

レビュー情報を用いた情報推薦

高橋将史[†] 青野雅樹[‡]

豊橋技術科学大学[‡]

1 はじめに

近年、家庭用ゲームは多様化しており、様々な種類のゲームが市場に出回っている。ゲームは、プレイする人によって好みが変わる傾向があり、アマゾンなどで採用されている協調フィルタリング手法による情報推薦^[1]では、個々のユーザの嗜好に沿った推薦は難しい。

そこで本論文では、ゲームのレビュー（感想）をインターネット上から収集、解析し、レビュー中に出現した特徴的な単語などを手掛かりにした情報推薦が有効と考えた。そして、レビュー文書内の手掛かりとなる特徴的な単語などから素性を定義し、個人の嗜好を的確に反映することを試みた。また、この素性をベースとし、ユーザにゲームの推薦を行えるシステムを開発し、実験したので報告する。

2 提案手法

2.1 概要

本提案システムの概観を図 1 に示す。情報推薦を行うための前処理として、全てのレビューをまとめたものを文書ベクトルとし、文書ベクトルに出現する単語集合を W とする。単語の重みづけモデルとして TF モデルを使用した。同時に、各ゲームのレビューをまとめたゲームベクトル、出現した各単語のみの単語ベクトルも作成しておく。

しかし、レビュー文書中には多種多様な単語が出現するため、非常にスパースなベクトル行列になってしまう。そこで、提案手法では、次元削減を行ってやることで、単語空間を本来の単語次元 $|W|$ からそれより十分小さい $k(k \ll |W|)$ 次元に次元削減する。

2.2 次元削減

本提案手法では、次元削減手法として LSI (Latent Semantic Indexing)^[2]を使用した。LSI とは、高次元の空間で定義される単語文書行列を射影により低次元空間の行列で近似する技術である。近似にあたっては行列の特異値分解

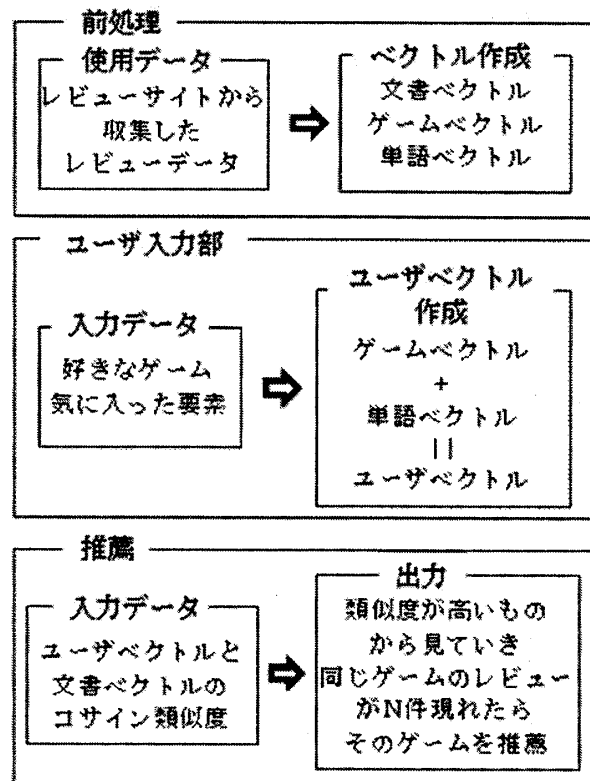


図1 提案システムの概観

(SVD: Singular Value Decomposition) を使用して、相互に関連のありそうな単語の次元を特定の次元に縮退させることで全体の次元を削減するという方針を採っており、これにより行列計算の計算量を下げつつ類義語を同じ次元に持ってくる事ができる手法である。本提案手法では、次元削減において、 $A = U \Sigma V^T$ で表わされる特異値分解の行列の V (右特異行列) を使用した。この処理を文書ベクトル、ゲームベクトル、単語ベクトルの全てのベクトルに対して前処理として V をかけてやっておくことで、次元削減を行っておく。

