

服装推薦のための気象特徴量に基づいた服装指数の推定

佐藤優多

Yuta Sato

法政大学情報科学部コンピュータ科学科

yuta.sato.9n@stu.hosei.ac.jp

Abstract

This paper aims to estimate a clothing index based on weather factors to help with clothing Index. Initially, the aim was to create a personalised system based on user preferences, but later the focus shifted to estimating the Clothing Index published by the Japan Weather Association. This index guides clothing selection based on weather conditions. The study developed an estimation equation involving variables such as temperature, humidity and wind speed, which was then subjected to regression analysis and comparison of individual sensory tendencies with the general index. The validity of the estimated index was verified through experiments and many values of error were confirmed to be less than 5.00. The results were then also obtained when adapted to the personal clothing index. The paper said that one year of individual data would be collected for future conclusions.

1 まえがき

元来、本研究では「ユーザー選好情報を用いたパーソナライズシステム」を掲げていた。この目標は、ユーザーとのやり取りからそのやり取りに関する情報を収集し、よりユーザーに特化した回答を行うシステムである。

ユーザーモデルによるユーザー選好情報を用いたパーソナライズシステムの具体的な例として同じ条件の天気や気温の日に、どのような服装で過ごすか質問をすると暑がりのユーザーには涼しい服装を提案し寒がりのユーザーには温かい服装を提案するようなシステムが挙げられる。

本研究を進めるにあたって、ChatGPTによる自然言語処理及びユーザーへ出力する文章の生成とユーザーモデルを用いたパーソナライズによるユーザーの嗜好管理を組み合わせたシステムを目指していた。

しかし、服装の選定に関して服装指数の算出が線形でないため算出が困難であったことや服装の評価を行う際にまずは服装指数について深く掘り下げる必要があると判断し、本研究では日本気象協会の発表する服装指数の推定を行う。

そのため、本研究の目標として「服装推薦のための気象特徴量に基づいた服装指数の推定」を定める。

服装指数の推定に際して、最初に日本気象協会の発表する服装指数を推定する算出式を作成する。はじめに体感温度に用いられる変数を個別に用いて服装指数の推定式を組む。その後、回帰によって重みを求める。さらに体感温度に重きを置いた場合の服装指数の算出式を作成する。その後、個人の体感の傾向を反映した状態の服装指数の算出式を回帰によって求め、両者の差を比較してどちらの式を提案手法で用いるのかを検討

する。

服装指数を推定出来た場合には、それらを評価して推定した式の妥当性を調べる。また、服装指数の推定を行う前に、服装指数を推定してそれを改良する必要があるかの予備実験を行う。

その後、算出した服装指数の妥当性を検証するために評価指標を用いた評価とそれに対する考察を行う。

2 現状の問題点の整理と従来研究の取り組み

本研究を進めるにあたって現状の問題点を把握する。はじめに、本研究では各個人に直接服装の提案を行うことはせずに、服装指数という指標を回答する事で服装の提案を行う。

このような処置をとった理由として、服装の数値化が難しい点と実際の服装を選ぶ際に回答先の人間に関する情報を扱う必要がある点が挙げられる。特に評価実験を行うにあたって、服装という主観的な評価に基づいて評価が分かれてしまう問題を扱うため、客観的根拠として数値を算出することが望ましい。数値を扱う事が出来た場合には提案手法が既存の手法と比べてどれくらい優れているのか、劣っているのかを示す事が容易となる。

また、服装の好みやにはその服を所持しているか、現在着用可能かどうか等の状況をモデリングする必要が出てくる。このような手段を用いることで、服装の提案をする際により膨大なデータが必要になってしまう。このような問題を回避するためにも服装指数という指標を提案することで、回答先の人間に関するモデリングを省略出来る。

2.1 現状の服装指数の問題点

ここで現状の問題点として、日本気象協会の発表する服装指数 [4] に従って服を着て出かけたが、実際にはその服装では暑すぎる、または寒すぎる問題が挙げられる。

これは、服装指数という万人に向けて発表された指標が個人の感覚に沿わないものとなってしまっている点が原因として考えられる。もちろん、この服装指数の通りに服を着ることで快適な一日を過ごす事が出来る人物も存在するであろうが、今回は服装指数が自身に合わないと感じている人物に焦点を置いて研究を進める。

