The Journal of Computer Science and Its Applications. An International Journal of the Nigeria Computer Society (NCS) Vol.32, No.1, June, 2026

Evolutionary Deep Learning for Social Media Keyword Optimization

V. Kaliappan
EY, Boston, USA

velumail@gmail.com

Abstract

This paper proposes an innovative method for intelligent and automated keyword selection tailored for Twitter data analysis, specifically addressing public health surveillance. It integrates deep learning with evolutionary computing to generate effective keyword sets for data collection. By optimizing against metrics like tweet retrieval power, this approach significantly enhances the relevance and reach of collected social media information. Empirical findings demonstrate that this combined learning strategy outperforms traditional human-driven keyword selection in real-world scenarios.

Keywords: Social Media Analytics, Keyword Optimization, Twitter Mining, Evolutionary Computation, Particle Swarm Optimization, Word Embeddings, GloVe, Syndromic Surveillance, Public Health Monitoring, Semantic Similarity.

1 Introduction

The evolution of social media platforms has drastically transformed the landscape of data acquisition and real-time analytics across various domains, particularly in public health monitoring. Among the plethora of platforms available today, Twitter emerges as a powerful tool due to its real-time nature, public accessibility, and vast volumes of user-generated textual data. With millions of users sharing personal thoughts, health experiences, and observations daily, Twitter offers an invalu-able source of information that can be mined for actionable insights.

In the domain of syndromic surveillance, which involves the continuous and near real-time collection, interpretation, and dissemination of health-related data, the integration of Twit-ter data introduces new dimensions. Syndromic surveillance aims to detect potential public health threats before clinical confirmation, thereby enabling timely intervention. Traditional surveillance systems primarily rely on formal healthcare data sources; however, these systems often suffer from reporting delays. In contrast, Twitter enables early detection by capturing spontaneous expressions of illness, symptoms, and health-related sentiments from the general public.

Prior research has demonstrated the feasibility of using Twitter for health monitoring tasks such as influenza tracking, mental health analysis, and allergen detection. Despite its util-ity, one of the most pressing challenges in leveraging Twitter for health analytics lies in the keyword selection process. Since Twitter users tend to use informal, colloquial, or slang expressions to describe their health experiences, identifying appropriate keywords becomes non-trivial. A poorly curated keyword list may result in either too few tweets (if the terms are too specific) or excessive irrelevant noise (if the terms are too broad). Hence, the effectiveness of tweet collection hinges significantly on the precision and

relevance of the selected keyword set.

To address this issue, we propose a novel, intelligent framework for automatic keyword selection that utilizes both deep semantic modeling and evolutionary computation. The approach begins with a domain-specific set of seed keywords related to a particular health condition—in our case, asthma and difficulty in breathing. These initial keywords are semanti-cally expanded using Global Vectors for Word Representation (GloVe) trained on a large corpus of Twitter data. GloVe embeddings allow the system to capture contextual similarities between words and phrases, thus generating a comprehensive candidate pool of keywords that better reflects the linguistic diversity found on social media.

Subsequently, the refined candidate set is subjected to optimization using Particle Swarm Optimization (PSO)—a population-based stochastic optimization technique inspired by social behavior in swarms. By modeling keyword combinations as particles within a high-dimensional search space, the PSO algorithm iteratively explores this space to discover subsets of keywords that maximize information retrieval performance. The optimization objective is designed to strike a balance between tweet relevance and tweet volume, ensuring that the collected data is both meaningful and sufficiently representative.

We evaluate our approach using a real-world use case in syndromic surveillance focused on respiratory syndromes in the United Kingdom. Our experimental setup compares the proposed automatic PSO-based method against a traditional baseline involving human-expert-curated keywords. Performance is assessed through both standard metrics such as precision and novel measures like Tweet Retrieval Power (TRP), which jointly consider relevance and reach. The results indicate that our method not only improves precision but also maintains satisfactory reach, outperforming the human baseline in overall utility.

In summary, the contributions of this work are threefold:

1) the development of a hybrid keyword selection system leveraging GloVe embeddings and PSO,
2) the creation of a robust evaluation framework using both traditional and problem-specific metrics, and
3) a validated case study that demonstrates the system's efficacy in real-time health monitoring.

