User Aware Edge Caching in 5G Wireless Networks

Asif Kabir†, M.Shahid Iqbal‡‡, Zain ul Abidin Jaffri†, Shoujat Ali Rathore††
Edvin J. Kitindi††† and Gohar Rehman‡‡‡
† College of Communication Engineering
‡‡ ‡‡ College of Computer Science, Chongqing University, Chongqing, 400044, P. R China
‡‡ ‡ School of Computer Science, Anhui University, Hefei P.R China
†††SM-College of Science and Education, Sokoine University of Agriculture, P.O Box 3038, Morogoro Tanzania

Summary
Wireless technology has become an ultimate weapon in today’s world. Caching has emerged as a vital tool in modern communication systems for reducing peak data rates by allowing popular files to be pre-fetched and then stored at the edge of the network. Caching at small cell base stations has recently been proposed to avoid bottlenecks in the limited capacity backhaul connection link to the core network. For predicting the popularity of the content, we need to analyze the behavior of the user, understanding collectively the behavior beneficial for content trend forecasting and improve network performance. The proposed model predicts the intensity of human emotions through social media (Twitter) and the classifier evaluates the features which are related to user behaviors and, finally, values of features are pushed to the user profile. We further demonstrate how emotions extracted from Twitter can be utilized to improve the forecasting, describing things in a new way which can further be exploited as an optimization basis for network performance enhancement.

Keywords:
Edge caching, emotions, popularity prediction, social networks, user aware contents and 5G

1. Introduction
The next generation networks are expected to support an increasing number of connected devices and diverse applications. This requires wireless communication systems to move towards a real-time information and user-based network. This phenomenon is further fueled by mobile video streaming that currently accounts for almost 50 percent of the mobile data traffic. It is predicted that it will increase 500 times during the next decade [1]. Social network at the same time is considered to be the biggest traffic contributor [2]. This new trend urges the mobile operators to boost their existing network, redesign and seek the most advanced and sophisticated techniques to expand coverage, network capacity and cost-effective content consumption close to the user. This traffic explosion impulses the service provider to enhance the network and brings content closer to the user. Various new technologies are studied to accommodate these challenges. To meet the emerging multimedia stringent quality of services (QoS) in the future fifth generation (5G) cellular network leads to introduction of a new decentralized architecture, based on the concepts of SBSs, like Femtocells, Picocells, or Microcells. SBSs are dedicated to deploying low operating costs with high QoS [3]. Deploying small cells can help to satisfy high-speed requirements and overcome the bottlenecks of limited backhaul. In a fifth generation (5G), SBSs will perform a fundamental role for coverage and capacity enhancements. Real-time data services and smart mobile application propagation tremendously will exert huge pressure on mobile networks. But on the other hand, the deployment of dense SBSs may put some extra burden on the backhaul links.

To reduce this extra burden, edge caching (EC) is one of the promising technologies. Local caching at the edge of the network can reduce network load especially backhaul traffic by using most frequently requested contents at local caches. It is an approach to strike a balance between data storage and data transfer. It has been introduced to exploit the storage capabilities of SBSs and user terminals [4]. To meet excessive traffic demand with limited network resources while keeping low price and providing good quality of services (QoS), edge content caching is a key technology. In fact, content caching is now considered to be basic network functionality in emerging network architectures such as Content-Centric Networking [2]. Moreover, cache reduces duplicate content transmission. Additionally, it improves the energy efficiency (EE) and spectrum efficiency (SE).

