A Unified Contextual Bandit Framework for Long- and Short-Term Recommendations
Maryam Tavakol and Ulf Brefeld

Motivation

Personalized recommendation:
Extracting user interests from long-term interactions of user with system

Followed by short-term zeitgeist

Unified model to capture
Long-term part + Short-term component
Framework: Contextual Multi-Armed Bandit e.g., LinUCB

Unified Model

Outcome
A Unified Contextual Bandit Framework for Long- and Short-Term Recommendations
Maryam Tavakol and Ulf Brefeld

Optimization:
• Gradient-based approaches (in dual or primal)
• Calculating the gradient depends on the loss function
• Model parameters, (θ, α), are obtained from α
• Kernel functions applicable

Algorithm:
for t = 1, 2, ..., T do
Observe user v_t and context x_t
for all α ∈ A_t do
Observe arm features a_t
r_t, α = mean reward + confidence bound
end for
Choose arm a_t = arg max α r_t, α and observe payoff r_t
Obtain α by optimizing the objective Computr (θ_t) and (α_t) from α
end for

Instantiation

Squared loss:
\[
\alpha^2 = x_t^\top (X_t^\top X_t)^{-1} a_t + z_t^\top (Z_t^\top Z_t)^{-1} z_t
\]

Logistic loss:
\[
\alpha^2 = x_t^\top (X_t^\top X_t)^{-1} a_t + z_t^\top (Z_t^\top Z_t)^{-1} z_t
\]

Empirical Study

• Using squared loss function
• Dataset: User transactions from Zalando
• Baseline: Matrix Factorization (MF)
• Performance measure: normalized average rank

Performance:

Conclusion

• Combining short- and long-term interests of users in one model
• Free choice of loss function and model complexity
• There is no one good model: the choice depends on the application

General Optimization

Objective function with arbitrary loss, \( V(\cdot, r_t) \)

\[
\sup_{\theta, \alpha} \inf_{\pi} \sum_{t=1}^{T} V(\pi_t, r_t) + \frac{1}{2} \sum_{t=1}^{T} (\theta_t - \theta_t^0)^2 + \frac{1}{2} \sum_{t=1}^{T} (\alpha_t - \alpha_t^0)^2
\]

Using the Fenchel-Legendre conjugate of loss function in the dual space

\[
\sup_{\theta, \alpha} \sum_{t=1}^{T} V^*(-(\pi_t, r_t)) + \frac{1}{2} \sum_{t=1}^{T} (X_t \delta_t)^\top (X_t \delta_t)^\top \alpha + \frac{1}{2} \sum_{t=1}^{T} (Z_t \delta_t)^\top (Z_t \delta_t)^\top \alpha
\]

Robustness of combined model in case of new user/item

Time Complexity:

The optimization time in combined model is exponential

The average term compensates for the new items

The average term compensates for the new users

Source code available at https://github.com/marytavakol/Bandits