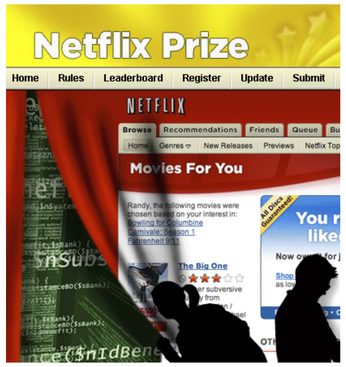
九州工業大学 情報工学部 システム創成情報工学科 廣瀬研究室

リコメンデーションアルゴリズムの研究

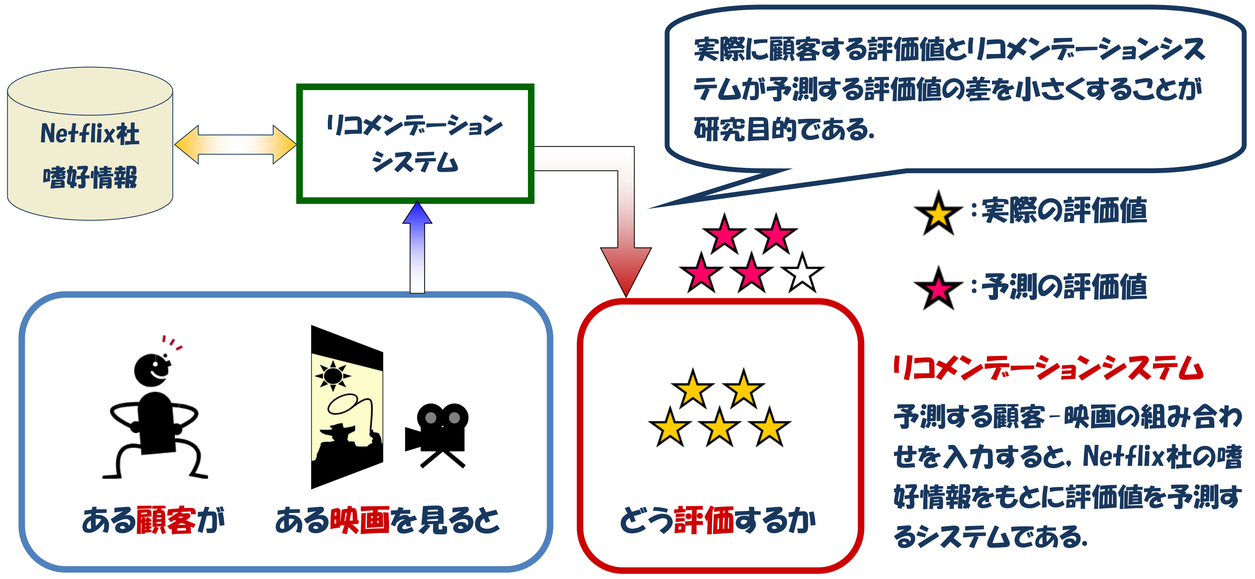
研究対象

オンラインDVDレンタルサービス会社のNetflix社は，膨大な情報の中からユーザの嗜好に合った情報を見つけ出し，その情報をユーザに能動的に提供する推薦技術の研究を行っている．また，自社の推薦システムの予測精度を高めるために，自社の顧客の嗜好データを公開し，多くの研究者と協調しながら研究を進めている．提供されたデータを研究対象とする．



研究目的

Netflix社に蓄積された顧客の嗜好情報をもとに，ある顧客がある映画を見るとどう評価するのかを予測する．この予測の**精度を高めること**が研究目的である．



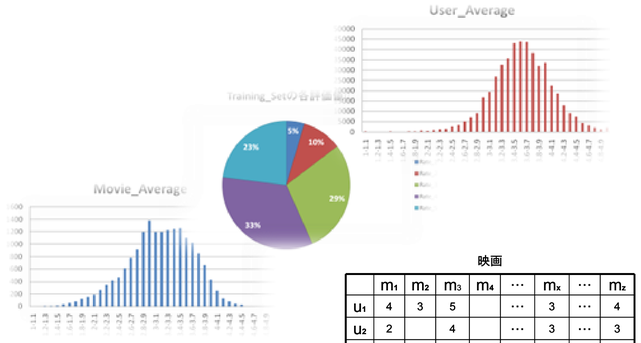
リコメンデーションシステム

予測する顧客－映画の組み合わせを入力すると，Netflix社の嗜好情報をもとに評価値を予測するシステムである．

予測手法

データ

Netflixデータは**478,615人**の顧客の**17,770本**の映画に対する**5段階評価**データである．評価されている項目は約**1億件**である． このデータを分析し，特徴をつかむ．



顧客 \ 映画	m_1	m_2	m_3	m_4	…	m_x	…	m_z
u_1	4	3	5		…	3	…	4
u_2	2		4		…	3	…	3
u_3	1		3	4	…	5	…	4
⋮	⋮	⋮	⋮	⋮		⋮		⋮
u_i	1		3	5	…	T_{ix}	…	5
⋮	⋮	⋮	⋮	⋮		⋮		⋮
u_n	2		2	3	…	4	…	4

協調フィルタリング

協調フィルタリング(Collaborative Filtering)とは，多くの顧客の嗜好情報を記録し，ある顧客と類似した嗜好を持つ他の顧客の嗜好情報から，当該の顧客の嗜好を推測する方法論である．

予測対象顧客u_iが見た映画の平均評価値T_iを求め，参考顧客u_jの平均評価T_jも求める．次に，参考顧客u_jとの相関係数C_{ij}を以下の式で求める．

$$C_{ij}=\frac{\sum_k (T_{ik}-\overline{T_i})(T_{jk}-\overline{T_j})}{\sqrt{\sum_k (T_{ik}-\overline{T_i})^2 \sum_k (T_{jk}-\overline{T_j})^2}}$$

C_{ij}を用いて未知の評価値 T_{ix}を以下の式から求める．

$$T_{ix}=\overline{T_i}+\frac{\sum_j C_{ij}(T_{jx}-\overline{T_j})}{\sum_j |C_{ij}|}$$

リコメンデーションアルゴリズムの予測精度の評価

リコメンデーションアルゴリズムの予測精度の評価は*RMSE*を用いて行う．予測の精度をあらわす指標である．値が0に近いほど正確で，精度の高い予測である．P_iを予測値，T_iを正解値，nを予測回数とする．

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