Although our focus is on asthma-related syndromes, the methodology is broadly applicable to other domains involving keyword-driven social media data collection. This establishes a scalable and adaptable foundation for future work in social media-based public health intelligence.

2. Literature Review

The task of keyword selection for social media data mining, particularly for platforms like Twitter, has been a subject of growing interest across various fields including information retrieval, public health, and machine learning. Traditional keyword selection approaches have relied heavily on domain experts who manually curate lists of relevant terms for specific use cases (Triple, 2011; De Quincey and Kostkova, 2009). However, these manually constructed lists are often limited in scope and fail to capture the evolving nature of language, particularly on informal platforms like Twitter (Morstatter et al., 2013; Liu et al., 2019).

Numerous studies have proposed semi-automated techniques for keyword identification based on lexical resources such as thesauri or medical ontologies (Kostkova et al., 2013; Collier and Kawazoe, 2008). While such methods provide more structure, they still lack the contextual understanding required to adapt to dynamic, user-generated language. Urban Dictionary and slang repositories have been utilized in conjunction with expert lists to improve coverage (Edo-Osagie et al., 2019), yet they often introduce ambiguities and irrelevant data.

In response, machine learning-based keyword selection techniques have emerged. Feature selection frameworks such as mutual information, information gain, and chi-square statis-tics have been used to identify high-impact terms for classification tasks (Kiritchenko and Jiline, 2008; Forman, 2003; Yang and Pedersen, 1997). Deep learning approaches further advanced the field by incorporating semantic representations using word embeddings like Word2Vec and GloVe (Liu et al., 2019; Pennington et al., 2014; Bojanowski et al., 2017). These embeddings provide vectorized representations of words that encode syntactic and semantic similarity, facilitating keyword expansion via similarity metrics (Bojanowski et al., 2017).

Several works have explored using embeddings for keyword identification in marketing (Lee and Kim, 2019; Li and Xu, 2020), sentiment analysis (Zhang et al., 2018), and misinformation detection (Shu et al., 2020). Social media applications have particularly benefited from such techniques. For example, Chen et al. (2018) applied temporal topic models to detect influenza phases from tweet content, while Paul and Dredze (2011) modeled health topics across social platforms.

Despite these advances, keyword selection still faces significant challenges in balancing relevance with coverage. This is particularly critical in public health surveillance where both the quality and quantity of data impact downstream analysis (Charles-Smith et al., 2019; Achrekar et al., 2011). Overly precise keywords might

miss useful data, whereas broad ones could collect excessive noise (Araujo et al., 2017). Optimization-based strategies have gained popularity as they provide a framework for balancing multiple conflicting objectives. Particle Swarm Optimization (PSO), in particular, has shown promise in text mining and feature selection tasks (Kennedy and R. Eberhart, 1995; Umapathy et al., 2010; Deb and Padhye, 2010). PSO has been applied to search engine marketing (Liang et al, 2019), power system optimization (Abido, 2002), and sentiment classification (Ma et al., 2017), demonstrating its robustness in complex search spaces. The integration of evolutionary computation and semantic models has also been explored. Zhang et al. (Zhang and Liu, 2015) combined genetic algorithms with semantic feature extraction to improve web crawling. Other researchers have proposed hybrid frame-works for social sensing that utilize neural models for semantic expansion and swarm-based methods for optimization (Gupta et al., 2018; Kang et al., 2017). Our approach builds upon this literature by introducing a two-stage hybrid system that uses GloVe embeddings for semantic keyword expansion and PSO for optimal keyword subset selection. Unlike prior works, we explicitly design the objective function to account for both the relevance (precision) and volume (reach) of retrieved tweets, which is particularly vital for syndromic surveillance tasks (Edo-Osagie et al., 2020; Edo-Osagie et al., 2020). In summary, the proposed framework contributes to the ongoing discourse on intelligent keyword selection by offering a domain-adaptive, scalable, and automated alternative that improves data quality and minimizes human intervention in keyword curation.

3. Methodology

In this section, we present our hybrid methodology for intelligent and automatic keyword selection, which integrates deep word embeddings with evolutionary optimization techniques. The approach unfolds in two key stages:

1) Semantic expansion of initial keywords using GloVe embeddings, and
2) Optimization of the expanded keyword set using Particle Swarm Optimization (PSO).

