光华管理学院 Guanghua School of Management

因思想 而光华

组合预测

Forecast Combination

李丰

人工智能驱动的预测

一罐糖豆：谁的猜测更可靠？

- 请每个人独立猜测罐中的糖豆数量。
- 你会相信某个人的答案，还是所有答案的平均？
- 如果大家都受到同一个错误线索影响，平均还有效吗？



组合预测：综合对同一未来的判断

- 对同一目标、同一步长，收集多个模型或专家的预测。
- 组合既可以整合不同信息，也可以分散选错模型的风险。
- Bates 与 Granger （1969）系统讨论了预测的最优组合。
- 能否改善预测，取决于成员的质量、偏差与误差相关性。

从两个预测开始

- 两个预测分别为 $\hat{y}_1$ 和 $\hat{y}_2$；第一个预测的权重为 w。

$$\hat{y}_c = w\hat{y}_1 + (1-w)\hat{y}_2$$

- 权重之和为 1；若权重都非负，组合就在两个预测之间。
- 等权组合取 $w = 1/2$，不需要从数据估计权重。

预测误差与拟合残差

$e_j = y - \hat{y}_j$，　　$\mathrm{MSE}_j = \mathrm{E}(e_j^2)$

$\mathrm{MSE}_j = \mathrm{Var}(e_j) + [\mathrm{E}(e_j)]^2$

- 下文假设成员无偏，即 $\mathrm{E}(e_j) = 0$。
- 此时记 $\sigma_j^2 = \mathrm{Var}(e_j) = \mathrm{MSE}_j$。
- 理论比较先把权重视为固定；实务中用样本外误差估计。

固定权重保留无偏性

$$e_c = y - \hat{y}_c = we_1 + (1-w)e_2$$

$$\mathrm{E}(e_c) = w\mathrm{E}(e_1) + (1-w)\mathrm{E}(e_2) = 0$$

- 无偏成员的固定线性组合仍然无偏，只要权重和为 1。
- 若每个成员都平均高估 10，等权组合仍平均高估 10。
- 思考：为什么“权重固定”是这里的重要条件？

组合误差方差：先展开平方

$$\mathrm{E}(e_c^2) = \mathrm{E}\{[we_1 + (1-w)e_2]^2\}$$
$$= w^2\mathrm{E}(e_1^2) + (1-w)^2\mathrm{E}(e_2^2)$$
$$+2w(1-w)\mathrm{E}(e_1e_2)$$

- 误差均值为零时，均方误差就是误差方差。
- 交叉项说明：只知道单个模型有多准，还不足以决定权重。

误差相关性与组合方差

$$\sigma_c^2 = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\rho\sigma_1\sigma_2$$

- 这里 $\rho = \mathrm{Corr}(e_1, e_2)$，是预测误差的相关系数。
- 固定 $0 < w < 1$ 且其他条件相同时，误差相关越低越有利。
- 预测曲线看起来相似，不代表预测误差没有互补性。

两模型的最优权重

- 对上页的误差方差求最小值，得到第一个预测的权重：

$$w^* = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

- 要求分母 $\mathrm{Var}(e_1 - e_2) > 0$；第二个权重为 $1 - w^*$。
- 这里暂不限制权重非负；最优权重可能小于 0 或大于 1。
- 若要求非负，在两模型情形可把该权重截到 [0, 1]。

误差不相关时的最优权重

- 如果 $\rho = 0$，则最优权重简化为：

$$w^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

- 例：$\sigma_1^2 = 4$、$\sigma_2^2 = 9$，第一个权重为 $9/13$。
- 组合方差为 $36/13 \approx 2.77$，小于 4 和 9。
- 注意：第一个模型的权重，分子使用第二个模型的方差。

讨论：最优权重如何变化？

- 误差不相关时，σ_1^2 增大，w^* 会怎样变化？为什么？
- 误差不相关时，σ_2^2 增大，w^* 又会怎样变化？
- 当误差相关性 ρ 上升时，最优权重一定上升吗？
- 若 $\sigma_1^2 = \sigma_2^2$，你能从公式中得到什么？

理论最优组合为何不会更差?

