

機械学習を用いた料理レシピ動画の難易度判定システムの設計

17se003 安藤佑基 17se079 住田彪迫

指導教員:野呂昌満

1 はじめに

近年レシピサイトや、YouTubeなどで料理レシピ動画が多く投稿されている。YouTubeなど、動画共有サービスの普及により、個人で撮影した料理レシピ動画を気軽に投稿できることが可能になった。料理レシピ動画は作り始めから出来上がりまでの調理工程を見ることができ、見るだけで調理の仕方がわかるものになっている。膨大に増えた料理レシピ動画によって、ユーザのどのような動画を見るかという選択肢は増えた一方でユーザは求める動画を探し辛くなったと考える。

料理レシピ動画が普及してきたが、現時点ではレシピ動画を見た際に、内容を見る前にわかりやすいか、わかりにくいと確認することができない。YouTube やクラシルやクックパッドなどの料理レシピ動画を扱っている動画サイトでは、ただ検索結果にヒットしたレシピ動画が並べられおり、本当にその動画が実際にユーザが料理を作るうえで、その動画がわかりやすいかどうかを動画を見る前に判断することができない。したがってユーザが料理レシピ動画をみる前にその動画がわかりにくい動画かわかりにくい動画を判断できればよいと考える。

本研究の目的は、料理レシピ動画を「わかりやすい動画」、「わかりにくい動画」に分類するニューラルネットワークの設計である。

本研究の技術的課題は以下の通りである。

1. ニューラルネットワークの設計における特徴量の決定
2. ニューラルネットワークの設計
3. 設計したニューラルネットワークの妥当性検証

定量的指標である特徴量を示し、機械学習を用いることで料理レシピ動画の難易度判定の精度を上げるニューラルネットワークの設計を目指す。

本研究より、レシピ動画から抽出した特徴量の入力から、料理レシピ動画の難易度判定をできる可能性を示唆できる。

2 先行研究

牧野 [1] は、調理動作の難易度に注目した。料理レシピの難易度を算出する方法を提案した。この研究では、小学校、中学校、高等学校、専門学校でそれぞれ使用されている家庭科の教科書に基づいて調理動作を定義した。この定義に基づいて料理レシピの難易度を計算した。

岩本 [2] らも調理動作の難易度に注目して料理レシピの難易度を算出する方法を提案した。この研究では、家庭料理技能検定に基づく調理動作の難易度別分類と、調理動作と重みによるスコア計算を用いて、料理レシピをランキン

グ化している。

矢嶋 [3] らは調理動作のみでなく食材の難易度も用いて料理レシピの難易度を算出する方法を提案した。この研究では、簡単に調理できるという視点から料理レシピを評価し、ユーザに簡単な料理レシピを推薦すつことを目的とした。

秋口 [4] ら料理番組をレシピ動画と定義した。テキストレシピから使用されている「食材」「調理器具」および「調理道具」の有無の3点を難易度の要素として抽出した。

秋口 [5] らは、料理番組をレシピ動画と定義した。レシピ動画内の「作業種類の多さ」「調理動作の並列性」「編集意図(カメラワーク)」3点を難易度の要素として抽出した。

3 ニューラルネットワークの設計

本章では、対象データと特徴量、ニューラルネットワークの設計について記述する。

3.1 システムの全体構成

本研究におけるシステムの全体図を以下の図1に示す。本研究において作成した、説明変数と正解データが入ったデータを入力する。レシピ動画を入力としない。

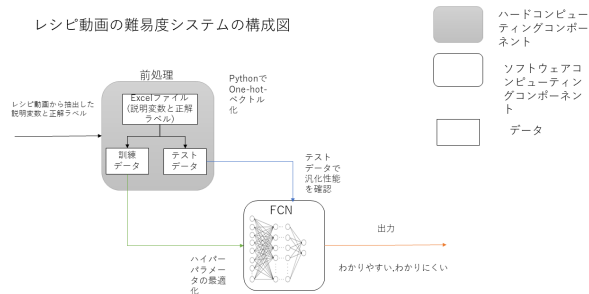


図1 全体の構成

前処理において、Excel ファイルに保存されている説明変数と正解ラベルを Python を用いて one-hot-ベクトル化し、学習データを訓練データとテストデータに分割する前処理を行う。FCN では訓練データを用いて、ハイパーパラメータの最適化を行う。テストデータを用いて汎化性能を確認する。動画が「わかりやすいか」、「わかりにくい」の分類結果を出力とする。緑色の線は訓練時のデータの流れ、青色の線はテスト時のデータの流れ、オレンジ色の線は判定時のデータの流れを表している。

この図における FCN は本研究における説明変数と正解ラベルを one-hot ベクトル化したものを入力とし、動画が「わかりやすいか」、「わかりにくい」を出力するニューラルネットワークである。本研究で抽出した特徴量は前処理

