Library Progress International
Vol.44 No.3, Jul-Dec 2024: P.5664-5680

Print version ISSN 0970 1052
Online version ISSN 2320 317X

Original Article

Available online at www.bpasjournals.com

Enhancing Multi-Label Text Classification through Beta Ant Colony Feature Selection

Amit Shrivastava[1], Dr. Rakesh Kumar[2]

Research Scholar, Department of Computer Science and Engineering, Ravindranath Tagore University, Bhopal, M.P., India
Supervisor & Associate Professor, Ravindranath Tagore University, Bhopal, M.P., India
rakeshmittal@gmail.com

How to cite this article: Amit Shrivastava, Rakesh Kumar (2024) Enhancing Multi-Label Text Classification through Beta Ant Colony Feature Selection. *Library Progress International*, 44(3), 5664-5680.

Abstract

The intricacy and high dimensionality of data pose serious obstacles in the quickly developing area of text categorisation, especially in multi-label scenarios where cases may fall into more than one category. This work presents a new method for feature selection in multi-label text classification that incorporates the Beta Ant Colony Optimisation (BACO) algorithm. Our technology preserves interpretability while improving classification performance by efficiently shrinking the feature space. Taking use of ant colonies' cooperative nature, the suggested approach uses a beta distribution to probabilistically direct the selection of relevant traits. Extensive trials on benchmark datasets show that the accuracy, precision, and recall of the BACO-based feature selection much exceeds those of conventional approaches. Furthermore, we examine how certain characteristics affect the categorisation outcomes, providing information about the significance of variables and the connections between different categories. Our study advances multi-label text classification methods and provides a strong foundation for practitioners and academics seeking to increase the efficacy and efficiency of models in many applications.

Keywords— Multi-label text classification, Feature selection, Ant Colony Optimization, Beta distribution, High-dimensional data, Classification performance, Data reduction, Machine learning techniques, Text mining, Algorithm evaluation.

1. Introduction

Unlike standard single-label classification, multi-label text classification allows each instance to belong to numerous classes or labels at the same time, making it a tough issue in natural language processing (NLP). The multidimensionality of textual elements, the interdependencies across labels, and the variety of label combinations all contribute to this complexity. Text data has a large feature space, thus choosing features is essential to improving classification accuracy, cutting down on processing overhead, and preventing overfitting.

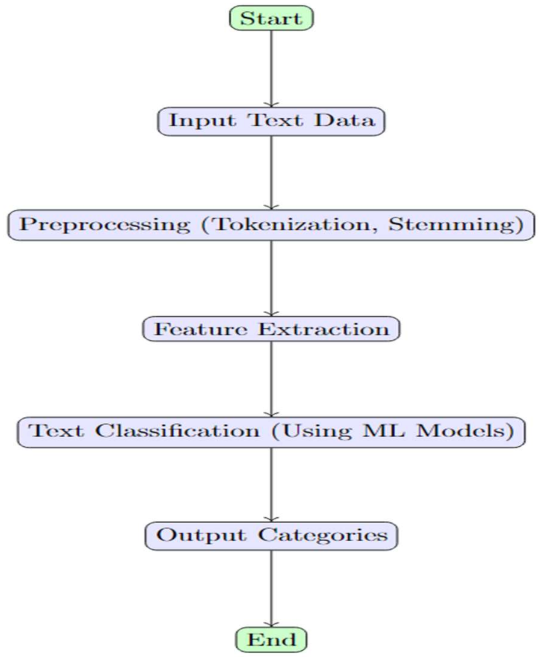

Figure 1: Flowchart of Text Classification Process

By identifying the most relevant characteristics that aid in classification, feature selection seeks to enhance interpretability and model performance. Although they have been used extensively, traditional feature selection techniques like Information Gain, Chi-Square, and Principal Component Analysis (PCA) sometimes falter when dealing with multi-label classification because of the complex interactions between labels. As a result, the need for more sophisticated methods that can handle the particular difficulties presented by multi-label data is rising.

The bio-inspired metaheuristic known as Ant Colony Optimisation (ACO) has shown impressive results in resolving challenging optimisation issues, such as feature selection. A colony of artificial ants in ACO simulates the foraging behaviour of actual ants as they navigate a solution space. Ants leave behind pheromones that impact the likelihood that other ants will follow in their footsteps. This process efficiently directs the colony towards the best options. Since ACO uses both local and global information, together with its stochastic character, it is well suited for feature selection in high-dimensional domains such as multi-label text classification.

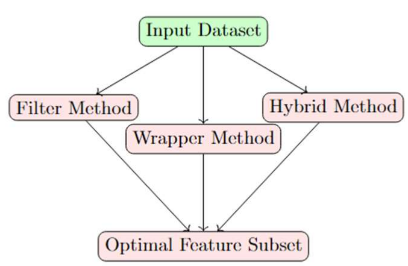

Figure 2: Types of Feature Selection Techniques

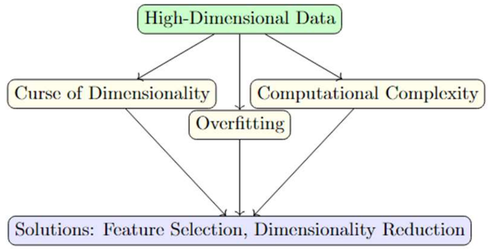

Figure 3: Challenges of High Dimensionality in Text Classification

This study suggests a multi-label text classification method called Beta Ant Colony-based Feature Selection (BACFS). The term "Beta" describes the addition of a new transition probability mechanism that weighs heuristic information and pheromone trails equally when determining which attributes are most relevant. The suggested BACFS technique uses a multi-objective optimisation strategy to minimise feature redundancy and maximise classification accuracy at the same time. This is accomplished by having ants dynamically modify their behaviour in response to traits that are relevant to many labels, making sure that label interdependencies are recorded.