Figure 1 provides a high-level view of the system pipeline, highlighting the flow from initial keyword seed to optimized keyword set selection.

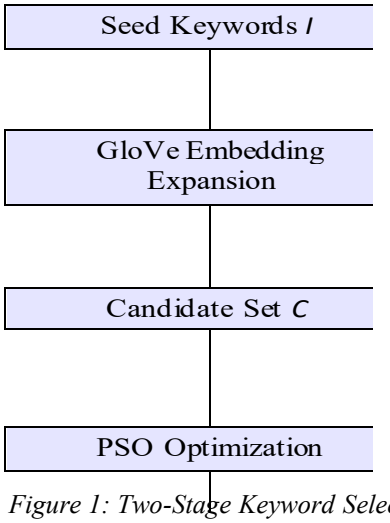


Figure 1: Two-Stage Keyword Selection System Architecture

3.1 Stage I: Semantic Expansion via GloVe Embeddings

The process begins with a curated list of seed keywords, denoted as $\boldsymbol{I} = \{w_1, w_2, \ldots, w_p\}$, selected based on domain knowledge (e.g., terms related to asthma and respiratory conditions). We use pre-trained GloVe embeddings trained on a corpus of five million unlabeled tweets, which capture both semantic and syntactic relationships among words in Twitter-specific language.

For each word $w_i \in I$, we compute the n most semantically similar words using cosine similarity. These similar terms are retrieved from the embedding space as follows:

$$S = \bigcup_{w_i \in I} MostSimilar(w_i, n) \quad (1)$$

The final candidate keyword set C is defined as the union of the original and similar keywords:

$$C = I \cup S \quad (2)$$

This expansion ensures the candidate set is comprehensive and accommodates informal language and synonyms common on social media.

3.2 Stage II: Keyword Optimization via PSO

The keyword selection task is framed as a combinatorial optimization problem, where the goal is to identify a subset of keywords from C that maximizes tweet relevance and collection volume.

Each particle in the PSO algorithm represents a potential keyword subset and is encoded as a vector of integers. A value of zero indicates the absence of a keyword in that position. The objective function Z is designed to minimize irrelevance while maximizing coverage:

$$Z = \frac{\propto+\beta}{2} \quad (3)$$

where α represents the irrelevance penalty and β captures retrieval deficiency based on collected tweet metrics. Particles update their positions and velocities iteratively based on their personal best (*pbest*) and global best (*gbest*) solutions:

$$v_i^{t+1} = wv_i^t + c_1 r_i(pbest_i - x_i^t) + c_2 r_2(gbest - x_i^t) \quad (4)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (5)$$

Here, ω is the inertia weight, c_1 and c_2 are acceleration coefficients, and r_1, r_2 are random values between 0 and 1.

1) Clarification of PSO Encoding Scheme

Each particle is encoded as a discrete vector where each integer corresponds to an index in the candidate keyword set *C*. Although PSO is originally formulated for continuous spaces, we apply a discrete variant by rounding updated positions to the nearest integer index. Velocity values influence the probability of keyword inclusion rather than literal continuous movement. This configuration follows discrete PSO practices commonly used in text mining applications.

3.3 Algorithm Configuration

Table I outlines the PSO parameters used in our experiments.

Table 1: PSO configuration parameter

Parameter	Value
Swarm Size	100
Max Iterations	50
Inertia Weight (ω)	0.8
Cognitive Coefficient (c_1)	1.0
Social Coefficient (c_2)	1.0
Particle Dimension (D)	10

4. Implementation

The implementation of our automatic keyword selection framework was carried out in Python, leveraging several open- source libraries for deep learning, text processing, and evolutionary computation. This section details the practical setup, toolchain, and data pipeline used to realize the methodology described earlier.

1) Data Collection and Preprocessing

To train our semantic embedding model, we collected ap-proximately 5 million unlabeled English tweets via the Twitter Streaming API. These tweets were acquired without keyword filtering to ensure linguistic diversity. The data was subjected to preprocessing steps such as:

- Tokenization using the NLTKtokenizer.
- Removal of stopwords and punctuation.
- Lowercasing and normalization of slang using a Twitter-specific dictionary.
- Filtering out non-English tweets using langdetect.

