ICC 2021 - IEEE International Conference on Communications | 978-1-7281-7122-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICC42927.2021.9500829

On Coded Broadcasting for Wireless Recommendation Systems

Rasmus Vestergaard*, Osama Hanna†, Linqi Song‡, Daniel E. Lucani*, and Christina Fragouli†
*Aarhus University †University of California, Los Angeles (UCLA) ‡City University of Hong Kong
rv@ece.au.dk, ohanna@g.ucla.edu, linqi.song@cityu.edu.hk, daniel.lucani@ece.au.dk, christina.fragouli@ucla.edu

Abstract—**This paper considers benefits of coding techniques in recommendation systems operating over wireless channels with erasures. We identify scenarios where coded broadcasting can increase the overall user satisfaction at a fixed channel utilization level. Such opportunities arise both when user preferences are unknown, and must be explored, and when they are known and must be exploited to recommend optimally. We determine the magnitude of the potential gains and show that coding is most beneficial if users have heterogeneous preferences. Finally, we provide inequalities that can be evaluated to determine whether coding would be beneficial for a certain reward structure.**

I. Introduction

Recommendation systems help make better choices for users who do not have sufficient personal experience with the alternatives [1]. For example, such systems allow for transfer of experiences between similar users under the assumption that their experience with the recommendations will be similar. Before we can generalize from our users' experiences, we first need to learn what they are. This is done by interacting with the users through Reinforcement Learning [2], a key component in recommendation systems.

In this paper, we study interactions between recommendation systems in a wireless medium and index coding [3], [4] techniques. The wireless medium implies that messages will be broadcast to several users, so for each transmission we should make the recommendation that best satisfies all recipients simultaneously. We assume that our system has continual interaction with the same set of users, but each transmission is not guaranteed to reach all users. Our goal is to quantify the advantages that arise if a coded transmission scheme is allowed in this setting. We make no assumption that users have initial side information, but instead utilize messages not received by all users as side information for coded messages. We show that coding techniques can maintain the same level of channel utilization while achieving a greater user satisfaction in several scenarios.

To envision such a wireless recommendation system, you may think of a shopping mall providing free WiFi to its patrons. This network also distributes advertisements to the visitors. To minimize the impact on the network, only enough time to transmit a single advertisement is allocated at specific intervals. Since users may move around inside the mall, they may be unable to receive every transmission, which is modeled as erasures. The recommendation system must aim to achieve as high a satisfaction with the advertisements as possible under the specified bandwidth constraints.

A. Related work

The trade-off between bandwidth and user satisfaction in wireless recommendation systems has only recently been explored. The bandwidth-aware recommendation problem was introduced in [5], where the problem was formulated with just a single recommendation round. The server is presented with a set of users who may have some side information and must find the optimal message to broadcast, such that the users achieve the highest possible level of satisfaction. The authors show the connection to index coding [3] and pliable index coding [6]. Like these problems, it is also NP-hard [5]. The work in [7] studies the trade-off that occurs if more bandwidth is allocated to transmit multiple messages at each recommendation time. They show that the learning speed is proportional to the square of the available bandwidth. The joint design of learning and broadcasting in recommendation systems is considered in [8], focusing on an online recommendation strategy that first learns (explores) user preferences and later recommends (exploits) based on the found preferences.

B. Contributions

We study potential advantages of coded broadcast transmissions in recommendation systems. Contrary to prior work, we include erasures in our broadcast channel model and utilize them as a source of side information. In this setting, we:

- discuss when coding may benefit the exploration phase;
- characterize how much coding could change the gain in the exploitation phase;
- determine which reward structure will benefit most from coding in the exploitation phase;
- provide explicit checks to determine whether coding is worthwhile in the exploitation phase.

II. Preliminaries

A recommendation system must recommend content of K types to N users. Users can have different preferences, so a recommendation from content type k to user n results in an (expected) reward $\mu_{n,k} \geq 0$. These define the preference matrix $\mu \in \mathbb{R}^{N\times K}$. We assume that any amount of content from each type can be generated, e.g., if users have a strong preferences for content about shoes, we can keep generating such advertisements. Each recommendation is considered a binary vector of l bits, which can be transmitted in one time slot. Coded messages will be the bitwise exclusive-or of messages. This setup fits the multi-armed bandit framework [2], with the

