ユーザの視聴操作と地理的関係性に基づく映像視聴支援システムの提案

王元元[1]　橋本樹[2]　河合由起子[3]　角谷和俊[4]

近年，動画配信サービスの普及により，映像に関連したさまざまな情報を推薦する研究が盛んに行われている．たとえば，NHKの放送通信連携サービス「NHK Hybridcast」 [23]では，テレビ番組の放送中に同画面上で関連情報の推薦を行っている．また，YouTubeではチャプター機能が実装され，ユーザの視聴操作の有用性が向上してきている．しかし，現在はユーザの視聴操作を考慮した補足情報推薦を行うサービスは実装されていない．そこで本研究では，ユーザが選択したシーンを抽出可能なユーザインタフェースを提案して利用することでユーザの視聴操作を抽出する．これにより，ユーザのシーン選択から得られた各シーンの地理的情報を持つキーワードと各シーンでのトピックを用いて空間的距離および意味的距離を算出することで，補足情報の検索クエリを生成する．本論文では，ユーザの視聴操作を用いたユーザの視聴傾向抽出および地理的関係性抽出に基づく映像視聴支援システムの提案とその検証を行う．

1　はじめに

TVer*1やGYAO!*2など見逃し配信を行う動画配信サービスの普及とNetflix*3やHulu*4などのサブスクリプションサービスの普及により，放送時間を気にして映像を視聴することが減っている．そのため，興味のあるシーンのみを視聴する「飛ばし見」を行うユーザが増え，映像全体を視聴する「流し見」を行うユーザが減ってきていると考えられる．本研究では，そのような研究背景に基づいて，主に「飛ばし見」を行うユーザを対象としている．上記の前提に基づき，ユーザの視聴操作を取得可能なユーザインタフェースにより，「飛ばし見」を行うユーザが選択したシーンの操作履歴を抽出する．抽出された選択シーンはユーザが求める情報であると仮定し，選択シーンの操作履歴内に出現するキーワードを用いてクエリ式を生成する．これにより，ユーザの視聴操作によって推薦する情報が変化するため，個人化された情報推薦を行うことが可能である．また，従来の映像を用いた研究ではクローズドキャプション内に出現するキーワードからトピックを抽出し情報推薦を行う場合が多い．しかし，本研究では，キーワード群より得られるトピックに加えて，地名や地域など地理的情報を持つキーワードを使用する．

図1　本研究で使用するユーザインタフェース

本論文では，地名など地理的情報を持つキーワードを空間的キーワードと定義する．空間的キーワードを使用することで地名や地域は集合記号を用いて表現することが可能であるため，シーンごとで得られる空間的キーワードの地理的関係性を用いて，推薦する情報を判定することが可能である．たとえば，比較対象のシーンで「兵庫県」に関する内容を放送し，視聴中のシーンで「関西学院大学」に関する内容を放送している場合，「兵庫県」に対して「関西学院大学」は包含関係にあり，「兵庫県∋関西学院大学」と表現される．この関係性により，シーン間の地理的関係性が「兵庫県」から「関西学院大学」に範囲が狭まっているため，番組内容では「兵庫県の関西学院大学」という流れで，より詳細な内容を放送されると考えられる．このように，シーン間の地理的関係性を使用することにより，推薦すべき情報の種類を判定することが可能である．本研究では，それらに対応したクエリ式の生成方式を提案し，これにより，ユーザの視聴傾向を考慮した補足情報として適切なWebページの推薦を目指す．

本論文の構成は以下のとおりである．2章では本提案システムのファイ用および関連研究について述べる．3章では，ユーザの視聴操作分析に基づく各シーンのトピック抽出について説明し，4章では，提案手法に基づく補足情報を検索するためのクエリ式の生成について述べる．5章では，ユーザインタフェースのユーザビリティに関する評価と補足情報を検索する際に使用するクエリ式の評価について述べる．最後に，6章でまとめと今後の課題について述べる．

2　システム概要と関連研究

2.1　映像視聴支援システム

本研究では，図1に示しているシーン選択可能なユーザインタフェースにより得られるユーザの視聴操作を抽出して，ユーザの視聴操作に基づく映像の補足情報提示手法を提案する．図2はユーザインターフェースの画面遷移図であり，ユーザが選択したシーンに基づいて表示する内容が変化する．図3に提案した映像視聴支援システムの概要図を示す．