The future popularity of content is not readily available at the caching decision time yet it needs to be forecasted. User device increasingly permeates the fabric of human lives. In daily routines the utilization of smart devices and content on demand increases. Content on demands leads to delay at the same time and also challenge the backhauling efficiency.
One of the promising ways to satisfy the user demand and tackling backhaul bottleneck is caching the content at the edge of networks. Caching the contents at the edge of the networks reduces the traffic burden on backhaul and the latency. There is a limited cache capacity at the edge nodes, so accurate content popularity prediction has a significant role in the efficient use of cache capacity. To exploit the full potential of the wireless edge cache, popularity prediction and user preferences are critical [5].
The online media ecosystem content popularity may vary across user preference, personal characteristics, device and external factors prediction [6]. There are a lot of predictive features which depends on variables used in the model and also on the performance of content creators (see detail in [7]). In wireless networks or SBSs, proactive content caching is assumed to be an appropriate method to increase the proficiency of network. Due to storage constraints in the local cache, the optimal content placement is crucial and also requires prior knowledge about the future popularity of content. Furthermore, the popularity of local content is conditional to fluctuations since the mobile operators with diverse interests join caching entity over time. In terms of wireless, content caching is projected to increase the efficiency of the network. It does not only reduce the cost of the storage but also, characteristically, only a small amount of popular contents explains majority of the data traffic.

Different users have different preferences, content popularity may vary across the users. Users preferences might be linked to various factors. We refer to them as context dimensions like personal characteristics which include age, gender, mood, personality, etc. [8], [9] or device characteristics and their location [10].

Human emotions mostly essential for our life can be contrastive or destructive, with the popularity of social media and smartphone devices, the problem shifted into the network word. A cloud base mobile user anger predict on social network is discussed in [11]. Emotion communication system based on non-line of sight model where emotions are considered as a kind of multimedia content transmitted over a long distance can be examined in [12].

The proposed solution predicts the intensity of human emotions through social media (Twitter). Our proposed approach can provide a better result in prediction, describe things in a new way and can be utilized as an optimization basis for network performance enhancement. Therefore, during caching, the popular content placement should be aware of context and consider the user’s preference. Content cache placement depends on specific objective and preference of users. We propose a model in which the content popularity is user’s context dependent, like personal physiognomies (behaviors) and also the proposed algorithm learn the personal behavior by social media. It is a user context-aware algorithm which will analyze emotion of user on social media using Twitter tweets.

The paper is organized as follows. The overview and related works are given in Section II, cache policies are highlighted in Section III. Section IV describes the popularity prediction by social networks. In Section V, we propose a prediction algorithm. Section VI gives results and discussion and section VII concludes the research work.

2. Related works

Online social network from several perspectives that describe the user behaviors has been studied in [13], [14]. References [11] shown that social networks can effectively predict the popularity of articles on news sites. In [12], the authors predicted movies rating on IMDb using Twitter tweets and comments on YouTube. The popularity of content is important to design more accurate prediction models to identify the factors and understand important variables that will impact and should be used to design a model. The popular content is still under discussion and some social impacts contribute significantly to the popularity of the content and the attention they received. So, emotion is one of the key elements of the online audience. The social videos are more likely to be the most powerful and positive or negative tool in the online community.

The content that generates high arousal emotions will spread faster on the internet and capture more user interest [15], which then gets the attention of additional catalyst for social sharing of users. The social network, blog or e-mail service through the video playback process in the short period has high diffusion. Similarly, social connections play a major role in enhancing the popularity of web content.

3. Caching policy

Caching contents at the edge of networks can reduce the backhaul links burden and latency. Cache in wireless networks while achieving the potential gain with an appropriate caching policy is critical nowadays. A caching policy can be either reactive or proactive. In the proactive caching policy, making cache decision before the cache request is based on the prediction. The reactive caching
policy makes a decision after it has been requested or according to replacement algorithms [16]. Prediction of the user’s context and giving a sense of user nature is a novel network paradigm. It is the fact that a huge number of user information is often available and also some human behavior can be predictable. Thus, the proactive caching mechanism is based on the idea to ensure popular content at the edge. In the machine learning technique, logs can be analyzed. The recent machine learning techniques predictive big data have received significant attention. Social networks carried out massive data traffic, which plays a vital role to shape the traffic. Therefore, social networking features and user’s behaviors can help better future network planning. By exploiting the data and user’s relations, the accuracy of predicting future events can be improved dramatically. Social media allows the users to create distributed contents and the explosive growth of information intensified online competition for user’s attention.