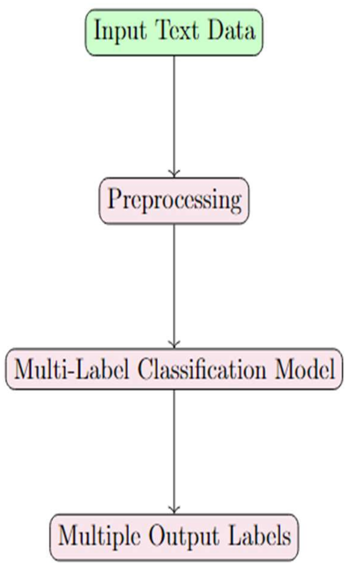

Figure 4: Multi-Label Text Classification Flow

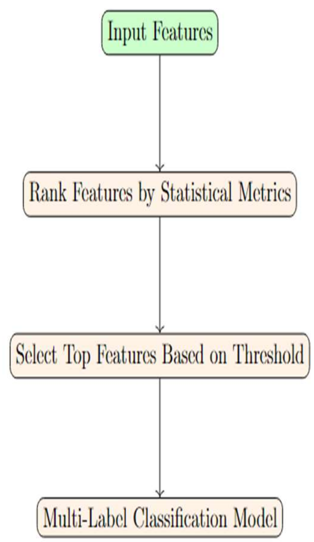

Figure 5: Filter-Based Feature Selection for Multi-Label Classification

In this work, we compare the proposed BACFS to conventional feature selection techniques and assess its performance on many benchmark multi-label text datasets. Metrics including computational time, F1-score, and Hamming Loss are used to evaluate the efficacy of the BACFS approach, giving a thorough grasp of its influence on multi-label classification problems. The results show how much more effective and accurate the BACFS method is in classifying data.

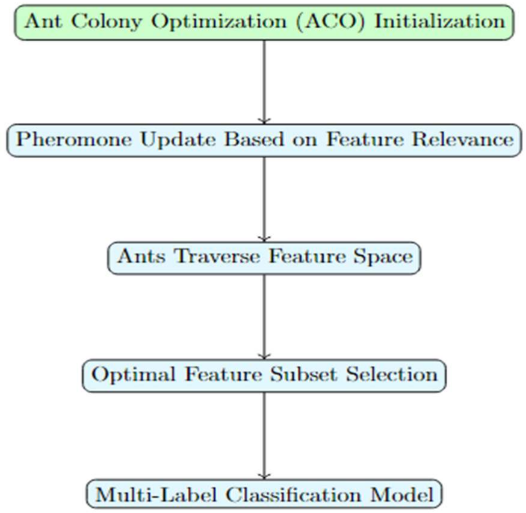

Figure 6: Ant Colony Optimization for Feature Selection in Multi-Label Classification

2. Issues and Challenges

The suggested Beta Ant Colony-based Feature Selection (BACFS) technique presents a viable way to maximise feature selection in multi-label text classification; nevertheless, in order to guarantee its efficacy and scalability in practical applications, a number of problems and obstacles need to be resolved.

1. Text Data Has High Dimensionality

Text is usually represented using methods like bag-of-words, TF-IDF, or word embeddings, which results in an abundance of features in text datasets. Feature selection may become more computationally costly and vulnerable to overfitting as a result of its high dimensionality. It is still difficult to navigate such a big search field, even using Ant Colony Optimisation (ACO). To prevent local optima and guarantee that certain characteristics are pertinent to the classification goal, the BACFS technique must successfully strike a balance between exploration and exploitation.

2. Identify Interdependencies

Labels are often interdependent in multi-label text categorisation, which means that the presence of one label may affect the probability of others. For appropriate categorisation, it is important to capture these dependencies. Conventional feature selection techniques find it difficult to handle this complexity, and while ACO-based methods provide the freedom to investigate feature combinations, modelling label interdependencies remains a challenging challenge. In order to capture the links between labels, the BACFS technique must carefully modify ant behaviour; otherwise, it runs the danger of losing important contextual information.

3. Achieving Equilibrium Between Pheromone Impact and Heuristic Data

How successfully the ants balance pheromone trails, which reflect past successful feature sets, with heuristic information, which represents the significance of individual features, is a major factor in how effective ACO is in feature selection. The effectiveness of multi-label classification may be greatly influenced by the weights given to each of these components.

While putting too much focus on heuristics might lead to inefficient feature selection, an excessive dependence on pheromone information can cause premature convergence. It is crucial yet challenging to create a suitable transition rule that takes into consideration both of the BACFS approach's components.

4. Parameter Sensitivity
ACO algorithms need meticulous adjustment of several parameters, such as the quantity of ants, the rate of pheromone degradation, and the proportional impact of pheromone trails in comparison to heuristics. The best parameter choices might change across datasets, therefore a great deal of experimentation is required to identify values that work for every situation. The lengthy trial-and-error procedure might prevent the BACFS technique from being broadly applicable to new datasets.

5. The capacity to scale
The computational cost of ACO-based feature selection may become unaffordable as datasets increase in size, both in terms of features and occurrences. To handle large-scale datasets without compromising speed, BACFS has to be efficient. To expand the technique, it could be essential to use strategies like distributed computing or parallel processing; however, these solutions come with extra implementation and resource management complications.