978-1-7281-7122-7/21/$31.00 ©2021 IEEE

system operating in two stages: exploration and exploitation. At each time slot, it chooses one message to broadcast and observes the users' reactions. In the exploration phase, the system serves users messages from all the content types to determine their preferences. When the system has obtained a satisfactory estimate of the preferences, it switches to the exploitation phase, where at each time slot it chooses the message that maximizes the expected reward. We treat the phases independently to determine the effects of coded broadcasting on each. The system's goal is to minimize the regret R over T transmissions, defined as the difference between each user always receiving their most preferred message and what is achieved with (coded) broadcasting[1]:

$$R(T) = \mathbb{E}_e\left[\sum_{t=1}^{T}\left(\sum_{n=1}^{N}\max_k \mathbb{E}\left[\rho_{n,k}\right]e_{n,t} - \sum_{(n,k)\in A_t}^{N}\mathbb{E}\left[\rho_{n,k}\right]e_{n,t}\right)\right] \quad (1)$$

where $\rho_{n,k}$ is the random variable for the reward from recommending type k to user n and $e_{n,t} \in \{0,1\}$ indicates whether user n experiences an erasure at time t or not. A value of 1 means that the user hears the broadcast. A_t captures the action taken by the system at time t. The regret can be split as

$$R(T) = R_L(T) + R_{BC}(T), \quad (2)$$

where R_L is the regret incurred due to the exploration stage and R_{BC} is the regret due to broadcasting, which largely depends on the rewards structure [7]. The best case is if all users desire the same content, since then a single broadcast transmission will achieve the maximum reward. On the other hand, if each user has zero reward for all but one content type and all users prefer different types, then the broadcast regret is large, since each recommendation can satisfy only one user.

Our system will only count messages that can be immediately decoded by the recipients, we do not utilize undecodable transmissions as side information: If a user receives a message that cannot be decoded immediately, it is discarded. We do not make restrictions to require fairness, so it is plausible that our system could keep recommending from a single content type due to the preferences of just one user, even if all other users are not interested in this content. The system only seeks to maximize the overall reward, not satisfy everyone equally.

For our channel model, we assume a broadcast erasure channel (BEC) with symmetric erasure probabilities. Thus, all users have probability $1-\varepsilon$ of receiving any broadcast transmission and probability ε of an erasure. Both the capacity region and achievability schemes for this channel is known [9], [10][2]. We restate the capacity results here. Let π be any permutation of the numbers from 1 to N, which we denote by $[N]$. A rate tuple is defined as $r = (r_1, r_2, ..., r_N, r_{BC})$, where we can select whichever content to recommend to user i at rate $0 \le r_i \le 1$ messages per transmission. r_{BC} is the rate at which it is possible to broadcast common information to all users. Any achievable rate tuple must satisfy $N!$ inequalities, each defined by a specific permutation π:

[1]This formulation is similar to [8], but averages over the erasure pattern.

[2]The capacity region is not known in general for non-symmetric erasures.

$$\mathcal{R}_\pi = \left\{ r \ge 0 : \frac{r_{\mathrm{BC}}}{1-\varepsilon} + \sum_{i\in[N]} \frac{r_{\pi(i)}}{1-\varepsilon^i} \le 1 \right\}. \quad (3)$$

The intersection of these $N!$ regions define the capacity region,

$$C_N = \bigcap_{\forall \pi \in \mathcal{P}_N} \mathcal{R}_\pi. \quad (4)$$

In particular, for two users the capacity region for the BEC with symmetric erasure probabilities is [11]:

$$\begin{aligned} C_2 = \Big\{ r \ge 0 : &\frac{r_1}{1-\varepsilon} + \frac{r_2}{1-\varepsilon^2} + \frac{r_{\mathrm{BC}}}{1-\varepsilon} \le 1, \\ &\frac{r_1}{1-\varepsilon^2} + \frac{r_2}{1-\varepsilon} + \frac{r_{\mathrm{BC}}}{1-\varepsilon} \le 1 \Big\}. \end{aligned} \quad (5)$$

This region is a polyhedron. For our use case, there are only two interesting vertices to consider for optimal operation:

$$V_{BC} = (r_1, r_2, r_{\mathrm{BC}}) = (0, 0, 1-\varepsilon), \quad (6)$$

$$V_C = (r_1, r_2, r_{\mathrm{BC}}) = (r_{\mathrm{C}}(2,\varepsilon), r_{\mathrm{C}}(2,\varepsilon), 0), \quad (7)$$

where

$$r_{\mathrm{C}}(2,\varepsilon) = \left(\frac{1}{1-\varepsilon} + \frac{1}{1-\varepsilon^2}\right)^{-1}. \quad (8)$$

The remaining nonzero vertices, $(1-\varepsilon, 0, 0)$ and $(0, 1-\varepsilon, 0)$, have smaller throughputs than V_{BC}, since they target only one user with the same rate. If a new message of the same content type is broadcast at every time slot, we expect each user to receive a fraction $1-\varepsilon$ of the recommendations. Thus, V_{BC} can be obtained without coding. This is not the case for V_C. This vertex corresponds to transmitting at rate $r_{\mathrm{C}}(2,\varepsilon)$ to both users simultaneously, where the content types may be different. This requires coding. When transmitting at rate $r_{\mathrm{C}}(2,\varepsilon)$ of content type k to user 1, we expect user 2 to overhear $(1-\varepsilon)r_{\mathrm{C}}(2,\varepsilon)$ of these transmissions. These will both be stored for use as side information and served to user 2, where they will also generate a reward if the user has a positive preference for that content type. One specific vertex is always associated with the largest gain in throughput when targeting individual users with independent information. For N users, it is the vertex where each user receives information at rate

$$r_{\mathrm{C}}(N,\varepsilon) = \left(\sum_{n=1}^{N} \frac{1}{1-\varepsilon^n}\right)^{-1}, \quad (9)$$

i.e., at $(r_1, \ldots, r_N, r_{BC}) = (r_{\mathrm{C}}(N,\varepsilon), \ldots, r_{\mathrm{C}}(N,\varepsilon), 0)$.

