

Q-learning を用いた精油ブレンド法の一考察

坂本萌佳¹ 高田雅美¹

概要: 精油は、植物から抽出される芳香物質である。数種類の精油をブレンドすることで、好みの香りを作ることが可能である。しかし、精油には多くの種類があるため、適切な精油を選択することは容易ではない。そのための方法として、強化学習の一種である Q-learning を採用して、個人の好みに合わせた精油のブレンドを自動的に選択する手法が提案されている。本稿では、提案された手法を用いて、異なる初期値で実行した場合に初期値が与える影響を検証する。

A Study of an Essential Oil Blending Method Using Q-learning

MOEKA SAKAMOTO^{†1} MASAMI TAKATA^{†1}

1. はじめに

精油とは、植物から抽出される揮発性の芳香物質である。精油は 200 種類以上の植物から作られ、抽出される植物の違いによって香りや効能が変わる。心身を癒すアロマセラピーは、精油の香りと効能を利用したものである。アロマセラピーの原点は、お香と呼ばれる芳香植物で、薬の原料として使われている[1]。アロマセラピーの代表的な効果としては、安眠、ストレス解消、リフレッシュ効果、頭痛や冷え性の改善などが挙げられる。アロマセラピーには、芳香浴、マッサージ、アロマバスなどがあり、それぞれに好ましい精油のブレンドがある。精油やアロマセラピーの研究により、日常生活に密着したものとなり、美容や健康増進、介護、医療、予防医学など、幅広い分野で注目されている。

ユーザはそれぞれ好きな香りが異なる。さらに、体調や気候などのさまざまな条件によっても好みの香りは異なる。例えば、夜、気持ちを落ち着かせたいときに好む香りと、湿度も気温も高い昼間に好む香りが同じとは限らない。このように、香りの好みは状況に応じて変化する。

日常生活における香りの利用に関する調査[2]では、被験者の約 90% が「癒し効果を感じる」と回答している。しかし、約 10% のユーザはアロマセラピーの効果を感じないと回答している。これらの結果から、アロマオイルの使用において、ユーザの好みに合わない香りを選択した場合、癒しを感じないだけでなく、頭痛などの健康への悪影響が出る可能性があることがわかっている。

精油は、複数の香りを組み合わせることで効果が変化する。また、精油をブレンドすることで、目的や好みに合ったより良い香りを作り出すことができる。しかし、精油に

は多くの種類があり、ユーザの心理的・身体的な状態によって香りの好み異なるため、精油の知識が乏しい、あるいは全くないユーザが自分の好みに合った最適な精油を選ぶことは容易ではない。自動的に効果的な精油を選択するためには、香りの好みの個人差や環境の変化に対応できることが必要である。そのための方法として、Q-learning を用いた精油ブレンド法[3]が提案されている。

本稿では、ユーザにとって最適な精油のブレンドを選択・提案する手法[3]をもとに、異なる初期値で実行した場合に、初期値が与える影響を検証する。

2. 精油ブレンド推奨法

強化学習とは、数値化された報酬を最大にするためにどのような動作をとるべきか学習する人工知能の手法の 1 つである。通常の機械学習のようにとるべき行動は教えられず、どのような行動をとれば最終的な価値を最大化できるのかを見つけ出す必要がある。強化学習は、学習に多くの試行を必要とするため、膨大な計算時間とコストを必要とする。そのため、部分的に強化学習を活用することが多い。強化学習は、不確実な環境とゴール探求型のエージェントとの相互作用の中に成り立つ。エージェントは、環境の様子を入力として取り入れ、自らの行動によって環境に変化を与える。

エージェントと環境の関係には、強化学習を構成する方策(policy)、報酬関数(reward function)、価値関数(value function)、環境のモデル(model)の 4 つの要素がある。

方策は、エージェントの行動を定義する。つまり、環境から得られたある状態において、その時とるべき行動を決めておくということである。

報酬関数は、学習問題における目標を定義する。報酬関数は、ある状態に対する行動の結果を報酬として表現することで、その状態の望ましさを表す。エージェントにとつ

¹ 奈良女子大学
Nara Women's University, Nara, Nara 630-8506, JAPAN

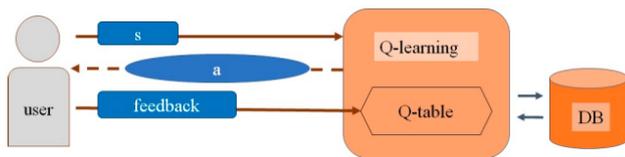


図 1 関係図

何が悪い出来事であるかを定義する。報酬関数が即時的なものであるのに対し、価値関数は最終的に何が悪いのかを決定するものである。報酬はある時点での本質的な望ましさを表している。報酬はある時点における固有の望ましさを表すが、価値はこれから得られるであろう報酬を考慮した長期的な望ましさを表すというものである。