[1] 正会員　山口大学大学院創成科学研究科
y.wang@yamaguchi-u.ac.jp
[2] 非会員　関西学院大学大学院総合政策研究科
ewd16738@kwansei.ac.jp
[3] 正会員　京都産業大学情報理工学部
kawai@cc.kyoto-su.ac.jp
[4] 正会員　関西学院大学総合政策学部
sumiya@kwansei.ac.jp
*1 https://tver.jp/
*2 https://gyao.yahoo.co.jp/
*3 https://www.netflix.com/jp/
*4 https://www.hulu.com/

DBSJ

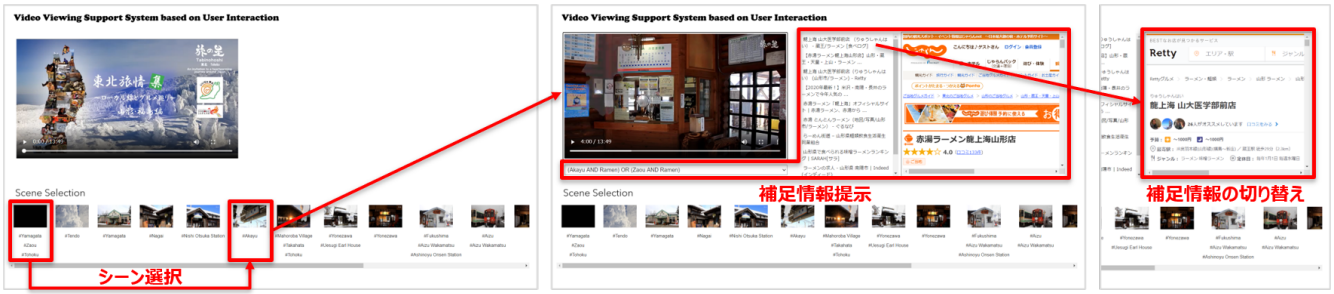

図2　ユーザインタフェースの画面遷移図

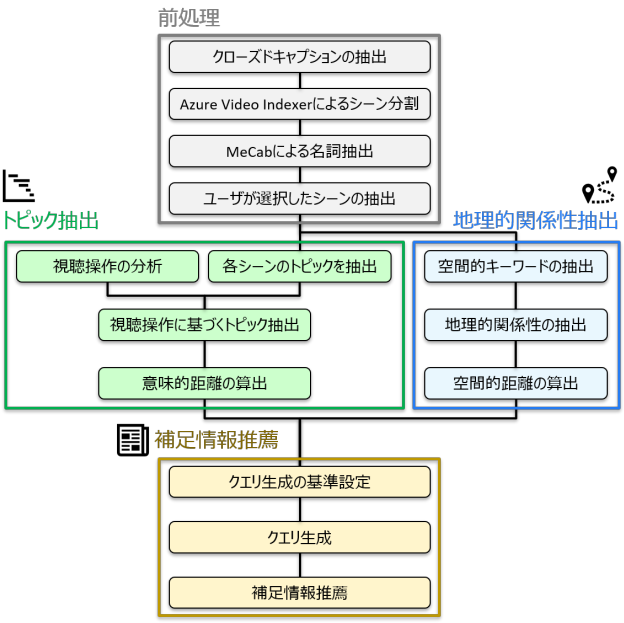

図3　システム概要図

本提案システムは，前処理として映像のクローズドキャプションを抽出し，MeCab [19]を用いて形態素解析を行い，映像のクローズドキャプションに出現している名詞のみを抽出する．シーン分割にはMicrosoft AzureのVideo Indexer [9]を使用し，分割されたシーンごとのサムネイルを図1に示しているユーザインタフェースの下部で表示する．サムネイルをクリックすると該当シーンの開始時間にスキップして映像が再生される．ユーザインタフェースにより得られるユーザのシーン選択をユーザの視聴操作と定義し，ユーザの視聴操作により得られる選択シーン内のキーワードと空間的キーワードを用いて，トピック抽出と地理的関係性抽出を行う．抽出された地理的関係性の種類によって，ブーリアン検索モデルに基づいて生成されるクエリの種類を決定する．地理的関係性を用いたクエリ生成手法については4章で述べる．生成されるクエリ式を図1に示しているユーザインタフェースの映像再生画面の下部に表示する．また，クエリ式を用いた検索結果の上位10件をユーザインタフェースの中央部に表示し，検索結果の上位1件のWebページを補足情報としてユーザインタフェースの右上部で表示する．