III. Coding in the Exploration Phase

In this section we discuss opportunities for coding to improve the exploration phase.

A. Potential for arbitrarily large gains

In this example, we see that the gains of employing coding can be theoretically unbounded. An example exploration phase for two users with two content types is shown in Table I, where the rewards are exactly the expected values. The cumulative reward for the four transmissions without coding is

$$\mu_{BC}^{\text{total}} = 2\mu_{1,1} + 2\mu_{2,2} + \mu_{1,2} + \mu_{2,1}, \quad (10)$$

TABLE I

EXAMPLE OF EXPLORATION WITHOUT CODING (LEFT) AND WITH CODING (RIGHT). $m_i^{(j)}$ IS MESSAGE (RECOMMENDATION) j OF CONTENT TYPE i. AN ERASURE PATTERN (e_1, e_2) INDICATES THAT USER i RECEIVES THE MESSAGE IF $e_i = 1$ AND IT WAS ERASED IF $e_i = 0$.

Known	$\begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}$	$\begin{bmatrix} \mu_{1,1} & ? \\ ? & ? \end{bmatrix}$	$\begin{bmatrix} \mu_{1,1} & ? \\ ? & \mu_{2,2} \end{bmatrix}$	$\begin{bmatrix} \mu_{1,1} & \mu_{1,2} \\ ? & \mu_{2,2} \end{bmatrix}$
Transmits	$m_1^{(1)}$	$m_2^{(1)}$	$m_1^{(2)}$	$m_2^{(2)}$
Erasures	(1, 0)	(0, 1)	(1, 1)	(1, 1)
Reward	$\mu_{1,1}$	$\mu_{2,2}$	$\mu_{1,1},\ \mu_{1,2}$	$\mu_{2,2},\ \mu_{2,1}$

Known	$\begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}$	$\begin{bmatrix} \mu_{1,1} & ? \\ ? & ? \end{bmatrix}$	$\begin{bmatrix} \mu_{1,1} & ? \\ ? & \mu_{2,2} \end{bmatrix}$	$\begin{bmatrix} \mu_{1,1} & \mu_{1,2} \\ \mu_{2,1} & \mu_{2,2} \end{bmatrix}$
Transmits	$m_1^{(1)}$	$m_2^{(1)}$	$m_1^{(1)} \oplus m_2^{(1)}$	$m_i^{(2)} : i = \operatorname{argmax}_j \mu_{1,j} + \mu_{2,j}$
Erasures	(1, 0)	(0, 1)	(1, 1)	(1, 1)
Reward	$\mu_{1,1}$	$\mu_{2,2}$	$\mu_{2,1},\ \mu_{1,2}$	$\mu_{1,i},\ \mu_{2,i}$

whereas with coding it is

$$\mu_C^{\text{total}} = \mu_{1,1} + \mu_{2,2} + \mu_{2,1} + \mu_{1,2} + \max_i(\mu_{1,i} + \mu_{2,i}). \quad (11)$$

This means that the regret from not coding is

$$\begin{aligned} R = \mu_{\text{C}}^{\text{total}} - \mu_{\text{BC}}^{\text{total}} &= \max_i(\mu_{1,i} + \mu_{2,i}) - \mu_{1,1} - \mu_{2,2} \\ &= \max\{\mu_{2,1} - \mu_{2,2},\ \mu_{1,2} - \mu_{1,1}\}. \quad (12) \end{aligned}$$

Letting $\mu_{2,1} \to \infty$ or $\mu_{1,2} \to \infty$, we see that the regret could be arbitrarily large. This implies an unbounded gain, as

$$\frac{\mu_{\text{C}}^{\text{total}}}{\mu_{\text{BC}}^{\text{total}}} = \frac{R + \mu_{\text{BC}}^{\text{total}}}{\mu_{\text{BC}}^{\text{total}}} \quad (13)$$

is also unbounded when $\mu_{2,1} \to \infty$ or $\mu_{1,2} \to \infty$. Similar examples exist for larger numbers of users and content types.