Using big data analytics, the demographic pattern of content requests is aggregated and then as per user basis, it can be predicted. Furthermore, prediction can perform an imperative role in caching design. To satisfy QoS and offload the backhaul burden, cache at the edge of the network, pre-store intelligently the strategic contents. So, considering the limited size of the cache and the dissimilarities in user behavior, the caching should be designed according to the behavior of users and location [6]. In many research areas, like network dimensioning, online marketing and real-world outcome predictions, the predicting popularity of content is very useful. Similarly, accurate content forecasting popularity distribution is significant in edge radio access network (RAN) to the effectual usage of cache capacity.

4. Popularity prediction by Social networks

Prediction of popular content is a major issue of forecasting, in order to maximize the rewards and accuracy, the forecast is performed on the basis of prior knowledge which does not have any training phase online. The forecasting content can be categorized as per contextual catalyst of the user on the social media. Content popularity sudden changes can be seen in social media, which evolving propagation patterns of content. The requested probability of a particular content is modeled normally by Zipf distribution and the most popular contents are cached in the network. Social media has been recently providing situational awareness and application requirements for events broadcast to inform predictors and decision-makers in various fields of security and monitoring [17], disaster management and economic forecasting. Twitter posts can be used to scale and forecast the revenue of box office movies [18]. Sentiment detection is investigated in [19], by exploring characteristics of the ways tweets are written and meta information of the words that compose these messages. In all of these applications, social networking is important for variety of reasons to predict the popularity of shared content across the network for service providers and mobile network operators(MNO). Basically, the prediction is dependent on the history of reliance. In contrast, social media users are proactive regarding the content they watch.
and are heavily influenced by their social media interactions. For instance, a particular content may be forwarded or not which does not only depend on its attractiveness or the contrary relevant conditions but also the contextual conditions propagated through social media. To improve the accuracy of prediction, popular methods such as machine learning, logarithmic linear correlation model, and evolutionary prediction were proposed earlier. By exploiting the predictive capabilities of 5G with leveraging the social networks features minimizes the peak cellular networks load. Predictive networking in SBSs will help to adequate storage capabilities at the edge of the networks [20]. Existing content prediction models retrieve form the big data, in SBSs population is very small, hence, it suffers poor accuracy and also it depends on social relationship [10].

5. System Model and Architecture

5.1 Local Cache

The idea of using edge caching to support mobile users in a cellular wireless network has been established in [21]. In RAN, local caching at the edge of the network has emerged as a promising technique for enhancing the QoS of user equipment’s (UEs). Cache capacity of local BS is a new type of resource besides time, frequency, and space. Moreover, cache reduces duplicate content transmission, additionally; edge cache (EC) improves the energy efficiency (EE) and spectrum efficiency (SE). Contents are shared and produced by users via social networks. Fig.1 enlightens the local cache and online user behavior. The caching is characterized by a consistent backhaul link to a core network, which has limited storage capacity. In the cache memory, the cache can store limited files. The users are positioned in coverage area so that they can easily join caching entity. Caching entity can be predicted periodically on the bases of users which are active on the social network. The statistics related to current cache content in which the user is interested can be found in caching entity store called local cache. Backhaul connection is used by the Macro Base Station (MBS) to download files from the core network although, in this situation, the load is on the macro cellular along with backhaul network.

To reduce the burden on backhaul networks, caching entity may wish to optimize cache content such that the traffic can be served and it corresponds to maximize a total number of cache hits. So, for this reason, the caching entity should be aware that which file gain supreme popularity over time.

5.2 Behavior Prediction Model and Algorithm

This paper describes the human behavior on the basis of social media. Data from Twitter as a data sample set was used. The user gets the access of local database and then system synchronizes the local database to the social media database. Systems retrieve necessary information from the social media database, save it and finally fetch this information from the local database. Furthermore, system functions in a local working memory and carry out analysis on it as saved data initially is tokenized. After that tagging is performed, system filters the required information compares this new data with existing data and concludes an emotion.

