

ユーザ間類似度と模擬育種法を用いた 受験校選定支援システムの提案

A proposal of support system for a choosing high schools
based on collaborative filtering and simulated breeding

飯沼浩之*1

Hiroyuki IINUMA

菱山玲子*1

Reiko HISHIYAMA

*1早稲田大学大学院創造理工学研究科経営システム工学専攻

Graduate School of Creative Science and Engineering, Waseda University

When students choose a high school, there are some problems that they don't have a clear reason for their choice and a multidirectional choosing high school support is a burden for teachers. Then I aimed at the construction of the system which put collaborative filtering and simulated breeding together. The purpose is reducing teachers' burden and considering students' taste. And I showed the possibility of the suggestion system by the experiment for active high school examinees and their teachers.

1. 研究背景と目的

中学を卒業するとき、すなわち義務教育課程を修了するとき、生徒は初めて自分の進路について考え自らその道を決める。そして、大半の生徒は高等学校への進学を選択している。進学においても、どの高校へ進学するか選択する必要がある。さらに選択した高校へ進学するためには入学試験を通過しなければならない。そういった意味では多くの人々にとって、高校受験が人生で最初の岐路ともいえる重要な節目となる。

これに際し受験校選定の基準として、各予備校などが開催している模擬試験を受験することは、もはや常識と言えよう。しかしながら、この模擬試験による合格判定は試験の点数から算出された偏差値のみによるものに留まっている。個人の嗜好が多様化してきている現代において、受験校選定もまた、多様な角度からの考察によるべきである。現状、生徒は受験校選定の基準として、模擬試験の結果に大きな比重を置いている。そのため、受験校が本当に自分の嗜好に合致した高校であるかが曖昧になり、受験勉強に対するモチベーションの低下や、受験校を確定させる時期の遅れに繋がることが多々ある。

本来必要とされる多角的な受験校選定支援は、教師が行っているといえる。教師は生徒と対話していく中で生徒の嗜好や実力を読み取り、それを自らの経験や知識と照らし合わせ、それぞれの生徒に適した受験校を推薦している。しかし、これには多大な時間が必要であり、生徒に対して効率の良い支援を行っているとは言い難い。また、教師にとっても生徒一人一人についてこうした推薦を行っていくことは、負担が大きい。

本研究では、こうした背景を踏まえ、生徒の嗜好や教師が蓄積した経験と知識を取り込んだ多角的な判断による受験校選定支援システムの提案を目的とする。

2. 関連研究

現状の教師による受験校選定支援や生徒の嗜好を取り込むことをシステムで表現するためには、過去の生徒データを参考にし、現役の生徒と対話を行いながら、レベル・嗜好の合致した高校を探していく過程を盛り込む必要がある。

他のユーザデータを参考にし、対象ユーザに推薦を行う研究としては、Resnickらによる GroupLens の実験 [Resnick 94]

連絡先: 飯沼浩之, 早稲田大学大学院創造理工学研究科,
h.inuma@asagi.waseda.jp

が挙げられる。これは過去の NetNews 記事への評価を基に協調フィルタリングを行い、ユーザが好むであろう NetNews 記事を推薦するという実験である。推薦を行うまでの過程は大きく2つに分けられる。

1 つ目は、推薦を行う対象ユーザと他ユーザとの類似度計算である。GroupLens では、この類似度計算にピアソン相関が用いられた。ユーザ A に対するユーザ B の類似度 S_{AB} は、以下の式 (1) で求められる。なお、記事 i はユーザ A、ユーザ B が共に評価値を付けた記事とし、 E_{Ai} はユーザ A の記事 i に対する評価値、 \bar{E}_A はユーザ A の記事に対する評価値の平均値とする。

$$S_{AB} = \frac{\sum_i (E_{Ai} - \bar{E}_A)(E_{Bi} - \bar{E}_B)}{\sqrt{\sum_i (E_{Ai} - \bar{E}_A)^2} \sqrt{\sum_i (E_{Bi} - \bar{E}_B)^2}} \quad (1)$$

2 つめは、評価値の予測である。推薦を行う対象ユーザが評価をつけていない記事に対して、式 (1) で求めた類似度による加重平均によって評価値予測を行う。ユーザ A の未評価記事 j に対する予測評価値 E_{Aj} は、以下の式 (2) で求められる。なお、ユーザ J は記事 j を評価済みの他ユーザとする。

$$E_{Aj} = \bar{E}_A + \frac{\sum_J S_{AJ}(E_{Jj} - \bar{E}_J)}{\sum_J |S_{AJ}|} \quad (2)$$