また、今回は気象情報を基に服装指数を算出するため、目的地の部屋や移動に用いた公共交通機関の空調の具合については考慮せずに研究を行う。そのため、本研究では初めに日本気象協会の発表する服装指数の算出式を推定し、その後個人毎の差異を反映できるような数式を組み立てる事を目標とする。

2.2 従来手法としての服装指数

そもそも服装指数とは日本気象協会の発表しているその日に着ていく服装を算出している指数のことを指す [4]。この服装指数は主に体感温度を基に決定される。

服装指数は、その日の朝、昼、夜の予想気温から、どんな服

装が適しているかを示すサービスである。服装指数は 10~100 までの 10 段階に分けた数値で発表される。服装指数の数値が低いほど寒くなるため厚着になり、数値が高いほど暑くなるため薄着になる。

服装指数はあくまでも目安であるため個人毎に感覚の違いが発生してしまう事は妥当である。

2.2.1 体感温度について

体感温度とは、人間がとある条件下で感じる温度感覚の事を示す。体感温度の算出にはミスナール体感温度の改良版である NET(net effective temperature)[1] を用いる。NET は以下の式 1 のように算出される。

$$NET = 37 - \frac{37 - t}{0.68 - 0.0014h + \frac{1}{1.76 + 1.4v^{0.75}}} - 0.29t(1 - \frac{h}{100}) \quad (1)$$

t = 温度 (°C), v = 風速 (m/s), h = 湿度 (%)

体感温度には温暖な地域の算出に特化したものや寒冷な地域の算出に特化した物が存在するが、日本のような寒暖差の激しい地域ではそのような体感温度を用いることは難しい。

しかし NET の場合はそのような地域でも体感温度を求められるように改良されたものであるため、日本のような地域でも比較的信頼して使える。

2.3 研究の流れの設定

本研究を進めるにあたって研究の順序を決めて研究を行う。これは、研究を進めるにあたって明確にしておくべき問題点がいくつか存在し、それを解消しておく必要があるためだ。主な研究順序は以下の通りである。

1. 服装指数推定の妥当性検証
2. 服装指数の推定
3. 予報と実測の差の検証
4. 推定式の誤差の検証
5. 個人向けの服装指数算出

3 服装指数の推定と実験

本研究では日本気象協会の発表する服装指数の算出式の推定を行う。服装指数の算出式の推定を行う前に予備実験として既存の服装指数を用いた実験を行い、服装指数の算出式を推定する必要があるかを検証する。服装指数の算出式の推定の必要性があることを確認できた後に、実際に服装指数の算出式の推定へ移る。

その後、推定された算出式から計算された服装指数と日本気象協会の発表する服装指数の誤差や決定係数による評価を行う。

3.1 予備実験

初めに、本研究を行うにあたって服装指数の算出式の推定の必要性を確かめるために予備実験を行う。予備実験の内容は日本気象協会の発表する服装指数に従って実際に服を着て出かけ、その時の気温の感じ方を主観的評価でまとめるものである。

主観的評価をする際には各段階で設定した服装を試し、それが快適なものになるかを検証する。この時、洗濯の都合などで服装が変わってしまう事を防ぐため、同じ服を必要数用意して実験に臨むこととする。

表 1: 日本気象協会が発表している服装指数と服装の関係

服装指数	服装指数の指す意味
10	ぶるぶる、何を着ても寒い！！
20	ダウンジャケットでしっかり防寒
30	コートを着ないと結構寒いなあ
40	裏地付トレンチコートがおすすめ
50	薄手のジャケットを羽織ろう
60	長袖シャツ・カットソーで快適に
70	半袖+カーディガンで温度調節を
80	半袖 T シャツ一枚で過ごせる暑さ
90	ノースリーブでもかなり暑い！！
100	暑さ対策必須！何を着ても暑い！

表 2: 服装指数と予備実験用に用意した実際に着て行く服装の関係

服装指数	選択した服装
10	長袖 T シャツ, ダウンジャケット, 手袋, ニット帽
20	長袖 T シャツ, ダウンジャケット
30	長袖 T シャツ, 暖かめなパーカー
40	長袖 T シャツ, ジャケット