We consider a single user interaction on social media. In system modules, a typical workflow is presented. The user requests for the content and system respond to that request. Firstly, system fetches tweets from the Twitter database between dates which are specified by the user, secondly, save the data in local database and finally filter the required information as shown in Figure 2. The key building blocks include cache management entity, local cache, cache decision, storage interface and user interface. The cache management comprises of cache controller. The decision module contains a decision engine. In the proposed system, workflow involves several steps as represented. The first step is to initialize learning module. In this process of caching, it first monitors content periodically; afterward, it refreshes cache and finally forwards the cache. When the user requests content subsequently, the user interface is forwarded to the management. The cache management keeps the request information, saves file from local cache and lastly performs the request. In the decision phase, upon completion, the decision handles the information about all the requests and management module as per requirement. So, decision module updates database according to the users.

6. Prediction Algorithm

The proposed algorithm is based on the prediction of user’s interaction on a social network. In proactive cache, the caching entity should be user’s context-specific and
popularity aware. In each context, the popularity of content will be learned per individual basis. The first step involves user’s request, after that the content application reads the input through algorithm, marking the posted emotion such as "[I] am [happy]". The second step is Part of Speech (POS) tagging, which identifies the words as "[Noun] [Adjective] [Verb]". We have used Hidden Markov Model(HMM) for POS tagging.

\[
\prod_{i=1}^{n} p(w_i / t_i) p(w_i / t_{i-1})
\]

HMM is a stochastic model, proposed system is a model using HMM, it is assumed to be a Markov process with observable output and has unobservable states. HMM, consists of probability system starting, transitioning and system emitting output. In the case of POS, tags are assumed to be the state and words are assumed to be the outputs. In our specific case and words tagger is interested in finding the most likely sequence of tags that generates a sequence of words. To gain this, we assume that words probability depends upon tags and the probability of tag depends only on its previous tag. In HMM the sequence of the words ‘w’ and sequence of tags ‘t’, we normally compute a most likely sequence of n tags by enumerating all possible sequence of tags. The third step determines the main part of the sentence. The fourth step is to identify the human psychological framework. Hence, cache content placement and popularity of content should depend on user-specific, that maximize the cache hits by such an approach. We optimize cache according to different prioritization user’s levels.

In machine learning and linguistics, the sentiment analysis is a well-known problem which is used in earlier work with different employed language and classifiers models. Commonly, it is expressed as a problem classifier for the given text which needs to be characterized as neutral, negative or positive. More significance is expected to be given to this segment. To capture the emotion, we define it as follows.

\[
Emotion = \frac{|Positive\ &\ & Neutral\ Tweets|}{| Neutral\ Tweets|}
\]

Here, we model that content popularity depending on user’s personal characteristics. Through the above principle, our method makes a decision whether a person is ‘happy’ or ‘sad’. In step 6, the system display results in a graphical form. Proposed algorithm learns content popularity according to the user’s specifications.

**Algorithm 1:** Proposed Emotion Prediction

1. BEGIN
2. \[x = \{x_1, x_2, ..., x_n\} \] //Get information from twitter
3. \[E = \{E_1, E_2, ..., E_n\} \] //Eight emotion condition
4. \[Z = \{Z_1, Z_2, ..., Z_n\} \] //emotion adjective
5. \[If \ Total X = \{D_1 = D_2\} \] //Dates of post on twitter
6. \[R/then, Record is found\]
7. \[end //then\]
8. 
   \[Tok _pt ([ ])\] //Tokenization the post
9. 
   \[Compute _Tag(x)\] //Given by (1) Markov Model
10. \[get_m \] (main part of x)//Identify the main part
11. \[max=Em//Get the emotion by (2)\]
12. \[if \{E_n, E_2, ..., E_n\} = \{A_1, A_2, ..., A_n\} \] // File learning the record find Adjective
13. 
   \[A_n = \{E_1, E_2, ..., E_n\} \]
14. \[A_m = Total E_m //Match emotion category\]
15. \[Output=max(Z)\]
16. \[if \{Pr_{bev}=Z\} \] //Classifier judge the features that are related to the user behavior
17. \[Update(Rd) //Update the user profile according the record\]
18. \[end if\]