2.2　関連研究

近年，映像の字幕やクローズドキャプションなどのメタデータを用いて，映像内容に関連する情報を推薦する研究が多く行われている．我々の先行研究は，テレビ番組のクローズドキャプション内に出現する地理的な情報を持つキーワードを用いて地理情報や関連する情報を推薦するシステムを提案している [6] [14] [15] [16]．これにより，地理的関係を考慮した情報推薦は有効であると考えられる．本研究も地理的関係を考慮した情報推薦を行うが，空間的距離の算出方法と検索クエリの生成基準が異なるまた，クリピンデルら [18]はクローズドキャプションに出現する単語ごとに重要度を付与する手法を提案し，手法として*TFiDF*を使用して重要度を算出している．三浦ら [20]は単語間の意味的関係の抽出手法を提案している．これらの研究では，映像のクローズドキャプションから単語間の意味的関係を考慮したクエリキーワードの抽出を行っている．本研究では，単語間の意味的関係性だけでなく，空間キーワード間の地理的関係性も重視した検索クエリの生成を行う．

また，映像を対象とした研究として，トピックの変化に基づくシーン分割に関する研究は盛んに行われている．Baraldiら [2]はビデオフレーム内に存在するトピックの時間的変化を利用したシーン分割手法を提案している．Sidiropoulosら [12]はBGMやナレーションなどの音声データを用いたシーン分割手法を提案している．ほかにも，Rasheedら [10]はビデオフレーム内のオープンキャプションやクローズドキャプションなどのテキストデータを用いて，映画やテレビ番組の構造を分析している．対照的に，Liuら [8]は動画のショットに注目し，ショットに表示されるオープンキャプションやナレーション，BGMなどのテキストデータと音声データの両方を用いたシーン分割手法を提案している．本研究は，Microsoft AzureのVideo Indexerを利用してシーンの分割を行っているため，その点が異なる．

さらに，ビデオフレームを用いた研究では，ビデオ要約に関する研究も盛んに行われている．ビデオ要約では，シーン分割と同様にトピックの変化や場面の切り替わりを判定し，個人化された短いビデオの推薦を行っている [4] [11] [13]．本研究では，ユーザのシーン選択を用いているため，映像全体をユーザが視聴しないという点などは内容的に近いと考えられる．しかし，映像の要約は映像制作者の意図を考慮する必要があるため，非常に困難であると考えられる．

表1　トピック抽出に使用したテレビ番組

番組	タイトル	放送時間	シーン数	平均シーン長	トピック数
A	そうだ旅に行こう	46分	22	2.1分	15
B	ブラタモリ #97	45分	28	1.6分	15
C	さわやか自然百景 #2	15分	8	1.9分	10

また，本研究では地理情報に関連する内容も含むため，地理情報の推薦に関する研究 [7] [21] [24]も参考にしている．これらの研究では，地理情報を推薦するために用いる特徴ベクトルの抽出が必要不可欠であり，特に倉田ら [21]は地理情報推薦にTopological Relations [5]を応用する手法を提案している．本研究においてもTopological Relationsを用いた手法は有効であると考え，地理的関係性抽出に本手法を使用している．

3　トピック抽出

本研究では，補足情報を推薦する際にトピックと空間的キーワードを用いてクエリ式を生成する．そのため，該当シーンにおけるクエリ式で使用するトピックの抽出手法について説明する．