環境のモデルは環境の挙動を模倣するような何かを指す。これは必ずしも全ての強化学習に存在するとは限らない。モデルは実際に起こる前の可能な将来の状況を考慮して動作を決定する方法を意味するプランニングのために用いられる。ある環境における状態 s において、行動 a を選択する価値（状態行動価値）を $Q(s, a)$ として、 (Q, s, a) の値が最大化されるよう行動を選択するよう制御する。

エージェントの現在の状態を t として、行動価値関数 $Q(s_t, a_t)$ は一般的に次の式で表される。

$$Q(s_t, a_t) = E_{s_{t+1}}(r_{t+1} + \gamma E_{a_{t+1}}(Q(s_{t+1}, a_{t+1}))) \quad (1)$$

ここで、 γ は将来の価値をどれだけ割り引くかを表す定数である。また、 r_{t+1} は報酬を表す。この Q 値を学習する基本的な手法として、動的計画法、モンテカルロ法、TD 学習の 3 つが挙げられる[4]。

図 1 は、実験方法の関係図である。エージェントはユーザとする。状態 s はユーザの幸福度を表す。アクション a として、推奨される精油のブレンドが返される。データベースには、ブレンドされる精油の情報を格納する。手順は以下の通りである。

- (1) Q-table の初期化
- (2) 状態 s をユーザが入力
- (3) Q-learning
- (4) アクション a として精油ブレンド情報をユーザに出力
- (5) ユーザからのフィードバック r_{t+1} を入力
- (6) Q-table の更新
- (7) n 回、手順 2 に戻る

Q-learning では、 Q の値を Q テーブルと呼ばれるテーブルに格納し、アクションを起こすたびに Q テーブルを更新することで学習を進めていく。Q-learning では、 Q の値は以下のように更新される。

$$Q(s_t, a_t) \leftarrow (1 - \alpha)Q(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1})) \quad (2)$$

(1)とは異なり、 Q 学習では期待値は次の状態に対する行動の最大値で表される。Q-learning では、 Q 値が最も大きい行動が優先的に選択される。しかし、学習の初期段階では、 Q 値がランダムに設定されるため、 Q 値の大きい行動だけが誤って選択されてしまう可能性がある。この問題を回避するために、 ϵ -Greedy 法[4]を採用する。 ϵ -Greedy 法では、定数 $\epsilon(0 \leq \epsilon \leq 1)$ を設定し、以下の手順で行う。

- (1) 選択肢 $p(0 \leq p \leq 1)$ をランダムに設定
- (2) $p \leq \epsilon$ の場合
アクションをランダムに選択
- (3) それ以外の場合
 Q 値の最も高いアクションを選択

これにより、最初の Q テーブルの影響を避けながら適切な学習を行うことが可能である。

3. 検証方法

3.1 概要

Q-learning を用いた精油ブレンド法[3]では、Q-learning を用いて、ユーザの幸福度に基づき精油のブレンドを推薦する手法を提案している。実験結果から、被験者ごとに異なる精油ブレンドが与えられていることが確認されている。本稿では、この実験で得られた結果を使用し、異なる初期値で実行した場合に初期値が与える影響を検証する。

Q テーブルは、幸福度と精油のブレンドを 2 次元配列で表したものである。 Q テーブルの初期値は、Q-learning を用いた精油ブレンド法で行われた実験の被験者が 100 回の学習を行った結果を使用する。ユーザは幸福度として 1 から 5 までの数字を入力する。ここでは、ブレンドの組み合わせとブレンド係数を精油のブレンドとして返す。

精油のブレンドについては 3.2 節で紹介する。フィードバックについては 3.3 節で説明する。

Q-learning の各反復では、フィードバックのために芳香浴を行う。連続して芳香浴をする場合、直前の芳香浴の影響を受けるため、香りの違いを識別することは困難である。そこで、連続して実験するのではなく、1 日に数回、決まった時間に Q-learning を行う。