The cleaned corpus served as the training input for building GloVe word embeddings.

2) GloVe Embedding Training

We trained 100-dimensional GloVe embeddings using the GloVetoolkit. The context window was set to 10, and only words with a minimum frequency of 5 were retained. This ensured that the resulting embeddings were rich in semantic relationships tailored to Twitter vernacular. The trained vectors were stored as a lookup table for similarity search during keyword expansion.

3) *Additional Details on GloVe Training*

The GloVe model was trained using the following hyperparameters: learning rate = 0.05, epochs = 50, context window size = 10, co-occurrence weighting parameter xmax = 100, and α = 0.75. We selected 100-dimensional vectors as a balance between computational cost and semantic richness. Alternative pretrained Twitter GloVe embeddings were evaluated, but domain-specific training provided better coverage of health-related slang and abbreviations.

4) Keyword Expansion and Candidate Generation

Given an initial seed list $\boldsymbol{I}$ of 20 domain-specific keywords, we used cosine similarity to retrieve the top-*n* semantically similar terms for each seed keyword. The value of *n* was set to 5 based on empirical tuning. The final candidate set $\boldsymbol{C}$ was constructed as a union of original and expanded terms, as defined in Equation (2) and (3).

5) Particle Swarm Optimization Implementation

We utilized the PySwarms library to implement the PSO component. The dimensionality D of each particle was fixed to 10, corresponding to a keyword set size limit. Each particle encoded a vector of integers, where each value indexed a keyword in the candidate pool C.

The objective function Z computed for each particle was based on live tweet collection statistics. Using the Twitter API, tweets were collected for each candidate keyword set. These tweets were manually annotated for relevance in a randomly sampled subset of 2000 tweets. From these annotations, the objective function metrics α

(irrelevance factor) and β (retrieval loss) were computed as described previously.

4) PSO Configuration and Execution

We initialized a swarm of 100 particles. The algorithm iterated for 50 generations. At each step:

- The relevance and reach of each keyword set were evaluated.
- Personal best (*pbest*) and global best (*gbest*) scores were updated.
- Each particle's velocity and position were updated using Equations (4) and (5).

The parameters used are summarized in Table I, previously presented.

5) Pseudocode for PSO-based Keyword Selection

[ht] PSO-based Keyword Optimization [1] Candidate key-word set C, embedding matrix X, particle count N Initialize swarm particles with random subsets of C each iteration each particle i Collect tweets using keyword subset k_i Evaluate objective function $Z(k_i)$ using relevance and volume Update $pbest_i$ if $Z(k_i) < Z(pbest_i)$ Update $gbest$ from all $pbest_i$ each particle i Update velocity v_i using Equation (5) Update position x_i using Equation (4) $gbest$

5. Results and Discussion

1) Reproducibility Enhancements

To improve reproducibility, we now include the random seeds used in all experiments, the exact GloVe and PSO hyperparameters, and pseudocode detailing the optimization pipeline. We also Provide a convergence plot for PSO and lightweight code snippets to facilitate replication of the retrieval and evaluation procedure.

In order to evaluate the performance of our proposed keyword selection framework, we conducted a comparative analysis against a traditional baseline that involved manually curated keywords by domain experts. The experiments were executed using real-world Twitter data related to asthma and difficulty breathing, aligning with our public health surveil-lance use case.

2) Experimental Setup

Two sets of keywords were tested during a seven-day Twitter data collection window. The first set comprised manually selected terms from public health experts, supplemented by terms from the Urban Dictionary. The second set was automatically generated using our GloVe + PSO framework. Tweets were retrieved in parallel using both keyword sets via the Twitter Streaming API from July 1 to July 8, 2019.

To validate keyword effectiveness, we performed random sampling of 2000 tweets from each set. These tweets were manually labeled as relevant or irrelevant to the surveillance context. The evaluation metrics were computed based on this annotated sample.