2.3 ユーザ入力部

ユーザの入力としては、好きなゲームと、そのゲームでユーザが気に入った要素について入力してもらう。そして好きなゲームで選ばれたゲーム g 個と要素 f 個を以下の式で組み合わせることでユーザベクトルとした。

$$Usr = \alpha \sum_i^g G_i + \beta \sum_j^f W_j \quad (1)$$

ここで、 G_i はゲームベクトル、 W_j は単語ベクトル、 α と β は重み定数である。

2.3 推薦

前処理、ユーザ入力部において作成した文書ベクトルとユーザベクトルのコサイン類似度を求めることによって推薦を行う。以下が推薦のモデルの式である。

$$\text{similarity}(d_i, Usr) = \frac{d_i \cdot Usr}{(|d_i| \times |Usr|)} \quad (2)$$

上式によって各レビューとユーザベクトル Usr の類似度を計算し、全てのレビューに対して計算が終了したら、類似度を降順にソートする。そして、上から見て同じゲームのレビューが N 件出た時点で、そのゲームを推薦する。

3 実験

3.1 使用データ

NintendoDS プレイヤー参加型レビューサイト NintendoDS mk2 からレビューデータを取得した。ゲームは 300 種類、レビューデータ総数は、18848 個、12009 単語である。

3.2 被験者実験による評価

次元削減数 k を、8, 16, 32, 64, 128 の 5 パターンとそれぞれ類似度の絶対値を使用するかどうかの組み合わせの全てで 10 パターンについて評価をしてもらった。評価人数は 17 人で、各パターンで、10 個の推薦されたゲームを評価してもらった。評価は、やってみたいか (やったことある人は楽しかったか) どうかを 5 段階で評価してもらった。また、同じゲームのレビューが $N=5$ 件出た時点で、そのゲームを推薦することとした。

3.3 実験結果

表 1, 2 に実験結果を示す。結果から、全ての場合において、類似度の絶対値を考慮しない方が、ユーザにとって嬉しいゲーム(評価値が 4 点か 5 点)を推薦できた。また図 1 で、最も高いスコアとなったのは 128 次元に次元削減して類似度の絶対値を考慮しない場合で、170 個のうち 95 個のゲームが嬉しいゲームと評価され、約 56% の確率で嬉しいゲームを推薦できている。結果を上位 3 件までで見ると、16 次元に削減して絶対値を考慮しない場合では、約 63% の確率で嬉しいゲームを推薦できた。また、次元削減で

表 1 結果 (上位 10 件中)

次元数	嬉しい	嬉しいくない
8	77	58
8_ABS	77	51
16	89	51
16_ABS	80	55
32	88	48
32_ABS	79	51
64	89	51
64_ABS	69	64
128	95	50
128_ABS	75	61

表 2 結果 (上位 3 件中)

次元数	嬉しい	嬉しいくない
8	30	12
8_ABS	24	16
16	32	11
16_ABS	23	18
32	28	13
32_ABS	23	16
64	30	12
64_ABS	20	21
128	29	14
128_ABS	23	19

次元数を下げすぎると図 1 では、精度が悪くなった。

4 まとめ

本稿では、ゲームのレビュー情報を用いて、文書ベクトルに対して、次元削減をすることで計算量的にも少ないリソースで推薦が可能となった。また、被験者実験から、LSI で次元削減を行っても、文書の特性は保持できており、推薦に利用できることが判明した。

今後は、推薦結果を次回の推薦に活かすためのフィードバック機能を付けることである。

参考文献

- [1] Greg Linden, Brent Smith, and Jeremy York. ; "Amazon.com recommendations: Item-to-item collaborative filtering.", IEEE Internet Computing, Vol. 07, No. 1, pp. 76-80, 2003.
- [2] Scott Deerwester, Susan T. Dumals, George W. Furnas and Thomas K. Landauer ; "Indexing by Latent Semantic Analysis", JOURNAL OF AMERICAN SOCIETY FOR INFORMATION SCIENCE, pp.391-407, 1990