これらの計算を行って未評価記事の評価値を予測し、その値が高い記事を推薦する。この手法は、単純かつ高精度であることから、映画の推薦やソフトウェア機能推薦 [大杉ら 04] など、幅広い領域に応用されている。

また、ユーザの嗜好を取り入れる手法としては、遺伝的アルゴリズムからの派生手法である模擬育種法 [畠見 94] が挙げられる。通常の遺伝的アルゴリズムでは、あらかじめ定めた評価関数を基に個体集合を選択し、交叉を行い、突然変異を加えて、次世代集合を生成するという操作を最終世代まで繰り返す。一方模擬育種法では、この遺伝的アルゴリズムの流れにおける選択のステップを評価関数ではなくユーザが行うことによって、ユーザの嗜好を次世代集合に反映させる。

この手法は人間の好みや主観に基づいた、設計や創造などを行う場合に有効と考えられ、その応用として石野ら [石野ら 97] は、一般消費財の新規商品設計時に実施されるアンケートデータから商品イメージと商品属性の関係を決定木によって明確に

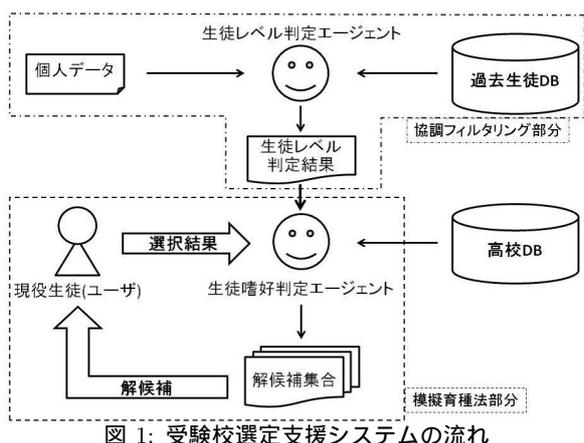


図 1: 受験校選定支援システムの流れ

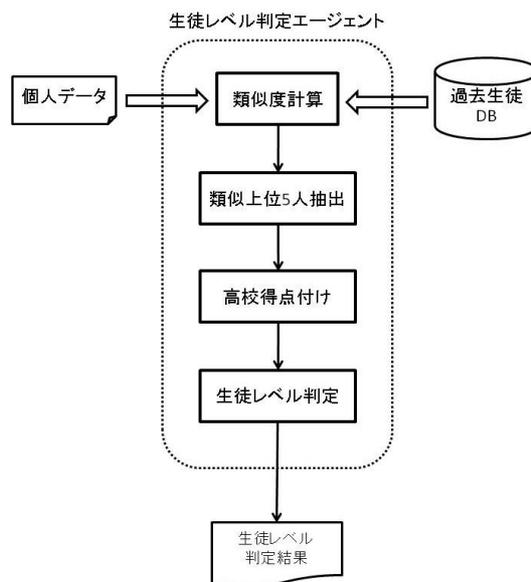


図 2: 生徒レベル判定エージェントの動き

するという研究に、この手法と帰納学習を組み合わせ用いた。その他にも、類似画像検索 [北本 97] やインテリアレイアウト [是永ら 00] など芸術分野や工業、教育などにも応用されている。

しかしながら、これらを組み合わせた手法や、それを受験校選定支援に適用した研究は未だ為されていない。

3. 提案

3.1 提案モデル

多角的な受験校選定支援を行うためには、生徒の嗜好や教師の経験と知識を取り入れることが必要となる。教師の経験と知識とは、過去の受験生指導から蓄積されるものであり、現役受験生がどの過去受験生に似ているのか、その過去受験生はどの高校に進学できたのか、といったものがそれにあたる。この経験と知識をシステムに取り込むために、類似ユーザを基に推薦を行う手法である協調フィルタリングを導入する。また、生徒の嗜好を盛り込んだ受験校選定支援を行うために、個人の主観を取り入れる手法である模擬育種法を導入する。つまり本研究では、協調フィルタリングと模擬育種法を組み合わせ手法を提案する。

提案システムではまず、協調フィルタリングを用い、現役生徒の総合的なレベルを判定する。次に、生徒のレベルにあった高校群（候補集合）を生成し、それらを個体として模擬育種法を行うことで、生徒のレベル・嗜好に合った高校を推薦する。その全体の流れを図 1 に示す。