3) Evaluation Metrics

We adopted both conventional and custom metrics to assess keyword performance:

- *Precision:* Precision measures the fraction of collected tweets that are relevant to the target topic:

$$Precision = \frac{|Relevant\ Tweets| \cap |Collected\ Tweets|}{|Collected\ Tweets|} \quad (6)$$

- *Reach:* Since recall cannot be accurately calculated (the total number of relevant tweets in the stream is unknown), we define a relative metric called *Reach* to measure the overall volume of tweets captured by a keyword set. It is calculated against a simple benchmark keyword $\hat{k}$, such as" asthma":

$$Reach = \frac{|CollectedTweets|_K - |CollectedTweets|_k}{|CollectedTweets|_k} \quad (7)$$

- *Tweet Retrieval Power (TRP):* To balance the trade-off between precision and reach, we define a new metric called Tweet Retrieval Power (TRP), which is the harmonic mean of both:

$$TRP = 2 \times \frac{Precision \times Reach}{Precision + Reach} \quad (8)$$

This metric ensures a fair balance between the quality and quantity of the collected tweets.

4) Statistical Significance and Robustness Testing

To strengthen the validity of our findings, we now include statistical significance testing. Bootstrapped confidence intervals were computed for precision, and paired statistical tests were applied to compare the PSO-based keyword sets with the manual baseline. Additionally, we evaluate the variation of Tweet Retrieval Power (TRP) across multiple days to assess temporal robustness.

5) Comparative Performance

Table 2 summarizes the comparative performance of the two keyword selection approaches.

Table 2: Performance Comparison of Keyword Selection Methods

Method	Precision	Reach	TRP
Manual Baseline	0.23	0.75	0.35
PSO-Based Approach	0.48	0.65	0.55

6) Analysis of Results

The automatic PSO-based approach significantly outperformed the manual baseline in terms of precision, doubling the relevance of the retrieved content. Although the reach of the manual approach was slightly higher, the trade-off was justified, as the PSO approach maintained a competitive volume while greatly improving relevance.

The resulting TRP score of 0.55 for the PSO-based method confirms its superiority in balancing both precision and volume, making it a highly effective method for keyword-based social media monitoring. These findings validate the suitability of the proposed framework for real-time public health surveil-lance and similar applications

7) Additional Baseline Methods

To provide a more comprehensive comparison, we incorporate two additional feature-selection-based keyword generation baselines: Mutual Information (MI) and Chi-square (χ^2) ranking. Both methods score candidate terms according to their discriminative power over a labeled tweet subset. These baselines allow comparison of evolutionary methods against classical statistical feature selection.

The experimental results highlight the practical advantages of using an evolutionary, machine-learning-based approach for automatic keyword selection in Twitter-based surveillance systems. This section explores the implications of these results, the observed trade-offs, and the broader context of their applicability.

1)Trade-off Between Precision and Reach

One of the most prominent observations from the results is the clear trade-off between tweet relevance (precision) and tweet volume (reach). The baseline, expert-defined keywords captured a broader tweet volume due to their generality, yielding a higher reach value. However, this came at the cost of collecting a large number of irrelevant tweets, leading to a low precision score of 0.23.

In contrast, the PSO-based keyword set significantly im-proved precision to 0.48, indicating a much higher proportion of relevant tweets. Although reach decreased slightly (from 0.75 to 0.65), the decrease was not significant enough to outweigh the benefits gained in relevance. This result confirms that intelligent optimization mechanisms can find a more effective balance between breadth and focus than manually curated keyword sets.

2) Effectiveness of Semantic Expansion

The use of GloVe embeddings proved critical in generating a rich candidate keyword set. This semantic expansion accounted for Twitter's informal language, colloquialisms, abbreviations, and alternate spellings—elements that traditional expert-derived lists often miss. By capturing semantic similarity in a continuous vector space, the embeddings allowed the algorithm to include relevant but non-obvious keywords, such as slang terms or phonetically misspelled variations, without requiring manual intervention.

3) Robustness of Particle Swarm Optimization

PSO demonstrated robust performance as an evolutionary optimization algorithm for discrete keyword selection. Despite being originally designed for continuous search spaces, its adaptation using integer encoding effectively explored combinations of keywords to minimize irrelevant tweet collection while maximizing data utility.

Furthermore, the swarm's ability to converge toward high-performing keyword sets within a limited number of iterations affirms its suitability for time-sensitive social media tasks, such as public health monitoring where rapid updates are essential.

