Social Aware Edge Caching in D2D Enabled Communication

Jingyi Chen[1,2,3](✉)

[1] School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
chenjytree@163.com

[2] Key Laboratory of Optical Communication and Network in Chongqing, Chongqing 400065, China

[3] Key Laboratory of Ubiquitous Sensing and Networking in Chongqing, Chongqing 400065, China

Abstract. As a promising architecture, mobile edge networks can effectively mitigate the load on backhaul links and reduce the transmission delay simultaneously. With the development of artificial intelligence (AI) and machine learning, how to reasonably combine AI as well as machine learning with communication is a hot topic. In this paper, considering the content features and user preference jointly, the projective adaptive resonance theory neural network (PART NN) is used to design the community architecture. After that, we can obtain the status table of communities. In order to reduce the redundant caching, the popular contents will be cached in the user equipment (UE) of center user in advance. The cache scheme of center user is adjusted according to the status table. Two transmission links are considered, i.e., cellular link and device-to-device (D2D) link, to reduce the transmission delay. Since the content preference of UE is time-varying and the migration patterns are various, the community architecture will be updated dynamically to further improve the cache hit rate. The migration patterns of UEs are affected by social factors as well as geographical factors. The simulation results show that the community construction scheme and cache scheme effectively improve the cache hit rate and reduce the transmission delay simultaneously.

Keywords: Device-to-Device (D2D) · Edge caching · Projective adaptive resonance theory neural network (PART NN) · Social attributes

This work was supported in part by the National Natural Science Foundation of China under Grants 61771082 and 61871062, and in part by the Program for Innovation Team Building at the Institutions of Higher Education in Chongqing under Grant CXTDX201601020.

S. Han et al. (Eds.): AICON 2019, LNICST 287, pp. 335–349, 2019.
https://doi.org/10.1007/978-3-030-22971-9_30

1 Introduction

According to the data released by Cisco in August 2017, the amount of data traffic will continue growing exponentially year by year [1]. The rapid development of data traffic has leaded to stringent requirements in not only communications but also computations, which will cost a mass of resources [2–6]. Compared to traditional centralized network architecture, the mobile edge networks utilize low-cost resources to provide computing and caching capabilities at the edge of cellular networks, which brings better performances in various aspects, i.e., the transmission delay, proximity services and so on [7]. Furthermore, edge caching is widely considered to effectively solve the large-scale content interaction of future networks and mitigate the backhaul loads.

Researches indicate that the data requested by user equipments (UEs) mainly conform to the Paretos principle, which means most of data requests result from less contents [8]. As a consequence, there may be a plenty of redundant repeated requests, which will cause ever-increasing backhaul loads. In order to reduce the backhaul loads, the consumption of community energy, and the request delay, the contents are cached on edge, i.e., the base station (BS) and the UE [9,10]. However, as a widely agreement, the caching capacities of BSs and UEs are limited. Hence, how to select the most useful or valuable data, which will be cached proactively, is critical to guarantee the quality of service (QoS) of different UEs. With the introduction of social attributes, the combination of social attributes and caching has stimulated growing interests. Generally, the social relationship, social trust, user mobility and other features are considered to build communities and select caching strategy [7,11–13]. However, the importance of content characteristic is ignored and the migration patterns of user are not fully considered. In addition, the communities are built based on the offline data and the community structure is fixed. Hence, the contents cached may not always meet the user demands, which will cause the decline of cache hit rate and the growth of the waste of resources.

With the developments of artificial intelligence (AI) and machine learning, many researches consider to incorporate them into communication systems and networks. Typically, to solve the problem mentioned above, part of researches preprocess the cached data using neural network, which can obtain the popular contents to improve the cache hit rate [14–16]. In [14], the preference list of each cache entity is derived from the data analysis of tweets during the 2016 U.S. presidential election using deep learning long short-term memory (LSTM) neural network. In [15], the convolutional neural network (CNN) is used to analyze the sentence and extract the features. In [16], the features of contents are analyzed using the extreme learning machine, and then the popularity of contents can be obtained. Generally, the popularity of content is defined as the requested ratio of particular content [13], and the data analysis is usually applied to offline data ignoring the fact that the data are updated constantly. The research in [17] shows the user preference is not always stable. On the contrary, it may change periodically according to different factors.

In addition, device-to-device (D2D) technology, which is one of the key technologies of 5G, has been validated to be able to guarantee the data forwarding performance well for short distance communications from the social point of view [18]. Hence, caching data in UEs and delivering data through D2D are promising ideas to further reduce the transmission delay.

Taking above situations into consideration, in this paper, a D2D enabled system model is shown where the UE of center user can cache contents. Under the system model, the requested contents can be delivered efficiently and fast to the UEs. A projective adaptive resonance theory neural network (PART NN) based community construction scheme is proposed considering social characteristics and content preferences of each UE. With such scheme, the content characteristics of each community become more and more distinct as the number of UEs increases, which will guide the caching of the center UE. Considering the case that the content preferences of the UEs may change varying with time and the migration patterns of UEs, the community architecture will be updated dynamically. The cache scheme of center UE will be adjusted accordingly.

The remaining of this paper is organized as follows. Section 2 introduces the system model environment. Section 3 illustrates the social aware community construction scheme. Section 4 studies D2D enabled caching scheme. In Sect. 5, simulation results are presented and discussed. Section 6 finally concludes the paper.