3.2 生徒レベル判定エージェント

生徒の総合的なレベルを判定するために、ユーザ間類似度を用いた協調フィルタリングを行う。生徒レベル判定エージェントは、入力された個人データと過去生徒 DB とを参照し、類似度の高い過去生徒を探す。その際、生徒に関するより多くのデータを要素に含めることで、総合的なレベルの判定を行うことができると考えられる。本研究では類似度として、それぞれ値を標準化した、偏差値 (2 項目)、内申点 (3 項目)、塾内模試平均点 (5 項目) の計 10 項目から計算したユークリッド距離を用いた。これに、過去生徒の卒業年度で重み付けを加えた。なお偏差値には、埼玉県で最も一般的な模擬試験である北辰テストのものを用いた。生徒レベル判定エージェントの動きを図 2 に示す。

Step1 入力された個人データと過去生徒データベースを参照し、過去生徒との類似度を計算する。類似度計算の式は

以下のとおりである。

$$S_{i,j} = \left\{ \sum (h_{i,k} - h_{j,k})^2 + \sum (n_{i,k} - n_{j,k})^2 + \sum (m_{i,k} - m_{j,k})^2 \right\}^{\frac{1}{2}} * \{1 + (10 - g_j)/10\} \quad (3)$$

$S_{i,j}$: ユーザ i のユーザ j に対する類似度

$h_{i,k}$: ユーザ i の 3 教科, 5 教科の平均偏差値

$n_{i,k}$: ユーザ i の各年次の内申点

$m_{i,k}$: ユーザ i の各教科の塾内模試平均点

g_j : ユーザ j の卒業年度

Step2 Step1 で計算された類似度の高いかつ公立高校に合格した過去生徒上位 5 人を抽出し、順位付けする。

Step3 類似度 1 位の過去生徒が合格した公立高校に 5 点, 2 位には 4 点, 3 位には 3 点, 4 位には 2 点, 5 位には 1 点とし、高校に得点付けする。

Step4 性別が適切かつ最も得点が高い高校を現役生徒の適正校とみなし、その高校の偏差値を現役生徒の学力レベルとして、生徒レベル判定結果を生成する。

3.3 生徒嗜好判定エージェント

生徒の嗜好を取り込むために、高校を個体とした模擬育種法を行う。「平成 24 年度用埼玉県高校受験案内」を基にした高校データを、遺伝子型で表し登録したものが高校 DB である。高校 DB には埼玉県にある全高校データ (430 件) が蓄積されており、それぞれ偏差値、塾最寄駅からの通学時間、盛んな部活動、設備などのデータが含まれている。

生徒嗜好判定エージェントは、生徒レベル判定結果と高校 DB とを参照し、生徒の嗜好に合った高校を探す。図 3 にそのアルゴリズムを示した。なお、本研究では突然変異率を 10%, 1 世代の個体数を 4 とし、最終世代を定めず、収束が見られるかどうかを終了基準とした。

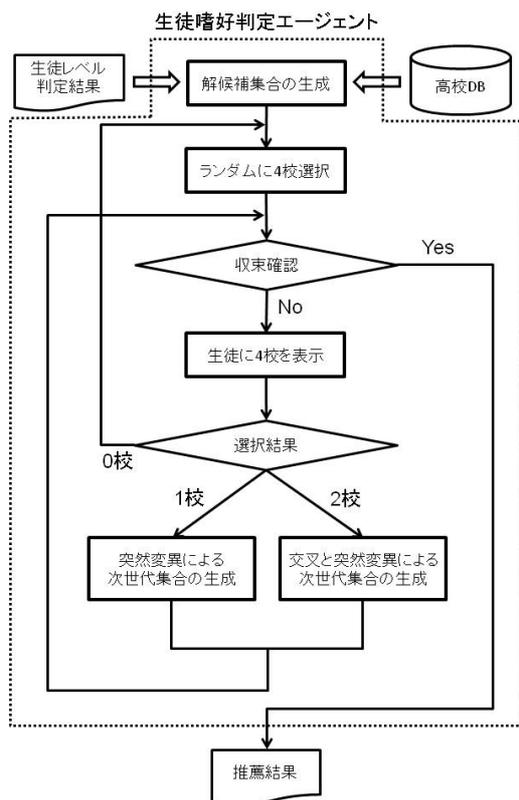


図 3: 生徒嗜好判定エージェントの動き



図 4: 嗜好診断画面

- 形式：1 人当たり 15 分程度の面談形式

実験結果を明白にするため面談前には、志望校と受験校選
びで重視する要素についての事前アンケートに回答してもら
った。更に、システム使用中や推薦結果が出た後にも、聞き取り
アンケートを行なった。また、現役受験生を担当している教師
には、提案システムによる推薦結果をまとめたものを見ていた
だき、アンケートに回答していただいた。