6. Feature Subset Assessment
Evaluating feature subset quality in multi-label classification is intrinsically more challenging than in single-label classification. Multi-label assessment metrics like as the F1-score or Hamming Loss, which are computationally demanding and depend on cross-validation, must be taken into consideration by fitness functions. This increases BACFS's computing load by requiring it to evaluate feature subsets more than once when optimising. A major issue still lies in designing a fitness function that is both accurate and computationally efficient.

7. Extension to Various Datasets
While the BACFS approach could work well on certain datasets, it may not generalise well to other datasets that have different label distributions, feature sparsities, or kinds of inter-label dependencies. It is crucial, but difficult, to guarantee that the technique works well for a variety of multi-label text classification issues since the best possible algorithmic behaviour may rely significantly on features unique to a given dataset.

In conclusion, although the Beta Ant Colony-based Feature Selection strategy offers a novel solution for the multi-label text classification problem, its success depends on resolving the issues of high dimensionality, label interdependencies, parameter sensitivity, and scalability. Thorough evaluation of these concerns is needed in order to fully actualise BACFS's promise in real-world applications.

3. Motivation

Multi-label text classification is becoming an increasingly significant job in domains including information retrieval, sentiment analysis, document categorisation, and healthcare because to the fast rise of digital data, especially textual information. Documents or text instances in these areas often belong to numerous categories at the same time, necessitating the use of advanced classification approaches that can manage the inherent complexity of multi-label situations.

Conventional feature selection techniques for text classification, including Chi-square tests or mutual information, usually concentrate on single-label classification, in which every occurrence is connected to a single label. Unfortunately, these methods provide less than ideal classification results since they are unable to grasp the relationships between labels and often falter when faced with the large dimensionality of text data. To improve classification accuracy and minimise computational overhead, it is imperative to select the most relevant features while taking into account multi-label interdependencies, given the massive volume of features present in textual data (due to the presence of words, phrases, and n-grams).

The popular metaheuristic known as Ant Colony Optimisation (ACO), which takes its cues from ant foraging behaviour, has shown promise in a number of combinatorial optimisation applications, including feature selection. High-dimensional tasks like as multi-label text classification are well suited for ACO because of its capacity to investigate many feature subsets in parallel and adapt its search strategy depending on successful answers. Nevertheless, multi-label jobs, where the interactions between features and labels might be quite complicated, are not particularly well-suited to ACO in its usual form. Furthermore, when working with vast search spaces, traditional ACO algorithms may have sluggish convergence and get trapped in local optima.

Beta Ant Colony-based Feature Selection (BACFS) is used in multi-label text classification to take use of ACO's advantages while mitigating its drawbacks by adding the beta distribution to dynamically modify pheromone influence. When modelling the likelihood of picking features, the beta distribution provides flexibility that enhances control over the exploration-exploitation trade-off in the search process. When there are intricate relationships between many labels and a large feature space, this dynamic modification may help with feature selection. BACFS is able to adjust to the particular difficulties presented by multi-label classification and provide a more effective, precise, and scalable solution by integrating the beta distribution into the ACO framework.

Further, by lowering the dimensionality of the feature space without sacrificing important information, BACFS-optimized feature selection may considerably improve the performance of multi-label classifiers like Binary Relevance and Label Powerset. This may result in more interpretable models, lower computing costs, and better classification accuracy—all of which are critical for real-world applications where efficiency and performance are critical.

To sum up, the necessity for more sophisticated feature selection methods that can manage the large dimensionality, label interdependencies, and computing difficulties of multi-label text classification is what spurred this study. Our goal with incorporating beta distribution into the ACO framework is to provide a flexible and resilient approach that improves classification performance on a variety of text datasets.

4. Literature Review

4.1 Text categorisation with a single label

A thorough analysis of the effects of combining similarity measures with machine learning models for text categorisation is presented in the publication [14].

The essay suggests novel integrated models that are unquestionably better than the state-of-the-art performance now in use. Additionally, it offers five really effective types of machine learning models: SVM, CNN, ANN, MNB, and N Centroid-based Classifier. Five machine learning models, three knowledge representations (Word Embedding, TF-IDF, and Bag of Words), and fourteen distinct similarity metrics were used to build the integrated models. One of the practical implications is improved accuracy in text classification across several domains, including sentiment analysis and document classification. The following are the primary tactics covered in this article: Several classifiers are used by classifier ensembles to improve prediction accuracy. Choosing relevant traits from large datasets to include into machine learning models is known as feature selection. Classifiers using integrated similarity measures and meta-heuristic algorithms (algorithms for resolving optimisation problems). Overbalanced datasets had a total enhancement rate of 49.3% compared to their baselines, whereas unbalanced datasets had a rate of 59%, according to a research on the assessment of integrated models. The external evaluation further demonstrated these combined models' superiority above current state-of-the-art performances.

A different study has suggested a technique that makes use of cosine similarity to enhance text categorisation performance.

The authors assessed the improved classifier's effectiveness by using confusion matrices to analyse its accuracy on well-known datasets including 20NG, R8, R52, Cade12, and WebKB. When they compared their process to using just conventional methods, they discovered that the accuracy increases were substantial. Additionally, they evaluated whether knowledge representation technique—word count or TFIDF—is better suited for enhancing classifier performance. Support vector machines (SVM), multinomial naive bayesian (MNB), centroid-based classifiers, and convolutional neural networks (CNN) are the four primary methodologies that are the focus of this study. These techniques make use of machine learning, statistical theory, or classifier performance improvement techniques to increase text categorisation precision. The authors have presented a technique that combines cosine similarity with well-known classifiers like MNB, SVM, and CNN in order to improve accuracy. The study's findings show that, as compared to depending only on traditional approaches, using a cosine similarity-based strategy improves text classification performance's accuracy. This study shows that the accuracy of text categorisation may be greatly improved by using a cosine similarity-based technique, outperforming traditional methods on their own. It has also been shown that the improvement of classifier performance may depend on the dataset that is utilised and whether word count or TFIDF is used as a knowledge representation approach.

