

[Log In](#)

[Join](#)

[Back To Module Home](#)

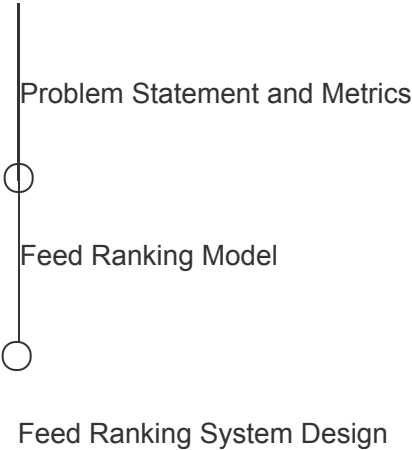
Machine Learning System Design

0% completed

Machine Learning Primer

Video Recommendation

Feed Ranking



Ad Click Prediction

Rental Search Ranking

Estimate Food Delivery Time

Machine Learning Knowledge

Machine Learning Model Diagnosis

Conclusion

Mark Module as Completed

Problem Statement and Metrics

Learn about the problem statement and metrics for building a Feed Ranking system.

We'll cover the following

- LinkedIn feed ranking
- 1. Problem statement
- 2. Metrics design and requirements
 - Metrics
 - Offline metrics
 - Online metrics
 - Requirements
 - Training
 - Inference
- Summary

LinkedIn feed ranking#

1. Problem statement#

Design a personalized LinkedIn feed to maximize long-term user engagement. One way to measure engagement is user frequency, i.e, measure the number of engagements per user, but it’s very difficult in practice. Another way is to measure the click probability or **Click Through Rate** (CTR).

LinkedIn feed ranking

On the LinkedIn feed, there are five major activity types:

- Connections (A connects with B)
- Informational
- Profile
- Opinion
- Site-specific
- Intuitively different activities have very different CTR. This is important when we decide to build models and generate training data.

Category	Example
Connection	Member connector follows member/company, member joins group
Informational	Member or company shares article/picture/message

Profile	Member updates profile, i.e., picture, job-change, etc.
Opinion	Member likes or comments on articles, pictures, job-changes, etc.
Site-Specific	Member endorses member, etc.

2. Metrics design and requirements#

Metrics#

Offline metrics#

- The Click Through Rate (CTR) for one specific feed is the number of clicks that feed receives, divided by the number of times the feed is shown.

$$CTR = \frac{number_of_clicks}{number_of_shown_times}$$

- Maximizing CTR can be formalized as training a supervised binary classification model. For offline metrics, we normalize cross-entropy and AUC.
- Normalizing cross-entropy (NCE) helps the model be less sensitive to background CTR.

$$NCE = \frac{-\frac{1}{N} \sum_{i=1}^n (\frac{1+y_i}{2} \log(p_i)) + \frac{1-y_i}{2} \log(1-p_i))}{-(p*\log(p)+(1-p)*\log(1-p))}$$

Online metrics#

- For non-stationary data, offline metrics are not usually a good indicator of performance. Online metrics need to reflect the level of engagement from users once the model has deployed, i.e., **Conversion rate** (ratio of clicks with number of feeds).

Requirements#

Training#

- We need to handle large volumes of data during training. Ideally, the models are trained in distributed settings. In social network settings, it’s common to have online data distribution shift from offline training data distribution. One way to address this issue is to retrain the models (incrementally) multiple times per day.
- Personalization: Support is needed for a high level of personalization since different users have different tastes and styles for consuming their feed.
- Data freshness: Avoid showing repetitive feed on the user’s home feed.

Inference#

- Scalability: The volume of users’ activities are large and the LinkedIn system needs to handle 300 million users.
- Latency: When a user goes to LinkedIn, there are multiple pipelines and services that will pull data from multiple sources before feeding activities into the ranking model. All of these steps need to be done within 200ms. As a result, the Feed Ranking needs to return within **50ms**.
- Data freshness: Feed Ranking needs to be fully aware of whether or not a user has already seen any particular activity. Otherwise, seeing repetitive activity will compromise the user experience. Therefore, data pipelines need to run really fast.

Summary#

Type	Desired goals
Metrics	Reasonable normalized cross-entropy
Training	High throughput with the ability to retrain many times per day

Inference	<div>Supports high level of personalization</div> <div>Latency from 100ms to 200ms</div> <div>Provides a high level of data freshness and avoids showing the same feeds multiple times</div>
-----------	--

Back

Video Recommendation System Design

Next

Feed Ranking Model

Mark as Completed

Report an Issue