2 System Model

The system model considered in this paper is depicted in Fig. 1, which contains three layers, i.e., the cloud layer, the access layer and the terminal layer. The whole network resources are controlled by the cloud servers in the cloud layer. Due to the long distances between the cloud servers and the UEs, the transmission delay is very high. Besides, the redundant requests cause huge traffic burden at the cloud servers and the resources can not be utilized rationally and efficiently. As a result, the computing and caching capabilities on the edge are used. In this paper, the BSs provide the network access and computing capabilities for UEs. However, in the hot spot region there are mass UEs within the coverage of a specific BS, which will cause vast repeated requests of the popular contents. Hence the construction of communities in the terminal layer is vital to improve the resource utilization rate. Furthermore, D2D technology can be used to reduce the transmission delay.

In detail, the communities are denoted by $C = \{C_1, C_2, \cdots, C_i, \cdots, C_{N_C}\}$, where N_C represents the number of communities. In each community, there are two types of UEs, i.e., one center UE and several normal UEs, which have similar content requests considering the social factors. The center UE will cache the popular contents in the community it belongs to in advance. The normal UEs will firstly build the D2D links with the center UE to obtain the contents. Due to the constraint of trust, normal UEs can not directly transmit data with each other via D2D links. Hence, if the center UE has not cached the contents

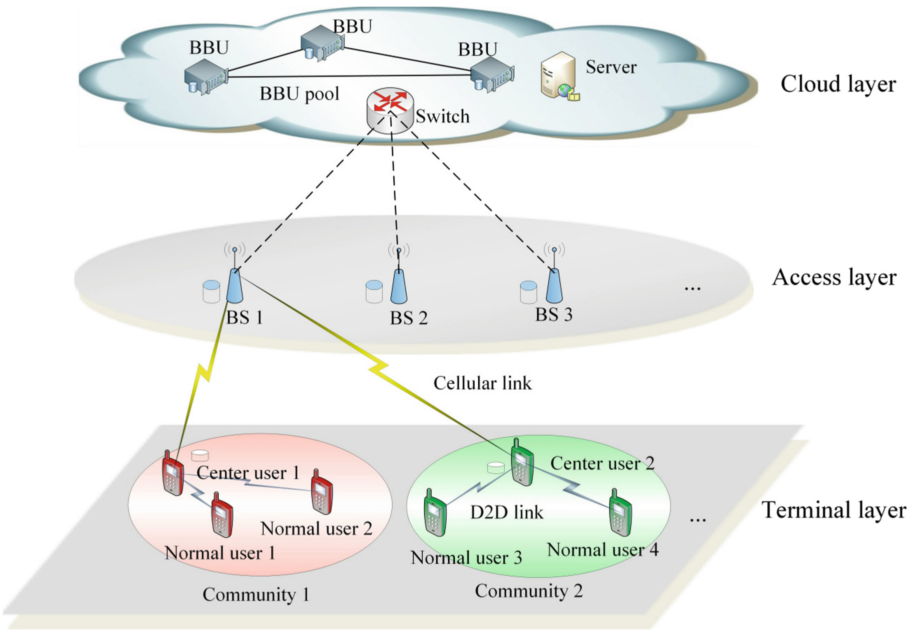

Fig. 1. System model.

requested by normal UE, the cellular link will be built between the normal UE and the corresponding BS to obtain the data. In this paper, UEs of both types are simply referred as UEs. Therefore, in an initial hot spot area, UEs served by the same BS are denoted by $U = \{u_1, u_2, \cdots, u_j, \cdots, u_{N_u}\}$, where N_u represents the number of UEs. Note that the belonging community of a particular UE may change dynamically due to the mobility as well as the time-varying content preference of UE. As a result, the community members are varying which will affect the contents cached in the center UE.

As mentioned above, the data delivery performance of the considered system model is dominated by the construction of community. Hence, how to build and update the community construction in real time, considering the social attributes, the content characteristics, and the migration patterns simultaneously, is critical to guarantee the content delivery performance.

3 Social Aware Community Construction Scheme

Generally, UEs in the same hot spot area may request similar popular contents at the same time period. To reduce the redundant data requests, the community should be built and the popular contents of community should be cached in the center UE in advance. However, most of existing community construction schemes just consider the effects of social attributes and geographical locations, ignoring the importance of content characteristics. Considering that the content

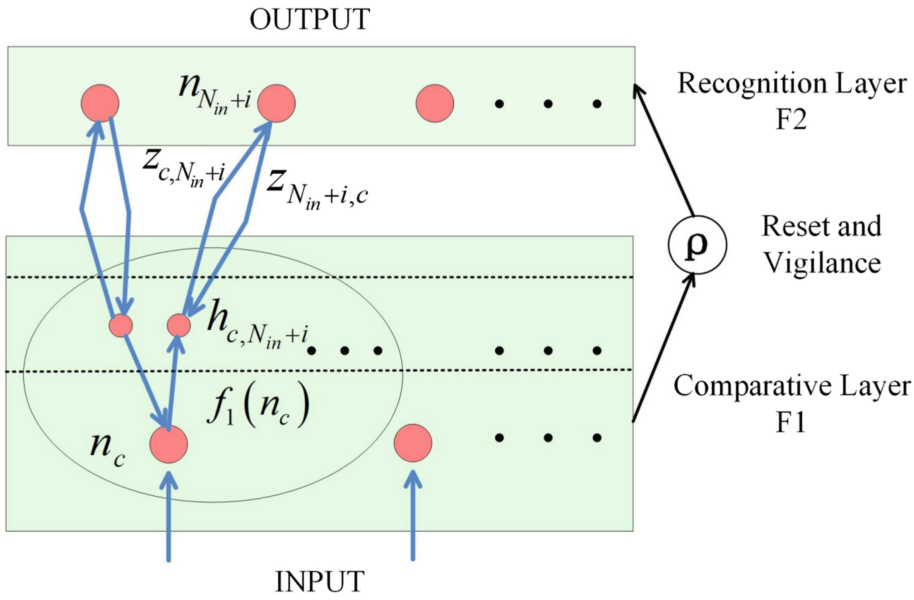

Fig. 2. PART NN model.

characteristics of UEs are different and change dynamically, the PART NN is used to construct and update the community dynamically.