3.1.1 特定個人向けの服装指数と服装の対応

服装指数に関する実験を行う際に、実際に所持している服装から服装指数の指示に可能な限り近いものを用意し、着服する。

この時、着ていく服装と服装指数の対応は日本気象協会より発表された指針があるためそれを参考にする。

日本気象協会が発表している服装指数についてのデータ [4] 及び服装指数と服装の対応をそれぞれ表 1、表 2 にまとめた。

また、実際の服装の決定に際して、夏服をすでに処分してしまった都合上冬服の対応できる範囲のみの定義となっている。

上記の関係性を特定個人向けの服装指数と服装の対応として定義し、予備実験に臨む。

3.1.2 予備実験の実験結果

予備実験は 1 月 10 日から 1 月 12 日の 3 日間で行われた。予備実験の際には先述した服装に従い外出し、その結果を記した。予備実験の結果として、服装指数と体感した服装の感想ではほとんどの場合で-10 程のズレが生じ、多い時で-20 程の差が生じていることが確認できたが、服装指数が 50 になると自身の感じた感覚と服装指数の差異がなくなっていることが確認できた。また、この値は服装指数から一律で数値を出し引きすることで実現できるものではない。

そのため、本研究ではこのような差異を埋め、より個人の好みに近い服装を算出できる服装指数を提案する。

3.2 服装指数の推定

本研究では服装指数を算出する都合上、日本気象協会が発表している服装指数の求め方を参考に必要がある。しかし、日本気象協会は服装指数の算出式を掲示していないため、これらは自力で推察する必要がある。

日本気象協会によると服装指数は体感気温を基に決められているとの記載がある [4] ため、体感温度を構成する気温、湿度、風速とその日の天気を用いて服装指数を推察できないかを試みる。

服装指数の推定には質的予想変数を用いた回帰分析を行い、

表 3: 2 変数と 4 変数の算出式の差異

日付	RMSE の二乗誤差	MAE の絶対誤差
1/10	0.291	1.617
1/12	-0.018	-0.069
1/15	-0.013	-0.057
1/24	-0.101	6.223
1/26	-0.081	2.665
平均	0.001	1.105

推定した式から算出した服装指数と日本気象協会の発表する服装指数の誤差や決定係数を基にその推定式が妥当なものかを評価する。

この時、服装指数に用いた天気に関しては実数として扱えるものではないため、ダミー変数を用いて数値化した。ダミー変数とは天気や性別などの数値によって表現する事が出来ない定性的なデータを説明変数に組み込む際に用いられる変数のことを指す

。そこで、本研究ではダミー変数に用いる天気を晴れ、曇り、雨、雪の四通りと定義し、それに従ってダミー変数を 2 つ用意した。また、それぞれの数値に以下のような意味を持たせた。

各変数は $Weather_0$ 及び $Weather_1$ がその時間の天気を表すダミー変数として用いられる。この二つのダミー変数は晴れ、曇り、雨、雪の 4 つの天気を数値としてあらわすために用いられる。

この際、 $Weather_0$ は大まかな天気の分類を行う。天気が比較的良い場合、つまり晴れと曇りの場合は $Weather_0$ の値を 1 とし、天気が比較敵悪い場合、つまり雨と雪の場合は $Weather_0$ の値を 0 とする。

次のダミー変数である $Weather_1$ は、 $Weather_0$ で分類した天気をさらに細分化する役割がある。 $Weather_0$ で分類した天気の中で状態の良いほうを 1、状態の悪いほうを 0 とする。

これにより、晴れの場合には $Weather_0 = 1, Weather_1 = 1$ といった具合に天気とダミー変数の対応が取れる。

ここで、より直感的なダミー変数の設定として晴れ、曇り、雨、雪の各天気にそれぞれ $Weather_0, Weather_3$ を割り当てる方法について検討する。この方法でダミー変数を設定すると晴れの場合には $Weather_0 = 1, Weather_1 = 0, Weather_2 = 0, Weather_3 = 0$ といった具合に天気とダミー変数の対応が取れる。

次に、ダミー変数を 2 つとる方法とダミー変数を 4 つとる方法の差を比べて、どちらの方法を本研究で採用するかを検討する。この比較を行う理由として、推定式内にダミー変数が多く入りすぎた場合には将来的な予測精度が下がってしまうことが懸念点であるためだ。

そこで、検討に用いる指標は後述する RMSE(式 5) と MAE(式 6) である。この 2 つの方法の差は RMSE の値の平均が 0.0010 MAE の値の平均が 1.105 となった。この結果を見るとダミー変数を 2 つとる方法とダミー変数を 4 つとる方法の両者の性能差に大きな差が観測できないことがわかる。これにより、今回は 2 つのダミー変数を用いる場合でも 4 つのダミー変数を用いる場合でも予測の精度に変わりがないことが確認できた。そのため、今回は 4 つのダミー変数を用いる方法を提案手法として採用する。

