

書誌情報に基づく書籍推薦システムの開発

担当者：大串 祐平

指導教員：長田 茂美 教授

1. まえがき

図書館に存在する書籍の中から、自分の嗜好に合った書籍を探す場合、過去に読んだ書籍に対する評価情報に基づいて、ユーザの嗜好に合った未読の書籍を推薦してくれるシステムがあれば、非常に便利である。

青空文庫というインターネット上の電子図書館には大量の書籍が蓄積されているが、ユーザの嗜好に合った書籍を推薦するシステムはない。

そこで本研究では、青空文庫の利便性の更なる向上を目指し、ユーザの嗜好の合う書籍の推薦システムの開発を行い、評価実験を行う。

2. システム概要

本研究で開発したシステムは、ユーザが書籍に評価を与える、評価入力モードと、ユーザの評価情報に基づいて、書籍の推薦を行う、書籍推薦モードの2つから成る。図1にシステムの概要図を示す。評価入力モードの流れを黒の矢印で示し、書籍推薦モードの流れを赤の矢印で示す。

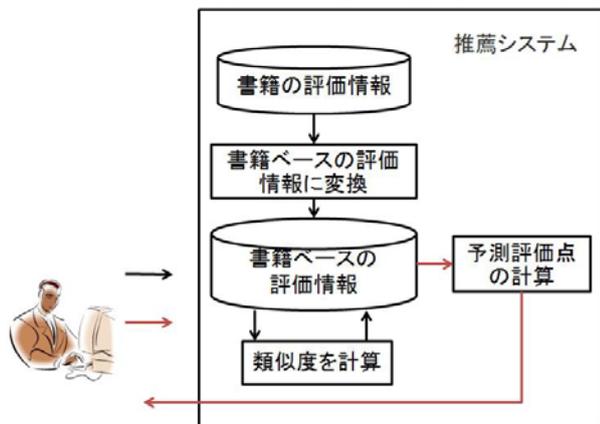


図1 システムの概要図

評価入力モードでは、ユーザはシステムに対して書籍の評価情報を入力する。評価情報は、ユーザ名をキーとして、ユーザ名、書籍名、書籍の評価点で構成されており、評価情報はシステムに蓄積される。

書籍間の類似度を推薦に使用するために、準備として評価情報のキーを、ユーザ名から書籍名に変換する。

類似度の計算を行う。評価情報に蓄積されている、ある書籍 A, B があるとす。2つの書籍 A と B 両方の書籍に評価情報を入力したユーザを n 人とし、A の書籍へ与えた評価点を (a_1, a_2, \dots, a_n) 、B の書籍へ与えた評価点を (b_1, b_2, \dots, b_n) とする。式(1)に A と B の2書籍間の類似度の計算式を示す。

$$\frac{1}{1 + \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}} \quad (1)$$

蓄積された全ての書籍間の類似度を計算した後、評価情報に書籍間の類似度を追加する。

書籍推薦モードは、評価情報に基づいて、システムが書籍推薦を行うモードである。

ユーザが評価した書籍数を m 、ユーザが評価入力した書籍の評価点を c_k 、ユーザが未読の書籍とすでに評価した書籍との類似度を e_k とする。式(2)に、予測評価点の計算式を示す。

$$\frac{\sum_{k=1}^m (c_k \times e_k)}{\sum_{k=1}^m e_k} \quad (2)$$

ユーザが未読の全書籍に対して、システムは予測評価点を計算する。未読の書籍を、計算した予測評価点と共に、ユーザに推薦する。

3. システム評価

評価実験として、ユーザの評価情報における書籍の評価点とシステムから推薦された書籍の予測評価点を比較し、違いがあるのか調査した。また、ユーザの評価情報において、評価した書籍数の違いにより、評価点と予測評価点に違いがあるのか調査した。

評価実験には、青空文庫が公開している17冊の書籍を使用した。書籍の評価点は、1点～5点の5段階評価とし、Amazon.com のカスタマーレビューにて書籍につけられた星の数を使用し、ユーザ83人を対象にして評価情報を作成した。

評価実験の結果、ユーザが評価入力する書籍の数が多いほど、システムの計算する予測評価点の数値は安定する。しかしユーザが高い評価点を与えた書籍であっても、システムが高い予測評価点で推薦することは少なかった。またユーザの評価入力点とシステムの予測評価点に差が見られた。

4. むすび

本研究では、青空文庫にある書籍の推薦システムを開発した。これによりユーザに未読の書籍を推薦できることを確認した。しかし評価実験の結果、ユーザが評価入力した評価点と、システムが計算した予測評価点に差が見られ、システムがユーザの嗜好に近いと予測する精度である、推薦精度に多くの改善の余地が残された。

参考文献

- [1] Dietmar Jannach 情報推薦システム入門 共立出版株式会社 (2012)
- [2] Toby Segaran 集合知プログラミング 株式会社オライリー・ジャパン (2008)