Evolved from the adaptive resonance theory neural network (ART NN), which is a representative of competitive learning, the PART NN is proposed to solve the problem that the subspace formed by clustering can not be determined in advance [19,20]. The traditional PART NN architecture is shown in Fig. 2, including the comparison layer (F1 layer) and the cluster layer (F2 layer). Furthermore, the actual implementation of PART NN includes five states, i.e., initialization stage, cognition stage, comparison stage, search stage and adjustment stage. The community construction scheme will be discussed in detail based on the PART NN.

The set $M = \{m_1, m_2, \cdots, m_c, \cdots, m_k\}$ denotes the contents considered in this paper, where k represents the number of content types. The preference degree set about contents M of UE u_j is denoted by $D^{u_j} = \{\mathrm{d}_{m_1}^{u_j}, \mathrm{d}_{m_2}^{u_j}, \cdots, \mathrm{d}_{m_k}^{u_j}\}$. Since various contents have different effects on UE u_j and the effects are varying with time, the value of preference degree $\mathrm{d}_{m_c}^{u_j}$ in D^{u_j} varies continuously and may be different. The value of $\mathrm{d}_{m_c}^{u_j}$ is an average value during a time period, and is standardized in the range of $[0, 1]$.

The input of PART NN is D^{u_j}. The nodes in F1 layer are denoted by $N_1 = \{n_1, n_2, \cdots, n_{N_{in}}\}$. That is, the input value of the node n_c is $\mathrm{d}_{m_c}^{u_j}$, where $c \in [1, k]$, and $N_{in} = k$. Additionally, the nodes in F2 layer, i.e., the output nodes are denoted by $N_2 = \{n_{N_{in}+1}, n_{N_{in}+2}, \cdots, n_{N_{in}+N_{out}}\}$, where the output node $n_{N_{in}+i}$ represents the community C_i.

In the PART NN, two weights are critical to construct the communities named bottom-up weight $z_{c,N_{in}+i}$ and the top-down weight $z_{N_{in}+i,c}$, which represent the features of corresponding node $n_{N_{in}+i}$ in F2 layer. At the beginning, no community is formed in the hot spot area. Hence, when the preference degree set of the first UE denoted by u_1, i.e., D^{u_1}, is input to the PART NN, the node $n_{N_{in}+1}$ will be assigned to u_1. That is, the u_1 is belong to new community C_1. In the whole process of community construction, the weights $z_{c,N_{in}+i}$ and $z_{N_{in}+i,c}$ associated with the node $n_{N_{in}+i}$ in F2 layer, which represents new community C_i, are calculated as

$$z_{c,N_{in}+i} = \frac{L}{L-1+k} \tag{1}$$

$$z_{N_{in}+i,c} = d_{m_c}^{u_j} \tag{2}$$

where L is the constant given at the initialization, and u_j is the first UE in the new community C_i.

On the other hand, if some communities have existed in the area, each existed community has at least one UE. For the specific C_i among the existed communities, after the current u_j is put in C_i, the weights $z_{c,N_{in}+i}$ and $z_{N_{in}+i,c}$ of corresponding node $n_{N_{in}+i}$ are calculated as

$$z_{c,N_{in}+i} = \begin{cases} \frac{L}{L-1+|x|} & if \quad h_{c,N_{in}+i} = 1 \\ 0 & if \quad h_{c,N_{in}+i} = 0 \end{cases} \tag{3}$$

$$z_{N_{in}+i,c} = (1-\alpha)\, z_{N_{in}+i,c}^{old} + \alpha d_{m_c}^{u_j} \tag{4}$$

where x is the current number of cached contents in C_i and α is the learning rate given at the initialization. $h_{c,N_{in}+i}$ is the selectable output signal which will be discussed later.

Based on $z_{c,N_{in}+i}$ and $z_{N_{in}+i,c}$, the communities are constructed as follows. As mentioned before, when the preference degree set D^{u_1} of the first UE u_1 is input to the PART NN, the first community C_1 will form and the corresponding weights are set as (1) and (2).