4) Comparison to Prior Work

Previous studies on keyword selection primarily focused on static or classifier-based methods [4], [8]. While effective to some extent, these methods often ignore the dynamic and con-textual nature of online discourse. Our approach extends the literature by combining semantic understanding with a real-time adaptive optimization process. Unlike static embedding-based classifiers, our method actively adjusts keyword subsets based on performance feedback gathered during tweet collection.

Additionally, most prior approaches fail to balance both volume and precision simultaneously, whereas our framework introduces a task-specific objective function that enforces this balance through the Tweet Retrieval Power (TRP) metric.

5) Error Analysis and Qualitative Insights

We performed a qualitative error analysis to better under-stand the limitations of the optimized keyword sets. False positives included tweets containing respiratory terms used metaphorically (e.g., "this game is breathtaking"). False negatives often involved misspellings not captured during se-mantic expansion.

We also analyzed keyword contribution through nearest neighbour embeddings and visualized high-impact terms using word clouds. These insights highlight opportunities for future improvement.

6) Generalizability

Although this framework was applied in the context of asthma-related tweet collection, the methodology is highly generalizable. The two-stage design—semantic expansion followed by evolutionary optimization—can be easily adapted for other syndromes, topics, or real-time event tracking. All that is required is a relevant set of seed keywords and an unlabeled tweet corpus from the new domain.

7) Limitations

Despite its promising performance, several limitations should be noted. First, the use of the Twitter Streaming API introduces sampling variability—two keyword sets may not be exposed to exactly the same tweets during live collection, affecting fairness in evaluation. Second, the reliance on manual labeling for relevance scoring introduces subjectivity, even though random sampling was employed. Moreover, the PSO optimization relies on static hyperparameters. While our chosen values (Table I) worked well empirically, adaptive tuning or hybrid

An inherent limitation is the variability introduced by the Twitter Streaming API, where two different keyword sets may not be exposed to identical tweet streams. This can introduce sampling bias in comparative evaluations. To mitigate this, we discuss alternative strategies such as replay datasets, static corpora, and fixed sampling windows for future work.

6. Conclusion

This paper presented a hybrid framework that integrates deep learning with evolutionary optimization for the intelligent and automatic selection of keywords in Twitter-based data collection systems. The approach was specifically designed for public health surveillance tasks, using asthma and respiratory syndromes as a case study. By combining semantically enriched word representations via GloVe embeddings with a Particle Swarm Optimization (PSO) algorithm, our method provides a scalable solution to an otherwise manual and heuristic-driven problem.

We demonstrated that the keyword selection process significantly impacts the quality and quantity of retrieved social media content. Through comprehensive evaluation, the PSO-based approach outperformed the manual baseline by achieving more than double the precision while maintaining a comparable reach. The newly introduced Tweet Retrieval Power (TRP) metric enabled an objective assessment that balanced both relevance and data volume—factors that are often in conflict within social sensing applications.

Moreover, the semantic expansion mechanism successfully captured the informal and evolving nature of language on Twitter, enhancing keyword diversity and robustness. The PSO algorithm effectively explored the high-dimensional candidate keyword space, converging toward optimal keyword subsets within reasonable computational timeframes. These results confirm that combining word embeddings with swarm intelligence can produce high-quality keyword sets automatically, reducing reliance on domain expertise and manual curation.

The framework not only improves data quality but also lends itself to broader applicability. While the implementation focused on respiratory health monitoring, the same methodology can be applied to a wide range of domains—such as disaster detection, sentiment analysis, or misinformation tracking—by simply adjusting the initial seed keyword list and retraining embeddings on the relevant corpus.

In essence, this work contributes to the advancement of adaptive, intelligent data acquisition techniques from social media streams. It introduces an effective and generalizable solution for optimizing keyword-driven monitoring systems and opens new directions for real-time, AI-powered syndromic surveillance using informal data sources like Twitter.

6.1 Future Work

While the proposed framework has demonstrated consider-able effectiveness in automating keyword selection for Twitter-based data collection, several promising directions can be explored to further enhance its capabilities and generalizability.