B. Exploring different content types for different users

Consider a preference matrix where the users' preferences are so different that each transmission can only satisfy one user. For instance, we might have expected rewards

$$\mu = \begin{bmatrix} \mu_{1,1} & 0 \\ \vdots & \vdots \\ \mu_{1,K} & 0 \\ 0 & \mu_{2,K+1} \\ \vdots & \vdots \\ 0 & \mu_{2,2K} \end{bmatrix}, \quad (14)$$

so we need to determine which of the first K content types to serve to user 1 and which of the following K content types to serve to user 2 or, indeed, if the benefits are largest by being unfair and just satisfying one of the users. Since rewards may be random, we need to see several realizations to reliably estimate the expected reward. We estimate that we need to see i realizations of a user's reward for each content type to have a good enough approximation of the preference matrix. For broadcasting, we expect to transmit $i/(1-\varepsilon)$ times before a user has received i messages. With K types for each of the $N = 2$ users, we thus need $NKi/(1-\varepsilon)$ transmissions. With coding, we can transmit at rate $r_{\text{C}}(N, \varepsilon)$ to each user simultaneously, giving us useful information at a rate of $r_{\text{C}}(N, \varepsilon)$ for each user. As we need to get Ki messages to each of the users, this means an expected $Ki/r_{\text{C}}(N, \varepsilon)$ transmissions.

To put this in perspective, if $\varepsilon = 1/2,\ K = 3,\ i = 5$, then to see the required $NKi = 30$ realizations, we expect to need 60 transmissions with pure broadcasting against 50 with coding, saving an expected 10 broadcast transmissions. If we need to distinguish more content types, the number of saved transmissions is greater. Thus, the learning phase will be shorter and an optimal recommendation strategy can be identified sooner.

C. Equal exploration of all combinations

Consider a simple exploration strategy that requires seeing at least i rewards for every content type for each of the users. It is straightforward to start the exploration by broadcasting uncoded content from each of the content types. Once any single user n has received the i realizations required of a particular content type, say type k, we may see opportunities for coding, since we are no longer interested in serving type k to user n, but we could still desire realizations of other users' rewards for type k. Plain broadcasting could then lead to over-exploration of type k at user n. In effect, the users' exploration preferences are now different, and an argument may be made as in the previous section. However, if i is large, then we can expect that the other users will already have received close to i messages of type k with high probability, since the erasure probability is the same for all users. This means that, even in the best case, only few coded transmissions would be beneficial near the end of the exploration phase, while the majority of transmissions can be uncoded.

D. Empowering exploration with coding

In essence, coding brings greater control over which content types different users receive, which is useful when we want to serve diverse content to the users. We expect that it is possible to design a sophisticated transmission strategy that reduces the regret of the exploration phase. In particular, it could be relevant to explore preferences for a particular content type unequally across users according to their relative magnitudes. Thus, the exploration strategy should aim to dynamically adapt, eliminating or reducing the exploration of certain user-content pairs once they are determined to be suboptimal. While coding does bring more control, we still cannot completely stop exploring a content type only for a subset of the users. This is due to the broadcast nature of the wireless channel: If we explore a content type for one user, then it is inevitable that other users will overhear some of the transmissions.

IV. CODING IN THE EXPLOITATION PHASE

For the analysis, we assume perfect knowledge of μ, i.e., that we have successfully learned the preferences of users similar to the current ones. If the users are the same as in the exploration phase, then they may have accumulated some useful side information that have not yet been utilized. This allows us to employ coding until all the side information is depleted, which we will do if our knowledge of μ indicates that this will result in a lower regret. This could be considered a separate state between exploration and exploitation, where the left over side information is utilized. This section considers steady state exploitation, i.e., when users do not start out with side information, so coding operations are only possible if some transmissions cause a subset of users to accumulate side information that can be utilized for future transmissions. We are now able to make a well-informed decision for every broadcast transmission. In general, coding will be worthwhile if the expected reward of coded broadcasting is greater than the expected reward of uncoded broadcasting, i.e.,

$$\mu_{\mathrm{BC}}^{\mathrm{total}} < \mu_{\mathrm{C}}^{\mathrm{total}}. \tag{15}$$

When broadcasting to N users, at least one content type will be optimal. Transmitting a new message from this content type at every opportunity results in a total expected reward of

$$\mu_{\mathrm{BC}}^{\mathrm{total}} = r_{\mathrm{BC}} \max_i \sum_{n=1}^{N} \mu_{n,i}. \tag{16}$$