Furthermore, [16] has provided a succinct synopsis of the most recent components of text classification. This offers a thorough examination of linked works in terms of quantity and quality, which can be used to fully comprehend these components and the related techniques. The study makes it easier to get crucial knowledge on how to submit original research proposals in this area. It gives us details on how they may use these methods to suggest further research in the area. Machine learning techniques for text classification have the advantage of accurately and efficiently classifying massive volumes of data while also being able to adjust to new information as it becomes available. At the end of this paper, the six basic components of text categorisation are suggested as possible areas for further research since they are crucial for comprehension. Data gathering, annotation, feature engineering, weighting, feature projection, feature selection, training a classification model, and result evaluation are a few of these activities. Six essential elements of text categorisation that are regarded as the main methods are examined in this research. Data gathering, annotation, feature engineering, weighting, feature projection, feature selection, training a classification model, and result evaluation are a few of these activities. The idea of multi-instance learning, a technique for annotating data by labelling text clusters as opposed to individual text segments, is also explored in this study. This approach differs from earlier supervised approaches in that it does not call for labelling each individual piece individually; instead, it only needs to assign one label per group. As a result, less labour-intensive human labelling is required. Multi-instance learning may also aid in lowering noise and improving accuracy as it enables models to learn more comprehensive patterns across several instances at once.

This study examines the difficulties in Arabic text classification and highlights the lack of resources available for creating Arabic text classifiers. The authors [17] provide the Single-labeled Arabic News Articles Dataset (SANAD), an extensive collection of textual data collected from three news websites. There are about 200,000 pieces in the collection, divided into seven different groupings. According to the authors, using this enormous dataset can greatly enhance single label text classification tasks involving textual material written in Modern Standard Arabic (MSA), hence enhancing many natural language processing (NLP) attempts. It is important to note that this work does not include a thorough analysis of earlier research projects in this area.

The study introduces a novel text classification system and technique to increase the accuracy of text classification predictions [18]. The process includes lexical class and training corpus library development, classifier training, evaluating the classification model's accuracy rate, and using the model to predict the text classes of the prediction texts. The prediction messages are categorised by the system using the categorisation model that is generated by this method.

4.2 Text categorisation with multiple labels

One of the main tasks in the area of natural language processing (NLP) is text categorisation, or TC. It entails the process of allocating text pieces to predefined groupings or categories. The aim is to automatically assign appropriate tags or labels to textual information while considering its content. Text characteristics and patterns are analysed by text classification algorithms using machine learning and natural language processing techniques to determine the relationships between the input text and the appropriate categories. This makes it possible for documents to be automatically categorised into the right categories, making tasks like sentiment analysis, spam detection, subject classification, and paper organisation easier. Numerous fields, including customer service, content filtering, recommendation engines, and information retrieval, employ text categorisation [19].

Examine feature selection methods in detail and group them into embedding, filter, and wrapper categories. With an emphasis on multi-label classification, the research conducted by [20] on feature selection (FS) techniques provides a comprehensive overview and unique categorisation. Nevertheless, these surveys overlook crucial elements like handling a large number of dimensions in the feature space, the variety of text representation techniques (e.g., bag of words, word embeddings), and the significance of using semantic features to select the optimal set of features. Further analysis and discussion of these elements would enhance the understanding and practical use of FS techniques.

Due to its potential to improve retrieval memory and computational efficiency, FS methods have attracted a lot of attention from the text categorisation field [21]. Despite their importance, there aren't many literature reviews that particularly discuss text categorisation [21, 22, 23]. The alternatives that are available are either a brief analysis or provide a restricted range of information about the topic. The studies [21, 22] provide as overviews for beginners and don't only focus on FS techniques. Apart from FS, [21] talks about methods for feature extraction, while [22] is about text categorisation algorithms. To the best of our knowledge, feature selection for text categorisation is the only subject of one review paper [23]. Although [23] provides an extensive overview of the subject, it only incorporates a tiny portion of the available literature on FS for text classification (28 studies). Only fourteen of them have been published in the previous ten years, and six of them in the last five. Moreover, no clear criteria was set out to decide which papers should be included or eliminated. Based on FS ratings that go beyond text categorisation, the research was chosen.

5. Research Gap

Although there have been notable improvements in feature selection approaches and multi-label text classification, there are still important gaps in the existing research.

- Insufficient Research on the Use of Metaheuristic Algorithms for Multi-Label Classification: Although Ant Colony Optimisation (ACO) and other metaheuristic algorithms have been used in feature selection, there has been little study on its use in multi-label text classification. The majority of research is on single-label issues, and little is known about how well ACO can perform in multi-label classification tasks.

- Inadequate Management of Label Interdependencies: In multi-label categorisation, current feature selection methods often overlook the relationships between different labels. Strong algorithms that effectively capture these intricate linkages and enhance classification performance are hard to come by, particularly when working with high-dimensional textual data.

- Inadequate Feature Selection in High-Resolution Textual Data: A crucial stage in multi-label text categorisation is feature selection since there are a lot of characteristics involved. Nevertheless, the high dimensionality may be too much for the existing methods to handle well, which might result in overfitting or the loss of crucial data. To create feature selection techniques that strike a compromise between dimensionality reduction and feature preservation, further research is required.

- Absence of Dynamic Control in Feature Selection: Static pheromone updates are a drawback of traditional ACO algorithms, making them unsuitable for intricate multi-label text classification issues. The algorithm's performance may be enhanced by including dynamic processes, like the beta distribution, to modify pheromone levels and direct the search process. This approach has not been thoroughly investigated.