また、 $Temperature$ はその時間の温度 [°C] を示し、

$Humidity$ はその時間の湿度 [%] を示す。そして $Wind$ はその時間の風速 [m/s] を示している。これらの値は実数であるため、ダミー変数を用いずにそのまま数値として扱う。

ここで、風の向く方向に注目し、風向きを変数として用いる。風向きも天気同様そのまま数値化することは難しい。そのため、東西南北のダミー変数を作成した。作成した変数はそれぞれ WD_N, WD_S, WD_E, WD_W で、それを組み合わせることで 16 段階の方角を設定する。ダミー変数に方角を落とし込む際には北のダミー変数 WD_N の場合、北=1, 北東=0.5, 東=0、逆に東のダミー変数 WD_E の場合には北=0, 北東=0.5, 東=1 といった具合に 0 から 1 の間の数を用いることで東北東などの細かい方角を表現する。また、方角のようなサイクルデータの回帰を行う方法についてまとめた論文 [2] も存在するが、初めに本研究の方法を用いて回帰を行うが、問題が発生した場合には参考研究 [2] の方法を用いて回帰を行うこととする。

このような各変数にそれぞれ重み変数 $w_0 \sim w_{10}$ を用いる。これらの重み変数の値を変更することで服装指数の式を推定する。服装指数の式を推定するために用いる式は以下の式 2 である。

$$C = w_0 * Weather_0 + w_1 * Weather_1 + w_2 * Weather_2 + w_3 * Weather_3 + w_4 * Temperature + w_5 * Humidity + w_6 * Wind + \beta \quad (2)$$

この数式を用いて日本気象協会の発表する服装指数との差異が小さくなる $w_0 \sim w_4$ を回帰によって求めたところ、 $w_0 = 0.547, w_1 = 0.565, w_2 = 0.393, w_3 = -1.505, w_4 = 1.844, w_5 = 0.064, w_6 = -2.812, w_7 = 2.007, w_8 = 2.608, w_9 = 3.088, w_{10} = 2.442, \beta = 17.815$ となり、決定係数は $R^2 = 0.923$ であった。

この式を見るに、服装指数の算出には体感温度の他に天気による影響も大きな割合を占める事が取れる。また、気温の影響度は気温が 1 °C 上がるにつれて服装指数が 1.844 上がる事が分かる。

また、個人的な感覚に基づいた服装指数の算出を行えるような式を作るためにも、日本気象協会の発表する服装指数以外にも自身の感じた服装指数を目的変数として算出した式を作成する。

さらに、服装指数は体感温度を基に算出しているという情報があるため、一般的な体感温度の算出式であるミスナル体感温度を改良した NET[5] を服装指数の算出式に適応した場合の式を作成する。

ここで、式 2 と式 3 のどちらのほうが精度がよくなるかを比べ、より精度が良くなるほうを実際の算出に用いることとする。比較する際に用いる式は気温、湿度、風速を式 1 に変更した式 3 である。

$$C = w_0 * Weather_0 + w_1 * Weather_1 + w_2 * NET + \beta \quad (3)$$

こちらの重みはそれぞれ $w_0 = 1.858, w_1 = 1.179, w_2 = 1.225, w_3 = -4.263, w_4 = 5.54, w_5 = 9.16, w_6 = 11.134, w_7 = 8.616, w_8 = 1.391, \beta = 31.52$ ととなり、決定係数は $R^2 = 0.899$ であった。

式 2 と式 3 の比較の結果、体感温度自身を用いるのではなく、体感温度の要素をそれぞれ推定式に用いたほうが誤差が少なく

なったため、服装指数の算出式は線形であることが分かった。そのため、式 2 を用いて服装指数の算出式を推定する。

4 データセットの作成

服装指数の算出式は回帰による重み付けにより行われる。そのため、回帰に用いるデータを収集する必要がある。本研究ではデータの収集方法として、その日の気象情報と服装指数を毎日記録する方法と過去の情報を収集する方法が挙げられる。そこで今回は、直近の気象情報と服装指数をテスト用データとして収集し、過去の気象情報や服装指数をトレーニング用データとして収集する。