When transmitting coded messages, we are able to target each user with messages of different content types. The first time a message is recommended it is broadcast uncoded, since no user can have it as side-information. When transmitting content type i to user n with rate r_n, we expect all other users to overhear the messages of type i with rate $r_n(1-\varepsilon)$, which both generates rewards and can be used as side information. Some users may have low rewards relative to the others, so it may be better to focus all transmissions on a subset of size $N^* \leq N$ of the users. Assuming, without loss of generality, that the N^* users with the greatest expected reward are users 1 to N^*, the expected total reward averaged over erasures becomes

$$\begin{aligned}\mu_{\mathrm{C}}^{\mathrm{total}} &= \sum_{n=1}^{N^*} r_n \max_i \left(\mu_{n,i} + (1-\varepsilon) \sum_{j=1, j\neq n}^{N} \mu_{j,i}\right) \\ &= r_{\mathrm{C}}(N^*, \varepsilon) \sum_{n=1}^{N^*} \max_i \left(\mu_{n,i} + (1-\varepsilon) \sum_{j=1, j\neq n}^{N} \mu_{j,i}\right),\end{aligned} \tag{17}$$

when we use coding to operate at the vertex of the rate region that enables the largest throughput to the N^* users,

$$\begin{aligned}&(r_1, \ldots, r_{N^*}, r_{N^*+1}, \ldots, r_N, r_{BC}) \qquad (18)\\ &\quad = (r_{\mathrm{C}}(N^*, \varepsilon), \ldots, r_{\mathrm{C}}(N^*, \varepsilon), 0, \ldots, 0, 0). \qquad (19)\end{aligned}$$

A. Analysis for 2 users

Without loss of generality, assume that 3 content types are relevant in the exploitation phase. One is optimal when transmitting to user 1, one optimal when transmitting to user 2, and one optimal when broadcasting. Their rewards are:

$$\mu_1' \triangleq \max_i \left(\mu_{1,i} + (1-\varepsilon)\mu_{2,i}\right) \tag{20}$$

$$\mu_2' \triangleq \max_i \left(\mu_{2,i} + (1-\varepsilon)\mu_{1,i}\right) \tag{21}$$

$$\mu_{\mathrm{BC}}' \triangleq \max_i \left(\mu_{1,i} + \mu_{2,i}\right). \tag{22}$$

We arrange these in the relevant preference matrix,

$$\mu' \triangleq \begin{bmatrix} \mu_{1,1} & \mu_{2,1} \\ \mu_{1,2} & \mu_{2,2} \\ \mu_{1,BC} & \mu_{2,BC} \end{bmatrix}, \tag{23}$$

so if we want to transmit to user 1 only, we would transmit from the type corresponding to the top row. If we want to target both users with the broadcast, we should transmit from the type corresponding to the bottom row. Rows are not guaranteed to be unique, and may indeed correspond to the same content type in practice.

Theorem 1. *For two users and a BEC with symmetric erasure probability ε, coding is worthwhile if and only if the following inequalities hold for the rewards:*

$$\begin{aligned}\mu_1' &< (1+\varepsilon)\mu_2' \\ \mu_2' &< (1+\varepsilon)\mu_1' \\ \mu_{BC}' &< \frac{1+\varepsilon}{2+\varepsilon}(\mu_1' + \mu_2')\end{aligned}$$

Proof. We formulate the following linear program (LP):

$$\begin{aligned}\underset{r}{\text{maximize}} \quad & r_1\mu_1' + r_2\mu_2' + r_{\mathrm{BC}}\mu_{\mathrm{BC}}' && (24)\\ \text{subject to} \quad & (r_1, r_2, r_{BC}) \in C_2 && (25)\end{aligned}$$

The dual of this program is

$$\begin{aligned}\underset{\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5}{\text{maximize}} \quad & -\lambda_1 - \lambda_2 && (26)\\ \text{subject to} \quad & \frac{1}{1-\varepsilon}\lambda_1 + \frac{1}{1-\varepsilon^2}\lambda_2 - \lambda_3 = \mu_1' && (27)\\ & \frac{1}{1-\varepsilon^2}\lambda_1 + \frac{1}{1-\varepsilon}\lambda_2 - \lambda_4 = \mu_2' && (28)\\ & \frac{1}{1-\varepsilon}\lambda_1 + \frac{1}{1-\varepsilon}\lambda_2 - \lambda_5 = \mu_{\mathrm{BC}}' && (29)\\ & \lambda_i \geq 0, \quad \forall i \in \{1,2,3,4,5\}. && (30)\end{aligned}$$

We are interested in determining whether coding is worthwhile. If this is the case, we must operate at the vertex where $r_{12} = 0$, $r_1 = r_2 \neq 0$. Thus, at this vertex $\lambda_3 = \lambda_4 = 0$ by complementary slackness [12]. Thus, we can solve the following three equations in three unknowns and check if $\lambda_i, i \in \{1,2,5\}$ are non-negative:

$$\mu_1' = \frac{\lambda_1}{1-\varepsilon} + \frac{\lambda_2}{1-\varepsilon^2} \tag{31}$$

$$\mu_2' = \frac{\lambda_1}{1-\varepsilon^2} + \frac{\lambda_2}{1-\varepsilon} \tag{32}$$

$$\mu_{\mathrm{BC}}' = \frac{\lambda_1}{1-\varepsilon} + \frac{\lambda_2}{1-\varepsilon} - \lambda_5. \tag{33}$$

