

ショート料理動画の字幕データに基づくレシピ補助情報の抽出

米澤 拓也[†] 王 元元^{††} 河合由起子^{†††,††††} 角谷 和俊[†]

[†] 関西学院大学総合政策学部 〒669-1337 兵庫県三田市学園2丁目1

^{††} 山口大学大学院創成科学研究科 〒775-8611 山口県宇部市常盤台2-16-1

^{†††} 京都産業大学情報理工学部 〒603-8555 京都府京都市北区上賀茂本山

^{††††} 大阪大学サイバーメディアセンター 〒567-0047 大阪府茨木市美穂ヶ丘5番1号

E-mail: [†]{fdu49780,sumiya}@kwansei.ac.jp, ^{††}y.wang@yamaguchi-u.ac.jp, ^{†††}kawai@cc.kyoto-su.ac.jp

あらまし Kurashiru や DELISH KITCHEN などの1分程度に編集されたレシピ動画は、短時間で料理の要点を概欄可能であり、レシピ内容の把握が容易である。しかしながら、一度の視聴だけでは調理動作を理解することが困難である。そこで、本研究では、既存のショート料理動画に含まれている、「食材や調味料」、「調理動作」から、それぞれの要素に重みづけを行い、補足情報を表示する箇所を分析することで、ユーザの料理レベルに適合するアドバイスや調理動作の詳細な説明や映像コンテンツを抽出するシステムの構築を目指す。

キーワード レシピ動画, テキストレシピ, 補助レシピ情報

1. はじめに

近年では、クックパッドや楽天レシピなどのようなユーザ投稿型レシピサイトが普及している。これらのサイトの利点は、無料で大量のレシピが閲覧可能であるため、利便性が高く、多くのユーザが利用している。また、ユーザ自身のオリジナルのレシピを投稿できるため、1つの料理においても複数のアレンジレシピを閲覧することも可能である。さらに、近年では上記のようなテキストレシピサイトの他に、料理動画も話題になっている。

料理動画は、テキストレシピと違い、文字ではなく、視覚的な情報から料理の要点を知ることが可能になるコンテンツである。主に難易度の高い料理は、料理番組でよく取り上げられており、プロの料理人のアドバイスとオープンキャプションの表示によって理解を容易にさせているため、多くの人が視聴している。また、比較的に調理が難しい料理に関しては、Kurashiru や DELISH KITCHEN などの1分程度に編集されたショート料理動画で多く取り上げられている。これらの動画は、料理の要点を映像とオープンキャプションを用いて解説されており、その手軽さから現在注目されつつある。これらのショート料理動画は、短時間で料理の要点を概欄可能であり、レシピ内容の把握を容易にすることができる。加えて、これらのサイトで用いられている料理動画は、ノーカットの映像を早送りで再生されているため、途中生成物の様子を確認することができるため、感覚的に料理のコツを掴むことができる。

しかしながら、ショート料理動画は、映像自体の展開が早いいため、映像を再生しながら調理を行うことは難しい。また、本来であれば詳細に説明を行うべき調理動作に対しても必要以上に情報の要約が行われている場合があるため、ユーザは、一度レシピ動画を視聴しただけでは、調理動作をすべて理解することは困難である。そこで、本研究では、ショート料理動画に含まれているオープンキャプションや、映像内に含まれている

「食材」「調理動作」「器具」などのコンテンツを基に分析を行う。動画内のコンテンツでは、主にそれぞれの出現頻度や使用頻度から重みづけを行うことにより、ショート料理動画内で、「より詳細に説明されるべき箇所」を判定する。

2. 関連研究

王ら[1]は、ユーザが視聴している映像コンテンツに各ユーザが求めているコンテンツ（興味や関心）を反映させることができていないことに注目し、上記の問題を解決するために、ユーザがどのレベルで映像コンテンツが含んでいる興味・関心を反映させた映像を構成するかをシーンの話題性をX軸、編集した映像の長さをY軸にした2軸で考察し、話題性と詳細度に基づく4種類の映像コンテンツの中から適切な映像を自動生成する手法を提案している。

秋口ら[2]は、料理番組をレシピ番組と定義したうえで、映像内の「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）」の3点の要素を抽出し、レシピ動画の時間的特性に基づいた難易度判定手法を提案した。料理動画を扱った研究として類似しているが、本研究では、料理番組ではなく、ショート料理動画を扱ったうえで、それらの映像コンテンツに対して、要約されている情報から補足情報を挿入する箇所を分析するという点で差異がある。