4.1 過去の天気予報の収集

過去の天気予報の収集にあたって、法政大学南館の図書館に保存されている朝日新聞の天気予報を収集する。天気予報の収集先として朝日新聞を選んだ理由としては 3 時間ごとの天気が記されている点が挙げられる。

3 時間ごとの天気予報というのは法政大学が取得している新聞社の中で最も細かいものであった。以上の理由により、朝日新聞の天気予報を取得した。また、天気予報の取得範囲は 6 月 16 日から 12 月 15 日までの 6 ヶ月間である。これは法政大学に保存されている新聞紙の最長保存期間が 6 ヶ月であるためである。

また、天気の予報については 3 時間ごとの予報であったが、気温に関しては最高気温と最低気温の情報しか発表されていないため実際の気温の変動を基に予報されていたであろう気温の推移を予測する。

気温の予測方法については以下の式 4 である。この式は実際に観測された最高気温を予報された最高気温に合わせ、その後実際に観測された最低気温を予報された最低気温に一致するようにスケールリングする。そうすることで、最高気温と最低気温の情報から実際に予報されていたであろう気温の推移を予測する事が出来る。これらの情報から、服装指数の回帰のための情報を収集していく。

$$\text{Temperature} = \alpha + (\beta + \gamma) - \alpha * \delta \quad (4)$$

α = 予報の最高気温

β = 実際の気温

γ = α - 実際の最高気温

$\delta = \frac{\text{実際の最低気温} + \gamma}{\text{予報の最低気温}}$

4.2 過去の服装指数の収集

過去の服装指数の収集にあたって、ウェブサイトのアーカイブを保存するサイトを用いる。このウェブサイトは過去のウェブサイトを保存しているウェブサイトである。

また、天気予報は「東京」の天気を取得しているが、この「東京」は東京都千代田区の北の丸公園で観測されたデータを基に予報しているため、服装指数の取得先は必然的に千代田区のものとなる。

ここでアーカイブによる服装指数の収集に関する問題が発生する。それが、アーカイブの保存間隔である。アーカイブの保存間隔は毎日ではないため、必然的にデータがまばらになってしまう。この時、千代田区以外の服装指数を取得する事で可能な限り穴を埋める事が出来るが、それではデータの信頼性が落ちてしまうため今回は千代田区のみ服装指数を用いる。

表 4: 日本気象協会の発表する服装指数と実測値を基にして算出された服装指数との各指標における誤差

対象	RMSE	MAE
1/10	1.360	5.763
1/12	1.401	5.259
1/13	1.156	4.325
1/15	0.953	3.438
平均	1.120	4.301

4.3 予報値からの服装指数と実測値からの服装指数の差

ここで、服装指数はすべて予報を基に算出されたものであることに注目する。これは先述したように服装の選出が基本的に朝に行われるためである。ここで、実際に測定された天気や気温などの情報を用いて服装指数の算出式を推定する。実際に測定された情報を基に推測された服装指数の算出式と予報を基に推定された算出式を比べる。

実測値を基に算出された誤差は以下の表 4 である。

表 4 と表 5 を比べると、予報の値から求めた服装指数と実測値から求めた服装指数の差はほとんど変わらないことがわかる。そのため、予測の範疇であった服装指数を実測値と組み合わせて算出することが可能であると示す事が出来た。

4.4 過去の気象情報の収集

先述したセクションで天気予報による服装指数の算出と実際の気象情報を下にした服装指数の算出に大きな差がないことが確認できた。そのため、過去の気象情報と過去に発表された服装指数を収集することでデータセットの拡大を行う。過去の気象情報と服装指数を収集することで、夏に発表された服装指数の情報を手に入れるられるため、回帰に用いるデータの幅が増える。

データ収集の結果、約 3000 時間分のデータを集められた。収集したデータは主に沖縄などの温暖な地域と北海道などの寒冷な地域である。これらを対象とした理由は、温暖な気候で発表される服装指数の情報が必要であった点と寒冷な気温及び雪の天気の際の情報が必要であった点が挙げられる。

温暖な気候の地域では夏に服装指数が 90 近くになることが多く、東京ではあまり観測できない服装指数になる。また、北海道などの寒冷な地域では雪の天気の際の情報が多く手に入る。これは、関東地方を中心にしたデータだと雪の天気の情報が手に入りにくいいため、雪のデータを増やす事に有効であるといえる。