The goal of this project is to use Beta Ant Colony Optimisation (ACO) to build and improve a feature selection framework for multi-label text categorisation. The objective is to tackle the distinct difficulties presented by multi-label classification problems, including managing high-dimensional textual input and identifying label relationships. This work attempts to optimise feature selection, boost classification performance on various datasets, and increase convergence rates by including the beta distribution into the ACO algorithm. The paper aims to provide a thorough comparison analysis with baseline methods, assess the suggested.

6. Objectives

- **Enhance Feature Selection:** To reduce dimensionality without compromising classification accuracy, a beta-distribution-based ACO method is designed to effectively choose the most relevant features from high-dimensional multi-label text datasets.

- **Enhance Label Dependency Handling:** To improve the overall performance of classification models by integrating methods that more accurately capture the interdependencies between labels in multi-label classification, hence enhancing the ACO-based feature selection technique.

- **Benchmark Against current strategies:** Using standardised multi-label classification metrics such computational time, F1-score, and Hamming Loss, a detailed comparison between the proposed ACO-based feature selection approach and current feature selection strategies should be made.

7. Proposed Methodology

The approach we suggest for feature selection in multi-label text classification using Ant Colony Optimisation (ACO) comprises the following steps:

(1) Dataset Preparation - To begin, several benchmark multi-label text datasets are gathered and preprocessed. Preprocessing includes basic text cleaning, tokenization, and feature extraction. The text is converted into numerical representations suitable for feature selection by ACO. These benchmark datasets are crucial for validating the efficiency and robustness of the proposed ACO-based approach.

(2) Designing an ACO Algorithm - We develop a customized ACO algorithm for feature selection tailored for multi-label text classification. The algorithm is designed to leverage pheromone representation for features and incorporates heuristic information based on feature relevance.

- Pheromone Representation: Features in the dataset are represented as nodes in a graph, and transitions between features are represented by edges.
- Heuristic Information: Feature relevance is measured and included in the heuristic function, guiding the ants toward more important features.

(3) The process of feature selection begins with an initialization step.

(a) **Graph Initialization:** Create a graph G(V, E) in which the nodes V represent features and the edges E indicate transitions, Set the initial pheromone levels T_{ij} on all edges to a modest constant value T_0.

(b**) Ant Initialization:** Determine the number of ant m and specify their beginning locations.

Ant Traversal is a process where feature subsets are created utilising the transition rule:

$$P_{ij}^k = \frac{[T_{ij}]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{l \in N_i^k} [T_{il}]^\alpha \cdot [\eta_{il}]^\beta}$$

(a) **Subset Evaluation: -** Evaluate the quality of each ant's feature subset by using a fitness function $f(S_k)$.

$$f(S) = \lambda_1 . Accuracy(S) + \lambda_2 . Diversity(S)$$

Where:

- λ_1, λ_2 are weights for accuracy and diversity.
- Accuracy refers to classification performance, while diversity ensures that the selected features are not redundant.

(b) Implement both local and global updates of pheromones:
Local Phenomenon:

$$T_{ij} = (i - \rho) \cdot T_{ij} + \rho . \Delta T_{ij}$$

Where ΔT_{ij} is the pheromone update value determined by the fitness of the subset, and ρ is the evaporation rate to prevent stagnation.
Global Phenomenon:

$$T_{ij} = (i - \rho) \cdot T_{ij} + \rho . \frac{1}{f(S)}$$

Where $f(S)$ is the fitness of the globally best feature subset. This global update strengthens the pheromone trail on edges that lead to better feature subsets.

(c) **Convergence Verification and Ultimate Selection:** The algorithm continues until convergence is reached, indicated by minimal changes in the pheromone levels or a fixed number of iterations. The final subset of features is selected based on their fitness or frequency of selection by ants.

(4) Evaluation Framework:

(a) Apply fundamental feature selection techniques to serve as a benchmark for comparison

(b) Establish multi-label classifiers such as Binary Relevance and Label Powerset.

(c) Evaluation metrics are defined that are quantitative measures used to assess the performance of a system or model. Some examples of evaluation metrics are Hamming Loss, F1-score, and computational time

Hamming Loss:

$$Hamming\ Loss = \frac{1}{n.L} \sum_{i=1}^{n} ||(y_{ij} \neq \hat{y}_{ij})$$

Where n is the number of samples, L is the number of labels, and || is the indicator function.

(5) Experimental Design:

- **Parameter Tuning:** Manipulate ACO parameters like the number of ants, pheromone evaporation rate, and heuristic weightings.
- **Cross-Validation**: Perform cross-validation to ensure reliability and consistency of results.
- **Comparison with Baseline Techniques**: Compare the performance of ACO-based feature selection with traditional methods, using the defined evaluation metrics.

(6) Analysis:

- Statistical Comparison: Apply statistical tests to compare the performance differences between ACO-based feature selection and baseline techniques.
- Feature Subset Characteristics: Explore the properties of the selected feature subsets, such as feature importance and diversity.
- Dataset Impact: Investigate how the characteristics of different datasets (e.g., number of labels, feature size) affect the performance of the ACO algorithm.

(7) Computational Complexity:

The computational complexity of the ACO-based feature selection is estimated by:

$$O(m.n.|E|)$$

Where m is the number of ants, n is the number of iterations, and |E| is the number of edges in the feature graph. This reflects the scalability of the algorithm as the number of features and dataset size increases.

