固定业务定价下考虑移动用户兴趣分析的能效多播推送研究

1. 移动用户兴趣图谱分析+用户兴趣指数条件概率分布函数拟合

这块不是项目的研究重点，我们假设通过移动用户兴趣分析与建模，网关处得到其覆盖范围内K个用户关于M个业务的兴趣图谱$\mathbf{Q}=[q_{k,m}]_{1\le k\le K,1\le m\le M}$

$$\mathbf{Q}=\begin{bmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,M} \\ q_{2,1} & q_{2,2} & \cdots & q_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ q_{K,1} & q_{K,2} & \cdots & q_{K,M} \end{bmatrix} \tag{1}$$

其中元素$q_{k,i}\in[0,1]$表示用户k对业务m的归一化兴趣指数。

通过实际测量，得到不同移动用户兴趣分析模型下，给定用户k对业务m的兴趣指数分析值$q_{k,i}$前提下，其实际兴趣指数的条件概率分布函数$f_{k,m}(x\,|\,q_{k,m})$。

一般来说，该条件概率分布函数应该满足:

- 定义域为$x\in[0,1]$： $f_{k,m}(x\,|\,q_{k,m})>0,\ 0\le x\le 1$； $f_{k,m}(x\,|\,q_{k,m})=0,\ x<0 \text{ or } x>1$
- 存在一个极值点$\mu_{k,m}=[q_{k,m}+\alpha_{k,m}]_0^1$，其中$\alpha_{k,m}$为极值点与分析值$q_{k,i}$间的偏差$\alpha_{k,m}$，可正可负
- 从极值点往端点$x=0$与$x=1$， $f_{k,m}(x\,|\,q_{k,m})$应单调减小
- 均值$E_{k,m}=\int_{x=0}^{1}xf_{k,m}(x\,|\,q_{k,m})\mathrm{d}x$逼近分析值$q_{k,i}$

为此，项目拟使用高斯分布来模拟该条件概率分布函数，具体表达式如下

$$f_{k,m}(x\,|\,q_{k,m})=\frac{1}{S}\times\frac{1}{\sqrt{2\pi}\sigma_{k,m}}\exp\left\{-\frac{(x-\mu_{k,m})^2}{2\sigma_{k,m}^2}\right\} \tag{2}$$

其中 $\mu_{k,m}=[q_{k,m}+\alpha]_0^1$ 是设置的极值点， $\sigma_{k,m}$ 式设置的标准差， $S=\frac{1}{\sqrt{2\pi}\sigma_{k,m}}\int_{x=0}^{1}\exp\{-(x-\mu_{k,m})^2\big/2\sigma_{k,m}^2\}\mathrm{d}x$ 为归一化参数。

在该拟合条件概率函数中，参数$\mu_{k,m}(\alpha_{k,m})$， $\sigma_{k,m}$是需要优化的变量。项目拟通过实际测量进行参数优化，具体步骤如下。

1> 根据业务特征将业务分为m类，根据用户特征将用户分为n类

我们假设同一类型的用户对同一类型的业务的兴趣指数条件分布具有类似特征，这样简化了拟合工作量，只需要拟合$n\times m$种条件概率分布函数。

2> 测试集获取

在完成移动用户兴趣分析建模后，邀请不同类型用户对不同类型业务进行兴趣打分，获取测试参数集合。

3> 参数优化

针对不同n，m以及$q_{k,m}$（$q_{k,m}$离散化处理），统计属于用户类型n的用户对业务类型m的兴趣打分，并利用 MMSE 准则，优化条件概率模型中的参数进行优化。

为了简化优化工作量，可以做一些假设，如条件概率的极值点与分析值的偏差$\alpha_{k,m}$可以设为常数或者简单分段设置。

2. 高效多播业务推送方案设计

目标：给定业务集合 $\mathcal{M}_{\text{tot}} = \{1,2,...,M\}$ 以及用户集合 $\mathcal{K}_{\text{tot}} = \{1,2,...,K\}$，基于上述移动用户分析模型（包括移动用户兴趣图谱 $\mathbf{Q} = [q_{k,m}]$ 以及实际兴趣指数条件概率分布 $f_{k,m}(x \mid q_{k,m})$），针对每个业务，选择合适的多播推送用户集合 $\mathcal{K}_m \subseteq \mathcal{K}_{\text{tot}}$，并联合进行基站端资源分配，最大化业务传输效率。

2.1 问题建模

假设系统采用 OFDM 频谱接入方式，子载波带宽为 Δf Hz，集合为 $\mathcal{N} = \{1,2,...,N\}$，基站端的子载波分配与功率分配分别为 $\rho_{m,n}$, p_n。