19. end

analyses the features related to the user behaviors. Those features are assigned to the user profile. Pearson correlation coefficient is used to measure the similarities. The recommendation engine on the basis of those calculation will recommend the content to the user. Propose algorithm for forecasting the cache content according to context specific. Generate content popularity and then push to the users according to the mood.

**7. Results and discussions**

Due to storage space limitation, a vast amount of multimedia contents request and also users have a different degree of attention towards the content at SBSs. Hence sophisticated cache forecast algorithms that learn the preference of users with different context are required. We describe content aware prediction process which depends upon a user interaction on social network. We have conducted two different types of analysis, through which we demonstrate the results of user’s emotions. The following are the analysis method to estimate the user’s emotion.

**7.1 First Analysis**

Here the system analyzes the tweets for ten days; first system extracts the user’s tweets within ten days from "April 20, 2016, to April 30, 2016". In ten days a user posts “118” tweets and "185" adjectives are found from the tweets after the POS tagging. At first, the text is retrieved from the Twitter database, system marks that text and then POS
tagging executes. Afterward, the system filters out the adjectives. From the above tweets, various adjectives are created as shown in table 1.

After analyzing the above emotional adjectives, the system displays the result which is depicted in figure 3. The figure 3 can also be used as an application to view the user’s emotion for the purpose of the analysis. In this analysis, different emotional words are used and find their adjective of the emotional words like (successful, tired and happy, etc.).

Table 1: Enlightens the abbreviation of emotions which was posted by a user on social media. Features include a unique lexical function.

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curiosity</td>
<td>CO</td>
</tr>
<tr>
<td>Urgency</td>
<td>UR</td>
</tr>
<tr>
<td>Confusion</td>
<td>CO</td>
</tr>
<tr>
<td>Angry</td>
<td>AG</td>
</tr>
<tr>
<td>Satisfied</td>
<td>SA</td>
</tr>
<tr>
<td>Happy</td>
<td>HA</td>
</tr>
<tr>
<td>Inspired</td>
<td>IN</td>
</tr>
<tr>
<td>Peaceful</td>
<td>PE</td>
</tr>
</tbody>
</table>

Fig. 3. Ten days tweets

In this experiment, we defined one hundred and thirty-nine adjectives. System displays different mood states of the user as shown in figure 3. The graph mentioned in figure 3 shows the number of adjectives in the category of emotional words like happy, unhappy etc.

In figure 4, the user emotions are shown. In ten days, the user posts numerous tweets which were related to the happiness. User behavior can also be seen on the system. On the basis of posted tweets of the user which were related to the happiness, we can assume that the user was happy in the last ten days. Four times the user posted ‘beautiful’, five times posted ‘best’, fourteen times posted ‘good’, two-time posted ‘great’, nine-time posted ‘happy’ and three-time posted ‘irritating’ etc. System displays maximum tweets posted by the user which is related to the happy emotional category. From the results, we conclude that the user is happy within ten days. System analysis the user’s emotion and then predicts the content. Decision making depends upon a principle that how many times a certain word is repeated.

7.2 Second Analysis

The system tries to make a 45-day analysis of the published tweets. The system acquires tweets from users within 45 days of the data range from “April 01, 2016 to May 15, 2016”. There is a total of 121 tweets. All tweets retrieved by the system are mentioned in table 1. The system executes the POS tag and finds the 217 adjectives mentioned from the tweets. After the POS tagging, the system filters out the adjectives. In figure 5, the system displays the results of the Twitter post. The figure shows the emotional words filtered from the user’s tweets. Users posted words include 7% curiosity, 3% emergency, 13% chaos, 11% rage, 13% satisfaction, 18% happy, 16% inspired and 9% peace related. After analyzing the above emotional adjectives, the system displays the results which are shown in figure 5.