When the preference degree set of the jth ($j > 1$) UE, i.e., D^{u_j}, is input to the PART NN, since some communities already exist, each existing community will be determined whether it is the most appropriate to u_j or not. If so, the u_j will be put in the most appropriate community. If not, a new community will be assigned to the u_j. To this end, for the input D^{u_j} of u_j, the selectable output signal should be calculated considering the weights of each node of N_2 as follows:

$$h_{c,N_{in}+i} = h_\sigma\left(f\left(n_c\right), z_{N_{in}+i,c}\right) l\left(z_{c,N_{in}+i}\right) \tag{5}$$

where $l\left(z_{c,N_{in}+i}\right)$ is a threshold function, and h_σ denotes whether the condition of similarity is satisfied. There holds

$$h_\sigma\left(f\left(n_c\right), z_{N_{in}+i,c}\right) = \begin{cases} 1 & if \quad d\left(f\left(n_c\right), z_{N_{in}+i,c}\right) \le \sigma \\ 0 & otherwise \end{cases} \tag{6}$$

where the signal equation $f\left(n_c\right)$ is generated at n_c, $d\left(f\left(n_c\right), z_{N_{in}+i,c}\right)$ means the quasi-distance, and σ is the distance parameter, which controls the degree of similarity between the input UE and current community. And then the forward output filter parameter $T_{N_{in}+i}$ can be calculated for all the nodes in F2 layer:

$$T_{N_{in}+i} = \sum_{n_c \in N_1} z_{c,N_{in}+i} h_{c,N_{in}+i} \tag{7}$$

Furthermore, the matching degree of content similarity between u_j and C_i needs to be calculated. The node whose matching degree Mat_{N_in+i} is less than the vigilance parameter ρ is considered as a candidate, i.e.,

$$Mat_{N_{in}+i} = \sum_{c} h_{c,N_{in}+i} \leq \rho \tag{8}$$

Finally, the most appropriate C_i corresponding to u_j should satisfy the following expression:

$$\begin{array}{ll} i = \arg\max & T_{N_{in}+i} \\ \quad \text{s.t.} & (8) \end{array} \tag{9}$$

When the most appropriate community C_i is found, the weights of corresponding node $n_{N_{in}+i}$ should be updated according to (3) and (4). If the optimal community is not existing, i.e., the constraint of vigilance parameter is not satisfied, a new community will be assigned to the current UE, and the weights associated with the node in the layer F2, which represents the new community will be updated according to (1) and (2).

When all the UEs' preference degree sets are put into the PART NN, the initial community construction is completed. The content characteristics of community C_i and UE u_j are denoted as $G^{C_i} = \left\{g_{m_1}^{C_i}, g_{m_2}^{C_i}, \cdots, g_{m_k}^{C_i}\right\}$ and $G^{u_j} = \left\{g_{m_1}^{u_j}, g_{m_2}^{u_j}, \cdots, g_{m_k}^{u_j}\right\}$, where

$$g_{m_c}^{C_i} = \begin{cases} 1 & \text{If } C_i \text{ has } m_c \\ 0 & \text{otherwise} \end{cases},$$

$$g_{m_c}^{u_j} = \begin{cases} 1 & \text{If } u_j \text{ has } m_c \\ 0 & \text{otherwise} \end{cases}.$$

The preference degree of content m_c in the community C_i is denoted as $\mathrm{d}_{m_c}^{C_i}$ and defined as

$$\mathrm{d}_{m_c}^{C_i} = \frac{\sum\limits_{\{j|u_j \in C_i\}} \mathrm{d}_{m_c}^{u_j} \cdot w_{u_j}}{|C_i|} = \frac{\sum\limits_{\{j|u_j \in C_i\}} \mathrm{d}_{m_c}^{u_j} \cdot \frac{1-E_j}{|C_i| - \sum\limits_{\{j|u_j \in C_i\}} E_j}}{|C_i|} \tag{10}$$

where $|C_i|$, w_{u_j}, and E_i denote the number of UEs in C_i, the weight and the information entropy of u_j, respectively. For the specific community, the less number of interests is, the smaller the E_j is, and the greater effect on the content feature of the community is. And there holds

$$E_j = -\ln\left(|C_i|\right)^{-1} \sum_{\{j|u_j \in C_i\}} \frac{sd_{m_c}^{u_j}}{\sum\limits_{\{j|u_j \in C_i\}} sd_{m_c}^{u_j}} \ln \frac{sd_{m_c}^{u_j}}{\sum\limits_{\{j|u_j \in C_i\}} sd_{m_c}^{u_j}} \tag{11}$$

where $sd^{u_j}_{m_c}$ is the standardized value of preference degree of the specific content m_c about u_j, which holds as

$$sd^{u_j}_{m_c} = \frac{d^{u_j}_{m_c} - \min D^{u_j}}{\max D^{u_j} - \min D^{u_j}} \tag{12}$$

Therefore, all the preference degrees of contents $d^{C_i}_{m_c}$ and the corresponding preference degree set $D^{C_i} = \left\{ \mathrm{d}^{C_i}_{m_1}, \mathrm{d}^{C_i}_{m_2}, \cdots, \mathrm{d}^{C_i}_{m_k} \right\}$ in C_i can be obtained. In detail, the status table of communities is shown as an example in Table 1. Besides, the status table of communities can be updated dynamically to represent the content preferences of the current users in each community, which may be affected by the changes of preferences and the migrations of UEs, and then the cache strategy of center UE will be updated correspondingly which will be introduced in the following.

Table 1. Status table of communities.

i	C_i	G^{C_i}	D^{C_i}
1	$(u_1, u_3, \cdots)$	$(0, 0, \cdots, 1, 1, 1)$	$(0, 0, \cdots, 0.1, 0.3, 0.04)$
2	$(u_2, u_7, \cdots)$	$(1, 1, \cdots, 0, 0, 1)$	$(0.03, 0.1, \cdots, 0, 0, 0.2)$
3	$(u_4, u_5, \cdots)$	$(1, 0, \cdots, 1, 0, 0)$	$(0.5, 0, \cdots, 0.1, 0, 0)$
...	...	...	...