此时，多播业务 m 的传输速率为

$$R_m = \Delta f \sum_{n=1}^{N} \rho_{n,m} \log_2(1 + p_n \min_{k \in \mathcal{K}_m} \gamma_{n,k}) \tag{3}$$

其中 $\gamma_{n,k} = H_{n,k} / N_0 \Delta f$。

系统的收益取决于两个因素：1）接收业务的用户数目，2）业务的传输速率，显然接收业务用户数越多，业务传输速率越大，系统收益越多。这里我们简单使用所有用户的和接收速率来表示系统的收益。

这样，业务 m 进行多播推送传输的平均期望收益可以表示为

$$U_m = R_m E[\sum_{k \in \mathcal{K}_m} s_{k,m}] = R_m \sum_{k \in \mathcal{K}_m} \varepsilon_{k,m} \tag{4}$$

其中 $s_{k,m} \in \{0,1\}$ 分别表示用户 k 是否愿意（有偿）接收业务 m，$E[\bullet]$ 表示期望运算，$\varepsilon_{k,m} = E[s_{k,m}] = \int_{x=0}^{1} x f_{k,m}(x \mid q_{k,m}) \mathrm{d}x$ 为用户 k 对业务 m 的兴趣指数条件期望值。

Remark 1：此时看似我们选择的多播推送集合越大越好，其实不然，因为有的用户信道差，将其加入多播推送集合会增加传输代价或者说降低单位代价可得到的收益。这点与多描述多播类似，不同的是每个用户关于业务接收都有一个兴趣指数分布函数。

进一步，我们分析用户接收不感兴趣业务的潜在收益损失。当业务推送不当，会造成用户对于系统用户兴趣分析算法准确性以及业务推送机制必要性的怀疑，降低用户对业务推送机制的好感，造成用户抵制系统的业务推送机制，严重降低系统的未来收益。因此收益函数可以修正为

$$U_m = R_m E[\sum_{k \in \mathcal{K}_m} s_{k,m}] - \beta_m E[\sum_{k \in \mathcal{K}_m} (1 - s_{k,m})] = \sum_{k \in \mathcal{K}_m} \left[R_m \varepsilon_{k,m} - \beta_m (1 - \varepsilon_{k,m}) \right] \tag{5}$$

这里 β_m 是由多播推送集合中用户不愿意接收该业务造成的收益损失，这里不考虑业务接收功率消耗与存储消耗，因为即使接收感兴趣业务，也需要消耗功率资源与存储资源，这些消耗将在后续业务传输代价中进行建模。

针对 β_m 的不同取值表达式，我们进行如下分析：

- β_m = 0：不存在惩罚，用户接收不感兴趣业务不会对自身造成损失

此时，基站传输业务 m 换取的收益取决于多播推送集合中用户实际愿意接收该业务创造的收益。从收益最大化角度出发，多播推送用户集合越大，收益越大。另一方面，传输业务 m 的可得传输速率只取决于播推送用户集合中用户的最差信道接收条件。

这样，考虑收益最大化原则，业务多播推送用户选择退化为多描述多播中多播用户选择问题，可以得到如下多播推送用户集合设置原则：

Lemma 1：

多播推送用户的选择只与用户的信道条件相关，与用户对业务的兴趣指数无关。换言之，最佳的多播推送用户集合一定是信道条件最好的前 k（$1 \le k \le K$）的用户集合。

- $\beta_m = \alpha R_m$：即用户接收不感兴趣业务的损失与接收数据大小成正比

该情况下，首先容易证明只有当 $\varepsilon_{k,m} \ge \alpha/(1+\alpha)$ 时，将用户 k 加入多播推送集合才会或者正收益，所以有

$$\mathcal{K}_m \subseteq \tilde{\mathcal{K}}_m = \{k \mid \varepsilon_{k,m} \ge \frac{\alpha}{1+\alpha}\} \tag{6}$$

这样，考虑收益最大化，业务多播推送用户选择退化为传统多描述多播中用户选择问题，Lemma 1 可以解决。

- β_m 为常数：即用户接收不感兴趣的业务造成的损失固定为常数 β_m

此时播推送用户选择比较复杂，无法得到较为明确的方法。

再进一步，考虑进行业务多播推送传输的功率代价，包括基站侧与用户侧，其中基站侧功耗建模与用户侧功耗建模分别为

$$P_{BS} = \sum_{n=1}^{N} p_n + P_{fix,BS} \tag{7}$$

$$P_{UE,k} = P_{fix,UE}\sigma(k \in \cup_{1\le m\le M} \mathcal{K}_m) + \xi \sum_{m=1}^{M} R_m \sigma(k \in \mathcal{K}_m) \tag{8}$$

$$P_{UE} = \sum_{k=1}^{K} P_{UE,k} = \sum_{k \in \cup_{1\le m\le M} \mathcal{K}_m} P_{fix,UE} + \sum_{m=1}^{M} \sum_{k \in \mathcal{K}_m} \xi R_m \tag{9}$$

