International Journal for Modern Trends in Science and Technology, 8(12): 24-28, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0812005
Available online at: http://www.ijmtst.com/vol8issue12.html

Local Conceptual Word Character level in Microblog Extraction of features Using Query Expansion (QE)

B.Veerendra[1] | Pulige Swamy Manideep[2]

[1]Associate Professor, Dept of CSE, Kakinada Institute of Technology and Science, Divili, A.P
[2]PG Student, Dept of CSE, Kakinada Institute of Technology and Science, Divili, A.P

To Cite this Article
B.Veerendra and Pulige Swamy Manideep. Local Conceptual Word Character level in Microblog Extraction of features Using Query Expansion (QE). International Journal for Modern Trends in Science and Technology 2022, 8(12), pp. 24-28. https://doi.org/10.46501/IJMTST0812005

Article Info
Received: 16 November 2022; Accepted: 27 November 2022; Published: 04 December 2022.

ABSTRACT

Traditional methods of information retrieval are severely hindered by the word mismatch problem when applied to microblog posts like tweets due to their fixed character limit of 140 characters. In order to enhance microblog retrieval performance, the author here employs local conceptual word embeddings. In order to better understand the information need, a novel technique called k-nearest Neighbor (QE) Query Expansion generates words by means of local word embeddings, thereby expanding the original query. It is possible that the most recent tweets related to a given topic will appear higher in search results if temporal evidence is included in the enlargement algorithm. Our method is shown to be significantly more effective than the gold standard approaches on the official TREC Twitter corpora.

***KEYWORDS:** Information Retrieval, Query Expansion, Word Embeddings, TREC.*

1. INTRODUCTION

As the popularity of the social media platform rises, Twitter-based information search has garnered the interest of academics. By typing "RTST" into Twitter's search bar, you may have the service provide the top-trending tweets on a certain subject at a given time (T). No one needs the added difficulty of real-time infrared when microblogging (IR). Linguistic discrepancy occurs often in online search tasks that need instantaneous results [2, 3, 4]. Because of their limited word count, tweets might be challenging to filter through in order to locate those that are relevant to your search (140 characters). This gets more challenging when trying to locate organisations that use a variety of aliases. In addition, a real-time search often exposes a demand for knowledge on a present incident. The most current tweets regarding the issue, therefore, must have priority in IR planning. So search engines need to weigh how recent a tweet is against how relevant it is in order to provide the best results in real time. The difficulties outlined above may be overcome in a number of ways; for example, by using [5], [6], or [7]. Search engine optimization (SEO) is the process of improving search engine results by making adjustments to individual queries to better match user intent (SEO). They believe the highest-ranked texts are the most relevant and best for expanding queries with related terms. The right language [9], [10], and [11] in the question is an example of this issue. Due to the casual character of the medium, even highly scored tweets may include many irrelevant