で扱った，説明変数と正解ラベルである．説明変数と正解ラベルはすべてカテゴリーデータであり，ニューラルネットワークへ入力する上で，one-hot ベクトル化する必要がある．

3.2 対象データと特徴量

本節では対象としたデータや設計したニューラルネットワークについて説明する．

3.2.1 対象データ

本研究ではレシピ動画を対象にし，難易度判定を行う．レシピにはテキストレシピとレシピ動画が存在する．テキストレシピは，料理の材料，調理工程，調理器具などが書かれているものである．レシピ動画は，テキストレシピの内容を動画化し，実際に調理する光景を撮影したものである．

レシピ動画と動画のわかりやすさの違いについて説明する．レシピ動画は上記で記述した通り，テキストレシピを動画化したものである．レシピ動画のわかりやすさは，正しくユーザに対して，調理方法が伝えられているかという観点で，その動画がわかりやすいかわかりにくいである．動画としてのわかりやすさは，動画の内容は関係なく，動画そのものが快適に見れるかどうかであると考えられる．

レシピ動画はレシピ通りに正しく調理が行われているかどうか重要であるが，見る人にとって見やすいことも重要である．レシピ通り調理を行っていたとしても，動画の見栄えが悪ければ，ユーザがそのレシピ動画の内容を理解することが難しい場合がある．

本研究では，レシピ動画と同じ難易度の要素を用いて，別々の料理を難易度判定することは難しいのではないかと考えた．別々の料理において，共通の難易度の要素は存在する．例えば，調理手順や扱う材料の g 数の説明があるかなどである．このように共通の難易度要素は確かに存在するが，料理ひとつひとつに特別な難易度の要素が存在するのではないかと考えた．例えば，チャーハンにおいて鍋を振ることである．鍋を振るといいう調理工程はチャーハン特有の調理方法であると考えられる．チャーハンのように固有の難易度が存在した場合，この要素を無視して難易度判定することは，レシピ動画の難易度判定の精度を下げることに繋がるのではないかと考える．よって，本研究では，対象とするレシピ動画の料理を限定する．本研究では，チャーハン動画を料理レシピ動画と定義した．

3.2.2 特徴量

チャーハン動画から抽出した特徴量は以下の 9 つである．チャーハン動画を複数見ることで，難易度判定に必要なであると予想される特徴量を抽出した．

- 主食材
- 食材の説明
- 何 g の説明

- 器具
- 鍋を振っているか
- 作り方の手順
- 動画を見て作りたいと思うか
- 食材の特異性
- チャーハン動画のわかりやすさ

各特徴量について説明する．主食材はチャーハンの目玉になる食材である．カニチャーハンであれば主食材はカニとなる．食材の説明は食材そのものの説明が動画でされているかいないかである．何 g の説明があるかは食材や調味料などが何 g 使用するか説明されているかどうかである．器具はチャーハン作成に使用している器具である．鍋を振っているかはチャーハンを作る際に鍋を振っているかどうかである．作り方の手順はチャーハンの基本的な作り方を定め，それ通りにチャーハンが作成されているかどうかである．動画を見て作りたいと思うかはチャーハン動画を見た人がそのチャーハンを実際に作ってみたいかどうかである．食材の特異性は主観でチャーハンに珍しい食材が使われているかどうかである．チャーハン動画のわかりやすさはチャーハン動画を見た人がチャーハン動画がわかりやすいかわかりにくいかどうかである．

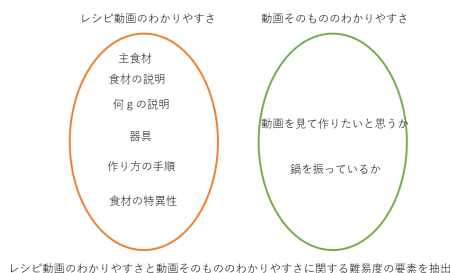


図 2 レシピ動画と動画そのものの特徴量

図 2 に本研究におけるレシピ動画と動画そのものの特徴量を図示する．本研究ではレシピ動画を対象としたことから，動画そのものの特徴量はあまり抽出していない．この中で動画としての特徴量を鍋を振っているか，動画を見て作りたいと思うかを抽出した．動画の内容だけで難易度判定は可能ではあるが，動画そのものの特徴量も抽出することで難易度判定の精度は上がると考える．

学習データの正解ラベル付けについて記述する．本研究では，チャーハン動画が「わかりやすい」か「わかりにくい」かを正解データとしている．人間が料理動画を見てわかりやすいかわかりにくいかを判断することは可能であるが，その判断は個人による判断が大きい．このことから他人とは判断結果が違う可能性がある．つまり，普遍的に料理レシピ動画を「わかりやすい動画」か「わかりにくい動画」かを判断することは簡単ではない．したがって複数人に動画を見てもらうことにより正解データに対して客観