其中 $\sigma(x)$ 为判别函数，当 x 为真，$\sigma(x)=1$；否则，$\sigma(x)=0$；$P_{fix,BS}$ 与 $P_{fix,UE}$ 分别为基站与用户进行业务发送与接收的固有功耗，ξ 表示用户没接收单位数据业务的额外功耗。

综上，系统进行业务多播推送传输的效用可建模为

$$\begin{aligned}
\eta &= \sum_{m=1}^{M} U_m - (\alpha_1 P_{BS} + \alpha_2 P_{UE}) \\
&= \sum_{m=1}^{M} \sum_{k \in \mathcal{K}_m} [R_m \varepsilon_{k,m} - \beta_m (1-\varepsilon_{k,m})] - \alpha_1 [\sum_{n=1}^{N} p_n + P_{fix,BS}] - \alpha_2 [\sum_{k \in \cup_{1\le m\le M} \mathcal{K}_m} P_{fix,UE} + \sum_{m=1}^{M} \sum_{k \in \mathcal{K}_m} \xi R_m] \quad (10) \\
&= \sum_{m=1}^{M} \sum_{k \in \mathcal{K}_m} [R_m (\varepsilon_{k,m} - \alpha_2 \xi) - \beta_m (1-\varepsilon_{k,m})] - \alpha_1 \sum_{n=1}^{N} p_n - \alpha_1 P_{fix,BS} - \alpha_2 \sum_{k \in \mathcal{K}} P_{fix,UE}
\end{aligned}$$

其中$\mathcal{K}=\bigcup_{m=1}^{M}\mathcal{K}_m$表示所有业务的接收用户集合，$\alpha_1$与$\alpha_2$分别为基站功耗与用户功耗对效用的权重。由于$P_{fix,BS}$为定值，与优化问题无关，$\sum_{k\in\mathcal{K}} P_{fix,UE}$虽然取多播推送用户选择有关，但是由于$P_{fix,UE}$一般很小，所以也可以近似舍去。这样，系统效用可简化为

$$\tilde{\eta}=\sum_{m=1}^{M}\sum_{k\in\mathcal{K}_m}[R_m(\varepsilon_{k,m}-\alpha_2\xi)-\beta_m(1-\varepsilon_{k,m})]-\alpha_1\sum_{n=1}^{N}p_n \tag{11}$$

这样，基于移动用户兴趣分析的能效多播推送问题建模如下

$$\begin{aligned}
\mathcal{Q}_0:\quad & \max_{\mathcal{K}_m,\rho_{n,m},p_n}\ \tilde{\eta}=\sum_{m=1}^{M}\sum_{k\in\mathcal{K}_m}[R_m(\varepsilon_{k,m}-\alpha_2\xi)-\beta_m(1-\varepsilon_{k,m})]-\alpha_1\sum_{n=1}^{N}p_n\\
& \mathcal{C}_1:\mathcal{K}_m\subseteq\mathcal{K}_{\text{tot}},\ \forall m\\
& \mathcal{C}_2:\rho_{n,m}\in\{0,1\},\ \forall n,m;\ \sum_{m=1}^{M}\rho_{n,m}\le 1,\ \forall n\\
& \mathcal{C}_3:p_n\ge 0,\ \forall n;\ \sum_{n=1}^{N}p_n\le P_{\max}\\
& \mathcal{C}_4:R_m\ge R_{m,\min},\ \forall m
\end{aligned} \tag{12}$$

其中$P_{\max}$表示基站端的最大发送功率，$R_{m,\min}$是业务m的最小传输速率。

在多播推送传输中，由于可以提前进行用户兴趣分析与业务推送，可以灵活选择传输周期以及断点续传，所以最小传输速率$R_{m,\min}$往往很低。

2.2 具体算法设计

针对$\beta_m=0$或$\beta_m=\alpha R_m$两种情况，问题可以简化，这里不做研究。我们主要关心β_m为常数的情况。

1） 问题上限设计

首先，我们给出一个上限分析求解方法，假设每个子载波上多播推送用户集合可以不一致（这可以通过使用多描述编码多播等高级编码方案实现），此时的效用函数可改写为

$$\tilde{\eta}=\sum_{n=1}^{N}\sum_{m=1}^{M}\rho_{n,m}\left\{\sum_{k\in\mathcal{K}_{n,m}}\left[(\varepsilon_{k,m}-\alpha_2\xi)\Delta f\log_2(1+p_n\gamma_{n,m})-\beta_{n,m}(1-\varepsilon_{k,m})\right]-\alpha_1 p_n\right\} \tag{13}$$

