

時間特性を考慮したマッチング型時空間コンテンツ検索手法

永島 薫[†] 妙中 雄三^{††} 塚本 和也[†]

[†]九州工業大学大学 〒820-8502 福岡県飯塚市川津 680-4

^{††}奈良先端科学技術大学院大学 〒630-0192 奈良県生駒市高山町 8916-5

E-mail: [†]nagashima@infonet.cse.kyutech.ac.jp, tsukamoto@cse.kyutech.ac.jp, ^{††}yuzo@is.naist.jp

あらまし 近年, IoT 技術の発達により, 異分野データ連携が注目されている. そこで先行研究では, 物理位置に基づく異分野データ連携を行うための通信基盤として, Geo-Centric 情報プラットフォーム (GCIP) を提案した. GCIP では各地域で生成された IoT データを収集し, その地域で様々な事業者が時空間コンテンツ (STC) を生成する. この環境では, 地域内で生成される IoT データの多様性に伴って STC の種類や生成タイミングが異なるため, ユーザは「どの STC」が「いつ」「どこで」生成されているか, を知ることができない. そこで先行研究ではユーザ要求と STC 生成時に取得可能な統計情報を用いてマッチングを行う検索手法を提案した. しかし, 先行研究はユーザ要求に一致する STC の検索に主眼を置いたため, STC の有効時間については考慮していなかった. そこで本研究では先行研究と同様にユーザに新たな気づきを与えつつ, STC の時間特性を考慮した新しいマッチング型検索手法を提案した. シミュレーション結果から提案手法は検索要求やコンテンツ生成環境によらず最適なサーバを確実に選択できることを示した. キーワード コンテンツ検索, IoT, 異分野データ連携, 地理指向

Matching based Spatio-Temporal Contents discovery method considering temporal characteristics

Kaoru NAGASHIMA[†], Yuzo TAENAKA^{††}, and Kazuya TSUKAMOTO[†]

[†] Kyusyu Institute of Technology

^{††} Nara Institute of Science and Technology

E-mail: [†]nagashima@infonet.cse.kyutech.ac.jp, tsukamoto@cse.kyutech.ac.jp, ^{††}yuzo@is.naist.jp

Abstract Cross-domain data fusion is becoming a key driver to growth of the numerous and diverse applications in IoT era. We have proposed a concept of new information platform, Geo-Centric information platform (GCIP), that enables IoT data fusion based on geolocation. GCIP produces new and dynamic contents by combining cross-domain data in each geographic area and provides them to users. In this environment, it is difficult to find appropriate contents requested by a user because the user cannot recognize what contents are created in each area beforehand. we proposed a content discovery method for GCIP in previous research. However, previous research is not considering temporal characteristics. In this paper, we propose new content discovery method considering temporal characteristics in STC. Simulation results showed that appropriate contents can be discovered in response to user's request.

Key words Contents discovery, IoT, Cross-Domain Data Fusion, Geolocation

1. 研究背景, 課題

近年, IoT 技術の発達により, 多種多様なモノ/ヒト等をセンシングして得られる異分野データの連携が注目されている [1]. 我々の研究グループでは, 地域に分散する IoT 機器の種類や IoT データの管理者に依存しない形での二次利用を目的として, 地理空間を意識した IoT データの収集, 処理, 取得を行う Geo-Centric 情報プラットフォーム (GCIP) [2] を提案している.

GCIP では, 図 1 のように空間を経度緯度に基づいて階層構造のあるメッシュ構造に分割しメッシュ構造に対応したネットワークを構成し, メッシュ内で収集した IoT データからコンテンツを生成する.

一つのメッシュ内にはデータ蓄積サーバ (DS サーバ) と異分野連携サーバ推定 (DF サーバ) の 2 種類のサーバを配置する (図 1). DS サーバには, メッシュに対応した地理空間範囲内の全ての IoT データが収集され, DF サーバは DS サーバから多種

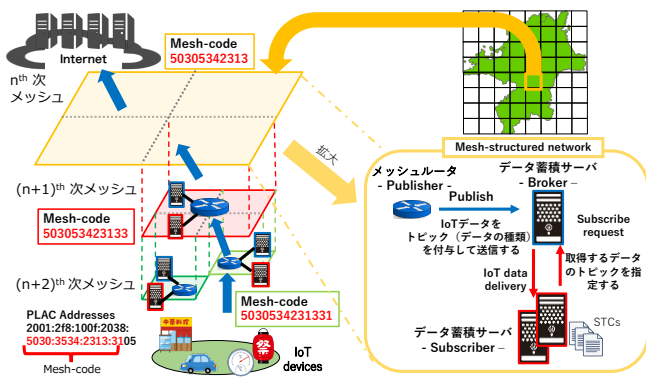


図1 GCIP の想定環境

多様な IoT データを取得し、組み合わせてその地域のためのコンテンツである時空間コンテンツ (STC) を生成する。

収集データは、IoT 機器の種類や移動によりその収集間隔やデータ量、存在自体までもが時々刻々と変化する。その結果、DF サーバは収集データに応じて動的に STC を生成する。よって、ユーザは DF サーバが保持する STC を事前に把握できないため所望するコンテンツの発見が困難となる。