3.1　トピック抽出手法

トピック抽出手法には，潜在的ディリクレ配分法（Latent Dirichlet Allocation, LDA） [3]を使用し，番組内に出現するキーワードに対して重み付けを行う．これにより，シーンごとでトピックとなるキーワードの判定を行うことが可能となる．次に具体的な処理の流れについて説明する．番組におけるカメラの切り替わりをショットと定義し，ショットの意味的な集合をシーンと定義し，各ショット内で発言されたナレーションやセリフを1つの文書として使用する．番組全体でのショットの集合を入力として，MeCab [19]による形態素解析を行い，それにより得られる名詞のみをキーワードとして使用する．また，形態素解析の際に不要語を除去するために数字1文字やひらがな1文字などをストップワードとして設定した．そして，抽出されたキーワードの出現頻度を用いて*TFiDF*によるベクトル化を行い，それらを用いて*LDA*によるトピック分布を求める．

ハイパーパラメータα，βの設定はZhaoら [17]に従い，$\alpha = \frac{50}{T}$，$\beta = 0.01$とした．また，トピック数の設定は渡邊ら [22]に従い，トピック数Tは，$T = 5$から5刻みで$T = 30$まで実行し，人手で調べた結果から最も適切にトピックが構成されたと判断されたトピック数を使用する．適切なトピック数は番組の長さによって異なるため，45分の番組ではトピック数$T = 15$，15分の番組ではトピック数$T = 10$とする．シーンごとのトピックを判定するために，該当シーン内に含まれるショットごとの各トピックに属する確率の平均値を用いて，シーンごとで上位3件に含まれるトピックを抽出する．そして，上位3件のトピックにおける上位キーワード10件を抽出し，各キーワードがトピックに属する確率の平均値を閾値として，閾値以上のキーワードをトピックワードとする．

3.2　視聴操作に基づくトピック抽出

本研究では，ユーザの視聴操作を考慮した補足情報推薦を目的としているため，トピック抽出においても視聴操作を考慮した手法として*LDA-U*を提案する．*LDA-U*では*LDA*と同様に番組全体でのショットの集合を入力として処理を行うが，シーンごとのトピックを判定する際に該当シーンと2つ前までのシーン選択履歴の3シーンを用いて各ショットごとの各トピックに属する確率の平均値を算出する．また，平均値の算出を行う際に定数を重みとして付与する．重みの設定としては該当シーンにおける確率には0.5，1シーン前のシーンにおける確率には0.3，2シーン前のシーンにおける確率には0.2を乗算する．これにより，視聴履歴でのトピックも考慮したトピック分布の構成が可能となる．

3.3　各手法におけるトピック抽出結果の比較

実際のテレビ番組（表1）を用いて*LDA-U*と*LDA*により抽出されるトピックワードの比較を行った結果を表2に示す．また，比較手法として，従来手法である*TFiDF*の結果も示す．*TFiDF*の計算は下記の式 (1) により算出され，番組全体における総シーン数をNとし，該当するシーンをs，該当するシーンに出現するキーワードをkと定義する．

$$TFiDF = TF(k, s) \times \log \frac{N}{DF(k)} \tag{1}$$

結果として，テレビ番組AとCにおける*LDA-U*を用いたトピック抽出では，1シーン前のシーンでトピックワードとして抽出された「温泉」や「シギ」などのトピックワードを抽出することが可能である．また，*TFiDF*により抽出されたトピックワードと*LDA-U*，*LDA*により抽出されたトピックワードを比較すると*TFiDF*では出現頻度のみを用いて計算されるため，該当シーンで出現頻度の高いキーワードのみが抽出されるが，*LDA-U*，*LDA*では該当シーンに出現していないが関連のあるキーワードも抽出することが可能である．以上の結果より，提案手法である*LDA-U*を用いることでユーザの視聴操作を考慮したトピック抽出が可能であるということが確認された．

4　クエリ生成

本章では，補足情報を推薦するためにWebページを使用する．よって，補足情報となるWebページを抽出可能なクエリ式を生成する．クエリ式の生成には情報検索システムで使用されているブーリアン検索モデルを使用する．

4.1　空間的・意味的距離を用いたクエリ生成の基準

シーン間の地理的関係性とトピックを用いたクエリ生成では，シーン間の内容が明らかに異なる場合や実距離が遠く離れている場合がある．よって，関係の無いシーン間でのクエリ生成を制御するためにクエリ生成の基準を設定する必要があると考えられる．本研究では，空間的キーワードの郵便番号を用いた空間的距離とシーン間の内容の類似度を用いた意味的距離の2つの距離を定義し，これらの距離に基づき，クエリ式を生成するための基準を表3に示す．