这里我们假设$\beta_{n,m}$满足$\sum_{n=1}^{N}\rho_{n,m}\beta_{n,m}=\beta_m$，这会导致新增加一个优化变量$\beta_{n,m}$，但是在优化中我们可以通过一定方法回避。

这样优化问题$\mathcal{Q}_0$放松为

$$\begin{aligned}
\mathcal{Q}_1:\quad & \max_{\mathcal{K}_{n,m},\rho_{n,m},p_n}\ \tilde{\eta}\\
& \mathcal{C}_1:\mathcal{K}_{n,m}\subseteq\mathcal{K}_{\text{tot}},\ \forall n,m\\
& \mathcal{C}_2\sim\mathcal{C}_4
\end{aligned} \tag{14}$$

使用拉格朗日对偶求解。对约束$\mathcal{C}_3,\mathcal{C}_4$定义对偶变量$\lambda,\mu_m$可得对偶函数表达式

$$\mathcal{L}=\tilde{\eta}-\lambda\left(\sum_{n=1}^{N}p_n-P_{\max}\right)+\sum_{m=1}^{M}\mu_m(R_m-R_{m,\min})=\sum_{n=1}^{N}\mathcal{D}_n+\lambda P_{\max}-\sum_{m=1}^{M}\mu_m R_{m,\min} \tag{15}$$

其中

$$\mathcal{D}_n=\sum_{m=1}^{M}\rho_{n,m}\mathcal{D}_{n,m}=\sum_{m=1}^{M}\rho_{n,m}[\sum_{k\in\mathcal{K}_{n,m}}\mathcal{D}_{n,k|m}-(\alpha_1+\lambda)p_n] \tag{16}$$

$$D_{n,k|m}=(\varepsilon_{k,m}-\alpha_2\xi+\mu_m)\Delta f\log_2(1+p_n\gamma_{n,m})-\beta_{n,m}(1-\varepsilon_{k,m}) \tag{17}$$

这样，对偶问题为

$$\begin{aligned}\mathcal{F}: \quad & \min_{\lambda,\mu_m}\max_{\mathcal{K}_{n,m},\rho_{n,m},p_n}\mathcal{L}\\ \text{s.t.}\ & \mathcal{K}_{n,m}\subseteq\mathcal{K}_{\text{tot}},\ \forall n,m\\ & \rho_{n,m}\in\{0,1\},\ \forall n,m;\ \sum_{m=1}^{M}\rho_{n,m}\le 1,\ \forall n\\ & p_n\ge 0\\ & \lambda\ge 0,\mu_m\ge 0\end{aligned} \tag{18}$$

其中外部 min 问题可以使用子梯度迭代，内部 max 问题求解算法如下。

内部 max 问题可以拆分为如下 N 个子问题，其中第 n 个子问题负责子载波 n 的业务选择、功分以及相应的多播推送用户选择，表述为

$$\begin{aligned}&\max_{\mathcal{K}_{n,m},\rho_{n,m},p_n}\sum_{m=1}^{M}\rho_{n,m}[\sum_{k\in\mathcal{K}_{n,m}}\mathcal{D}_{n,k|m}-(\alpha_1+\lambda)p_n]\\ &\text{s.t.}\ \mathcal{K}_{n,m}\subseteq\mathcal{K}_{\text{tot}},\ \forall m\\ &\rho_{n,m}\in\{0,1\},\ \forall m,\ \sum_{m=1}^{M}\rho_{n,m}\le 1\\ &p_n\ge 0\end{aligned} \tag{19}$$

对于上述问题，分下面两步完成：

应该是没有删除已选业务的过程，这样能得到上限

把每个子载波单独拿出来分配，对每个子载波，通过遍历的方法，找出能使效益最大的业务，并把这个业务分配给这个子载波。

1> 假设子载波 n 分配给业务 m，计算此时最佳的功分与推送用户集合

问题表述如下：

$$\begin{aligned}&\max_{\mathcal{K}_{n,m},p_n}\sum_{k\in\mathcal{K}_{n,m}}\mathcal{D}_{n,k|m}-(\alpha_1+\lambda)p_n\\ &\text{s.t.}\ \mathcal{K}_{n,m}\subseteq\mathcal{K}_{\text{tot}},p_n\ge 0\end{aligned} \tag{20}$$