terms. By using the k-Nearest Neighbor method, we suggested a new approach to query expansion (QE) that places emphasis on semantic similarity rather than proximity (kNN). It's possible that users' information needs may be better expressed if the original query was expanded using expansion words derived from local word embeddings. Word embeddings are learned not at the level of individual words but at the level of concepts, and they are learned not on huge, thematically unrestricted corpora but on smaller, topic-specific ones. Both string (using tools like the Longest Common Subsequence [12]) and context (using tools like WordNet) similarity is analysed (e.g., word embedding [13], [14]). Contextual similarity attempts to record the context in which a term appears in the corpus. However, context is crucial in microblogs due to the prevalence of non-standard word representations. The cosine similarity between the term-vectors of the two phrases may therefore be used to evaluate the degree to which they are semantically related. All the words in a given lexicon may be represented in a two-dimensional space using continuous word embeddings (CWE). A reasonable assumption is that the distance between two words, a and b, is a quantifiable measure of their semantic similarity. Recently, scholars in the field of natural language processing (NLP) have shown an interest in these methods, but their potential use in international relations (IR) has received less attention. Since the purpose of word embeddings is to discover words that are semantically relevant to a user's query, they should be able to enhance QE. For the simple reason that it is intuitive to estimate semantic similarity in semantic space. Word embeddings that are trained on the whole corpus may provide only naive representations of important concepts. This strategy proposes that local (topic-specific) representations of language should be superior than global (task-general) representations for many applications. The query-related texts are used to retrain the word embedding learnt from the topic-specific corpus. Produce a novel technique for query expansion that can identify phrases that are similar to the query by using local word embeddings. Use the temporal evidence available in the microblogosphere by including a distribution on document recency into the aforementioned design, as suggested in [18]. Applying sigmoid mapping is a great way to increase the discriminatory power of the similarity measurement. A new k-nearest neighbours (KNN)-based Query Expansion (QE) method with local word embeddings is proposed, which better captures the nuances of topic-specific language than global embeddings; (ii) the temporal evidence is incorporated into QE method to trade-off between relevance and recency; and (iii) the well-known sigmoid funct is proposed for use in QE method. Experimental findings show that the suggested method improves retrieval performance, and that for microblog retrieval tasks, global word embeddings are superior than topic-specific embeddings.

2. PROPOSED SYSTEM

We introduce a novel kNN-based Query Expansion (QE) method that uses local word embeddings, temporal evidence, and the well-known sigmoid function in place of global embeddings to better comprehend the information requirements of our customers. Finally, we discuss potential developments in this area. Experiments using the proposed method reveal that word embeddings, when trained globally, perform poorly in comparison to topic-specific embeddings when it comes to microblog retrieval tasks.

3. MODULE DESCRIPTION

Admin

In this module, the administrator needs a working login name and password to access the system. After he logs in, he has access to a number of features, including Login, Make a User/Document Ranking/Ratio Chart and View, Add, or List Everyone's Positions.

Viewing and Authorizing Users

In this module, in order to verify a user's identity, the administrator will review their profile. The full e-mail address and phone number of the user must be entered.

User

In this module, Right now, there are n concurrent users online. The user has to log in before they can do anything. After a user registers, their details are added to a database. After signing up, you'll be asked to enter your username and password to access the site. ' If a user's login attempt is successful, they will be able to register and log in. You may try inputting a keyword or title to see what comes up. Use either the Top k List or a Specific Domain to Find What You Need.

Viewing Profile Details

In this module, Information such as the user's name, address, email, and phone number is displayed in the user profile.

Search Friends, Request, and View Friend Requests, View all Friend Details

In this, simply type a user's username into the search bar to find them. Every user's pal can have a profile with a photo and a brief biography viewed right on their device.

Search Query by keyword

In this, an individual's post search results will be split into two columns when employing a query. There are two kinds of entirely matched postings and two kinds of perfectly matched ones. Users may "like" or "dislike" a post to indicate their opinion on it and so decide whether to share it with their friends.

4. ALGORITHM

Algorithm for Fuzzy Query Expansion:

//Considering the initial query entered by user having terms I where 1≤i≤n .and value of maximum distance"L" considered as L=3

Step 1: Query is entered by the user having the terms 1≤i≤n.

Step 2: Use Tag POS to identify adjectives, adverbs, verbs and nouns by removing stop words. The original query only includes these POS-tagged terms.

Step 3: Fuzzy WordNet Graph G is formed by exploring the two semantic relations between words in WordNet.

Step 4: Create a sub graph from WordNet Graph G by setting the SubGraph G to NULL and then creating the sub graph

Step 5: Using word embedding models on the nodes, give each edge in G a weight in G.

Step 6: Measuring Graph Centrality Calculate the Fuzzy Centrality Measures for the graph for each node in Subgraph G as follows. fuzziness Soft Page Rank It's a haz Uncertainty Fuzzy HIT

Step 7: Each fuzzy centrality metric is sorted, and the top 10 nodes are taken from each metric.