4.1.1　空間的距離の算出

空間的距離の算出には，上3桁と下4桁で構成される郵便番号を使用し，実距離の近似値を算出する．これにより，実距離を使用するコストを大幅に削減することが可能である．また，郵便番号を使用することにより，空間的キーワード間の地理的関係性を判定することが可能である．空間的距離の算出には下記の式 (2) に

表2　*LDA-U*, *LDA*, *TFiDF*を用いたトピック抽出の例

番組	シーン選択	*LDA-U*	*LDA*	*TFiDF*
A	7 → 9 → 13	ベッキー，後ろ，急， 一，花月亭，海鮮，温泉	ベッキー，後ろ，急， 一，花月亭，海鮮，本	証明書，焼却炉， 最終目的地，予約
B	12 → 16 → 22	有馬，温泉，街，海水， 調査，一緒，モフモフ	有馬，温泉，街，セント，海水， 調査，中央，一緒，モフモフ，温水	高速道路，大通り
C	4 → 3 → 5	鳥，藤前，干潟，食べ物， 泥，栄養素，場所，シギ，羽	鳥，藤前，干潟，食べ物， 泥，栄養素，場所	魚

表3　空間的・意味的距離を用いたクエリ生成の基準

地理的関係性	Disjoint	Equal	Inside	Contain	Disjoint	Equal	Inside	Contain
空間的距離 / 意味的距離	近い				遠い			
近い	生成	生成	生成	生成	生成	生成しない	生成	生成
遠い	生成	生成	生成	生成	生成しない	生成しない	生成しない	生成しない

より算出される．

$$GeoD(s_i, s_j) = |Zipcode(s_i) - Zipcode(s_j)| \quad (2)$$

4.1.2　意味的距離の算出

意味的距離を算出するために，シーン間での類似度をユークリッド距離により算出する．これにより，関連の無いシーン間での補足情報の推薦を防止する．ユークリッド距離の算出には特徴ベクトルとしてシーンごとの*TFiDF*値を使用する．また，意味的距離の計算は下記の式 (3) と (4) により算出される．

$$SemD(s_i, s_j) = \sqrt{\sum_{i=1}^{N}(s_j - s_i)^2} \quad (3)$$

$$s_i = \{TFiDF(k_1), TFiDF(k_2), ..., TFiDF(k_j)\} \quad (4)$$

4.2　地理的関係性に基づくクエリ生成

本手法では，該当シーンに対して1シーン前のシーン履歴を比較シーンと呼ぶ．また，該当シーンと比較シーンに出現する空間的キーワードを用いて，シーン間での地理的関係性を抽出し．生成されるクエリの種類を判定する．使用される地理的関係性のパターンとしてTopological Relations [5]の9種類の空間的関係より，4種類の関係性を使用する．使用する関係性は**Disjoint**，**Equal**，**Inside**，**Contain**の4種類である．以下に地理的関係性によって推薦される情報を説明する．

Disjoint　該当シーンに出現する地名と比較シーンに出現する地名が分離しているため，地域の比較情報を推薦する．

Equal　該当シーンに出現する地名と比較シーンに出現する地名が一致しているため，トピックの比較情報を推薦する．

Inside　該当シーンに出現する地名が比較シーンに出現する地名に対して内在しているため，地域の詳細情報を推薦する．

Contain　該当シーンに出現する地名が比較シーンに出現する地名に対して内包しているため，地域の概要情報を推薦する．

それぞれの地理的関係性の例は図4に示す．また生成されるクエリ式は表4に示す．

5　評価実験

本章では，本研究で使用するユーザインタフェースのユーザビリティに関する評価と補足情報を抽出する際に使用するクエリ式の評価を行う．ユーザインタフェースの評価には，システムのユーザビリティを評価する際に使用される指標であるSUS (System Usability Scale) スコア [1]を用いる．また，クエリ式の評価には実際の番組を用いてクエリ生成を行い，生成されたクエリ式により得られるWebページを*TFiDF*と*TF*により得られるWebページと比較することで評価を行う．