上述问题可以通过定义域优化基准用户完成，基准用户的定义如下：

假设用户 k 为集合 $\mathcal{K}_{n,m}$ 的基准用户，则集合 $\mathcal{K}_{n,m}$ 可以表示为

$$\mathcal{K}_{n,m}=\{j\,|\,\gamma_{n,j}\ge\gamma_{n,k},\mathcal{D}_{n,k|m}\ge 0\} \tag{21}$$

按顺序依次选择用户作为基准用户，然后选出能使效益最大的用户最为最佳基准用户。

当基准用户为 k 时，最佳的功分为

$$p_{n,k}=[\frac{\Delta f\sum_{k\in\mathcal{K}_{n,m}}(\varepsilon_{k,m}-\alpha_2\xi+\mu_m)}{\ln 2(\alpha_1+\lambda)}-\frac{1}{\gamma_{n,k}}]^{+} \tag{22}$$

联合迭代式(21)与(22)可求出最佳的功分 $p_{n,k}$ 与推送用户集合 $\mathcal{K}_{n,m}$。

问题是，把每个子载波单独研究，和联合优化相比，这种方法求出来的是不是上限？求出来的值和子载波优化顺序有没有关系？

2> 遍历 M 种可能的子载波分配策略，选出最佳的子载波分配

遍历 M 种可能的子载波分配策略，并将每次的结果代入表达式 $\sum_{k\in\mathcal{K}_{n,m}}\mathcal{D}_{n,k|m}-(\alpha_1+\lambda)p_n$，选出是的表达式最大的业务作为子载波 n 的承载业务。

Remark 2：在每次子梯度迭代中，优化内部 max 问题需要知道 $\beta_{n,m}$，这与每个子载波

上的多播推送用户集合选择息息相关，间接影响了最终功分。为此，在迭代初始 $t=0$，我们假设 $\beta_{n,m}(0)=M\beta_m/N$，即假设每个业务占据的子载波数目相等；接着在每次迭代，我们利用上一次的优化结果更新 $\beta_{n,m}$，规则如下

$$\boxed{\beta_{n,m}(t+1)=\beta_m\Big/\sum_{n=1}^{N}\rho_{n,m}(t)} \tag{23}$$

2） 次优算法设计

我们的目的是设计低复杂度次优算法，核心思想是采用（迭代）分步优化方法。

(1) **迭代优化方法 1**：多播推送用户集合选择 + 子载波与功率联合分配

该方案的迭代环节分为两步：

➢ 给定多播推送用户集合下，子载波与功率联合分配算法

此时，问题 $\mathcal{Q}_0$ 简化为

$$\mathcal{Q}_{2-1}: \quad \max_{\rho_{n,m},p_n}\sum_{m=1}^{M}\omega_m R_m-\alpha_1\sum_{n=1}^{N}p_n-\delta_1$$

$$\mathcal{C}_2:\rho_{n,m}\in\{0,1\},\ \forall n,m;\ \sum_{m=1}^{M}\rho_{n,m}\le 1,\ \forall n$$

$$\mathcal{C}_3:p_n\ge 0,\ \forall n;\ \sum_{n=1}^{N}p_n\le P_{\max}$$

$$\mathcal{C}_4:R_m\ge R_{m,\min},\ \forall m$$

其中 $\omega_m=\sum_{k\in K_m}(\varepsilon_{k,m}-\alpha_2\xi)$，$\delta_1=\sum_{m=1}^{M}\sum_{k\in K_m}\beta_m(1-\varepsilon_{k,m})$ 是固定常数。

上述问题可以使用对偶分解+子梯度进行求解。

对偶函数为：

$$\begin{aligned}L&=\sum_{m=1}^{M}\omega_m R_m-\alpha_1\sum_{n=1}^{N}p_n-\lambda\Big(\sum_{n=1}^{N}p_n-P_{\max}\Big)+\sum_{m=1}^{M}\mu_m(R_m-R_{m,\min})\\&=\sum_{m=1}^{M}\omega_m R_m-\alpha_1\sum_{m=1}^{M}\sum_{n=1}^{N}\rho_{n,m}p_{n,m}-\lambda\Big(\sum_{m=1}^{M}\sum_{n=1}^{N}\rho_{n,m}p_{n,m}-P_{\max}\Big)+\sum_{m=1}^{M}\mu_m(R_m-R_{m,\min})\\&=\sum_{m=1}^{M}\left((\omega_m+\mu_m)R_m-(\alpha_1+\lambda)\sum_{n=1}^{N}\rho_{n,m}p_{n,m}\right)+\lambda P_{\max}-\sum_{m=1}^{M}\mu_m R_{m,\min}\end{aligned}$$

