Predictive & Classification Accuracy Metrics

Analytics Vidhya

Predictive Accuracy Metrics

- Datasets with items rated by users
  - MovieLens datasets 100K-10M ratings
  - Netflix 100M ratings
- Historic user ratings constitute ground truth
- Metrics measure error rate
  - Mean Absolute Error (*MAE*) computes the deviation between predicted ratings and actual ratings

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | p_i - r_i |$$

  - Root Mean Square Error (*RMSE*) is like *MAE*, but places more emphasis on larger deviation

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (p_i - r_i)^2}$$

Analytics Vidhya

Predictive Accuracy Metrics – Pros & Cons

- **Advantages**
  - Mechanics of the computation are simple and easy to understand
  - These metrics have well studied statistical properties that provide for testing the significance of a difference between the mean absolute errors of two systems.
- **Disadvantages**
  - These metrics may be less appropriate for tasks such as find good items where a ranked result is returned to the user, who then only views items at the top of the ranking.
  - Mean absolute error may be less appropriate when the granularity of true preference is small

Classification Accuracy Metrics: Precision and Recall

- Measure the frequency with which recommender system makes correct or incorrect decisions about whether the item is good

- **Precision:** a measure of exactness, determines the fraction of relevant items retrieved out of all items retrieved
  - E.g. the proportion of recommended movies that are actually good

- **Recall:** a measure of completeness, determines the fraction of relevant items retrieved out of all relevant items
  - E.g. the proportion of all good movies recommended

Analytics Vidhya

Precision@K

Evaluates the list of recommendations

- Precision at k is the proportion of recommended items in top-k set that are relevant

$$Precision@k = \frac{\#\ of\ relevant\ items\ in\ the\ top\ k\ positions}{k}$$

- Suppose that my precision at 10 in a top-10 recommendation problem is 80%, this means that 80% of recommendations I make are relevant to the user

Analytics Vidhya

Recall@K

Evaluates the list of recommendations

- Recall at k is the proportion of relevant items found in the top-k recommendations

$$Recall@k = \frac{\#\ of\ recommended\ items\ @k\ that\ are\ relevant}{total\ \#\ of\ relevant\ items}$$

- Suppose that we computed recall at 10 and found it is 40% in our top-10 recommendation system. This means that 40% of the total number of the relevant items appear in the top-k result

Analytics Vidhya

Example for Precison@K & Recall@K

item	user1 (Actual/Predicted)
item1	4/2.3
item2	2/3.6
item3	3/3.4
item4	?/4.3
item5	5/4.5
item6	?/2.3
item7	2/4.9
item8	?/4.3
item9	?/3.3
item10	4/4.3

Let's say relevant items are ones with rating greater than 3.5

- Relevant items:

{item5, item10 and item1}

total # of relevant items = 3

- Recommended items @ 3:

{item7, item5, item10}

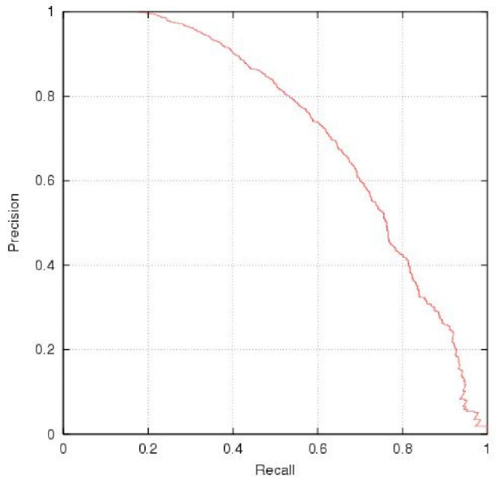
# of recommended items at 3 = 3

- Precision@3 $= \frac{2}{3}$
- Recall@3 $= \frac{2}{3}$

Analytics Vidhya

Precision & Recall

- Typically when a recommender system is tuned to increase precision, recall decreases as a result (or vice versa)



F1 Metric

- The **F_1 Metric** attempts to combine Precision and Recall into a single value for comparison purposes.
  - May be used to gain a more balanced view of performance

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

Analytics Vidhya