5 評価実験

本研究で推定された服装指数が日本気象協会の発表する服装指数と比較して妥当な式であるかを評価する。また、日本気象協会の発表する服装指数以外にも自身の感じた服装指数を目的変数として算出した式、NET を基に算出した式も同様に評価する。これは日本気象協会の発表する服装指数以外にも算出できると検証する事で個人に特化した式の推定を行う事が目的である。

5.1 算出した服装指数の評価方法

推定した服装指数の評価方法として指標を用いた評価を行う。評価に用いる指標は RMSE,MAE[3] である。

表 5: 日本気象協会の発表する服装指数と算出された服装指数 (オリジナル) との各指標における誤差

対象	RMSE	MAE
1/10	0.785	2.964
1/12	0.545	5.357
1/15	0.996	3.964
1/24	2.190	2.357
1/26	1.054	1.178
平均	1.172	3.554

また、評価指標はそれぞれ RMSE が Root Mean Squared Error、MAE が Mean Absolute Error を意味する。

5.1.1 RMSE

RMSE は残差の二乗の平均を意味する Mean Squared Error に平方根をつけたものである。MSE は実際の値とモデルによる予測値との誤差の平均値を示している。しかしこの算出方法には欠点が存在する。計算結果が二乗されているために直感的な理解が難しい点である。計算結果が二乗されているため、誤差を測る前後の値の単位が変わってしまい、そのまま比較することが難しくなる。

その問題を解決する方法として平方根を付ける事が挙げられる。平方根をつけることにより、Mean Squared Error では値が二乗された状態となったためにほかの数値と比較しづらい問題を解消している。MAE の計算式は以下の式 5 である。

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (5)$$

5.1.2 MAE

次に MAE は残差の絶対値の平均を示す。残差とは予測値と実際の値との差の絶対値の平均を示す指標である。

この MAE は、他の複雑な誤差指標と比較してよりシンプルな方法で誤差の大きさを算出するものとなっている。MAE は、各データポイントにおける予測値と実際の値との差異を絶対値に変換しそれらの平均を取ることで計算する。

この手法により、誤差の平均的な大きさを直感的に理解しやすく、モデルの予測精度を評価する際に有用な情報を提供する。また、MAE は外れ値の影響を受けにくいという特性も持っておりデータに大きな外れ値が含まれる場合でも安定した誤差評価が可能である。MAE の計算式は以下の式 6 である。

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (6)$$

5.2 評価の結果

上記の評価方法を用いて 1 月 10 日から 1 月 13 日までの日本気象協会が発表した服装指数と推定した服装指数との誤差を表 5 にまとめた。同様に自身の基準から測定した服装指数の算出も同時に行い、2 つの服装指数の算出について比較する。自身の基準から測定した服装指数は表 6 にまとめた。

この結果、ほぼ全ての日付で誤差が低くなる事が確認でき、推定された式の妥当性を証明できた。特に 1 月 10 日や 1 月 11 日は RMSE の値が 0.785, 0.777 と 1.0 を下回る数値となっている。RMSE は値が 0 に近づくほど誤差が少なくなるため、本研究で提案した服装指数の算出式は比較的精度の高い状態であると捉えられる。また、1 月 12 日の誤差が RMSE の場合

表 6: 自身の基準から測定した服装指数と算出された服装指数 (オリジナル) との各指標における誤差

対象	RMSE	MAE
1/10	1.816	7.083
1/12	1.502	4.583
1/15	1.178	3.333
1/24	2.357	9.166
1/26	1.102	2.916
平均	1.526	4.833

表 7: 日本気象協会の発表する服装指数と自身の体感した服装指数を基に算出された服装指数との各指標における誤差

対象	RMSE	MAE
1/10	1.033	4.245
1/12	2.333	7.487
1/15	0.479	1.875
1/22	2.403	8.307
1/26	1.212	5.294
平均	1.308	4.898

0.545、MAE で 5.357 といった具合に他の日程よりも大きくなってしまっている。

5.3 服装指数のパーソナライズ

これまでの記述から、服装指数を求める式の算出が出来ることが示された。そのため、次は個人に合わせた服装指数の算出を行う。これは対象となる個人が日々実際に感じた服装指数を記録し、その服装指数と気象情報を基に回帰を行うことで個人の服装指数を算出できるようにする。

個人向けの服装指数の算出式が推測され次第、その算出式を基にその日個人の服装の指数を算出し、それに従って出かけて体感した服装指数と比べる。その結果は表 7 である。