代入：

$$R_m=\Delta f\sum_{n=1}^{N}\rho_{n,m}\log_2\left(1+p_n\min_{k\in K_m}\gamma_{n,k}\right)$$

得到：

$$L=\sum_{n=1}^{N}\sum_{m=1}^{M}D_n^m+\lambda P_{\max}-\sum_{m=1}^{M}\mu_m R_{m,\min}$$

其中：

$$D_n^m=\rho_{n,m}\left(\left(w_m+\mu_m\right)\Delta f\log_2\left(1+p_{n,m}\min_{k\in K_m}\gamma_{n,k}\right)-\left(\alpha_1+\lambda\right)p_{n,m}\right)$$

注：$\sum_{n=1}^{N}$ 应该放在最外，因为公式变换的目标是把每个子载波独立出来研究功率分配。

OP4 转化为：

$$\min_{\lambda,\mu_m}\max_{\rho_{n,m},p_n} L$$

$$C_2:\rho_{n,m}\in\{0,1\},\forall n,m;\sum_{m=1}^{M}\rho_{n,m}\le 1,\forall n$$

$$\lambda\ge 0;\mu_m\ge 0,\forall m;$$

假设第 n 个子载波已经分配给了多播组 m，即 $\rho_{n,m}=1$ 且对于 $\forall i\in M,i\ne m$ 有 $\rho_{n,i}\ne 1$，根据 KKT 条件，最优功率分配应该满足:

$$\frac{\partial L\left(p_n,\lambda,\mu\right)}{\partial p_n}=\frac{\partial\left(\left(w_m+\mu_m\right)\Delta f\log_2\left(1+p_{n,m}\gamma_{n,m}^{\min}\right)-\left(\alpha_1+\lambda\right)p_{n,m}\right)}{\partial p_n}$$

$$=\frac{(\mu_m+w_m)\Delta f}{(1+p_{n,m}\gamma_{n,m}^{\min})\ln 2}-(\alpha_1+\lambda)=0$$

得到:

$$p_{n,m}=\left[\frac{(\mu_m+w_m)\Delta f}{\left(\alpha_1+\lambda\right)\gamma_{n,m}^{\min}\ln 2}-\frac{1}{\gamma_{n,m}^{\min}}\right]^+$$

对第 n 个子载波，计算该子载波被分配到哪一个多播组的时候，所得到的效益最大（D_n^m 取到最大），并把该子载波分配给这个多播组:

$$m_n=\arg\max_{1\le m\le M} D_n^m$$

$$\rho_n^m=\begin{cases}1, & m=m_n\\ 0, & \text{其他}\end{cases}$$

$$p_n=p_n^m$$

$$\lambda(l+1)=\left[\lambda(l)+c_1\left(\sum_{n=1}^{N}p_n-P_{\max}\right)\right]^+$$

$$\mu(l+1)=\left[\mu(l)-c_2\left(R_m-R_{m,\min}\right)\right]^+$$

➢ 给定子载波分配与功率分配下，多播推送用户集合选择

此时问题简化为

$$\mathcal{Q}_{2-2}:\qquad \max_{\mathcal{K}_m}\tilde{\eta}=\sum_{m=1}^{M}\sum_{k\in\mathcal{K}_m}[R_m(\varepsilon_{k,m}-\alpha_2\xi)-\beta_m(1-\varepsilon_{k,m})]$$

$$\mathcal{C}_1:\mathcal{K}_m\subseteq\mathcal{K},\ \forall m$$

$$\mathcal{C}_2:R_m\ge R_{m,\min},\ \forall m$$

为此，我们设计如下多播推送用户选择方案。

首先，定义矩阵收益矩阵 $\mathbf{A}=[a_{k,m}]$，其中 $a_{k,m}$ 初始定义为

$$a_{k,m}=\Delta f\,(\varepsilon_{k,m}-\alpha_2\xi)\sum_{n\in\phi_m}\log_2(1+p_n\min(\gamma_{n,m},\gamma_{n,k}))-\beta_m(1-\varepsilon_{k,m})\qquad(24)$$

其中 $\phi_m=\{n\,|\,\rho_{n,m}=1\}$，$\gamma_{n,m}$ 是多播推送用户集合 $\mathcal{K}_m$ 中最差接收 SINR。

其次，我们依据收益矩阵进行多播推送用户选择。我们将之命名为基于收益矩阵更新的多播推送用户选择算法。

初始化:

1> 初始业务多播推送集合$\mathcal{K}_m=\varnothing,\forall m$，每个集合中初始最差接收 SINR $\gamma_{n,m}=\max_{k\in\mathcal{K}_{\text{tot}}}\gamma_{n,k}$，初始收益矩阵$\mathbf{A}=[a_{k,m}]$，初始多播业务的收益$U_m=0,1\le m\le M$。

2> 初始化每个多播业务的最大可选推送用户集合$\mathcal{K}_m^a$

$\mathcal{K}_m^a=\{k\in\mathcal{K}\mid R_{k,m}=\Delta f\sum_{n\in\phi_m}\log_2(1+p_n\gamma_{n,k})\ge R_{m,\min}\}$。

迭代用户选择:

1> 选择收益最大的用户—业务对(k^*,m^*)，即$(k^*,m^*)=\arg\max_{k\in\mathcal{K}_m^a,1\le m\le M}a_{k,m}$

2> 若$a_{k^*,m^*}>0$，进入步骤 3>进行判决；否则，用户选择结束。

3> 假设将用户k^*加入业务m^*的多播推送用户集合，从而更新该多播推送用户集合中最差接收 SINR $\tilde{\gamma}_{n,m^*}=\min(\gamma_{n,m^*},\gamma_{n,k^*})$，并计算该多播业务的整体收益

$$\tilde{U}_{m^*}=\sum_{k\in\mathcal{K}_{m^*}\bigcup\{k^*\}}\left[R_{m^*}\varepsilon_{k,m^*}-\beta_m(1-\varepsilon_{k,m^*})\right];$$

4> 若$\tilde{U}_{m^*}\ge U_{m^*}$，将用户$k^*$加入业务$m^*$的多播推送用户集合，即$\mathcal{K}_{m^*}=\mathcal{K}_{m^*}\bigcup\{k^*\}$，并转入步骤 5>进行更新；否则，令$a_{k^*,m^*}=0$，跳回步骤 1>。

5> 更新多播推送用户集合中最差接收 SINR 以及收益矩阵:

- 令$\gamma_{n,m^*}=\min(\gamma_{n,m^*},\gamma_{n,k^*})$，$a_{k,m^*}=\Delta f\sum_{n\in\phi_m}\log_2(1+p_n\min(\gamma_{n,m^*}+\gamma_{n,k}))$ $-\beta_{m^*}(1-\varepsilon_{k,m^*})-tP_{fix}^{\text{UE}}$，$\forall k\ne k^*$，$U_{m^*}=\tilde{U}_{m^*}$；
- 令$a_{k^*,m^*}=0$，且若$k^*\notin\bigcup_{m\ne m^*}\mathcal{K}_m$，更新$a_{k^*,m}=a_{k^*,m}+tP_{fix}^{\text{UE}}$，$\forall m\ne m^*$，

并跳转到步骤 1>。

(2) **分步优化方法 2**: 多播推送用户集合选择与子载波分配 + 功率分配

该方案的迭代环节分为两步:

➢ 给定多播推送用户集合与子载波分配下，功率分配算法

此时，问题$\mathcal{Q}_0$简化为

$$\begin{aligned}\mathcal{Q}_{3-1}:\quad &\max_{p_n}\sum_{m=1}^{M}\omega_m R_m-\alpha_1\sum_{n=1}^{N}p_n-\delta_1\\ &\mathcal{C}_1: p_n\ge 0,\ \forall n;\ \sum_{n=1}^{N}p_n\le P_{\max}\\ &\mathcal{C}_2: R_m\ge R_{m,\min},\ \forall m\end{aligned}$$

其中$\omega_m=\sum_{k\in\mathcal{K}_m}(\varepsilon_{k,m}-\alpha_2\xi)$，$\delta_1=\sum_{m=1}^{M}\sum_{k\in\mathcal{K}_m}\beta_m(1-\varepsilon_{k,m})$是固定常数。

上述问题可以使用对偶分解+子梯度进行求解，求解过程略。

➢ 给定功率分配下，多播推送用户集合选择与子载波分配

此时问题简化为

$$
\begin{aligned}
\mathcal{Q}_{3-2}: \quad & \max_{\rho_{n,m},\mathcal{K}_m} \tilde{\eta} = \sum_{m=1}^{M}\sum_{k\in\mathcal{K}_m}[R_m(\varepsilon_{k,m}-\alpha_2\xi)-\beta_m(1-\varepsilon_{k,m})] \\
& \mathcal{C}_1: \mathcal{K}_m \subseteq \mathcal{K},\ \forall m \\
& \mathcal{C}_2: \rho_{n,m}\in\{0,1\},\ \forall n,m;\ \sum_{m=1}^{M}\rho_{n,m}\le 1,\ \forall n \\
& \mathcal{C}_3: R_m \ge R_{m,\min},\ \forall m
\end{aligned}
$$