Therefore, to have $\lambda_1, \lambda_2, \lambda_5 \geq 0$, we must have:

$$\mu_1' < (1+\varepsilon)\mu_2' \tag{34}$$

$$\mu_2' < (1+\varepsilon)\mu_1' \tag{35}$$

$$\mu_{\mathrm{BC}}' < \frac{1+\varepsilon}{2+\varepsilon}(\mu_1' + \mu_2'). \tag{36}$$

When these inequalities are satisfied, we will have $\lambda_i \geq 0, \forall i$, i.e., a proof of optimality for the LP, which is a necessary and sufficient condition for the coding vertex to be optimal. □

From the equations of Theorem 1, we can gain some insight about when coding is worthwhile. The first two equations requires balance between the value of transmitting to each user. The third equation concerns the relationship between the best broadcasting and individual preferences. If one of the inequalities does not hold, then plain broadcasting can achieve at least the same reward as with coding.

Theorem 2. *Using coding to transmit at the coding vertex for two users, V_C, can change the reward by at most*

$$r_C(2,\varepsilon)\frac{2-\varepsilon}{1-\varepsilon} \le \frac{r_1\mu_1' + r_2\mu_2'}{r_{BC}\mu_{BC}'} \le \frac{1+\varepsilon}{1+\varepsilon/2}.$$

Proof. By definition, we must have

$$\mu_1' = \mu_{1,1} + (1-\varepsilon)\mu_{2,1} \ge \mu_{1,\mathrm{BC}} + (1-\varepsilon)\mu_{2,\mathrm{BC}} \quad (37)$$
$$\mu_2' = \mu_{2,2} + (1-\varepsilon)\mu_{1,2} \ge \mu_{2,\mathrm{BC}} + (1-\varepsilon)\mu_{1,\mathrm{BC}}, \quad (38)$$

so we can lower bound the change as:

$$\begin{aligned}
&\frac{r_1\mu_1' + r_2\mu_2'}{r_{\mathrm{BC}}\mu_{\mathrm{BC}}'} && (39)\\
&\ge \frac{r_1\left(\mu_{1,\mathrm{BC}} + (1-\varepsilon)\mu_{2,\mathrm{BC}}\right) + r_2\left(\mu_{2,\mathrm{BC}} + (1-\varepsilon)\mu_{1,\mathrm{BC}}\right)}{r_{\mathrm{BC}}\mu_{\mathrm{BC}}'} && (40)\\
&= \frac{r_{\mathrm{C}}(2,\varepsilon)(2-\varepsilon)\left(\mu_{1,\mathrm{BC}} + \mu_{2,\mathrm{BC}}\right)}{r_{\mathrm{BC}}\mu_{\mathrm{BC}}'} && (41)\\
&= \frac{r_{\mathrm{C}}(2,\varepsilon)(2-\varepsilon)}{r_{\mathrm{BC}}} && (42)\\
&= r_{\mathrm{C}}(2,\varepsilon)\frac{(2-\varepsilon)}{(1-\varepsilon)}. && (43)
\end{aligned}$$

The upper bound is obtained by noting that $r_{\mathrm{BC}} \ge r_i,\ i \in \{1,2\}$ and applying (36). Indeed,

$$\begin{aligned}
\frac{r_1\mu_1' + r_2\mu_2'}{r_{\mathrm{BC}}\mu_{\mathrm{BC}}'} &\le \frac{r_{\mathrm{BC}}\mu_1' + r_{\mathrm{BC}}\mu_2'}{r_{\mathrm{BC}}\mu_{\mathrm{BC}}'} && (44)\\
&= \frac{\mu_1' + \mu_2'}{\mu_{\mathrm{BC}}'} && (45)\\
&\le \frac{1+\varepsilon}{1+\varepsilon/2}, && (46)
\end{aligned}$$

concluding the proof. □

Theorem 3. *For* 2 *users and a symmetric BEC, coding provides the greatest benefits when the users' preferences are as different as possible, i.e., the rewards form a (scaled) identity matrix.*

Proof. We formulate an optimization problem to maximize the gain of coding. We know that we have at most three relevant content types to broadcast, one for targeting each of the users and one to use when broadcasting, as collected in μ', so

$$\begin{aligned}
\underset{\mu}{\text{maximize}} \quad & r_{\mathrm{C}}(2,\varepsilon)\left(\mu_1' + \mu_2'\right) - r_{\mathrm{BC}}\mu_{\mathrm{BC}}' && (47)\\
\text{subject to} \quad & \mu_{1,BC} + \mu_{2,BC} \ge \mu_{1,1} + \mu_{2,1} && (48)\\
& \mu_{1,BC} + \mu_{2,BC} \ge \mu_{1,2} + \mu_{2,2} && (49)\\
& 0 \le \mu_{i,j} \le 1, \quad i,j \in \{1,2\}. && (50)
\end{aligned}$$