3.2 精油

本実験では、アロマセラピー検定 2 級[1]の香りテストに採用されている以下の 9 種類の精油を用いる。

- Sweet orange
- Lemon
- Peppermint
- Rosemary
- Geranium

表 1 Fragrance1

essential oil	Sweet	Spicy	Bitter	Sour
Sweet orange	0.7	0.0	0.0	0.3
Lemon	0.0	0.0	0.1	0.9
Peppermint	0.0	1.0	0.0	0.0
Rosemary	0.0	0.8	0.2	0.0
Geranium	1.0	0.0	0.0	0.0
Lavender	0.5	0.5	0.0	0.0
Frankincense	0.0	0.8	0.2	0.0
Eucalyptus	0.2	0.4	0.4	0.0
Tea tree	0.5	0.0	0.5	0.0

表 2 Fragrance2

essential oil	Clean	Fresh	Flowely	Sharp	Soft
Sweet orange	0.0	0.5	0.0	0.0	0.5
Lemon	0.5	0.5	0.0	0.0	0.0
Peppermint	0.8	0.0	0.0	0.2	0.0
Rosemary	0.4	0.0	0.0	0.6	0.0
Geranium	0.2	0.0	0.6	0.0	0.2
Lavender	0.0	0.0	0.5	0.5	0.0
Frankincense	0.0	0.0	0.0	0.4	0.6
Eucalyptus	0.2	0.1	0.0	0.7	0.0
Tea tree	0.3	0.0	0.0	0.7	0.0

- Lavender
- Frankincense
- Eucalyptus
- Tea tree

表 1 は、香りを味覚的表現で表す。表 2 は、直感的な表現で香りを表す。

精油のブレンドは、9 種類の精油から 3 種類を選んだその組み合わせとする。精油ブレンドの香りを表す数値はその 3 種類の数値を足し合わせたものとする。

ブレンド方法には芳香浴を用いる。それぞれの精油を染み込ませたコットンをガラス瓶に入れ、提示される 3 種類の精油を同時に嗅ぐことで香りを評価する。3 つのコットンを使うことで、精油を混ぜる順番や混ぜることによる影響を避けることができる。精油のブレンド比率にはブレンドファクターと呼ばれる比率を採用する。ブレンドファクターは、香りの強さと成分を考慮した数値で、各精油につけられている。ブレンドファクターは、値が小さいほど香りが強く、大きいほど香りが弱い。本実験では、この数値を用いて、精油の滴数を決定している。

3.3 フィードバック法

提案手法では、ユーザからのフィードバックを報酬として採用している。フィードバックは香りに関する 2 つの質問で構成されている。

F1: 香りについてどう感じますか？

- － 好き
- － 普通

－ 嫌い

F2: どのような香りが好きですか？

- － フローラル系 (甘め)
- － 柑橘系 (酸っぱい)
- － ウッド系 (苦さ)
- － ミント系 (清涼感)

F1 は、ブレンドした精油の感想を聞くシンプルな質問である。それぞれの選択肢に対して、「好き」は+0.3、「普通」は+0.1、「嫌い」は-0.3 の報酬を設定する。

F2 は、F1 で「嫌い」を選択した場合のみ質問される。F2 は、香りの好みを問う項目である。ユーザが直感的に選択できるように、4 つの選択肢を用意する。つまり、F2 では、ユーザに理想的な香りを選んでもらうことになる。Q テーブルはこれらの回答に基づいて更新される。F2 では、関連する表 1 の値に 0.3 をかけた値を報酬とする。この報酬は、各精油を含むブレンドの Q 値を更新するために使用する。

4. 関連研究

本節では、学習による定式化の最適化に関する先行研究を紹介する。

[5]では、潤滑油添加剤の最適な配合を得るために、最適な誤差逆伝播法・ニューラルネットワークモデルと遺伝的アルゴリズムの組み合わせを提案した。誤差逆伝播法・ニューラルネットワークモデルの予測データと実際のテストデータを比較したところ、90%以上の相関係数が得られた。したがって、このモデルは潤滑油の性能を正確に予測することができる。遺伝的アルゴリズムと組み合わせた場合、すべての性能誤差は 5%以下であった。したがって、誤差逆伝播法・ニューラルネットワークモデルを遺伝的アルゴリズムと組み合わせで最適化することにより、正確な予測モデルを構築することができた。潤滑油の性能を予測する正確な複合モデルが得られた。このようにして、このモデルが最適な添加剤の配合を得ることができると実証された。