性を持たせる。よって本研究では、まず人間が料理レシピ動画を見て、その動画を「わかりやすい動画」か「わかりにくい動画」かを判断し、その結果を学習させる。これより普遍的に「わかりやすい動画」か「わかりにくい動画」か分類できるニューラルネットワークの作成を目指す。

本研究は料理レシピ動画をわかりやすいか、わかりにくいかに分類するニューラルネットワークを設計することが目的である。よって本研究における説明変数を主食材、食材の説明、何gの説明、器具、鍋を振っているか、作り方の手順、動画を見て作りたいと思うか、食材の特異性とした。また目的変数をチャーハン動画のわかりやすさとした。

3.3 ニューラルネットワークの構造

本研究では、料理レシピ動画を「分かりやすい動画」、「わかりにくい動画」に分類するニューラルネットワークの設計をする。ニューラルネットワークの構造は以下の図3の通りである。図3のようにニューラルネットワークを作成した理由を以下に示す。

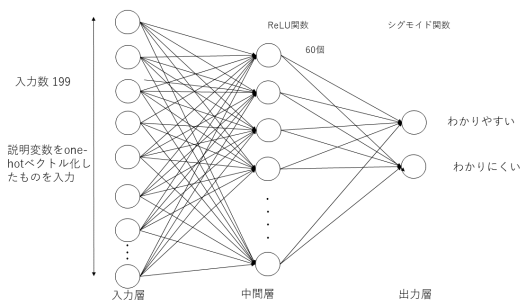


図3 ニューラルネットワークの構造

1. 入力ユニットが199

上記で説明した説明変数はすべてカテゴリーデータであることから、説明変数に対し one-hot ベクトル化を行った結果、入力数が199となった。

2. 中間層の総数、ユニット数と活性化関数がReLU関数の理由

実験により中間層の層数、ユニット数を決定する。中間層の層数とユニット数はハイパーパラメータである。実験にて最も分類精度が高い中間層の層数とユニット数を決定する。パラメータ推定における勾配の計算は中間層の活性化関数を微分した値を乗算するので、正の領域では微分した値が1になるReLU関数を用いることで勾配が小さくなることを防ぐ。

3. 出力層のユニット数とシグモイド関数を用いる理由

チャーハン動画が「わかりやすい」と「わかりにくい」の2分類をすることから出力ユニットは2つである。シグモイド関数は出力の値が0~1になることより確率として扱うことができ、2つのクラスに分類するとき出力層の活性化関数として用いられる。このこと

から本研究に適していると考え出力層の活性化関数にシグモイド関数を用いた。

4. FCNを用いる理由

入力として用いる、主食材、食材の説明、何gの説明、器具、鍋を振っているか、作り方の手順、動画を見て作りたいと思うか、食材の特異性は時系列データではなく、さらに局所特徴抽出可能であるかどうか不明であることから、汎用的なFCNを選択した。

4 実験

ニューラルネットワークを設計する上でハイパーパラメータを決める必要がある。

4.1 実験の概要

本研究では、ニューラルネットワークを設計するにあたり中間層のユニット、バッチサイズ、中間層の数を実験にて決定した。実験は指定した中間層のユニット数とバッチサイズ、中間層の層数を組み合わせて行った。学習データの総数が736であることからバッチサイズは32に固定した。

- 中間層のユニット数：70,100,150,200,300
- バッチサイズ：32
- 中間層の層数：1層、2層

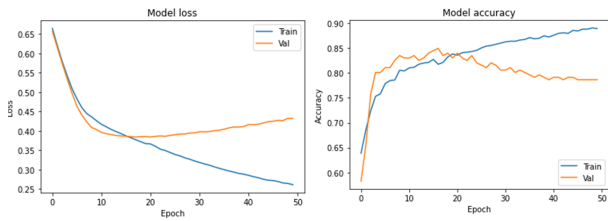
上記のハイパーパラメータの実験に加えて、主食材を説明変数に含めて行う実験と含めずに行う実験の2パターンの実験を行った。学習の精度を確かめる指標として、訓練データとテストデータに対する損失関数と評価関数の値を組み合わせ毎に導出した。2パターンの実験を行った理由は、データ数が少なくまた主食材における入力数が多いことから学習が収束しない可能性があったからである。主食材なしの場合ではニューラルネットワークに対する入力数が少なくなることより、中間層のユニット数を20,40,60,70,80,100,200の順に実験を行う。

4.2 実験結果

主食材ありの実験結果と主食材なしの実験結果は以下図4、図5の通りである。図3は主食材ありの場合で中間層1層、中間層のユニット数70の場合の図である。図4は主食材なしの場合で中間層1層、中間層のユニット数60の場合の図である。