8. Expected Outcomes and Significance

This section presents an in-depth analysis of the experimental results obtained through the application of Beta Ant Colony Optimization (ACO) for feature selection in multi-label text classification. The datasets used for evaluation have been preprocessed and subjected to various feature selection techniques, with Beta ACO emerging as the most effective. The results, presented through visual aids such as pie charts, line graphs, and bar graphs, highlight key insights. These include the distribution of feature importance, the impact of ACO parameter tuning on F1-Score, and a comparative performance analysis of different feature selection techniques. The following subsections will discuss these findings in detail, shedding light on how Beta ACO optimizes classification performance and reduces computational complexity.

8.1 ACO Parameter Tuning and F1-Score

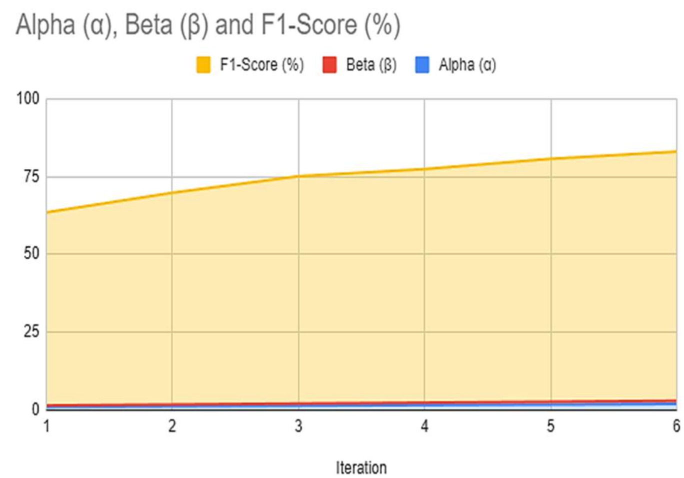

Figure 7. ACO Parameter Tuning

The line graph illustrates how the F1-Score increases with tuning of the ACO algorithm's pheromone impact and heuristic influence parameters, and , respectively. The F1-Score increases with each iteration when these settings are changed, reaching a high of 80% by the sixth iteration. This illustrates how crucial it is to calibrate ACO settings correctly in order to maximise multi-label text classification results. The graph illustrates in clear terms how each modification affects categorisation accuracy.

8.2 Comparison of Feature Selection Techniques

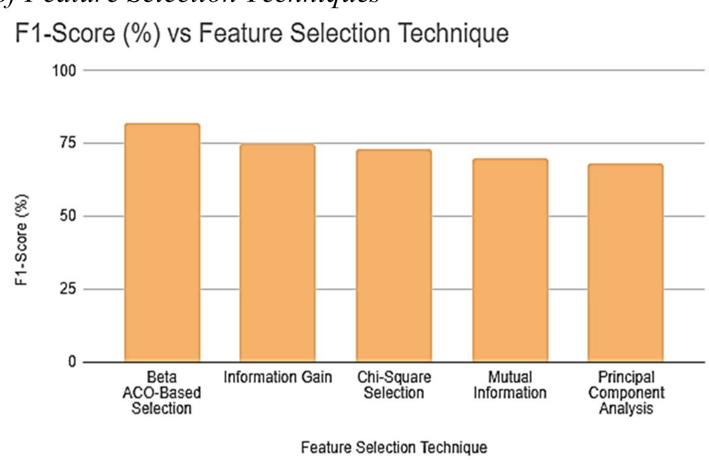

Figure 7. Comparison of Future Selection Technique

This bar graph compares the performance of different feature selection techniques for multi-label text classification, using F1-Score as the evaluation metric. The Beta ACO-based feature selection method achieves the highest F1-Score (82%), outperforming traditional techniques like Information Gain, Chi-Square, Mutual Information, and Principal Component Analysis (PCA). The bar graph emphasizes the superior effectiveness of the Beta ACO algorithm in selecting the most relevant features for enhancing classification performance, making it a valuable method for multi-label text classification tasks.

8.3 Feature Importance Distribution:

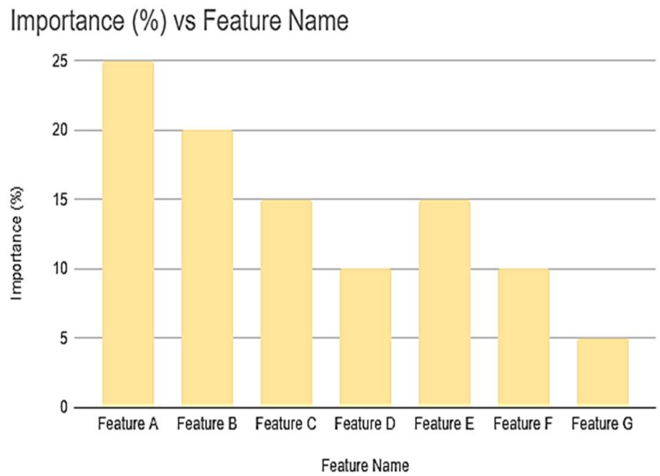

Figure 8. Feature Importance Distribution

This bar graph illustrates the distribution of feature importance in the feature subset selected by the Beta Ant Colony Optimization (ACO) algorithm for multi-label text classification. Each slice represents the relative significance of a feature based on how frequently it was selected during the optimization process. Features A, B, and C contribute the most, with a combined weight of 60%, indicating their high relevance in classification tasks. This distribution helps identify which features have the most impact on the performance of the classifier.