牧野ら[3]は、小学校、中学校、高等学校などで扱われている家庭科の教科書にて出現する調理動作に着目し、客観的に調理動作の難易度を定義する研究を行った。また、ここで定義した難易度を基に料理レシピの難易度を算出する手法を提案している。この研究では、テキストに記載されている調理動作からの難易度判定を行っているが、本研究では、映像コンテンツから重みづけを行うことで、重要度を判別するという点で差異がある。

蒯ら[4]は、料理番組映像を料理動画と定義し、映像に含まれている調理動作の中でも、繰り返し行われている動作を解析し、調理動作の分類を行う手法を提案している。この研究では、料

理動画を扱った研究であるが、本研究で扱っている料理動画は料理番組ではなく、ショート料理動画を扱っているという点で差異がある。

大滝ら[5]は、料理レシピ動画に含まれている調理動作・器具・食材の3点に着目し、それぞれの出現回数や料理の移動回数などを考慮し、調理にかかるコストを算出する手法を提案している。本研究では、ショート料理動画を扱い、出現頻度や出現時間から重みづけを行ったという点で差異がある。

3. 提案手法

まず、ショート料理動画内に映されている「食材と調味料」「調理動作」それぞれの出現頻度や、出現時間・映像全体の時間を算出する。次に、それぞれの要素の出現頻度からTF値を求め、その値に出現時間と映像時間を使って4種類の計算を行う。これらの計算を行った結果、算出された数値と補足されるケースの高いと考えられる調理動作を照らし合わ、最も適応していると考えられる計算式を判別する。また、今回の実験に際して使用した映像は、ショート料理動画を取り扱った動画レシピサイトであるKurashiruから、無作為に20件の映像を選択した。これらの映像から、今回は全て手作業で「食材と調味料」「調理動作」の出現頻度や出現時間を求め、それぞれの計算を行った。

3.1 実験

実験では、まず前処理としてKurashiruから選択した20件の映像データから「食材と調味料」「調理動作」などの出現頻度や映像内での出現時間を、手作業で書き出したのちに、以下の式でそれぞれのTF値を求めた。

TF(t,d) = n_{t,d} / \sum_{s \in d} (n_{s,d})

TF値を求める際、tは食材キーワード、もしくは調理動作キーワードとし、dは対象のショート料理動画の集合とする。また、これらの数値の他に、映像全体の長さをAとし、出現時間をaとしたうえで計算を行う。

実験で行う計算は、以下の4種類の計算式を用いて、どの計算式が有用であるかを判別する。

- (1) 出現頻度が多く、出現時間が長い
TF × (a/A)
- (2) 出現頻度が多く、出現時間が短い
TF × (A/a)
- (3) 出現頻度が少なく、出現時間が長い
1/TF × (a/A)
- (4) 出現頻度が少なく、出現時間が短い
1/TF × (A/a)

上記の計算を、無作為に抽出した20件のショート料理動画で行うと図1と図2のような結果になった。食材で4種類の計算を行った結果、計算式(1)が適合率の高い計算式である事が判別できた。この計算で抽出された上位の食材は、主に、そのレシピの中で扱っている食材の中でも重要度の高いものが抽出されたからである。また、計算式(4)では、主に調味料を中心に上位へ抽出された。また、調理動作では、計算式(4)が適

	計算方法(1)		計算方法(2)		計算方法(3)		計算方法(4)	
チキン南蛮	鶏ささみ	0.447761	塩こしょう	7.444444	卵	3.044776	砂糖	150.75
	きゅうり	0.358209	マヨネーズ	4.1875	しょうゆ	2.134328	塩こしょう	93.8
	小麦粉	0.134328	酢	3.722222	酒	1.791045	塩	67
エビのチリソース	エビ	0.507692	ラー油	5.909091	サラダ油	3.723077	塩	97.5
	片栗粉	0.223077	塩	2.708333	水	3.2	しょうゆ	56.33333
	にんにく	0.192308	ケチャップ	2.5	しょうゆ	3	水	52.8125
ローストビーフ丼	牛もも肉	0.638298	料理酒	3.916667	サラダ油	2.574468	砂糖	70.5
	塩こしょう	0.264438	みりん	2.238095	塩こしょう	1.439716	しょうゆ	33.94444
	長ねぎ	0.180851	ごはん	1.678571	しょうゆ	1.244681	ごはん	26.85714