- 当真实误差协方差已知，最小化的是同一个总体 MSE。
- 选择 $w = 1$ 或 $w = 0$，就分别退化为两个单独预测。
- 因此，在包含这两个选择的可行集合中，最优值不会更大。
- 但改善未必严格；估计权重在新样本中也没有这种保证。
- 思考：如果两个成员的误差完全一样，组合会发生什么?

多个预测的最优权重

- 有 K 个无偏预测；$\boldsymbol{\Sigma}$ 是它们的误差协方差矩阵。

$$\min_{\mathbf{w}} \ \mathbf{w}^{\top}\boldsymbol{\Sigma}\mathbf{w} \quad \text{s.t.} \quad \mathbf{1}^{\top}\mathbf{w} = 1$$

$$\mathbf{w}^{*} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}}$$

- 其中 $\mathbf{1}$ 为全 1 列向量；这里假设协方差矩阵正定。
- 增加非负约束后，通常需要用数值优化求解。

逆误差方差加权

- 若各成员误差不相关，协方差矩阵为对角矩阵。

$$w_j = \frac{1/\sigma_j^2}{\sum_{i=1}^{K} 1/\sigma_i^2}$$

- 误差方差越小，分配的权重越大。
- 这减少了需要估计的参数，但忽略了成员之间的冗余。
- 成员有偏时，用逆 MSE 加权可以作为启发式方法。

真实误差方差未知，怎样估计？

- 在多个历史预测时点，保存当时真正做出的预测。
- 等真实值公布后，计算每个成员的预测误差。
- 用这些误差估计方差、协方差或 MSE，再确定组合权重。
- 模型预测与专家判断都适用；不要用样本内拟合值替代。
- 每个预测步长可以有不同权重；在独立测试期比较表现。

权重随时间更新，也需要平滑

- 先用近期误差得到新权重 $\tilde{w}_T$，再平滑更新：

$$w_T = \gamma w_{T-1} + (1-\gamma)\tilde{w}_T, \qquad 0 \le \gamma < 1$$

- 也可让近期误差在方差估计中占更大比重：

$$\tilde{\sigma}^2_{j,T} = \sum_{t \in \mathcal{V}_T} \alpha_{T,t} e^2_{j,t}, \qquad \sum_t \alpha_{T,t} = 1$$

- 近期权重大，响应更快，也可能更不稳定；参数在验证期选择。

组合预测之谜：等权为何难胜？

- 理论最优权重通常不等于 $1/K$。
- 但在许多实证研究中，等权常常优于估计出来的复杂权重。
- 这称为“组合预测之谜”(Forecast Combination Puzzle)。
- 关键区别：真实最优权重已知，与最优权重需要估计。
- 等权应是基准；加权是否更好，要交给样本外证据。