Step 8: Nodes V should not be the same as any Query Terms that were used in the original query, therefore choose them. More than one metric of centrality places Node V as one of the top ten most important nodes.

Step 9: Add the above-selected keywords to the Original Query for the expansion of the search.

5. PROPOSED SYSTEM ARCHITECTURE

The administrator need a login and password to access the system. After signing in, the user has access to the ability to add/view/list all people, as well as rating documents and ratio charts. A system administrator may preview a user's information before they've given their consent. Don't forget to provide the whole email address and phone number. n individuals are online right now, which means Getting logged in is a must for doing anything. After enrolling, users' data is saved in a database After registering, you'll need to log in using your username and password. ' Then they may register and log in. Look for a certain phrase or book. Find what you're looking for in a certain domain or in the top K. User information such as name, address, email, and mobile phone number are viewable. To find someone, type their username into the search bar. The user's friend has a profile and photo. A user may publish a search by entering a query term and seeing the results in two columns. It's not impossible to find a perfect correspondence between two advertisements. Through the use of likes and dislikes, users are able to recommend content to their social networks.

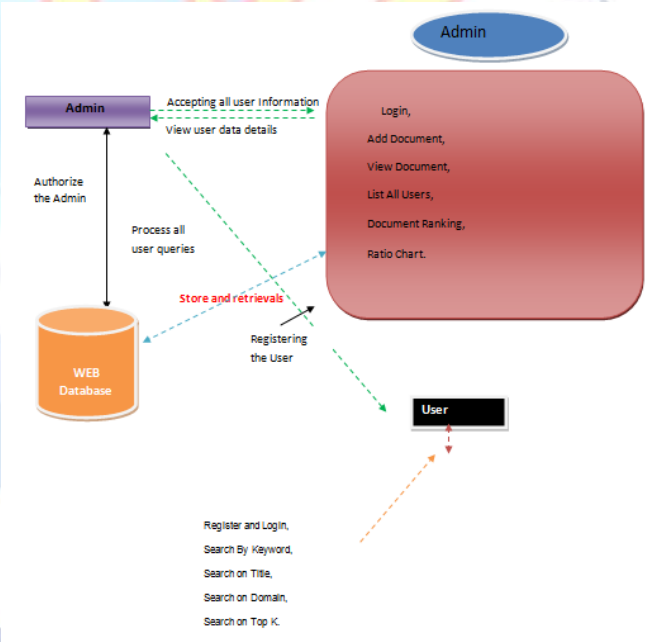

Fig1: Proposed System Architecture

6. EXPERIMENTAL RESULTS

The system requires a user ID and password to enter. Logging in grants access to the whole database of users, documents, and ratios for viewing, editing, and listing. To examine user data, an administrator must first provide access. The user's email address and phone

number must be entered in full. There are now n users logged in, but you cannot join them until you do so. When a user registers, their information is added to a database. Creating an account is required to use this website. ' They may sign up and enter after that. Locate a suitable phrase. Whether you're looking for a domain or the top k sites, Identifying information such as username, email address, and phone number are all publicly displayed. In such case, you might try searching for them by their username. Pictures and biographies of all your pals in one place. A user may share a query-based search that produces results in two columns. Two postings may be exactly the same. Posts are shared with friends using likes and dislikes.