4 D2D Enabled Caching Scheme

In order to fully utilize the resources on the edge, the caching capacity of UE is considered in this paper to decrease redundant data requests and transmission delay. Hence, for each community, the UE which is outgoing as well as reliable is selected as the center UE considering the duration of stay simultaneously. The center UE in the community C_i is denoted by cu^i and will cache the popular contents in advance. The remaining UEs in C_i are normal UEs denoted by nu^i_s which represents the sth normal UE. Because the incentive mechanisms have been considered in many literature to motivate users to cache content for others [21,22], we assume that the center UE cu^i is selfless. Hence the entire caching space of cu^i will be used to cache the popular contents of C_i.

Due to the limited caching capacity of cu^i, not all the popular contents whose $g^{C_i}_{m_c}$ equals to one in the content characteristics G^{C_i} of C_i can be put in the caching space of cu^i. Hence the preference degree of content $d^{C_i}_{m_c}$ is crucial to determine whether the content will be cached by cu^i or not. As a result, the content m_c in G^{C_i} will be put into the caching space of cu^i according to the descending order of $d^{C_i}_{m_c}$. And the constraint of caching capacity should be satisfied simultaneously as follow

$$\sum_{c=1}^{k} I \cdot g^{cu^i}_{m_c} \leq Q_i \tag{13}$$

where

$$g_{m_c}^{cu^i} = \begin{cases} 1 & \text{if } cu^i \text{ caches } m_c \\ 0 & \text{otherwise} \end{cases}$$

The set $G^{cu^i} = \left\{ g_{m_1}^{cu^i}, g_{m_2}^{cu^i}, \cdots, g_{m_c}^{cu^i}, \cdots, g_{m_k}^{cu^i} \right\}$ denotes popular contents set of cu^i and Q_i is the caching capacity of center UE cu^i. The content size of each type is assumed to be the same size denoted by I.

Since the content preference of UE is time-varying and the caching capacity of the center UE is limited, the contents cached on cu^i in advance may not always be consistent with the contents requested by normal UEs in the same community. Hence, the contents cached on the cu^i should be updated to guarantee the cache hit rate. Impelled by the contents desired, social trust, and time, the normal UEs may migrate to other communities. The migration of normal UEs will change the status table of communities. Furthermore, the caching scheme of cu^i will be changed accordingly. In order to improve the cache hit rate and decrease the transmission delay, as a important factor, the migration patterns of UEs deserve to be discussed in this section.

The migration patterns are divided into two categories, i.e., purposeful migration and purposeless migration. The purposeful migration means the purpose of migration is impelled by the content preferences or social relationship. For the first case, the popular contents requested by nu_b^i are known and not included in the current community C_i. Hence, nu_b^i will migrate to another community in terms of the similarity between the updated content characteristic set $G^{nu_b^i}$ of nu_b^i and G^{cu^x} of the center UE cu^x. And then the nu_b^i belongs to the community C_{win} which has the maximal similarity, i.e.,

$$\begin{aligned} \text{win} &= \arg\max_x \left(sim \left(G^{nu_b^i}, G^{cu^x} \right) \right) \\ &= \max \left(\frac{\sum_{c=1}^{k} \left(g_{m_c}^{nu_b^i} - \overline{g_{m_c}^{nu_b^i}} \right) \left(g_{m_c}^{cu^x} - \overline{g_{m_c}^{cu^x}} \right)}{\sqrt{\sum_{c=1}^{k} \left(g_{m_c}^{nu_b^i} - \overline{g_{m_c}^{nu_b^i}} \right)^2} \sqrt{\sum_{c=1}^{k} \left(g_{m_c}^{cu^x} - \overline{g_{m_c}^{cu^x}} \right)^2}} \right) \end{aligned} \tag{14}$$

where x is an integer and $x \in [1, |C|]$. $|C|$ denotes the number of communities. The set $G^{nu_b^i} = \left\{ g_{m_1}^{nu_b^i}, g_{m_2}^{nu_b^i}, \cdots, g_{m_c}^{nu_b^i}, \cdots, g_{m_k}^{nu_b^i} \right\}$ denotes the content characteristic set of nu_b^i, and

$$g_{m_c}^{nu_b^i} = \begin{cases} 1 & \text{if } nu_b^i \text{ wants } m_c \\ 0 & \text{otherwise} \end{cases}.$$

For the second case, we suppose there is a social relationship between the normal UE nu_a^x and the particular center UE cu^i, i.e., classmate relations, colleague relations and so on. In order to exchange data with cu^i, nu_a^x will migrate to C_i directly where cu^i belongs to. Due to the selflessness of cu^i, the contents are cached for the normal UEs in the community C_i. Hence the purposeful migration will not affect the caching scheme of center UE.

The purposeless migration means the migration of normal UE is impelled by the time. For example, the preference content set $G^{nu_s^i}$ of normal UE nu_s^i is changing with varying time or caused by the change of location. In this case, the attribution community of normal UE and the preference content set of community should be adjusted dynamically. Hence, the new preference content sets of normal UEs should be input into the PART NN which is discussed in the 3th section. Under the construction and update of communities, the center UE can cache data in a highly efficient way. The data transmission is mainly resorted to the D2D links, which not only reduces the traffic load but also improves the content delivery performance.