1) Adaptive and Real-Time Keyword Evolution

One limitation of the current system is its static nature—keywords are optimized once based on past data. In future work, we aim to implement a real-time, adaptive mechanism that continuously refines the keyword set as new tweets arrive. This dynamic adaptation could be achieved using online learning models or reinforcement learning frameworks that adjust the optimization objective based on feedback from tweet retrieval performance over time.

2) Hybrid Evolutionary Models

Although PSO provided strong results, future efforts may investigate alternative or hybrid evolutionary techniques such as Genetic Algorithms (GA), Differential Evolution (DE), or Ant Colony

Optimization (ACO). These methods may offer additional search space diversity and avoid premature convergence issues often seen in standard PSO. A comparative analysis of these optimization algorithms may reveal more effective strategies for high-dimensional keyword space exploration.

3) Context-Aware Embedding Models

Currently, the keyword expansion process relies on GloVe embeddings, which are static and context-independent. Integrating contextualized embeddings such as BERT, RoBERTa, or Twitter-specific transformer models may further enhance se-mantic similarity estimation, especially for polysemous terms or rapidly evolving language trends. This could enable the system to better differentiate between contextual uses of ambiguous words and improve the relevance of the candidate keyword pool.

4) Automated Tweet Relevance Labeling

Manual annotation of tweet relevance remains a bottleneck for large-scale evaluation. Future implementations may lever-age weak supervision, semi-supervised learning, or human-in-the-loop labeling systems to automate the relevance classification process while maintaining acceptable accuracy. This would enable real-time feedback loops for optimization with-out extensive human effort.

5) Cross-Domain and Multilingual Adaptation

Another valuable extension is to evaluate the generalizability of the framework across domains and languages. For instance, applying the method to domains such as climate monitoring, political sentiment, or financial forecasting can assess its robustness. Additionally, incorporating multilingual embeddings and language detection tools can facilitate deployment in diverse linguistic and geographical contexts, particularly in global health surveillance scenarios.

6) Integration with Visualization Dashboards

To increase usability and adoption in operational settings, the keyword selection framework can be integrated into a real-time data dashboard for public health or event monitoring. This would allow non-technical users to visualize keyword impact, tweet volume trends, and topic shifts over time, enabling faster decision-making and policy response.

7) Scalability and Cloud-Based Deployment

Finally, deploying the entire system in a scalable, cloud-based architecture (e.g., using Docker, Kubernetes, or server-less functions) would allow it to handle large-scale social media streams efficiently. Integration with big data platforms such as Apache Kafka and Elasticsearch could facilitate continuous ingestion, storage, and querying of tweet data, supporting long-term monitoring pipelines.

References

S. Triple, “Assessment of syndromic surveillance in europe,” *Lancet*, vol. 378, no. 9806, p. 1833, 2011.

E. De Quincey and P. Kostkova, “Early warning and outbreak detection using social networking websites: The potential of twitter,” *International Conference on Electronic Healthcare*, pp. 21–24, 2009.

F. Morstatter, J. Pfeffer, H. Liu, and K. Carley, “Is the sample good enough? comparing data from twitter’s streaming api with twitter’s firehose,” *ICWSM*, pp. 400–408, 2013.

A. Liu, M. Srikanth, N. Adams-Cohen *et al.*, “Finding social media trolls: Dynamic keyword selection methods for rapidly-evolving online debates,” *arXiv preprint arXiv:1911.05332*, 2019.

P. Kostkova, M. Szomszor, and C. St. Louis, “Syndromic surveillance for public health using web 2.0 data,” *IEEE Internet Computing*, vol. 14, no. 6, pp. 10–16, 2010.

N. Collier and A. Kawazoe, “Towards open access public health litera-ture based on biomedical ontologies,” *Journal of Biomedical Informatics*, vol. 41, no. 4, pp. 706–721, 2008.

O. Edo-Osagie, B. De La Iglesia, I. Lake, and O. Edeghere, “Intelligent keyword selection for twitter data collection in public health surveil-lance,” *Proceedings of the International Conference on Healthcare Informatics*, 2019.

S. Kiritchenko and M. Jiline, “Keyword optimization in sponsored search via feature selection,” *Feature Selection in Data Mining*, pp. 122–134, 2008.

G. Forman, “An extensive empirical study of feature selection metrics for text classification,” *Journal of Machine Learning Research*, vol. 3, pp. 1289–1305, 2003.