4.2 生徒に対する実験の結果と考察

表 1 は、提案システムの対象生徒総数と嗜好診断実施数、事
前アンケートで答えた受験校選びで重視する要素、事前アン
ケートでは挙がらなかったが嗜好診断中には重視した要素の生
徒 1 人当たりの数をまとめたものである。

表 1: 嗜好診断実施数と受験校選びで重視する要素数

対象生徒 総数	嗜好診断 実施数	要素数 (個/人)	新たな要素数 (個/人)
23 人	12 人	3.57	1.75

嗜好診断実施生徒数は、受験校がすでに確定しており意思
の堅い生徒や、大学附属高校（主に東京都）志望の生徒、シス
テム自体に興味を示さなかった生徒などには嗜好診断を使用し
なかったため 12 人となった。意思の堅い生徒や大学附属高校
志望の生徒は、自らの嗜好に従い主観的に受験校を選定してい
るため、嗜好診断を使用する必要がないと判断した。システム
自体に興味を示さなかった生徒は、現状の志望校にも明確な志
望理由がなく、自ら受験校を選ぼうとする意志が弱かった。意
志の弱い生徒に自身の嗜好を考慮した主観的な受験校選定を促
すことは今後の課題といえる。

受験校選びで重視する要素数は、事前アンケートでは 1 人当
たり平均 3.57 個であり、嗜好診断中には新たに平均 1.75 個増
加した。事前アンケート結果の 3.57 個と比較すると、約 49%
要素数が増加したことになる。増加した要素としては、修学
旅行先や制服の様式など、嗜好が分かれる要素があった。修学
旅行先に関しては海外の方がいいといった意見が多く、それに
伴って海外研修制度の有無が新たな要素になることがあった。
これらのことより、提案システムの嗜好診断を使用することが
より多角的に学校を選ぶことに繋がると考えられる。

次に、推薦結果に対するアンケートをまとめると表 2 のよ
うになった。

- Step1 生徒レベル判定結果に合ったレベルの高校全てを解候補集合として生成する。
- Step2 解候補集合のなかからランダムに 4 校を選択する。
- Step3 生成された高校に収束が見られたら終了とし、収束が見られない以外は Step4 へ移動する。
- Step4 生徒に 4 校のデータを表示する。
- Step5 生徒に個体を選んでもらい、その選択校数によって以下のように移動する。

- 0 校 Step2 へ移動する。
- 1 校 それを親として突然変異のみを行い、次世代の 4 校を生成して Step3 へ移動する。
- 2 校 それを親として一点交叉と突然変異を行い、次世代の 4 校を生成して Step3 へ移動する。

4. 実験

4.1 実験内容

提案システムの有用性を確認するため、受験校選定支援シ
ステムを実装し、実際に受験生に提案システムを使用してもら
う実験を行った。主観診断のインターフェースを図 4 に示した。
また実験内容は以下の通りである。

- 人数：現役高校受験生 23 人，受験生担当教師 5 人
- 期間：2011 年 11 月 30 日～12 月 14 日
- 場所：ラサール進学教室、ラサールスタディールーム

表 2: 推薦結果に対するアンケート

受験しようと思わない	興味有り	受験しようと思う
45%	32%	23%

「受験しようと思う」と回答された高校数は全体の 23% であり、これは提案システムが受験校選定を支援しているといえ、非常に良い推薦を行えた割合だと考えられる。また、「興味有り」と回答された学校数は全体の 32% であり、これは生徒に気付きを与え、選択の幅を広げる推薦を行えた割合だと考えられる。この点に関しては実験時期が関わると考えられる。実験時期がより早ければ、「興味有り」と答えた高校を生徒が更に調べることができ、受験校となる可能性が出てくるからである。今回の実験では、「受験しようと思う」「興味有り」を合わせた 55% の推薦に成果があったといえる。

しかしながら、45% の推薦結果は「受験しようと思わない」と回答された。この要因として、名前でも高校を選んでおり聞いたことのない高校には興味を示さなかった生徒がいたことや、実験を行うより前に、実際に高校に行ってみて自分に合わない雰囲気の高い高校だと感じた生徒がいたことなどが挙げられる。前者は主観的に選択する意志が比較的弱く、誰もが知っている高校という選択肢の中での選択に留まっている。これらの生徒への対策としては、早期から主観的に選択する練習をさせることが挙げられる。雰囲気に関してはデータとして表現できる範囲に限られるため、後者への対策は今後の課題としたい。