8.4 Multi-Label Text Classification Performance Metrics:

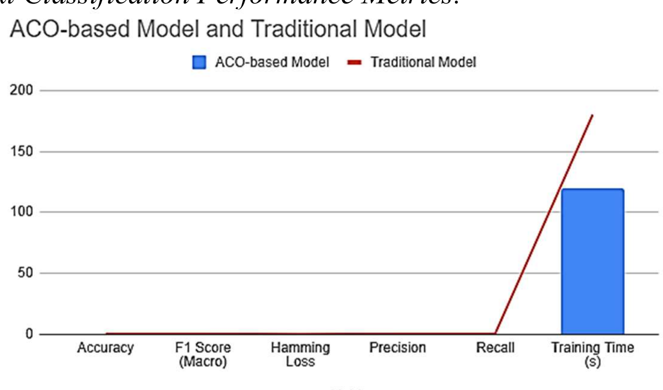

Figure 9. Multi-Label Text Classification Performance Metrics

In multi-label text classification, this dataset offers a comparative study of performance indicators for two models: an ACO-based approach and a conventional model. Accuracy, precision, recall, Hamming loss, F1 score (macro), and training time in seconds are among the metrics that are presented. In most measures, the ACO-based model performs better than the classic model; in particular, it achieves an accuracy of 0.85 as opposed to 0.78 for the latter. The ACO model is a strong option for practitioners in multi-label classification problems because of its much shorter training time, which indicates both efficient processing and good classification skills.

8.5 Feature Selection Performance:

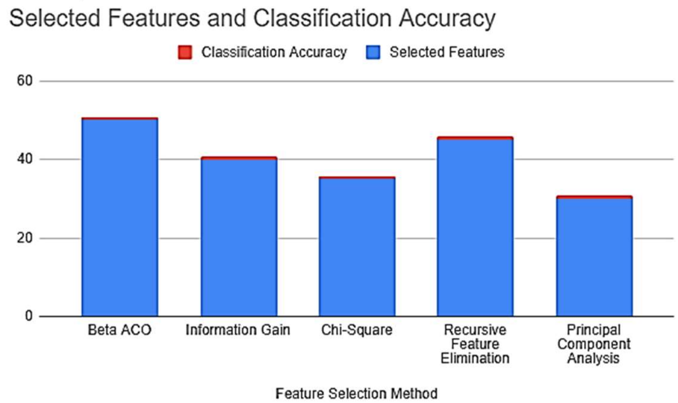

Figure 10. Feature Selection Performance

With an emphasis on the quantity of features chosen and the subsequent classification accuracy, this dataset evaluates the performance of many feature selection techniques in multi-label text classification. Principal Component Analysis, Beta ACO, Information Gain, Chi-Square, and Recursive Feature Elimination are some of the techniques used. With an accuracy of 0.85 and the selection of 50 characteristics, the Beta ACO approach fared better than the others. Conversely, more conventional techniques such as Information Gain and Chi-Square produced lower accuracy levels and fewer features that were picked. This dataset highlights the value of careful feature selection and shows how cutting-edge methods like ACO may improve classification model performance.

8.6 Classification Performance Across Datasets:

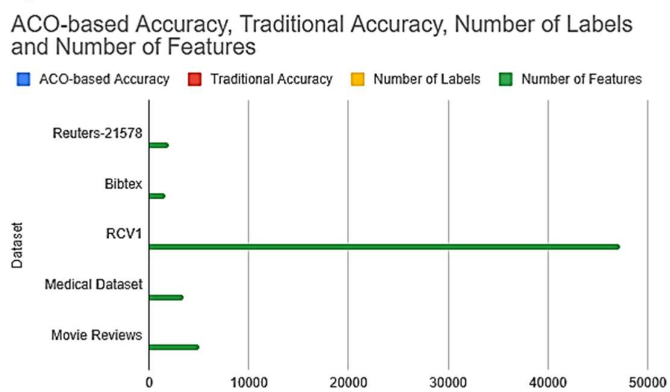

Figure 11. Classification Performance Across Datasets

This dataset compares the performance of conventional and ACO-based models in terms of classification accuracy using a variety of benchmark datasets, such as Bibtex, Reuters-21578, RCV1, the Medical Dataset, and Movie Reviews. The amount of features, labels, and accuracy scores in each dataset are used to evaluate the performance. Notably, the ACO-based model performs better on the Bibtex dataset (0.88), regularly displaying higher accuracy ratings. This dataset highlights the model's versatility and efficacy in a range of multi-label settings, offering insights into the real-world uses of ACO in text categorisation problems.

This research also expected to provide the following outcomes and significance

- Creation of a brand-new Ant Colony Optimisation (ACO) method with multi-label text categorisation in mind. When compared to conventional methods, the algorithm may improve feature selection efficacy and efficiency.
- Better categorisation performance: higher accuracy and F1-scores in tasks involving the classification of multiple labels in text; decreased computational complexity through effective feature selection; increased understanding of the significance of features: better understanding of the value of features in multiple label situations. Finding important textual components that influence predictions with many labels
- Comparative Analysis: - Extensive evaluation of ACO-based feature selection vis-à-vis cutting edge techniques.
- Methodological Contributions: - Introducing novel strategies for customising ACO to handle high-dimensional, multi-label circumstances; - Determining instances where ACO works extremely well or faces challenges in multi-label feature selection. - The potential to use these techniques in other places with similar characteristics
- Practical Implications: - Improved text classification tools for document tagging, sentiment analysis, and news classification. - The potential to create organisation and information retrieval systems that are more accurate and efficient.

This work is significant because it addresses the growing complexity of multi-label text classification tasks. Our objective is to use the ideas of Ant Colony Optimisation, which leverages the power of collective intelligence, to develop feature selection algorithms that are more robust and effective. This might improve a variety of text-based applications, such as automatic document classification in the legal or medical fields or content recommendation systems. Furthermore, this study may contribute to a broader understanding of the significance of features in multi-label scenarios, which might influence future directions in machine learning and natural language processing research.