Y. Yang and J. O. Pedersen, “A comparative study on feature selection in text categorization,” *ICML*, pp. 412–420, 1997.

T. Mikolov *et al.*, “Efficient estimation of word representations in vector space,” *arXiv preprint arXiv:1301.3781*, 2013.

J. Pennington, R. Socher, and C. Manning, “Glove: Global vectors for word representation,” *EMNLP*, pp. 1532–1543, 2014.

P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword

information," *Transactions of the ACL*, vol. 5, pp. 135–146, 2017.

D. Lee and K. Kim, "Web site keyword selection method by considering semantic similarity based on word2vec," *Journal of Society for e-Business Studies*, vol. 23, no. 2, 2019.

Q. Li and Z. Xu, "Embedding-based keyword extraction for online advertisement targeting," *ACM Transactions on the Web*, vol. 14, no. 1, pp. 1–19, 2020.

L. Zhang, S. Wang, and B. Liu, "Deep learning-based sentiment analysis using twitter data," *IEEE Transactions on Affective Computing*, vol. 9, no. 3, pp. 362–373, 2018.

K. Shu, D. Mahudeswaran, S. Wang, and H. Liu, "Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for fake news research," *Big Data*, vol. 8, no. 3, pp. 171–188, 2020.

X. Chen *et al.*, "Flu detection from tweets using temporal topic models," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 3, pp. 564–573, 2018.

M. Paul and M. Dredze, "You are what you tweet: Analyzing twitter for public health," *ICWSM*, 2011.

L. Charles-Smith, T. Reynolds, and M. Cameron, "Public health and social media: Using twitter for disease surveillance," *American Journal of Public Health*, vol. 109, no. 9, pp. 1307–1314, 2019.

H. Achrekar, A. Gandhe, R. Lazarus, S. Yu, and B. Liu, "Predicting flu trends using twitter data," *Computer Communications Workshops*, pp. 702–707, 2011.

M. Araujo, Y. Mejova, and I. Weber, "Using digital traces for public health surveillance: An overview," *IEEE Pervasive Computing*, vol. 16, no. 4, pp. 64–72, 2017.

J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proceedings of ICNN*, vol. 4, pp. 1942–1948, 1995.

P. Umapathy, C. Venkataseshaiah, and M. Arumugam, "Particle swarm optimization with various inertia weight variants for optimal power flow solution," *Discrete Dynamics in Nature and Society*, pp. 1–15, 2010.

K. Deb and N. Padhye, "Improving a particle swarm optimization algorithm using an evolutionary algorithm framework," *KanGAL Report*, vol. 3, 2010.

J. Liang, H. Yang, J. Gao *et al.*, "Mopso-based cnn for keyword selection on google ads," *IEEE Access*, vol. 7, pp. 125 387–125 400, 2019.

M. Abido, "Optimal power flow using particle swarm optimization," *International Journal of Electrical Power & Energy Systems*, vol. 24, no. 7, pp. 563–571, 2002.

Z. Ma, X. Jin, and Y. Cheng, "A pso-based approach for sentiment classification," *Applied Intelligence*, vol. 47, no. 4, pp. 1061–1073, 2017.

W. Zhang and H. Liu, "An evolutionary approach for web crawler optimization using semantic features," *Information Sciences*, vol. 301,
pp. 123–135, 2015.

R. Gupta, P. Kumaraguru, and S. Saha, "Social media analytics for public health: A hybrid approach," *ACM Transactions on Management Information Systems*, vol. 9, no. 4, pp. 1–27, 2018.

B. Kang, S. Jo, and J. Park, "Evolutionary keyword optimization in social data streams," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, pp. 607–617, 2017.

O. Edo-Osagie, B. De La Iglesia, I. Lake, and O. Edeghere, "Twitter syndromic surveillance: Evolutionary optimization for keyword detec-tion," *Health Informatics Journal*, vol. 26, no. 2, pp. 1082–1099, 2020.

Edo-Osagie, O., Iglesia, B. D. L., Lake, I., & Edeghere, O. (2020, November). An evolutionary approach to automatic keyword selection for twitter data analysis. In *International Conference on Hybrid Artificial Intelligence Systems* (pp. 160-171). Cham: Springer International Publishing.