5.1　ユーザインタフェースのユーザビリティ評価

ユーザインタフェースのユーザビリティに関する評価実験では，図1に示しているプロトタイプシステムのユーザインタフェースを使用し，ユーザインタフェースの操作の例は図2に示す．上記のユーザインタフェースを用いて，シーン選択有の場合とシーン選択無の場合の2つのパターンでどちらのほうは有用性が高いかを検証する．また，SUSスコアの算出に使用する質問は以下の項目で行う．

Q1 このシステムをしばしば利用したいと思う
Q2 このシステムを利用するには説明が必要となるほど複雑であると感じた
Q3 このシステムは容易に使いこなす事ができると思った
Q4 このシステムを利用するのに専門家のサポートが必要だと感じる
Q5 このシステムにあるコンテンツやナビゲーションは十分に統一感があると感じた
Q6 このシステムでは一貫性のないところが多々あったと感じた
Q7 たいていの人は，このシステムの利用方法をすぐに理解すると思う
Q8 このシステムはとても操作しづらいと感じた
Q9 このシステムを利用できる自信がある

p_{wi} **Disjoints** p_{wj} | p_{wi} **Equals** p_{wj} | p_{wi} **is Inside** p_{wj} | p_{wi} **Contains** p_{wj}

推薦情報

地域の比較情報 | トピックの比較情報 | 地域の詳細情報 | 地域の概要情報

クエリ式

$(P_{wi}$ AND $t_i)$ OR $(P_{wj}$ AND $t_i)$ | $(P_{wi}$ AND $t_i)$ OR $(P_{wi}$ AND $t_j)$ | $(P_{wi}$ AND $t_i)$ NOT P_{wj} | $(P_{wi}$ AND $t_i)$ NOT P_{wj}

クエリ式の例

(下田 AND 宿) OR (熱海 AND 宿) | (下田 AND 宿) OR (下田 AND 旅行) | (蓮台寺温泉 AND 宿) NOT 下田 | (下田 AND 宿) NOT 蓮台寺温泉

t_i:宿 t_j:旅行
p_{wi} = 該当シーンに出現する空間的キーワード p_{wj} = 1シーン前に出現する空間的キーワード
t_i = 該当シーンに出現するトピック t_j = 1シーン前に出現するトピック

図4 地理的関係性の例

表4 地理的関係性に基づくクエリ生成

地理的関係	推薦情報	クエリ式	クエリ式の例
(1) Disjoint	地域の比較情報	$(p_{w_i}\ AND\ t_i)\ OR\ (p_{w_j}\ AND\ t_i)$	(下田 *AND* 宿) *OR* (熱海 *AND* 宿)
(2) Equal	トピックの比較情報	$(p_{w_i}\ AND\ t_i)\ OR\ (p_{w_i}\ AND\ t_j)$	(下田 *AND* 宿) *OR* (下田 *AND* 旅行)
(3) Inside	地域の詳細情報	$(p_{w_i}\ AND\ t_i)\ NOT\ p_{w_j}$	(蓮台寺温泉 *AND* 宿) *NOT* 下田
(4) Contain	地域の概要情報	$(p_{w_i}\ AND\ t_i)\ NOT\ p_{w_j}$	(下田 *AND* 宿) *NOT* 蓮台寺温泉

表5 SUSスコアを用いたユーザインタフェースの評価結果

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUSスコア	グレート	Adjective Rating
シーン選択有	3.7	2.7	3.9	2.8	4.2	1.9	4.1	2.0	4.0	2.9	**69.2**	C	Okay
シーン選択無	3.2	2.9	3.7	2.0	4.1	2.3	3.3	2.0	3.8	3.1	64.4	D	Poor

表6 クエリ生成の評価実験で使用するテレビ番組

番組	タイトル	放送時間	シーン数	出現割合	回答者数	視聴操作
旅番組	旅の星	14分	17	24%	11人	2 → 6 → 11 → 13 → **14**
ニュース番組	ニュースきょう一日	15分	29	7%	11人	1 → 2 → 6 → 7 → 13 → **19**