Fig2: Main Page

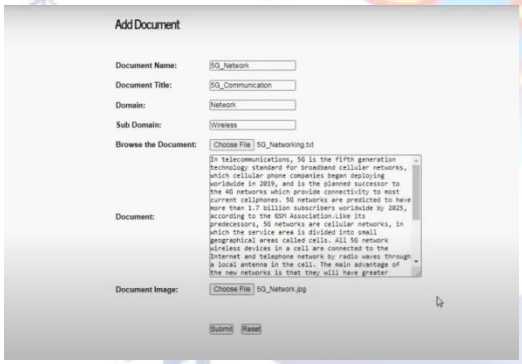

Fig3: Add Documents

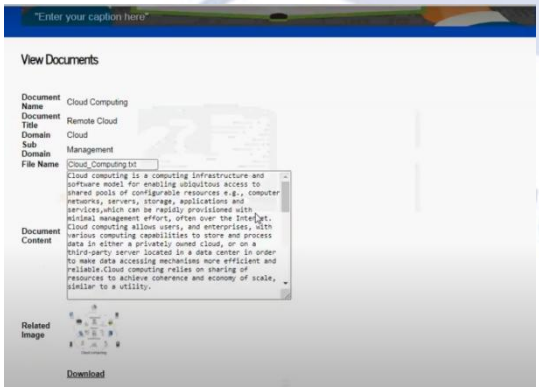

Fig4: View Documents

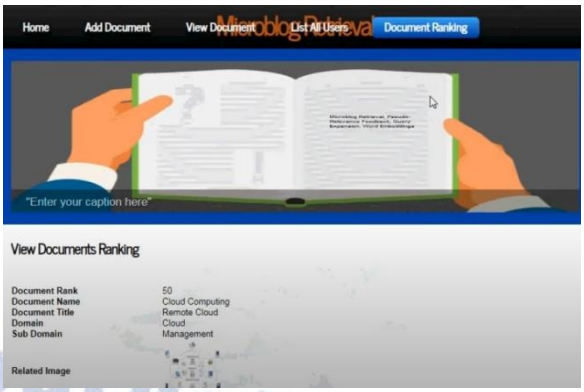

Fig5: View Documents Ranking

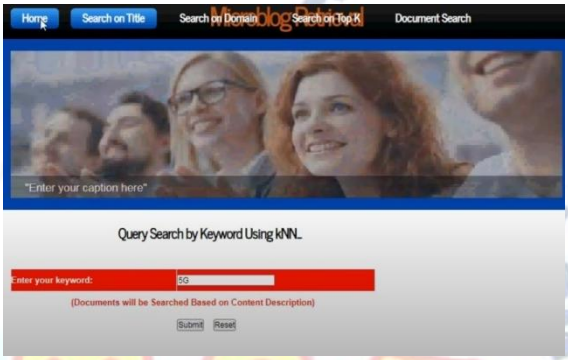

Fig6: Quick Serch Keyword Using KNN

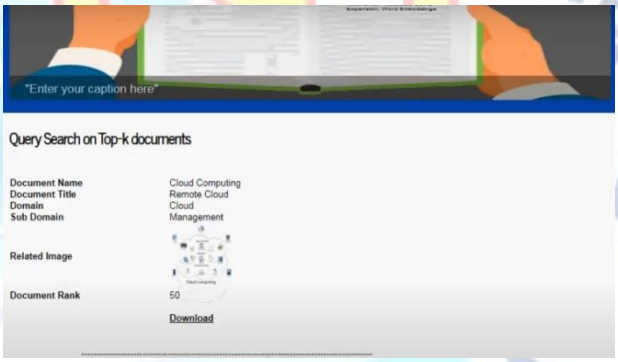

Fig7:Query Search on Top-k documents

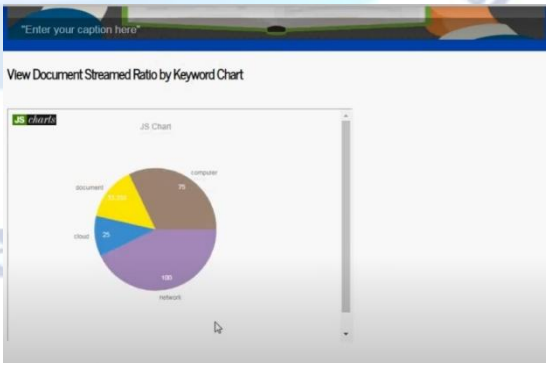

Fig8: View Document Keyword Chart

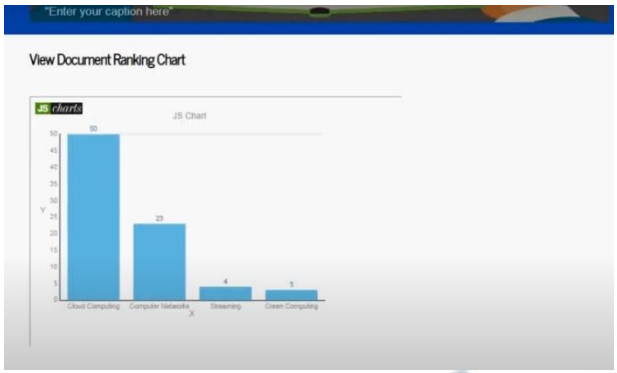

Fig9: View Documents Ranking Chart

7. CONCLUSION

Several different IoT use cases may be accommodated by the Fog Computing idea (IoT). Have Created a Resource Aware Scheduler for Deploying Application Modules to Fog Clouds (RACE). The incoming applications and their accompanying modules are what ultimately define how much work an application will need. In order to test the efficacy of the RACE algorithm, this method used three distinct ways for deploying workloads: deploying on the Cloud, deploying in the Fog layer, and deploying from the Fog layer to the cloud layer. The RACE(CFP) optimises execution time, bandwidth use, and overall cost for all workloads. Increasing the number of fog devices in use may improve RACE (FOP) performance in certain circumstances. When compared to RACE, the results of Race (CFP) are better in terms of execution time, bandwidth utilisation, and cost (CFP).

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1]] J. Manyika et al., "McKinsey Global Institute: Unlocking the potential of theInternet of Things," McKinsey Glob. Institute., pp. 1–144, 2015.