估计权重带来额外不确定性

- 如果 $\hat{w}$ 来自样本，它也是随机变量：

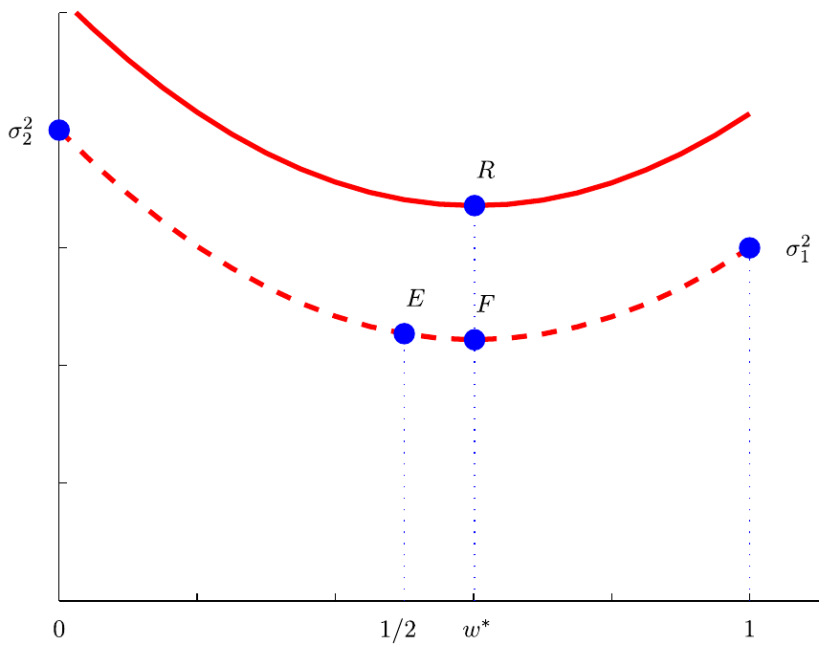
$$e_c = \hat{w}e_1 + (1-\hat{w})e_2$$

$$\mathrm{E}(e_c) = \mathrm{Cov}(\hat{w}, e_1) - \mathrm{Cov}(\hat{w}, e_2)$$

- 即使成员无偏，权重与误差相关也可能产生组合偏差。
- 权重估计误差也会改变 MSE；影响取决于其方差和相关结构。

图解组合预测之谜

- F： 把权重当作固定常数时的理论最优点。
- E： 固定等权组合； R： 图示随机权重机制下的最优点。
- 为什么固定权重下最优，未必意味着估计后表现最好？



Claeskens, Magnus, Vasnev & Wang （2016）， Fig. 1

把组合写成一个回归问题

$$y_t = \sum_{k=1}^{K} w_k \hat{y}_{k,t} + \varepsilon_t$$

$$\mathrm{RSS}(\mathbf{w}) = \sum_{t \in \mathcal{V}} \left(y_t - \sum_{k=1}^{K} w_k \hat{y}_{k,t} \right)^2$$

- 用验证期的真实值回归各成员当时做出的预测。
- 可要求权重和为 1，或进一步要求权重非负。
- 成员很多且高度相似时，直接估计回归权重容易不稳定。

向等权收缩： Ridge

- 在拟合误差之外，惩罚权重偏离等权的程度：

$$\min_{\mathbf{w}} \left\{ \mathrm{RSS}(\mathbf{w}) + \lambda \sum_{k=1}^{K} \left(w_k - \frac{1}{K} \right)^2 \right\}$$

s.t. $\mathbf{1}^{\mathrm{T}}\mathbf{w} = 1, \qquad \lambda \geq 0$

- λ 越大，通常越接近等权；越小则越依赖样本拟合。
- 用时间顺序验证选择收缩强度，而不是凭测试结果调参。

向等权收缩： Lasso

- 把平方惩罚改成绝对值惩罚：

$$\min_{\mathbf{w}} \left\{ \text{RSS}(\mathbf{w}) + \lambda \sum_{k=1}^{K} \left| w_k - \frac{1}{K} \right| \right\}$$

s.t. $\mathbf{1}^{\top}\mathbf{w} = 1, \qquad \lambda \geq 0$

- 有些权重可恰好等于等权值；这不等于把成员权重压到零。
- Diebold 与 Shin（2019）进一步讨论筛选后向相等权重收缩。

怎样检验加权组合是否有效?

- 先建立等权组合，再考虑逆 MSE 、协方差加权与收缩。
- 验证期选择成员、估计方法、窗口长度和惩罚强度。
- 测试期比较单模型、等权组合与加权组合的误差和稳定性。
- 误差相关强、历史短或环境变化时，尤其要检查权重波动。
- 讨论：更低的验证期误差，足以支持更复杂的权重吗?