表7 3手法を用いて抽出されたWebページの評価結果

手法	評価値（旅番組）	評価値（ニュース番組）
提案手法	**3.56**	**3.00**
TFiDF	2.38	2.75
TF	1.76	2.36

Q10 このシステムを利用し始める前に知っておくべきことが多くあると思う

以上の質問に対して，被験者は5段階のリッカート尺度（1：全くそう思わない，2：あまり思わない，3：どちらともいえない，4：ややそう思う，5：とてもそう思う）で回答を行い，SUSスコアの評価は，A) Excellent: >80.3, B) Good: >68.0-80.3, C) Okay: >68.0, D) Poor: >51.0-68.0, E) Awful: >51.0の5つのグレートで評価する．本評価実験では，シーン選択の有無に関わらず，SUSスコアを使用し，各パターンを被験者9名ずつ，総被験者数18名（男性：7名，女性：11名）の20代大学生によって評価する．各質問の評価値や各パターンのSUSスコアの平均値を表5に示す．表5の結果より，シーン選択有の評価は C) Okay，シーン選択無の評価は D) Poorであることが確認された．また，SUSスコアの平均値は比較手法であるシーン選択無に比べて，提案手法であるシーン選択有のほうは評価値が高いため，提案手法の有用性が高いことを確認した．

5.2 補足情報の検索クエリ式の有効性検証

クエリ式の評価実験では，提案手法により生成されるクエリ式と比較手法である*TFiDF*と*TF*により抽出されるキーワードを用

いて比較を行い，クエリ式の有効性の検証を行う．データセットには，表6に示している旅番組とニュース番組の2種類のテレビ番組を使用し，各手法により生成されるクエリを用いて上位10件のWebページを抽出する．

旅番組では，提案手法により (喜多方 *AND* ラーメン) *NOT* 会津というクエリ式が生成され，比較手法である*TFiDF*では「喜多方」，*TF*では「ラーメン」というクエリキーワードが生成された．ニュース番組では，提案手法により (中国 *AND* 護衛艦) *OR* (アメリカ *AND* 護衛艦) というクエリ式が生成され，比較手法である*TFiDF*では「護衛艦」，*TF*では「防衛省」というクエリキーワードが生成された．

以上のクエリ式とクエリキーワードを用いて，上位10件のWebページを抽出し，抽出されたWebページを被験者5名ずつにより評価を行った．評価には5段階のリッカート尺度（1：適切だと思わない，2：あまり適切だと思わない，3：どちらともいえない，4：やや適切だと思う，5：適切だと思う）を利用し，得られた評価値の平均を表7に示す．表7の結果より，提案手法により得られたWebページは比較手法である*TFiDF*と*TF*に比べて，補足情報として適切なWebページを推薦することが可能であるということが確認された．

6 おわりに

本研究では，視聴操作を考慮した地理的関係性に基づく映像視聴支援システムを提案した．提案手法では，ユーザの視聴操作に基づく抽出されるトピックワードと空間的キーワード間の地理的関係性を用いた補足情報の検索クエリを生成し，補足情報となるWebページを提示した．また，ユーザインターフェースのユーザビリティと生成された検索クエリ式の有効性に関する評価を行った．その結果，比較した手法よりも高い評価を得ることが確認できた．

今後，空間的キーワードのようにキーワード同士の関係性を抽出可能なキーワードがあれば，本提案手法はeラーニングのような他分野に応用できると考えられる．そのため，別のジャンルの映像を用いた評価実験も検討する予定です．