First note that at least one of the two first constraints must hold with equality, since otherwise the objective could be improved by reducing the value of the broadcast message. Thus, the broadcast-preferred message must have the same reward as broadcasting one of the two content types. We assume without loss of generality that this is content type 1. This allows us to optimize over a 2×2 matrix only, as the third row will be a copy of the first. This simplifies the problem to

$$\begin{aligned}
\underset{\mu}{\text{maximize}} \quad & r_{\mathrm{C}}(2,\varepsilon)\left(\mu_1' + \mu_2'\right) - r_{\mathrm{BC}}(\mu_{1,1} + \mu_{2,1}) && (51)\\
\text{subject to} \quad & \mu_{1,1} + \mu_{2,1} \ge \mu_{1,2} + \mu_{2,2} && (52)\\
& \mu_{i,j} \ge 0, \quad i,j \in \{1,2\} && (53)\\
& \mu_{i,j} \le 1, \quad i,j \in \{1,2\} && (54)
\end{aligned}$$

from where it is easy to arrive at the equivalent dual problem,

$$\begin{aligned}
\underset{\lambda}{\text{maximize}} \quad & -\lambda_6 - \lambda_7 - \lambda_8 - \lambda_9 && (55)\\
\text{subject to} \quad & -r_{\mathrm{C}}(2,\varepsilon) + r_{\mathrm{BC}} - \lambda_1 - \lambda_2 + \lambda_6 = 0 && (56)\\
& -(1-\varepsilon)r_{\mathrm{C}}(2,\varepsilon) + r_{\mathrm{BC}} - \lambda_1 - \lambda_3 + \lambda_7 = 0 && (57)\\
& -(1-\varepsilon)r_{\mathrm{C}}(2,\varepsilon) + \lambda_1 - \lambda_4 + \lambda_8 = 0 && (58)\\
& -r_{\mathrm{C}}(2,\varepsilon) + \lambda_1 - \lambda_5 + \lambda_9 = 0 && (59)\\
& \lambda_i \ge 0,\ \forall i. && (60)
\end{aligned}$$

If the reward matrix is the identity, we have equality in the first primal constraint as well as the constraints $\mu_{i,j} = 1$ for $i = j$ and $\mu_{i,j} = 0$ for $i \ne j$, while the remaining constraints remains inactive. Thus, our dual must have $\lambda_i = 0$ for $i \in \{2,5,7,8\}$ by complementary slackness. We now look for a certificate of optimality. The following configuration satisfies all the constraints for any ε:

$$\begin{aligned}
\lambda_1 &= r_{\mathrm{C}}(2,\varepsilon) && (61)\\
\lambda_3 &= r_{\mathrm{BC}} - (2-\varepsilon)r_{\mathrm{C}}(2,\varepsilon) && (62)\\
\lambda_4 &= \varepsilon r_{\mathrm{C}}(2,\varepsilon) && (63)\\
\lambda_6 &= 2r_{\mathrm{C}}(2,\varepsilon) - r_{\mathrm{BC}} && (64)\\
\lambda_9 &= 0 && (65)
\end{aligned}$$

so it can serve as a proof of optimality for the vertex corresponding to the identity reward matrix. Thus, we conclude that this reward structure achieves the largest gain from coding. □

B. N users

We now turn our attention to recommendation systems that are broadcasting to more than two users simultaneously. First we note that, although we did not prove it here, Theorem 3 can be extended to the case of N users.

Theorem 4. *A coding-based transmission scheme can reduce the regret per transmission of the plain broadcasting by as much as a factor of* $1/(1-\varepsilon)$.

Proof. We know from Theorem 3 that the greatest difference between broadcasting with and without coding is when we have an identity reward matrix for two users. This result can be extended to N users. Let N' be the value such that $\varepsilon^{N'} \approx 0$

with the approximation as good as desired. We can find the greatest reduction in regret as the number of users grows large:

$$\sum_{n=1}^{N} r_C(N,\varepsilon)\mu_n' = N r_C(N,\varepsilon) \tag{66}$$

$$= N\left(\sum_{n=1}^{N}\frac{1}{1-\varepsilon^n}\right)^{-1} \tag{67}$$

$$= N\left(\sum_{n=N'}^{N}\frac{1}{1-\varepsilon^n}+\sum_{n=1}^{N'}\frac{1}{1-\varepsilon^n}\right)^{-1} \tag{68}$$

$$\approx N\left(\sum_{n=N'}^{N}1+\sum_{n=1}^{N'}\frac{1}{1-\varepsilon^n}\right)^{-1} \tag{69}$$

$$= N\left(N-N'+\sum_{n=1}^{N'}\frac{1}{1-\varepsilon^n}\right)^{-1} \tag{70}$$

$$= \frac{N}{N+C(N',\varepsilon)} \xrightarrow{N\to\infty} 1, \tag{71}$$

where $C(N',\varepsilon)$ collects the terms not changing with N. For broadcasting, the reward obtained with an identity reward matrix is $1-\varepsilon$, no matter which content type is chosen. Thus, the largest possible gain for coding instead of plain broadcasting is $1/(1-\varepsilon)$ for each transmission. This factor remains the same if the rewards are scaled. □

