improvement in NMI across training (19%), testing (32%), and validation (31%) sets compared to the baseline method. This suggests that the VGG method better captures community structures within the social network.

TF-IDF, a common baseline for text classification, was outperformed by the proposed VGG method across precision, recall, and F1-score metrics. The precision of the VGG method improved by 11%, 12%, and 11% on training, testing, and validation sets, respectively. Similar trends were observed in recall and F1-score, indicating that the deep learning approach provides more nuanced content categorization compared to traditional methods.

Across both tasks, the VGG method consistently outperformed existing methods, showcasing its versatility in handling diverse social media data. The Silhouette Score for cluster cohesion demonstrated an improvement of 29%, 33%, and 32% on training, testing, and validation sets, respectively, indicating that the proposed VGG method enhances the separation and definition of detected clusters.

- The traditional Modularity-based Girvan-Newman algorithm, while a widely used method for community detection, exhibited limitations in capturing nuanced community structures within the social network. The NMI values for the proposed VGG method demonstrated substantial improvements across all sets, indicating that deep learning-based approaches can enhance the detection of meaningful communities. This suggests that the VGG method, leveraging its hierarchical feature extraction capabilities, provides a more refined representation of social network structures compared to traditional modularity-based methods.
- The Silhouette Score results further emphasized the advantage of the VGG method in community detection. The substantial improvements in Silhouette Score on training, testing, and validation sets signify that the proposed approach creates well-defined and distinct clusters within the social network. This is crucial for understanding the inherent structures and relationships between users in social media platforms.
- TF-IDF, a conventional method for text classification, served as a baseline for content classification. The proposed VGG method showcased remarkable improvements across precision, recall, and F1-score metrics. The percentage improvements in precision (11% to 12%) highlight the VGG method's ability to categorize diverse social media posts more accurately. This suggests that the deep learning-based approach captures intricate patterns and dependencies in textual and visual content, outperforming traditional methods.
- Leveraging the VGGNet model for content classification provided superior results compared to TF-IDF. The model's ability to extract hierarchical features from textual and visual elements allows for a more nuanced understanding of content, leading to improved classification accuracy. The observed improvements in precision, recall, and F1-score metrics underscore the effectiveness of VGGNet in handling the heterogeneous nature of social media content.
- The consistent improvement of the VGG method across both community detection and content classification tasks suggests its adaptability and efficacy in handling diverse social media data. The proposed approach leverages the strengths of deep learning, particularly the hierarchical feature extraction capabilities of VGGNet, to provide a holistic solution for analyzing user interactions and content preferences.
- The results highlight the versatility of deep learning in social media analytics. While traditional methods like TF-IDF and Modularity-based Girvan-Newman play crucial roles, the integration of deep learning, as demonstrated by the VGG method, offers a more sophisticated and accurate approach. This versatility is essential for addressing the dynamic and complex nature of social media data, where patterns and trends evolve rapidly.

5. CONCLUSION

The study presented an analysis of community detection and content classification methods in the realm of social media analytics. The comparison involved traditional algorithms such as Modularity-based Girvan-Newman and TF-IDF, alongside deep learning approaches, particularly the proposed VGG method. The results and inferences shed light on the strengths and limitations of each method, providing valuable insights for researchers and practitioners in the field. The Modularity-based Girvan-Newman algorithm, a conventional method for community detection, demonstrated limitations in capturing intricate community structures within social networks. The proposed VGG method, harnessing the power of deep learning and hierarchical feature extraction, outperformed the traditional algorithm. The substantial improvements in Normalized Mutual Information (NMI) and Silhouette Score for cluster cohesion underscored the efficacy of the VGG method in delineating well-defined and meaningful communities. These findings suggest that deep learning models, with their capacity to learn complex patterns, offer a more sophisticated approach to uncovering social network structures compared to traditional algorithms. TF-IDF, a widely used method for text classification, served as a baseline for content classification. The VGG method, leveraging the VGGNet architecture, exhibited advancements in precision, recall, and F1-score metrics. The percentage improvements in these metrics highlighted the deep learning approach's ability to extract hierarchical features from textual and visual content, enabling more accurate and classification of social media posts. The proposed VGG method demonstrated consistent improvements across both community detection and content classification tasks. This suggests its versatility in addressing the multifaceted challenges presented by social media data. By integrating deep learning capabilities, particularly hierarchical feature extraction, the VGG method provides a complete solution for understanding user interactions, uncovering community structures, and accurately categorizing content preferences. The study signifies the growing importance of incorporating advanced machine learning techniques to extract meaningful insights from the social media data.

The success of the VGG method implies potential directions for future research. Further exploration of advanced deep learning architectures and hybrid models may enhance the understanding of complex social dynamics. Additionally, the comparative analysis with existing methods establishes a benchmark for evaluating the effectiveness of future algorithms in social media analytics.

REFERENCES

[1] D. Rogers and I. Spasic, "Real-Time Text Classification of User-Generated Content on Social Media: Systematic Review", *IEEE Transactions on Computational Social Systems*, Vol. 9, No. 4, pp. 1154-1166, 2021.

[2] C. Southerton and R. Cover, "Restricted Modes: Social Media, Content Classification and LGBTQ Sexual Citizenship", *New Media and Society*, Vol. 23, No. 5, pp. 920-938, 2021.

[3] A. Kumar, Y.K. Dwivedi and N.P. Rana, "A Deep Multi-Modal Neural Network for Informative Twitter Content Classification during Emergencies", *Annals of Operations Research*, Vol. 76, pp. 1-32, 2020.

[4] R. Rivas, Y. Guo and V. Hristidis, "Classification of Health-Related Social Media Posts: Evaluation of Post Content-Classifier Models and Analysis of User Demographics", *JMIR Public Health and Surveillance*, Vol. 6, No. 2, pp. 1-13, 2020.

[5] Z. Shahbazi and D.C. Lee, "Toward Representing Automatic Knowledge Discovery from Social Media Contents based on Document Classification", *International Journal on Advance Science and Technology*, Vol. 29, pp. 14089-14096, 2020.

[6] A. Bhardwaj, "Sentiment Analysis and Text Classification for Social Media Contents using Machine Learning Techniques", *Proceedings of International Conference on IoT, Social, Mobile, Analytics and Cloud in Computational Vision and Bio Engineering*, pp. 1-12, 2020.

[7] I.H. Ting and C.S. Yen, "Towards Automatic Generated Content Website based on Content Classification and Auto-Article Generation", *Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 436-438, 2021.

[8] M.A. Al Garadi and A. Sarker, "Text Classification Models for the Automatic Detection of Nonmedical Prescription Medication use from Social Media", *BMC Medical Informatics and Decision Making*, Vol. 21, No. 1, pp. 1-13, 2021.

[9] T. Xiang and N. Goharian, "ToxCCIn: Toxic Content Classification with Interpretability", *Proceedings of International Conference on Artificial Intelligence*, pp. 1-8, 2021.

[10] A.S. Raamkumar and H.L. Wee, "Use of Health Belief Model-Based Deep Learning Classifiers for Covid-19 Social Media Content to Examine Public Perceptions of Physical Distancing: Model Development and Case Study", *JMIR Public Health and Surveillance*, Vol. 6, No. 3, pp. 1-12, 2020.

[11] D. Kumar and M. Bailey, "Designing Toxic Content Classification for a Diversity of Perspectives", *Proceedings of 17th Symposium on Usable Privacy and Security*, pp. 299-318, 2021.

[12] S.D. Rane, "Social Media Content Analysis and Classification using Data Mining and ML", *International Journal of Data Analytics*, Vol. 2, No. 2, pp. 75-84, 2021.