テストした結果、主食材ありの場合、どの組み合わせにおいても損失関数が収束しなかった。損失関数のグラフにより過学習を起こしていると考えられる。

主食材無しの場合、どの組み合わせにおいても損失関数、評価関数の値に大きな変化は見られなかった。損失関数が減少し、ある値で収束したことより、学習は収束したと考えられる。実験結果より、中間層1層ユニット数60の場合にテストデータにおける評価関数の値が大きくなった。このことから本研究におけるニューラルネットワークの中間層は1層、中間層のユニット数を60個とした。



主食材あり中間層1層 ユニット数70 バッチ数32

図4 学習の損失と正答率 (主食材あり)

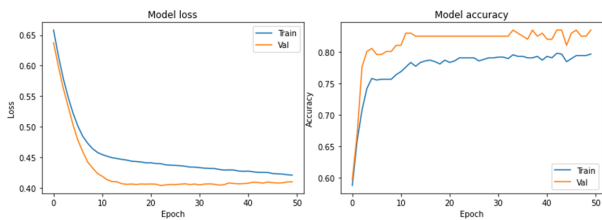


図5 学習の損失と正答率 (主食材なし)

5 考察

5.1 ニューラルネットワークについて

実験結果より主食材ありの場合過学習を起こした。過学習を起こした原因はデータの数に対して主食材種類が多いことから主食材一つあたりのデータ数が少ないことであると考えられる。データ総数が736個にも関わらず、主食材の入力数が169であった。多くデータを取ることができれば主食材ありの場合においても学習は収束すると考えられる。

主食材無しの場合、実験結果より設計したニューラルネットワークは妥当であるといえる。実験を通して、損失関数、評価関数の推移を確認した結果、主食材のなしの場合多くのハイパーパラメータの組み合わせにおいて、学習は収束しており、汎化性能を確認できた。中間層ユニット数60、中間層1層の場合81%の精度を得ることができた。これらのことより主食材なしの場合、設計したニューラルネットワークは妥当であると言える。

5.2 技術課題に関する考察

本研究では、特徴量を主食材、食材の説明、何gの説明、器具、鍋を振っているか、作り方の手順、動画を見て作りたいと思うか、食材の特異性、チャーハン動画のわかりやすさの9つを用いた。実験結果より主食材ありの場合において学習が収束しなかったことから、主食材は本研究の特徴量としては向いていなかったと考えられる。今回チャー

ハン動画から抽出した特徴量は動画内容による特徴量を多く用いた。動画そのものの特徴量を追加すれば、より分類の精度を上げることができると考える。例えば、動画の明るさであったり、動画の場面転換数などである。

ニューラルネットワーク妥当性の検証について記述する。妥当性を検証するのにテストデータを用いて、テストデータ時の評価関数の値を比較した。前述した通り、主食材なしの場合設計したニューラルネットワークは妥当であるといえる。

6 おわりに

本研究の目的は料理レシピ動画を「わかりやすい動画」、「わかりにくい動画」に分類するニューラルネットワークの設計であった。技術課題をニューラルネットワークの特徴量の決定、ニューラルネットワークの設計、設計したニューラルネットワークの妥当性の検証とした。上記の技術課題を達成することができたことから本研究の目的を達成できた。

本研究で決定した特徴量以外の新たな特徴量を追加することで、精度がどのように変化するか確認する必要がある。動画そのものの特徴量を追加することで、分類の精度を上げることができると考える。今回の学習では、十分なデータを収集することができず主食材ありの場合では学習を収束させることができなかつたことから多くのデータを集めて再度検証をし主食材が本研究の特徴量として適しているか確認する必要がある。

参考文献

- [1] 牧野望, 塩井隆円, 楠和馬, 波多野賢治: “調理動作に基づく料理レシピ検索のための難易度算出法の提案”, 第9回データ工学と情報マネジメントに関するフォーラム, C4-5 (2017).
- [2] 岩本純也, 宮森恒: “調理の難易度を考慮したレシピ検索システムの提案”, 第4回データ工学と情報マネジメントに関するフォーラム, E1-3 (2012).
- [3] 矢嶋亜紗美, 小林一郎: 個人の状況を考慮した“かんとん”なレシピの推薦, 第1回データ工学と情報マネジメントに関するフォーラム, E5-6 (2009).
- [4] 秋口いくみ, 林利憲, 角谷和俊: “レシピ動画の調理区間に基づく難易度判定とその応用” Vol.2017-DBS-166 No.3 2017/12/22
- [5] 秋口いくみ, 王元元, 河合由起子, 角谷和俊: “料理レシピ動画の時間特性抽出による難易度判定” DEIM Forum 2018 C4-4