Conceptually, it is clear that the gain depends on the erasure probability: more erasures provide greater opportunity for users to overhear messages, which can be utilized as side information. Theorem 4 implies that the regret over T time slots can be improved by at most $T\varepsilon$ for the identity reward matrix, increasing the cumulative reward from $T(1-\varepsilon)$ to T.

A concrete way to check whether coding will improve the exploitation phase for N users for a particular preference structure unfortunately does not have as concise a representation as the two-user case of Theorem 1. Our best option is to go back to Eq. (15). It is always easy to find the best value of plain broadcasting, $\mu_{\text{BC}}^{\text{total}}$, by checking what the expected reward would be for each content type. On the other hand, since nothing requires the recommendation system to be fair, it has no incentive to code to serve information to all N users. Indeed, it might be the case that it is better to ignore some users and simply generate recommendations that only is intended to satisfy a subset of the users. To determine whether coding is worthwhile or not, we first determine the value of targeting each user. That is, we find the value

$$\mu_n' = \max_i\left(\mu_{n,i}+(1-\varepsilon)\sum_{j=1,j\neq n}^{N}\mu_{j,i}\right) \tag{72}$$

for each user. We then sort the N users according to this, and use the top $N^* \leq N$ of these to calculate the expected reward of targeting N^* users according to Eq. (17). If any choice of N^* results in $\mu_{\text{C}}^{\text{total}} > \mu_{\text{BC}}^{\text{total}}$, then we may conclude that coding is beneficial and should be done targeting the $N^\star = \operatorname{argmax}_{N^*}\mu_{\text{C}}^{\text{total}}$ users with highest expected rewards.

V. Conclusion

In this paper, we have explored the interaction between coding techniques and recommendation systems using reinforcement learning over wireless broadcast channels with erasures. In particular, we have shown that coding can provide gains in both the exploration and exploitation phase, but this is not certain to be the case for all systems. These opportunities arise because coding enables the recommendation system to operate in more of the broadcast erasure channel's capacity region. We have presented methods for analyzing the preference matrix of a particular recommendation system to assess whether coding is worthwhile or not, in form of a closed form expression for recommendations to two users and a procedure for an arbitrary number of users. We have also shown that coding improves the exploitation phase the most when users' preferences are as different as possible, which is also when the broadcast regret is greatest. In theory, the improvements could be arbitrarily large in the exploration phase, but are limited in the exploitation phase. We leave system designers to determine whether the advantages will be worth the added complexity of coding.

Acknowledgments

This work was partially financed by the SCALE-IoT project (Grant No. DFF-7026-00042B) granted by the Danish Council for Independent Research, the AUFF Starting Grant AUFF-2017-FLS-7-1, and Aarhus University's DIGIT Centre. This work was also supported in part by the Hong Kong RGC grant ECS 21212419. Finally, the work was supported by NSF Award 2007714 and UC-NL grant LFR-18-548554.

References

[1] P. Resnick and H. R. Varian, "Recommender Systems," *Commun. ACM*, vol. 40, no. 3, pp. 56–58, 1997.
[2] R. S. Sutton and A. G. Barto, *Reinforcement Learning - An Introduction*, 2018.
[3] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index Coding With Side Information," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1479–1494, 2011.
[4] S. El Rouayheb, A. Sprintson, and C. Georghiades, "On the Index Coding Problem and Its Relation to Network Coding and Matroid Theory," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3187–3195, 2010.
[5] L. Song and C. Fragouli, "Making Recommendations Bandwidth Aware," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7031–7050, 2018.
[6] S. Brahma and C. Fragouli, "Pliable Index Coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 6192–6203, 2015.
[7] L. Song, C. Fragouli, and D. Shah, "Interactions Between Learning and Broadcasting in Wireless Recommendation Systems," in *IEEE ISIT*, 2019.
[8] ——, "Recommender Systems over Wireless: Challenges and Opportunities," in *IEEE ITW*, 2018.
[9] M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Multiuser Broadcast Erasure Channel With Feedback—Capacity and Algorithms," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5779–5804, 2013.
[10] C. Wang, "On the Capacity of 1-to-K Broadcast Packet Erasure Channels With Channel Output Feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 931–956, 2012.
[11] L. Georgiadis and L. Tassiulas, "Broadcast erasure channel with feedback - Capacity and algorithms," in *Workshop on Network Coding, Theory and Applications*, 2009.
[12] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.