目前，想到的方法是参考文献[1]中给出的二进制量子行为粒子蜂群 BQPSO 算法，将多播推送用户选择与子载波分配表示为如下位置向量

$$
\boldsymbol{\theta}=[c_{1,1},c_{1,2},...,c_{1,K},...,c_{M,1},c_{M,2},...,c_{M,K},d_{1,1},d_{1,2},...d_{1,\log_2 M},...,d_{N,1},d_{N,2},...d_{N,\log_2 M}] \qquad (25)
$$

然后使用 BQPSO 进行最佳向量搜索。

3. 能效多播推送方案设计

目标：给定业务集合 $\mathcal{M}_{\text{tot}}=\{1,2,...,M\}$ 以及用户集合 $\mathcal{K}_{\text{tot}}=\{1,2,...,K\}$，基于上述移动用户分析模型（包括移动用户兴趣图谱 $\mathbf{Q}=[q_{k,m}]$ 以及实际兴趣指数条件概率分布 $f_{k,m}(x\,|\,q_{k,m})$），针对每个业务，选择合适的多播推送用户集合 $\mathcal{K}_m\subseteq\mathcal{K}_{\text{tot}}$，并联合进行基站端资源分配，最大化业务传输能效。

3.1 问题建模

分析与上面类似，可得系统进行业务推送的收益与能耗为

$$U_m=R_mE[\sum_{k\in\mathcal{K}_m}s_{k,m}]-\beta_mE[\sum_{k\in\mathcal{K}_m}(1-s_{k,m})]=\sum_{k\in\mathcal{K}_m}\left[R_m\varepsilon_{k,m}-\beta_m(1-\varepsilon_{k,m})\right] \tag{26}$$

$$P_{\text{tot}}=P_{BS}+P_{UE}=\sum_{n=1}^{N}p_n+P_{fix,BS}+\sum_{k\in\mathcal{K}}P_{fix,UE}+\sum_{m=1}^{M}\sum_{k\in\mathcal{K}_m}\xi R_m \tag{27}$$

这样，系统进行业务多播推送传输的能效用可建模为

$$\eta=\frac{\sum_{m=1}^{M}U_m}{P_{BS}+P_{UE}}=\frac{\sum_{m=1}^{M}\sum_{k\in\mathcal{K}_m}\left[R_m\varepsilon_{k,m}-\beta_m(1-\varepsilon_{k,m})\right]}{\sum_{n=1}^{N}p_n+P_{fix,BS}+\sum_{k\in\mathcal{K}}P_{fix,UE}+\sum_{m=1}^{M}\sum_{k\in\mathcal{K}_m}\xi R_m} \tag{28}$$

基于移动用户兴趣分析的能效多播推送问题建模如下

$$\begin{aligned}\mathcal{Q}_0:\quad &\max_{\mathcal{K}_m,\rho_{n,m},p_n}\ \eta\\ &\mathcal{C}_1:\mathcal{K}_m\subseteq\mathcal{K}_{\text{tot}},\ \forall m\\ &\mathcal{C}_2:\rho_{n,m}\in\{0,1\},\ \forall n,m;\ \sum_{m=1}^{M}\rho_{n,m}\le 1,\ \forall n\\ &\mathcal{C}_3:p_n\ge 0,\ \forall n;\ \sum_{n=1}^{N}p_n\le P_{\max}\\ &\mathcal{C}_4:R_m\ge R_{m,\min},\ \forall m\end{aligned} \tag{29}$$

3.2 具体算法设计

我们主要关心 β_m 为常数的情况，为了与上面对应，这里舍去用户侧固有功耗 $\sum_{k\in\mathcal{K}}P_{fix,UE}$。

1） 问题上限设计

假设每个子载波上同一业务的推送用户集合可以不同，即放松优化变量 $\mathcal{K}_m$ 为 $\mathcal{K}_{n,m}$。

使用分式规划进行问题转换，然后求解与上面相同。

2） 次优算法设计

使用分式规划，其余与上面相同。

参考文献

[1] Quansheng Xu, Xi Li and Hong Ji, et al., “Energy-Efficient Resource Allocation for Heterogeneous Services in OFDMA Downlink Networks Systematic Perspective,” IEEE Trans. Veh. Technol., vol. 63, no. 5, pp: 2071-2082, 2014.