The bar chart option can also depict the user’s emotional state in the application. After performing the analysis from the tweets described above, the system displays the morphological state in the graph as depicted in figure 5. In figure 6, the tweets above are examined; the system indicates that the user twitter post is most relevant to happy mood category. The user posted following tweets four times (personal, bad & beautiful), five times (best), two-times (wrong, sure, severe, Secret, red, privileged, own, excessive, confidential, capable, argumentative, authoritative, angry, adaptable) thirty times (difficult), fourteen times (good), ten times (happy) etc. and all other emotions are posted just one time. So, from the above results, we observed that the person in 45 days was ‘happy’. On the basis of the sentimental words, cache management learning module
8. Conclusion

In this article, we have discussed the limitations of networks and proposed a novel proactive networking prediction paradigm, where social media plays an imperative role. Peak data traffic demands can be substantially reduced by proactively predicting users demands via strategic caching. The proposed cache dynamic mood based proactive prediction model can analyze the human emotion by observing the twitter posts. We have shown how social media can be utilized to forecast future outcomes. Our algorithm on a real-world dataset designed system can find lexical features and perform operations using an artificial intelligence technique such as natural language processing. Under the scope of complete analysis, the system will take a decision. Forecasting content according to the user profile will bring substantial improvement regarding content availability and cache storage capacity at the edge of network. At the deeper level, there is a need to improve text analysis algorithms to read pictures, videos and also build the system which can perform text analysis on multiple social networks.

References


ASIF KABIR is a Ph.D. candidate in the college of communication engineering at Chongqing University. He received his Master of Sciences degree in the field of Computer Science from the University of Azad Jammu & Kashmir Pakistan. He worked as a “Web Manager” in Information Technology Board Muzaffarabad AJ&K from January 2010 to December 2010. Afterwards, he joined University of Azad Jammu & Kashmir from December 2010 to May 2014 as a “System Engineer/ Network Administrator”. From May 2014 to date he is working as a “System Engineer/ Network Administrator” in University of Management Sciences and Information Technology Kotli” Azad Kashmir. His main research areas are in mobile wireless communication networks, wireless sensor networks, social networks and big data.

Muhammad Shahid Iqbal is currently working towards his Ph.D. at the School of Computer Science and Application Technology at Anhui University Hefei China. His research interest includes computing vision, image processing, machine learning, automated software engineering and software testing.

Zain ul Abidin Jaffri was born in Pakistan on Mar 01, 1988. He received his bachelor degree in electronics engineering from COMSATS Institute of Information Technology Abbottabad, Pakistan in 2010. Later on, he came to Chongqing University China for Master’s degree in electronics & communication engineering in 2012. He worked on Wireless Sensor Networks in the past and is now working on Wireless and Antenna Propagation in Chongqing University, China for completion of his Ph.D. research work.

Shoujat Ali Rathore is working as a lecturer at University of Kotli since 2007. He received bachelor’s degree in information technology from the University of Sindh Jamshoro and masters from Ghulam Ishaq Khan Institute in 2004 and 2007 respectively. Currently, he is pursuing his Ph.D. at Chongqing University, Peoples Republic of China. His research interests include but are not limited to Image processing, Face recognition and Machine learning.

Edvin J. Kitindi received the B.Sc. degree in ICT-management from Mzumbe University Morogoro Tanzania and the M.Eng. degree in electronics and communication engineering with Chongqing University, Chongqing, China, in 2008 and 2012, respectively. He is currently pursuing the Ph.D. degree in communication and information systems with Chongqing University, China. He has been with the Sokoine University of Agriculture, Tanzania, as an Academic Staff of informatics, since 2008. His current research interests include wireless network virtualization, mobile networks’ resources management, and ICT for Development.