In this paper, since the caching capacity of the center UE is limited, the desired data of normal user may not be cached in the caching space of the center UE in advance. Therefore, there are two kinds of links for data transmission, i.e., D2D link and cellular link. When the center UE cu^i has the data requested by the user nu_s^i, and the D2D link constraints are satisfied, a D2D link will be built for data transmission between cu^i and nu_s^i. The transmission rate R_{cu^i,nu_s^i} is

$$R_{cu^i,nu_s^i} = B_{dl}\log_2\left(1 + \frac{P_{cu^i}h_{cu^i,nu_s^i}}{N_0}\right) \tag{15}$$

where B is the channel bandwidth, P_{cu^i} denotes the transmitting power of cu^i and h_{cu^i,nu_s^i} denotes the channel gain, N_0 denotes the additive Gaussian white noise.

When the all the center UEs within the BS do not have the data required by nu_s^i, the cellular link is built to transmit data, and also the cu^i obtain the data from the BS through the cellular link. The corresponding transition rate $R_{BS,UE}$ holds as

$$R_{BS,UE} = B_{cl}\log_2\left(1 + \frac{P_{BS}h_{BS,UE}}{N_0}\right) \tag{16}$$

where P_{BS} denotes the transmission power of the BS, $h_{BS,UE}$ denotes the channel gain.

Furthermore, the minimum total content delivery delay holds as

$$T_{total} = \frac{\sum\limits_{\left\{m_c \middle| g_{m_c}^{cu^j}=1 \ \& \ g_{m_c}^{nu_s^j}=1,\ j=1,2,\cdots,N_C\right\}} I_{m_c}}{R_{cu^j,nu_s^j}} + \frac{\sum\limits_{\left\{m_c \middle| g_{m_c}^{cu^j}=0 \ \& \ g_{m_c}^{nu_s^j}=1,\ j=1,2,\cdots,N_C\right\}} I_{m_c}}{R_{BS,nu_s^j}} \tag{17}$$

5 Performance Analysis

In this section, we evaluate the performance of the PART NN based community construction scheme (PNCCS) and that of the social aware caching scheme (SACS) with the help of MATLAB. The PNCCS in this paper is compared with two related strategies. The first one uses the two-step coalitional game

(2SCG) to build the community architecture considering the payoff [23]. The second one mainly considers the mobility and localization of users in both geographical and virtual communities (MLGVC) to obtain the community architecture [24]. Besides, two caching schemes are compared with the SACS. The first one caches the most popular contents in the world on the UEs which is called as most popular caching scheme (MPCS). The second one caches the content on the UEs randomly which is called as random caching scheme (RCS). In this section, the transmission bandwidth is 10 MHZ and the noise spectral density is -174 dbm/Hz. The transmit power of D2D user and cellular user are 17 dbm and 23 dbm respectively. Besides, the max D2D transmission distance is 50 m and the radius of cell is 500 m.

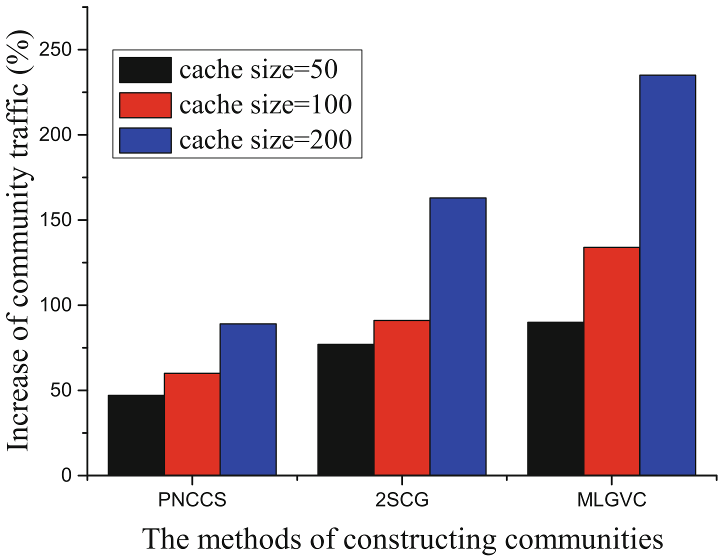

Fig. 3. Increase of community traffic.

Figure 3 depicts the increases of community traffic of different community construction strategies. For the specific cache size of center UE, the increase of community traffic in PNCCS is the minimum compared to other strategies. The reason is that the PNCCS fully utilizes the content characteristics by PART NN. Furthermore, considering the social characteristics and migration patterns, community architecture is updated dynamically to guarantee the cache hit rate of center UEs and efficiently decreases the similar requests simultaneously. Differently, the 2SCG ignores the importance of content characteristics and the MLGVC only considers the location and mobility. Since the factors are not considered completely, the performances of 2SCG and MLGVC are both not optimal. Additionally, Fig. 3 also shows that with the rise of cache size, the increase of community traffic increases accordingly. In detail, the increase of community traffic of PNCCS compared with 2SCG and MLGVC is decreased by 40.7% and 57.2% on average.