[6]では、機械学習とメタ・ヒューリスティックアルゴリズムに基づいて、コンクリートの混合比を最適化する多目的最適化手法を提案した。コンクリートの混合比を最適化するためには、多くの変数を持つ複数の目的を、非線形性の高い制約条件の下で同時に最適化する必要がある。ここでは変数はコンクリートの成分を表し、目的には強度、コスト、スランプなどである。この研究では、誤差逆伝播法・ニューラルネットワーク、サポートベクトル回帰、回帰木アルゴリズムなどの複数の機械学習モデルの性能を評価した。ベクトル回帰、回帰木アルゴリズム、ランダムフォレストアルゴリズム、k-近傍分類器、ロジスティック回帰、重回帰など複数の機械学習モデルの性能を比較しました。比較の結果、強度のような連続的なデータでは誤差逆伝播

法・ニューラルネットワークの方が優れた性能を発揮し、ランダムフォレストアルゴリズムは、スランプのような離散的なデータに対して高い予測精度を示した。次に、多目的粒子群最適化アルゴリズムを用いて複数の機械学習モデルの混合比率を最適化するために使用した。その結果、高性能コンクリートの2目的混合物最適化問題と、プラスチックコンクリートの3目的混合物最適化問題のパレートフロントが得られることが確認された。したがって、この手法は、建設段階でのフェーズでの意思決定を容易にするための設計ガイドとしての役割を果たすことができる。

また、[7]では、自然を模倣した最適化アルゴリズムである遺伝的アルゴリズムと、機械学習を用いた予測モデルを組み合わせた手法が提案されている。材料科学の分野では、適切な特性や性能を持つ材料を設計することが求められている。

[8]では、再生可能な代替燃料として、廃植物油をディーゼルに混合して圧縮点火エンジンに混合した。この目的のために、3因子3水準の実験配列設計を用いた応答曲面法を用いて、エンジン性能を予測した。予測の結果、廃油を混合した燃料を使用することで、ディーゼル単独よりもCOとNOxの排出量が少なく、経済的であることが示された。このブレンド燃料は、既存のブレンド燃料とは異なる新しいアプローチである。

5. まとめ

強化学習の一種であるQ-learningを採用して、個人の好みに合わせた精油のブレンドを自動的に選択する手法が提案されている。本稿では、提案された手法を用いて、異なる初期値で実行した場合に初期値が与える影響を検証している。初期値を変えても、問題なく個人の好みに合わせた精油ブレンドを提案することができる。

参考文献

- [1] “(公社) 日本アロマ環境協会” .
<https://www.aromakankyo.or.jp/>, (参照 2021-11-12).
- [2] 川本利恵子, 阿南あゆみ, 長聡子, 中尾久子, 宮園真美, 木下由美子, 金岡麻希, 潮みゆ. 日常生活における香りに関する影響要因の検討. 応用心理学研究, 2013, Vol. 39, no.1, p. 25-32.
- [3] Kyoko Nakano, Masami Takata. Proposal of an Essential Oil Blending Method using Q-learning. *Advances in Parallel & Distributed Processing, and Applications*. 印刷中
- [4] 三上貞芳, 皆川 雅章. 強化学習. 森北出版, 2000.
- [5] Yu, T., Yin, P., Zhang, W., Song, Y. and Zhang, X.: A compounding-model comprising back propagation neural network and genetic algorithm for performance prediction of bio-based lubricant blending with functional additives, *Industrial Lubrication and Tribology*, Vol.73 No.2, pp.246-252 (2021).
- [6] Zhang, J., Huang, Y., Wang, Y., and Ma, G.: Multi-objective optimization of concrete mixture proportions using machine learning and metaheuristic algorithms, *Construction and Building Materials*, Vo.253 (2020)
<https://www.journals.elsevier.com/construction-and-building-materials> Last accessed 1 November 2021
- [7] Kim, C., Batra, R., Chen, L., Tran, H., and Ramprasad, R.: Polymer design using genetic algorithm and machine learning, *Computational Materials Science*, Vol.186 (2021)
<https://www.journals.elsevier.com/computational-materials-science> Last accessed 1 November 2021
- [8] Dey, P., and Ray, S.: Optimization of Waste Vegetable Oil 菜油 Blends for Engine Performance: A Response Surface Approach, *Arabian Journal for Science and Engineering*, Vol.45, pp.7725-7739 (2020).