謝辞

本研究の一部は，JSPS科研費JP19H04118，JP21K17862の助成を受けたものである．ここに記して謝意を表す．

参考文献

[1] Aaron Bangor, Philip T Kortum, and James T Miller. An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction*, 24(6):574–594, 2008.
[2] Lorenzo Baraldi, Costantino Grana, and Rita Cucchiara. Scene segmentation using temporal clustering for accessing and re-using broadcast video. In *Proceedings of the 2015 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2015.
[3] David Meir Blei, Andrew Yan-Tak Ng, and Michael Irwin Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
[4] Shayok Chakraborty, Omesh Tickoo, and Ravi Iyer. Adaptive keyframe selection for video summarization. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 702–709. IEEE, 2015.
[5] GITTA. Topological relations. `http://www.gitta.info/SpatialQueries/en/html/TopoBasedOps\_learningObject1.html` (参照 2021-01-11).
[6] Honoka Kakimoto, Yuanyuan Wang, Yukiko Kawai, and Kazutoshi Sumiya. Query generation for web search based on spatio-temporal features of tv program. In *International MultiConference of Engineers and Computer Scientists*, pages 82–92. Springer, 2018.
[7] Kaori Kobayashi, Daisuke Kitayama, and Kazutoshi Sumiya. Cinematic street: Automatic street view walk-through system using characteristics of modified maps. In *International Symposium on Web and Wireless Geographical Information Systems*, pages 142–158. Springer, 2011.
[8] Cailiang Liu, Dong Wang, Jun Zhu, and Bo Zhang. Learning a contextual multi-thread model for movie/tv scene segmentation. *IEEE Transactions on Multimedia*, 15(4):884–897, 2013.
[9] Microsoft. Azure video analyzer for media. `https://vi.microsoft.com/ja-jp/` (参照 2021-01-11).
[10] Zeeshan Rasheed and Mubarak Shah. Detection and representation of scenes in videos. *IEEE Transactions on Multimedia*, 7(6):1097–1105, 2005.
[11] Aidean Sharghi, Boqing Gong, and Mubarak Shah. Query-focused extractive video summarization. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 3–19, Cham, 2016. Springer International Publishing.
[12] Panagiotis Sidiropoulos, Vasileios Mezaris, Ioannis Kompatsiaris, Hugo Meinedo, Miguel Bugalho, and Isabel Trancoso. Temporal video segmentation to scenes using high-level audiovisual features. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(8):1163–1177, 2011.
[13] Ba Tu Truong and Svetha Venkatesh. Video abstraction: A systematic review and classification. *ACM Trans. Multimedia Comput. Commun. Appl.*, 3(1):3–es, feb 2007.
[14] Yuanyuan Wang, Daisuke Kitayama, Yukiko Kawai, and Kazutoshi Sumiya. Automatic street view system synchronized with tv program using geographical metadata from closed captions. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, AVI '14, page 383–384, New York, NY, USA, 2014. Association for Computing Machinery.
[15] Yuanyuan Wang, Daisuke Kitayama, Yukiko Kawai, Kazutoshi Sumiya, and Yoshiharu Ishikawa. An automatic video reinforcing system for tv programs using semantic metadata from closed captions. *International Journal of Multimedia Data Engineering and Management (IJMDEM)*, 7(1):1–21, 2016.
[16] Yuanyuan Wang, Maho Nishizawa, Yukiko Kawai, and Kazutoshi Sumiya. Location-based image viewing system synchronized with video clips. In *Proceedings of the 13th International Conference on Location Based Services, Vienna, Austria*, pages 14–16, 2016.
[17] Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. Comparing twitter and traditional media using topic models. In *European conference on information retrieval*, pages 338–349. Springer, 2011.
[18] クリピングデルサイモン, 高橋正樹, 山内結子, and 苗村昌秀. 視聴者の興味内容推定における epg や字幕中の単語への重みの付与. 映像情報メディア学会年次大会講演予稿集, 2015:12C–2, 2015.
[19] 工藤拓. Mecab. `https://taku910.github.io/mecab/` (参照 2021-01-11).
[20] 三浦菊佳, 山田一郎, 宮崎太郎, 加藤直人, and 田中英輝. 単語間の意味的関係を用いたテレビ番組マップ生成. 第*76*回全国大会講演論文集, 2014(1):37–38, 2014.
[21] 倉田陽平. 地理情報科学における位相的空間関係の研究と今後の課題. *GIS*-理論と応用, 18(2):149–159, 2010.
[22] 渡邊恵太 and 加藤昇平. トピックモデルと協調フィルタリングに基づくユーザ興味を反映した情報推薦システム. In 第*28*回人工知能学会全国大会論文集, pages 2M34–2M34. 一般社団法人人工知能学会, 2014.
[23] 日本放送協会. Nhk hybridcast. `https://www.nhk.or.jp/hybridcast/online/` (参照 2021-01-11).
[24] 北山大輔, 宮本節子, and 角谷和俊. ユーザ操作履歴と地名出現パターンに基づくオンライン地図の動的カスタマイズ方式. 情報処理学会論文誌データベース *(TOD)*, 3(4):65–81, 2010.