図1 食材と調味料の結果

	計算方法(1)		計算方法(2)		計算方法(3)		計算方法(4)	
コロッケ	型を作る	0.0461538	型を作る	21.666667	混ぜる	2.4923077	切る	286
	炒める	0.0273504	潰す	7.2222222	炒める	2.2153846	混ぜる	73.125
	付ける	0.0205128	レンジで加熱	5.4166667	切る	1.6923077	揚げる	65
天津飯	ほぐす	0.0666667	ほぐす	15	煮る	4.0444444	切る	198
	混ぜる	0.0345679	焼く	0.6428571	混ぜる	2.8	焼く	64.285714
	煮る	0.0239316	切る	0.4090909	切る	2.4444444	煮る	41.785714
ピビンバ	茹でる	0.1621622	和える	3.0833333	炒める	2.4324324	切る	101.75
	和える	0.0810811	切る	0.8409091	切る	1.1891892	炒める	33.3
	炒める	0.03003	混ぜる	0.4567901	茹でる	0.3648649	和える	12.333333

図2 調理動作の結果

合率の高い計算式である事が判別できた。この計算で抽出された上位の調理動作では、レシピ内に含まれている調理動作の中で、一度失敗すると料理として破綻してしまうといった、重要度が高いと考えられる調理動作が抽出する事ができたため、計算式(4)が適合していると判別した。

4. 考察

4章で述べた結果から、「食材と調味料」では、計算式(1)が高い適合率を示す事ができたが、補足情報を挿入する対象を「メインとなる食材」ではなく、「調味料の分量」を対象として補足情報を挿入するのであれば、計算式(4)も扱えると考えられる。「調理動作」では、計算式(4)が適合率の高い計算式だと考えられる結果となった。理由としては、動画全体が早送りになっているため、重要度の高いものと低いものの映像時間がアンバランスという特徴が起因したと考えられる。

今後の予定としては、補足情報を挿入する箇所の分析の精度向上のために、ユーザのビデオ視聴履歴および視聴動作を含めた分析を行い、本研究で得た結果との比較などを行う。また、補足情報として他のレシピ動画を利用することも検討する。この他にも、今後の検討事項としてあげられることは、ユーザインタフェースに関連する内容である。ユーザにとって必要な補足情報をどのような形で表示するか言及ができていないため、今後の研究を進めていくうえで、検討していく。

現在検討している対応策としては、映像内部にて料理の補足情報を既存のオープンキャプションに追加するほか、音声によって伝達を行うことも検討している。また、ユーザにとって理解が困難であった調理動作の箇所を、他の料理動画で、同じ調理動作を行っているものを挿入することも検討している。これによってカメラのアングルの変化で、調理動作に対する理解度を向上させることを狙いとしている。

5. おわりに

本研究では、既存のショート料理動画に付与されているオープンキャプションを元に、料理で使用される食材や調味料、専

専門用語や調理動作から調理動作の重みづけを行い、補足情報を表示する箇所を分析する。これらの工程を行った後に、ユーザーの料理レベルに適合するアドバイスや調理動作の詳細な説明や映像などのコンテンツを抽出する手法を提案した。今後は、ユーザインタフェースについて考慮に入れながら、補助レシピ情報として追加するコンテンツを検討する。

文 献

- [1] 王元元, 渡部雅俊, 河合由起子, 角谷和俊: 映像における話題シーン検出と話題性に基づく受動的映像視聴システム, 第7回データ工学と情報マネジメントに関するフォーラム, F2-2 (2015).
- [2] 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権限特性抽出による難易度判定, 第10回データ工学と情報マネジメントに関するフォーラム, C4-4 (2018).
- [3] 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).
- [4] 蒯承穎, 志土地由香, 高橋友和, 井手一郎, 村瀬洋: 料理動画における調理動作の解析, 第4回デジタルコンテンツシンポジウム講演予稿集, 2014.
- [5] 大滝健太郎, 鷹野孝典: 調理動作・器具・食材に注目した調理動画からの調理コスト判定手法の検討, 情報処理学会第79回全国大会, Vol. 2017.