先行研究 [3] はユーザが STC を適切に発見するための検索手法として、DF サーバが STC 生成時に要求するデータ属性の統計情報とユーザ要求の優先度をマッチングし、適切な DF サーバを検索する手法を提案した。しかし、先行研究はユーザ要求を完全に満たす STC が保持する期待値が最大の DF サーバを優先していたため、各 STC に有効時間を考慮していない。

そこで本研究では多種多様な STC が生成される環境下でユーザ要求を満たし、STC の有効時間を考慮した検索手法を提案する。2 節ではコンテンツ検索に関する関連研究を紹介し、3 節は、先行研究や想定環境を述べる。4 節では提案手法について説明し、5 節で比較手法や評価指標を説明する。6 節でシミュレーションの考察を行い、7 節でまとめる。

2. 関連研究

本節では既存のコンテンツ検索手法について述べる。参考文献 [4] ではコンテンツ検索に焦点を当てた既存の研究をまとめている (location-based [5], metadata-based [6], event-based [7] など)。しかし、GCIP でのコンテンツ検索に必要な時間、場所、コンテンツの中身の 3 つのパラメータすべてを使用した手法はなかった。さらに、収集した IoT データから動的に生成されたコンテンツを検索する手法についても言及されていなかった。本稿ではこの 3 つに着目した検索手法を提案する。

次に既存のコンテンツ検索手法として、データ指向型ネットワーク (ICN:Information-CentricNetworking) のコンテンツ検索手法について述べる。ICN は、効率的なコンテンツの検索・流通を実現する有望な概念である。ICN は IP ベースではなくコンテンツベースで運用されているので、ユーザはコンテンツの位置を知らなくても、コンテンツ名を使って直接コンテンツを検索することができる [8]。しかし、GCIP ではコンテンツが、その時点で収集された IoT データから作成されるため、ICN で

の検索タイミングではユーザはコンテンツの名前を知ることができず、コンテンツを検索することが非常に困難である。本稿では、動的なコンテンツ生成に対してコンテンツを構成するトピックを用いることで検索を実現する。

3. Geo-Centric Information Platform: GCIP

3.1 物理位置調和型通信

先行研究 [2] では、図 1 のメッシュに対して地理ごとに一意の ID を埋め込んだネットワークアドレスを割り当てることで物理位置に基づく通信経路制御を実現する物理位置調和型通信を提案した。この物理位置調和型通信を利用することで時空間コンテンツ検索時に、位置情報を意識した検索が可能となる。

3.2 GCIP 内での STC 生成方法

各メッシュ内の DS サーバは各都道府県や市町村などの自治体が設置し、DF サーバはその地域でユーザに STC を提供したいコンテンツプロバイダが設置することを想定する。本研究では GCIP のあるメッシュ内には DS サーバが 1 台、DF サーバが複数台設置されていると想定する。

STC 生成を実現するために、各メッシュ内に設置されているメッシュルータと DS サーバ、DF サーバ間の通信はデータの 2 次利用を行うため、メッシュルータと DF サーバの非同期通信を実現する必要がある。そのため、Publish/Subscribe (Pub/Sub) モデルを使用する (図 1)。メッシュルータを Publisher、DS サーバを Broker、DF サーバ DF サーバを Subscriber とする。Publisher であるメッシュルータは IoT 機器がクラウドに送信したデータを途中で複製し、データの種別を示すトピックを付与して DS サーバに Publish することで DS サーバにデータを蓄積できる。一方で STC 生成のために DF サーバは取得したいデータのトピックを複数指定して DS サーバに Subscription を送信し、受け取ったデータから処理を行い STC を生成する。例えばトピック A, B, C の 3 つを用いて STC を作成したい場合はトピック A, B, C を Subscription に含めて送信する。

なお、各 STC の有効時間は DF サーバによって設定されており、有効時間を超過した STC には価値はないものとする。本研究では 1 回の Subscription で 1 つの STC が生成されるものの、同一のトピックの組み合わせで構成される Subscription が同一の STC なのか、異なる STC を生成しているか DS サーバでは判断できない。