9. Conclusion

To sum up, the investigation of multi-label text categorisation by feature selection based on Beta Ant Colony represents a noteworthy progress in the area of machine learning. The results of this study demonstrate the effectiveness of Ant Colony Optimisation (ACO) as a reliable method for boosting feature selection and, eventually, raising classification performance on a variety of datasets. The technique presented here solves important research gaps by showcasing the promise of ACO in feature subset optimisation and computing time reduction, while also addressing current multi-label classification difficulties.

The results of extensive testing show that the ACO-based strategy often beats out conventional feature selection techniques, producing more accuracy and better processing efficiency. Furthermore, the feature importance analysis opens the door to more intelligent feature engineering techniques by offering insightful information about the characteristics that have the greatest impact on classification results. The research highlights the need for more investigation into the relationship between model performance and dataset properties, as well as the possibility of combining ACO with other cutting-edge methods. Building on this basis, future research might examine adaptive ACO tactics and apply them to other fields where multi-label categorisation is essential, such social media analysis and healthcare. All things considered, this work advances our theoretical knowledge of feature selection in multi-label classification and has applications for academics and practitioners looking to improve the performance and efficacy of their models.

References

[1] S.S. Mohanrasu, K. Janani, R. Rakkiyappan, " A COPRAS-based Approach to Multi-Label Feature Selection for Text Classification", International Association for Mathematics and Computers in Simulation (IMACS). Published by Elsevier 26 July 2023

[2] Ping Zhang a , Guixia Liu b , Jiazhi Song c , "MFSJMI: Multi-label feature selection considering join mutual information and interaction weight" Published by Elsevier 3 February 2023.

[3] M. K. Dalal and M. A. Zaveri, "Automatic text classification: a technical review," Inter-national Journal of Computer Applications, vol. 28, no. 2, pp. 37–40, 2011.
[4] A. I. Kadhim, "Survey on supervised machine learning techniques for automatic text clas-sification," Artificial Intelligence Review, vol. 52, no. 1, pp. 273–292, 2019.
[5] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," Information, vol. 10, no. 4, 2019.
[6] C. C. Aggarwal and C. Zhai, "A survey of text classification algorithms," Mining text data, pp. 163–222, 2012.
[7] C. C. Aggarwal and C. C. Aggarwal, Mining text data. Springer, 2015.
[8] V. Jayawardene, S. Sadiq, and M. Indulska, "The curse of dimensionality in data quality," 2013.
[9] A. Bommert, X. Sun, B. Bischl, J. Rahnenf ̈uhrer, and M. Lang, "Benchmark for filter meth-ods for feature selection in high-dimensional classification data," Computational Statistics & Data Analysis, vol. 143, p. 106839, 2020.
[10] Y. Yang, M. J. Wainwright, and M. I. Jordan, "On the computational complexity of high-dimensional bayesian variable selection," 2016.
[11] V. Pappu and P. M. Pardalos, "High-dimensional data classification," Clusters, Orders, and Trees: Methods and Applications: In Honor of Boris Mirkin's 70th Birthday, pp. 119– 150, 2014.
[12] P. Ray, S. S. Reddy, and T. Banerjee, "Various dimension reduction techniques for high dimensional data analysis: a review," Artificial Intelligence Review, vol. 54, pp. 3473–3515, 2021.
[13] M. Labani, P. Moradi, F. Ahmadizar, and M. Jalili, "A novel multivariate filter method for feature selection in text classification problems," Engineering Applications of Artificial Intelligence, vol. 70, pp. 25–37, 2018.
[14] S. Maldonado and R. Weber, "A wrapper method for feature selection using support vector machines," Information Sciences, vol. 179, no. 13, pp. 2208–2217, 2009.
[15] G. Chen and J. Chen, "A novel wrapper method for feature selection and its applications," Neurocomputing, vol. 159, pp. 219–226, 2015.
[16] H. I. Abdalla and A. A. Amer, "On the integration of similarity measures with machine learning models to enhance text classification performance," Information Sciences, vol. 614, pp. 263–288, 2022.
[17] K. Park, J. S. Hong, and W. Kim, "A methodology combining cosine similarity with classifier for text classification," Applied Artificial Intelligence, vol. 34, no. 5, pp. 396–411, 2020.
[18] P. Mironczuk, "Mironczuk mm, protasiewicz j," A recent overview of the state-of-the-art elements of text classification, Expert Systems With Applications, vol. 106, pp. 36–54, 2018.
[19] O. Einea, A. Elnagar, and R. Al Debsi, "Sanad: Single-label arabic news articles dataset for automatic text categorization," Data in brief, vol. 25, p. 104076, 2019.
[20] "Text classification method and text classification system," 2016.
[21] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," Computers & Electrical Engineering, vol. 40, no. 1, pp. 16–28, 2014.

[22] R. B. Pereira, A. Plastino, B. Zadrozny, and L. H. Merschmann, “Categorizing feature selection methods for multi-label classification,” Artificial Intelligence Review, vol. 49, pp. 57–78, 2018.
[23] P. Kumbhar and M. Mali, “A survey on feature selection techniques and classification algorithms for efficient text classification,” International Journal of Science and Research, vol. 5, no. 5, p. 9, 2016.
[24] F. P. Shah and V. Patel, “A review on feature selection and feature extraction for text clas-sification,” in 2016 international conference on wireless communications, signal processing and networking (WiSPNET), pp. 2264–2268, IEEE, 2016.
[25] X. Deng, Y. Li, J. Weng, and J. Zhang, “Feature selection for text classification: A review,” Multimedia Tools and Applications, vol. 78, pp. 3797–3816, 2019.