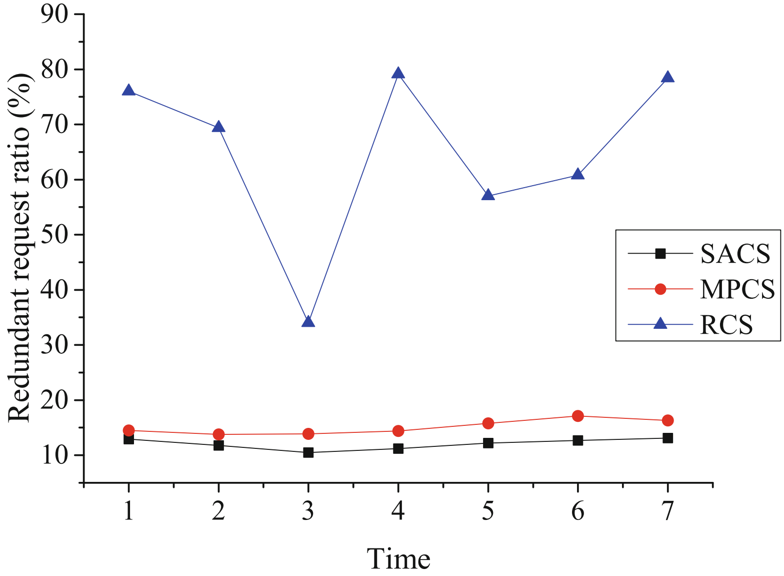

Fig. 4. Redundant request ratio.

Figure 4 depicts the redundant request ratio varying with the period of time under different caching schemes. In RCS, due to the randomness of contents cached, the contents cached in the center UE in advance may not always satisfy the demands of normal UEs within the same community. Hence, the redundant request ratio changes irregularly. Besides, the contents cached in MPCS are globally popular which may not be the popular contents in local area. Hence, normal UEs need to obtain data from the BS. Since the popular contents in a particular area are similar, redundant requests will exist. Considering different factors jointly to accurately build and update the community architecture, Fig. 4 validates that the performance of SACS is the best.

Figures 5 and 6 exhibit the variation trends of the cache hit rate and the average total delay of UEs in different number of UEs. The cache hit rate is positively correlated with the number of UEs. In SACS, as the number of people increases, the community characters will be more and more obvious based on PART NN. Compared with MPCS and RCS, the contents cached on center UE in SACS are more targeted. Considering the social characteristics and migration patterns of UEs, the community architecture is adjusted dynamically to better meet the data requests of UEs. Hence, Fig. 5 shows that the cache hit rate of SACS is always higher than that in MPCS and RCS. In detail, numerical results show that the cache hit rate of SACS is improved by 31.7% compared with MPCS on average. Besides, Fig. 6 shows that as the number of people increases, all the average total delays of three schemes decrease. Moreover, the average total delay of UEs in SACS is the lowest. In addition to the factors considered above, another reason is that the popular contents are cached on center UE in advance, and normal UEs can directly build D2D links with center UE for data transmission. Hence the transmission delay can be reduced effectively.

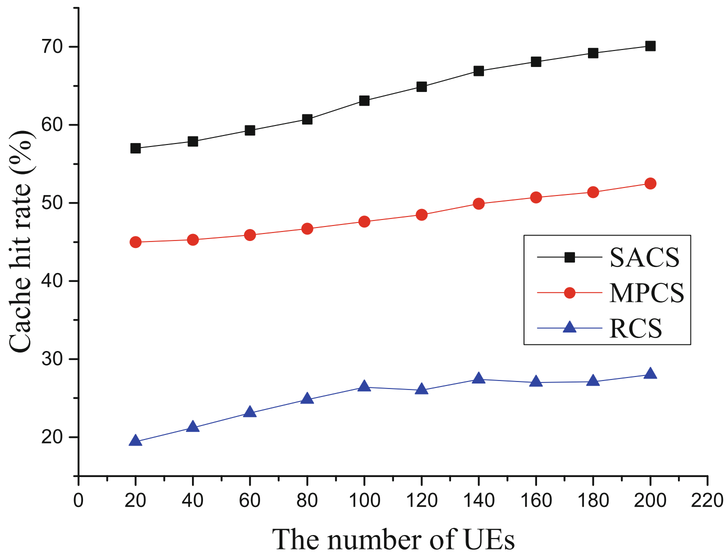

Fig. 5. Cache hit rate.

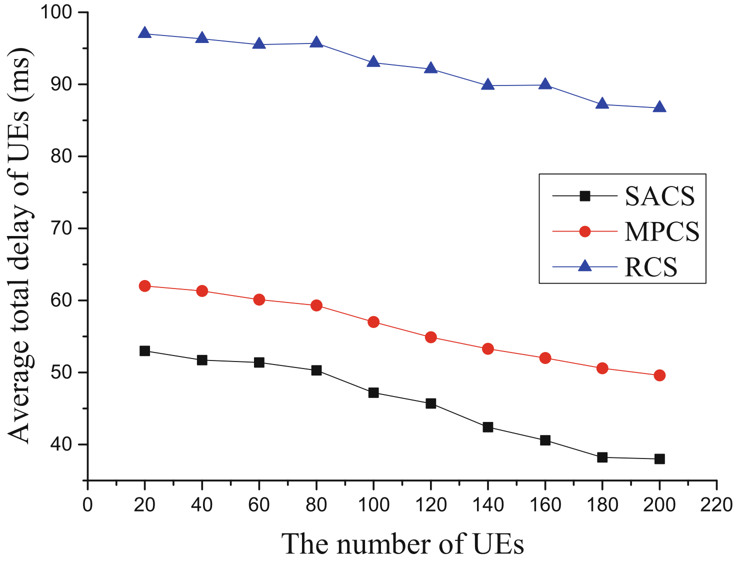

Fig. 6. Average total delay of UEs.