[2] F. S. Collins,E. D. Green, A. E. Guttmacher, and M. S. Guyer, "A vision for the future of genomics research," Nature, vol. 422, no. 6934, pp. 835–847, 2003, doi:10.1038/nature01626.

[3] S. (2015). Smith, "'Internet of Things' Connected Devices to Almost Triple to Over 38 Billion Units by 2020 - Juniper," http://www.juniperresearch.com/press/press-releases/iot-connected-devices-to-triple-to-38-bn-by-2020, 2016.

[4] B. P. Rimal, E. Choi, and I. Lumb, "A taxonomy and survey of Cloud Computing systems," in INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on, 2009, pp. 44–51.

[5] M. Nabi, S., Ahmed, "OG-RADL: Overall Performance Based Resource Aware Dynamic Load-balancer for Deadline Constrained Cloud Tasks," J. Supercomput., 2020.

[6] M. Ibrahim et al., "An In-Depth Empirical Investigation of State-of-the-Art Scheduling Approaches for Cloud Computing," IEEE Access, vol. 8, pp. 128282–128294, 2020, doi: 10.1109/ACCESS.2020.3007201.

[7] A. Hussain, M. Aleem, A. Khan, M. A. Iqbal, and M. A. Islam, "RALBA: a computation-aware load balancing scheduler for Cloud Computing," Cluster Comput., vol. 21, no. 3, pp. 1667–1680, 2018, doi: 10.1007/s10586-018-2414-6.

[8] S. Sarkar and S. Misra, "Theoretical modelling of Fog Computing: A green computing paradigm to support IoT applications," IET Networks, vol. 5, no. 2, pp. 23–29, 2016, doi: 10.1049/iet-net.2015.0034.

[9] Sandvine, "Sandvine Global Internet Phenomena Report 2H-2015," Sandvine Corp., p. 17, 2015, doi: 10.1007/s13398-014-0173-7.2.

[10] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog Computing and its role in the internet of things," Proc. first Ed. MCC Work. Mob. cloud Comput. - MCC '12, no. August 2012, p. 13, 2012, doi: 10.1145/2342509.2342513.