4.3 教師に対するアンケートの結果と考察

教師に対するアンケート項目は、「推薦結果は適切だと思うか」「このシステムは教師としての仕事の負担軽減に繋がると思うか」とした。それぞれ、思う場合は 5、思わない場合は 1 とし、5 段階で評価していただいた。表 3 は、教師に対するアンケート結果をまとめたものである。

表 3: 教師に対するアンケート結果

教師名	A	B	C	D	E
結果は適切か	4	3	3	2	2
軽減に繋がるか	1	3	3	4	3

まず、推薦結果に対する評価として、「判定が厳しい」との意見が多数あった。この要因として、レベル判定に使用した偏差値データが、10 月までの結果であったことが挙げられる。教師の立場としては、入試直前まで生徒の成績を伸ばさなければならぬということを考慮すると、10 月時点での判定は厳しいものになってしまった可能性がある。この対策として、個人データ入力の際に生徒の受験までの伸びを考慮に入れることが挙げられる。教師間では生徒の伸びについて普段から議論していることが多く、受験までの伸びを考慮し入力する個人データを変動させることは可能と考えられる。

次に、負担軽減に繋がるかに対する評価としては、「受験校選定の判断要因になり得る」という意見が複数挙げられた。これらのことから、生徒の可能性を考慮し判定精度を高めることが、教師の負担軽減に繋がると考えられる。

その他の意見としてはまず、嗜好診断の際に親の意見も取り入れてはどうかといったものがあった。生徒には親を交え教師と三者面談をする機会がある。これら場面において提案システムを使用することで、親の意見も取り入れることは可能である。また、受験校についてあまり考えていない中学 1,2 年生に対して提案システムを使用するのはどうかといった意見も見られた。現状では、入力する個人データは固定されており、3 年生になってから得られるデータも多いため、1,2 年生には適用できない。しかし、類似度計算に使用するパラメータを変更することで、これらの生徒にも対応することは可能となる。

5. まとめと今後の課題

本研究では、ユーザ間類似度による協調フィルタリングと模擬育種法を用い、教師の知識や生徒の主観を取り入れた多角的な受験校選定支援システムを提案した。そして、現役受験生とその教師による実験によって、提案システムの有用性を検証した。その結果、生徒にとっては自身の嗜好に気付きを与え主観的な受験校選定を促し、教師にとっては業務の負担を減らし効率化を図るシステムの可能性を示した。

今後の課題としてまず、提案システムを使用する時期の検討が挙げられる。生徒が受験校を変えることができるほど気持ちに余裕があり、かつ教師が生徒の受験までの伸びを予想できる時期が思わしい。あるいは受験生になる前、つまり中学 1, 2 年生の段階で使用してみることも新たな可能性につながると思われる。次に、嗜好診断において、強い嗜好を持った生徒ほど収束が遅いまたは収束しないといった事象があったことから、それら生徒に対応するため嗜好診断における学校選択時の強制の導入も課題として挙げられる。また、現状では個人データが全て揃っている必要があるので穴抜けデータへの対応や、教師に対するアンケートで挙げられた判定精度の向上、他分野への応用として就職活動に適用することなども課題といえる。

参考文献

- [Resnick 94] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl: GroupLens: an open architecture for collaborative filtering of netnews, *Proceedings of ACM conference on Computer supported cooperative work (CSCW '94)*, pp.175-186, 1994.
- [大杉ら 04] 大杉直樹, 門田暁人, 森崎修司, 松本健一: 協調フィルタリングに基づくソフトウェア機能推薦システム, *情報処理学会論文誌*, Vol.45, No.1, pp.267-278, 2004.
- [畷見 94] 畷見達夫: 遺伝的アルゴリズムとコンピュータグラフィクスアート, *人工知能学会誌*, Vol.9, No.4, pp.42-47, 1994.
- [石野ら 97] 石野洋子, 寺野隆雄: 模擬育種法と帰納学習を適用したマーケティング情報分析, *人工知能学会誌*, Vol.12, No.1, pp.121-131, 1997.
- [北本 97] 北本朝展: パイプライン型遺伝的アルゴリズムによる類似画像検索パラメータの対話的な最適化, *情報処理学会第 55 回全国大会*, pp.457-458, 1997.
- [是永ら 00] 是永基樹, 萩原将文: 対話型進化計算法によるインタリレイアウト支援システム, *情報処理学会論文誌*, Vol.41, No.11, pp.3152-3160, 2000.