6 Conclusions

In this paper, we first propose a social-aware community construction scheme based on PART NN to improve the cache efficiency. Considering the content preferences, the initial community architecture is built and the status table of communities can be obtained. Due to the time-varying characteristics of users's data requests and user mobility, the community architecture should be updated dynamically. Hence we discuss the migration patterns of users which are affected by the social characteristics, content characteristics as well as mobility. In order to make full use of resources on the edge, the center UEs are selected to cache the popular contents of communities in advance. The caching scheme of center UE is guided by the status table of communities. To further reduce the transmission delay, the D2D links can be built between UEs. Numerical results show that the proposed schemes can not only improve the cache hit rate significantly but also decrease the redundant request ratio and the average total delay of UEs.

References

1. Cisco Systems Inc.: Cisco visual networking index: global mobile data traffic forecast update, White Paper, March 2017
2. Zhou, L., Wu, D., Chen, J., Dong, Z.: When computation hugs intelligence: content-aware data processing for industrial IoT. IEEE Internet Things J. **5**(3), 1657–1666 (2018)
3. Wu, J., Wu, J., Cui, H., Luo, C., Sun, X., Wu, F.: DAC-Mobi: data assisted communications of mobile images with cloud computing support. IEEE Trans. Multimedia **18**(5), 893–904 (2016)
4. He, C., Wang, H., Hu, Y., Chen, Y., Fan, X.: MCast: high-quality linear video transmission with time and frequency diversities. IEEE Trans. Image Process. **27**(7), 3599–3610 (2018)
5. Wu, D., Shi, H., Wang, H., Wang, R., Fang, H.: A feature-based learning system for Internet of Things applications. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2018.2884485
6. Tan, B., Lu, J., Wu, J., Zhang, D., Zhang, Z.: Toward a network slice design for ultra high definition video broadcasting in 5G. IEEE Wirel. Commun. **25**(4), 88–94 (2018)
7. Wang, S., Zhang, X., Zhang, Y., Wang, L., Yang, J., Wang, W.: A survey on mobile edge networks: convergence of computing, caching and communications. IEEE Access **5**, 6757–6779 (2017)
8. Bastug, E., Bennis, M., Debbah, M.: Living on the edge: the role of proactive caching in 5G wireless networks. IEEE Commun. Mag. **52**(8), 82–89 (2016)
9. Gregori, M., Gomez-Vilardeb, J., Matamoros, J.: Wireless content caching for small cell and D2D networks. IEEE J. Sel. Areas Commun. **34**(5), 1222–1234 (2016)
10. Liu, A., Lau, V.K.: Asymptotic scaling laws of wireless ad hoc network with physical layer caching. IEEE Trans. Wireless Commun. **15**(3), 1657–1664 (2016)
11. Li, Y., Wu, T., Hui, P., Jin, D., Chen, S.: Social-aware D2D communications: qualitative insights and quantitative analysis. IEEE Commun. Mag. **52**(6), 150–158 (2014)
12. Li, Y., Su, S., Chen, S.: Social-aware resource allocation for device-to-device communications underlaying cellular networks. IEEE Wirel. Commun. Lett. **4**(3), 293–296 (2015)
13. Wang, Z., Sun, L., Zhang, M., et al.: Propagation- and mobility-aware D2D social content replication. IEEE Trans. Mob. Comput. **16**(4), 1107–1120 (2017)
14. Tsai, K.C., Li, W., Zhu, H.: Caching for mobile social networks with deep learning: Twitter analysis for 2016 U.S. election. IEEE Trans. Netw. Sci. Eng. (2018)

15. Tsai, K.C., Li, W., Zhu, H.: Mobile social media networks caching with convolutional neural network. In: IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Barcelona (2018). https://doi.org/10.1109/WCNCW.2018.8368988
16. Tanzil, S.M.S., Hoiles, W., Krishnamurthy, V.: Adaptive scheme for caching YouTube content in a cellular network: machine learning approach. IEEE Access **5**, 5870–5881 (2017)
17. Golder, S.A., Wilkinson, D.M., Huberman, B.A.: Rhythms of social interaction: messaging within a massive online network. Chaos Interdisc. J. Nonlinear Sci. **93**(10), 098701–66 (2015)
18. Mach, P., Becvar, Z., Vanek, T.: In-band device-to-device communication in OFDMA cellular networks: a survey and challenges. IEEE Commun. Surv. Tutor. **17**(4), 1885–1922 (2015)
19. Cao, Y., Wu, J.: Projective ART for clustering data sets in high dimensional spaces. Neural Netw. **15**(2), 105–120 (2002). The Official Journal of the International Neural Network Society
20. Cao, Y., Wu, J.: Dynamics of projective adaptive resonance theory model: the foundation of PART algorithm. IEEE Trans. Neural Networks **15**(2), 245–260 (2004)
21. Yi, C., Huang, S., Cai, J.: An incentive mechanism integrating joint power, channel and link management for social-aware D2D content sharing and proactive caching. IEEE Trans. Mob. Comput. **17**(4), 789–802 (2017)
22. Shi, L., Zhao, L., Zheng, G., Zhu, H.: Incentive design for cache-enabled D2D underlaid cellular networks using stackelberg game. IEEE Trans. Veh. Technol. **68**(1), 765–779 (2019)
23. Wang, F., Li, Y., Wang, Z., et al.: Social-community-aware resource allocation for D2D communications underlaying cellular networks. IEEE Trans. Veh. Technol. **65**(5), 3628–3640 (2016)
24. Filiposka, S., Mishev, A., Gilly, K.: Community-based allocation and migration strategies for fog computing. In: IEEE Wireless Communications and Networking Conference (WCNC), Barcelona (2018). https://doi.org/10.1109/WCNC.2018.8377095