これにより、ほとんどの日付において個人に特化した服装指数の推定が可能であることが示されたが、一部誤差が大きくなる日付もあった。そのうちのひとつとして 1 月 22 日の誤差が、RMSE で 2.403 となり、MAE で 8.307 となってしまっている。

5.4 考察

今回の実験では、提案した推定式による服装指数の推定においてはおおむねの誤差が RMSE の場合に 1 付近に収める事が出来た。

ここで、日本気象協会の発表する服装指数が 10 から 100 までの 10 毎の離散的な値である点に注目する。算出した服装指数は小数を含むものであるに対して評価する対象が離散的な値を持つため、これらの違いによる評価値の差異がどれほどの影響を与えるかを考慮しなくてはならない。

今回は評価のもととなった日本気象協会の発表する服装指数が間隔が 10 の指標であるため、誤差が 5 以下であれば許容範囲内であるとする。その理由として、離散的な値であるため四捨五入をすると、-5 から +4 までの値は同一の値として丸め込む事が出来る。そのため、四捨五入で同一の値になる範囲に閉じては十分な許容範囲内であると言える。

逆に、誤差が +5 を上回ると指数が 1 段階ずれることになるため、誤差が +5 を上回る地点にこそ問題が存在していると

考えられる。今回そのような誤差が目立った部分は服装指数が30から50付近の比較的寒いが暖かくもあるような気象の部分である。この部分は元来、寒さや暑さに振り切れている気象とは異なり、気温や風向きが少し変わるだけでも温度の感じ方が変わってしまう繊細な部分である。

また、今回は服装指数の推定に用いた要素として天気、気温、湿度、風速、風向きを用いたが、更なる精度向上を目標にするならば、これらのパラメーターに観測可能な気象情報を加える事でさらに細かく調整を効かせる事が可能になると考えている。

5.4.1 服装指数のパーソナライズに関する考察

また、服装指数のパーソナライズについてはデータの収集を継続して行わない限りは性能の向上は見込めないと考える。これは、データの収集している時期が冬のみであるため、暖かい時期の個人の服装指数の感じ方のデータが無いことが挙げられる。これは回帰に用いるデータに暖かくなった際の情報が少ないまま回帰してしまったため、暖かい気象状況に適応しない重み付けとなってしまった。これにより、温暖な気象条件の場合の算出精度が低くなるのが要因である。これを解決するためには主に2つの手段がある。

1つ目は1年分の服装指数の感じ方を記録し、それを回帰用のデータとして扱う方法である。こちらのやり方は本研究で行った様に気象情報と個人の感じた服装指数の関係を回帰によって用いる方法である。

今回の研究でパーソナライズした服装指数の性能が日本気象協会の発表する服装指数の推定式の性能に劣る原因として、データの数の少なさと収集した範囲の狭さが挙げられる。日本気象協会の発表する服装指数の回帰に関するデータ数は4000時間程であったが、パーソナライズ用の服装指数のデータ数は200程であった。

ここでパーソナライズ用の服装指数のデータの問題点として、パーソナライズ用の服装指数のデータは実際に自分が感じた値を基に作成するため、作成できるデータ量が1日につき1日分が限度である点、データの収集を始めた時期から現在までのデータしか扱う事が出来ない点が挙げられる。

このような問題があるため、正確な回帰を行うためには日数をかけてデータを取得する必要がある。データを収集して4000時間分のデータ数に追いつくためには、1日24時間全てのデータを使うと仮定しても166日程の間データを取得する必要がある。これは、2日に1回の頻度で服装指数に対する感じ方を記録する必要があることを示している。そのため、1年分の服装指数の感じ方を記録するためには最低でも2日に1回の頻度で感じ方を記録する必要がある。

2つ目は一定期間の服装指数の感じ方を集め、その情報から個人の傾向を推測する方法である。こちらは1つ目の方法とは異なり、日本気象協会の発表する服装指数に対する個人の感じ方のずれを予測し、それによって個人の服装指数の感じ方をある程度測定するものである。

この方法は、日本気象協会の発表する服装指数と個人の服装指数に対する感じ方が、どちらも寒くなれば服装指数は下がり暑く成れば服装指数が上がる特徴をもつ事があげられる。そこで、個人の服装指数の感じ方が日本気象協会の発表する服装指数からどれくらいズレているのかがわかれば、1年文のデータを集めずともパーソナライズ用の服装指数のデータが作成できるのではないかと考えられる。