预测不确定性

区间、模拟与组合预测

预测步长与区间宽度

- 平稳 AR(1)：$y_t = c + \phi y_{t-1} + \varepsilon_t$，且 $|\phi| < 1$。

$$\mathrm{Var}(e_{T+h|T} \mid \mathcal{I}_T) = \sigma_\varepsilon^2 \sum_{i=0}^{h-1} \phi^{2i}$$

$$= \sigma_\varepsilon^2 \frac{1-\phi^{2h}}{1-\phi^2} \longrightarrow \frac{\sigma_\varepsilon^2}{1-\phi^2}$$

- 平稳模型的误差方差逐渐趋于有限值；随机游走则持续增大。
- 这里假设参数已知、创新独立同分布；暂不计参数估计误差。

正态误差下的预测区间

- 按预测步长收集已实现的样本外误差；先检查系统性偏差。

$$\hat{\sigma}_h^2 = \frac{1}{N_h} \sum_{t \in \mathcal{V}_h} e_{t+h|t}^2$$

95% PI： $\hat{y}_{T+h|T} \pm 1.96\hat{\sigma}_h$

- 这是无偏、误差分布稳定且近似正态时的近似预测区间。
- 样本外误差均方不机械使用“样本数减参数数”的自由度修正。

用 Bootstrap 模拟未来

$$y_t = \hat{y}_{t|t-1} + e_t$$

- 如果有一组能代表未来的历史创新，就可以有放回地抽样。
- 把抽到的创新加到一步预测上，递推产生一条未来路径。
- 重复多次，用模拟路径的分位数构造预测区间。
- 简单残差 Bootstrap 假设创新近似独立同分布；相关或异方差时需调整。

Bootstrap 的递推步骤

- 从中心化历史残差抽样；对每条路径 $b = 1, \ldots, B$ 逐步递推。

$$e_{T+h}^{(b)} \sim \{\tilde{e}_1, \ldots, \tilde{e}_N\}$$

$$\tilde{y}_{T+h}^{(b)} = \hat{y}_{T+h|T+h-1}^{(b)} + e_{T+h}^{(b)}$$

- 将刚模拟的值作为历史，更新状态，继续直到所需步长。
- 每个步长的 2.5% 与 97.5% 经验分位数给出近似 95% 区间。

混合分布与加权和

- 混合分布：先按成员权重抽一个模型，再从它的预测分布抽样。

$$F_c(z) = \sum_{k=1}^{K} w_k F_k(z), \qquad w_k \geq 0, \quad \sum_k w_k = 1$$

- 加权和预测：把各成员的模拟结果加权，必须刻画成员间的依赖。
- 两种操作得到的分布通常不同；混合分布的分位数也不能直接平均。
- 讨论：你希望表达“哪种模型可能正确”，还是一个加权和的波动？

用 Winkler 分数评价区间

- 对名义覆盖率 $1-\alpha$ 的区间 $[l,u]$，Winkler 分数为：

$$W=(u-l)+\frac{2}{\alpha}(l-y)\mathbf{1}\{y<l\}$$
$$+\frac{2}{\alpha}(y-u)\mathbf{1}\{y>u\}$$

- 区间越窄越好，但真实值落在区间之外会受到额外惩罚。
- 分数越低越好；比较时保持覆盖率、预测步长和测试任务一致。

回到糖豆：你现在会怎样组合？

- 先检查成员预测是否针对同一目标，以及是否共享偏差。
- 用误差方差与协方差理解互补；区分理论权重与估计权重。
- 以等权为基准，检验加权和收缩是否带来稳定的样本外收益。
- 需要区间时，明确分布假设、模拟机制和区间评价指标。
- 讨论：哪些证据，会让你偏离简单平均？

主要阅读与原课件

- 原课件：李丰， Forecasting Combinations and Forecasting Uncertainty 。
- Bates & Granger （1969）， The combination of forecasts 。
- Claeskens et al. （2016）， The forecast combination puzzle 。
- Diebold & Shin （2019）， Partially-egalitarian LASSO 。
- Wang, Hyndman, Li & Kang （2023）， Forecast combinations: An over 50-year review. IJF, 39(4), 1518–1547 。