[11] A. V. Dastjerdi, H. Gupta, R. N. Calheiros, S. K. Ghosh, and R. Buyya, "Chapter 4 – Fog Computing: principles, architectures, and applications," Internet of Things, pp. 61–75, 2016, doi: 10.1016/B978-0-12-805395-9.00004-6.

[12] R. Lavanya, "Fog Computing and Its Role in the Internet of Things," Adv. Consum. Fog Comput. Archit., p. 63, 2018.

[13] J. Cao, Q. Zhang, and W. Shi, "Challenges and Opportunities in Edge Computing," in Edge Computing: A Primer, Springer, 2018, pp. 59–70.

[14] S. Sarkar, S. Chatterjee, and S. Misra, "Assessment of the Suitability of Fog Computing in the Context of Internet of Things," IEEE Trans. Cloud Comput., vol. 6, no. 1, pp. 46–59, 2018, doi: 10.1109/TCC.2015.2485206.

[15] A. T. Thien and R. Colomo-Palacios, "A Systematic Literature Review of Fog Computing," Nokobit, vol. 24, no. 1, pp. 28–30, 2016.

[16] P. Hu, S. Dhelim, H. Ning, and T. Qiu, "Survey on Fog Computing: architecture, key technologies, applications and open issues," J. Netw. Comput. Appl., vol. 98, pp. 27–42, 2017, doi: 10.1016/j.jnca.2017.09.002.

[17] M. Hajibaba and S. Gorgin, "A review on modern distributed computing paradigms: Cloud Computing, jungle computing and Fog Computing," J. Comput. Inf. Technol., vol. 22, no. 2, pp. 69–84, 2014, doi: 10.2498/cit.1002381.

[18] S. Yi, Z. Hao, Z. Qin, and Q. Li, "Fog Computing: Platform and applications," in 2015 Third IEEE Workshop on Hot Topics in Web Systems and Technologies (HotWeb), 2015, pp. 73–78.

[19] L. F. Bittencourt, J. Diaz-Montes, R. Buyya, O. F. Rana, and M. Parashar, "Mobility-Aware Application Scheduling in Fog Computing," IEEE Cloud Comput., vol. 4, no. 2, pp. 26–35, 2017, doi: 10.1109/MCC.2017.27.

[20] M. Taneja and A. Davy, "Resource aware placement of IoT application modules in Fog-Cloud Computing Paradigm," in Integrated Network and Service Management (IM), 2017 IFIP/IEEE Symposium on, 2017, pp. 1222–1228.

[21] X. Q. Pham and E. N. Huh, "Towards task scheduling in a cloud-Fog Computing system," 18th Asia-Pacific Netw. Oper. Manag. Symp. APNOMS 2016 Manag. Softwarized Infrastruct. - Proc., 2016, doi: 10.1109/APNOMS.2016.7737240.

[22] V. Cardellini, V. Grassi, F. Lo Presti, and M. Nardelli, "On QoS-aware scheduling of data stream applications over Fog Computing infrastructures," in Computers and Communication (ISCC), 2015 IEEE Symposium on, 2015, pp. 271–276.

[23] J. Fan, X. Wei, T. Wang, T. Lan, and S. Subramaniam, "Deadline-aware task scheduling in a tiered IoT infrastructure," in GLOBECOM 2017-2017 IEEE Global Communications Conference, 2017, pp. 1–7.

[24] J. Xu, B. Palanisamy, H. Ludwig, and Q. Wang, "Zenith: Utility-Aware Resource Allocation for Edge Computing," Proc. - 2017 IEEE 1st Int. Conf. Edge Comput. EDGE 2017, pp. 47–54, 2017, doi: 10.1109/IEEE.EDGE.2017.15.

[25] L. Ni, J. Zhang, C. Jiang, C. Yan, and K. Yu, "Resource allocation strategy in Fog Computing based on priced timed petri nets," IEEE Internet Things J., vol. 4, no. 5, pp. 1216–1228, 2017.