ここで、データを収集する際のポイントについて考える。最初に、服装指数のズレ方のタイプである。服装指数のズレ方は図1の3つのタイプに分けられる。

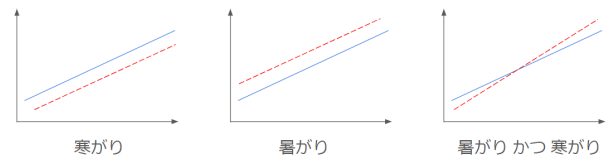


図1: 寒がり、暑がり、寒がりかつ暑がりのタイプ別による服装指数の感じ方のズレのイメージを共有するための簡易的な図

これらのタイプは服装指数の値に一括でマイナスされるか、一括でプラスされるか、どこかを境にプラスとマイナスが入れ替わるかによって分けられている。これらのタイプを判断するためには2つの方法がある。寒い気象条件の場合の服装指数と暑い気象条件の場合の服装指数を収集し、それらと日本気象協会の発表する服装指数の関係を調べる方法と、春や秋などの季節の服装指数から、日本気象協会の発表する服装指数を跨るかを調べる方法である。

前者の方法では、必然的に高い服装指数と低い服装指数に関する感じ方を記録する事が出来るため、回帰する際に必要になるデータの幅を確保することがメリットとして存在するが、デメリットとして服装指数の感じ方に関する記録を夏と冬の両方で行わなくてはならない点が挙げられる。そしてこの方法では夏から冬、または冬から夏の期間を要するため、この方法をとるなら前述した方法である2日に1回のペースで服装指数の感じ方を記録したほうが回帰における信頼性も向上するという問題点も存在する。

次に、後者の方法では収集する機関が四季の内の1つで済むため、前者の方法のように長期間のデータ収集期間を設けなくて済むメリットがあるが、デメリットとしてその季節以外の情報がないため、夏や冬などは必然的に回帰に用いたデータの範囲外になってしまう点が挙げられる。

これらのことから、一部の記録から服装指数のズレを予測する際にはあくまで予測によるものであるため、精度を保証できなくなる可能性があることを念頭に置く必要がある。

6 あとがき

本研究では服装指数の推定を行った。本研究では服装指数の推定を行う前に予備実験として既存の服装指数を用いた実験を行い、服装指数を推定する必要があるかを検証した。本研究では推定の必要性があることを確認した際には日本気象協会の発表する服装指数と自身の基準から測定した服装指数に-10から-20程のズレがあることが確認できた。

次に服装指数の推定を行った結果、服装指数推定に大きな影響のある変数が天気、気温、風速であり、それぞれの値として晴れの値が0.547、雪の値が-1.505、気温が0.393、風速の東が2.007、南で3.088となるなど、影響の大きい要素の重みが増加する事が確認できた。

最後に、推定された服装指数の算出式が妥当なものであるかを確認するために評価を行った。評価を行う際には推定された数式から算出された服装指数と日本気象協会の発表する服装指数の誤差や決定係数をRMSE、MAEといった指標を用いて評価した。評価の結果、大抵の場合においてRMSEの値が1

前後に収まることが確認できたため、推定された服装指数の式がある程度妥当なものであると示せた。

今後の課題としての1年分の個人の服装指数の感じ方のデータを集めることを示す。

参考文献

- [1] P.W. Li and S.T. Chan. Application of a weather stress index for alerting the public to stressful weather in hong kong. *Meteorological Applications*, 7:369–375, 2000.
- [2] Graham J. G. Upton, Kenneth A. Hickey, and Aaron Stallard. Regression models for cyclic data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 52(2):227–235, 2003.
- [3] Nikhil Mehta Maheswaran Sathiamoorthy Lichan Hong Ed Chi Wang-Cheng Kang*, Jianmo Ni* and Derek Zhiyuan Cheng. Do llms understand user preferences? evaluating llms on user rating prediction. *arXiv:2305.06474v1*, page 11, 2023.
- [4] 日本気象協会. 服装指数-日本気象協会. <https://tenki.jp/indexes/dress/>, 2023.
- [5] 伴野 明 and 山本 修平. 心理的要因による体感温度への影響の数値評価法. *電気学会論文誌E (センサ・マイクロマシン部門誌)*, 133(6):190–198, 2013.