4. 提案手法

4.1 マッチング検索手法の概念設計

各 DF サーバは異なる事業者が管理しコンテンツを生成しているため、各 DF サーバで生成コンテンツの種類や生成タイミングが異なるため、ユーザはどこでどんな STC が生成されているのかを知ることができず、DF サーバへ直接検索要求することができない。また、各メッシュ内には複数の DF サーバが存在し、各 DF サーバで生成コンテンツの種類や生成タイミングが異なるため、STC が生成される時刻が分からない。この環境ではユーザは生成コンテンツの名称や明示的なキーワードを指定して STC 検索をできないため、新たな検索方法が必要とな

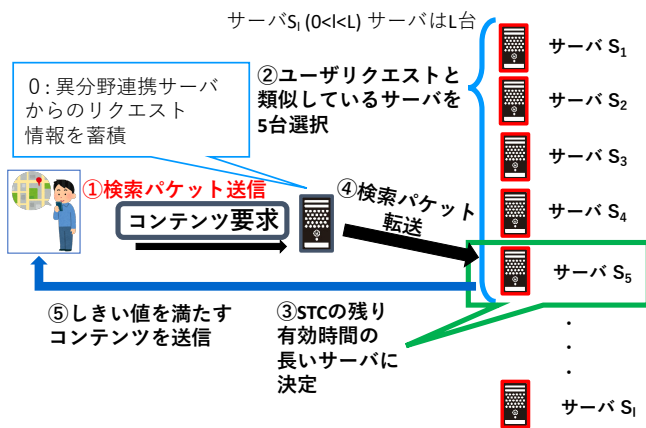


図2 マッチング型検索手法概要

る。そこで、DSサーバはメッシュ内に存在する全てのDFサーバからSTC生成時にSubscription要求を受信するため、統計情報を活用できる点に着目した。

図2にマッチング型検索の概要を示す。ユーザはまず、各メッシュ毎に設置され、全てのデータを保持しているDSサーバへ検索パケットを送信する。検索パケットを受け取ったDSサーバはDFサーバからのSubscription履歴をユーザリクエストとマッチングして、最適なDFサーバを推定する。その後、DSサーバは検索パケットを推定したDFサーバに転送し、サーバはユーザ要求を満たすSTCをユーザに送信する。以降は、まずユーザリクエスト及び、マッチング手法について説明する。

4.2 ユーザリクエスト

ユーザは生成されるSTCを予め知ることはできないため、検索パケットにはコンテンツ名やキーワードを含める事ができない。そこで、STC生成時に利用されるトピックをユーザリクエストで指定する。しかし、時々刻々とSTCが生成される環境でユーザが存在するSTCを表現する複数のトピックを指定することは困難である。そこで、ユーザは3つの情報を自身の端末に入力する。(1)位置情報: ユーザが特定の物理位置に存在するコンテンツを要求する場合、要求する地理範囲に対応したmesh-codeを、物理位置調和型アドレス内のmesh-codeフィールドに埋め込み、検索パケットを送信する。これにより、指定した検索パケットを送信できるため、その後、指定空間内に存在するコンテンツを検索できる。(2)検索キーワード: ユーザが所望するSTC表現するためにキーワード形式で指定する。(3)あいまい度: ユーザはどれくらい曖昧さを許容するかを0~100の数値で指定する。あいまい度が低いほうがより曖昧さを許容した検索を行うことになる。

上記の項目をユーザが入力した後で、検索キーワードをトピックに変換した上で、検索パケットが指定空間内のDSサーバに送信される。具体的には、(1)トピックの生成: ユーザが入力した検索キーワードを意図抽出技術[9],[10]に入力し、トピックに分解する。トピックに分解することでDFサーバがSTC生成時に使用しているトピックとの対応付けが可能となる。端末は日々のユーザの入力情報からユーザの嗜好を知っている事が前提となる。(2)各トピックの重要度の設定: 前述した意図抽出技

術を用いて各トピックに設定する重要度を自動で推定する。この際、全トピックの重要度の合計が100%になるように設定する。(3)あいまい度: ユーザが入力したしきい値 α をそのまま送信する。ユーザがどこまで曖昧な検索を許容するかのしきい値として α を使用する。

ユーザがキーワードとして「快適な作業環境を知りたい」と入力した場合、意図抽出によって、「騒音」、「Wifi」、「気温」トピックが決定され、重要度が(騒音, Wifi, 気温)=(35,35,30)のように設定される。また、トピックと重要度に加え、設定したあいまい度も検索パケットに含まれる。

4.3 最適な異分野連携サーバの定義

本研究での最適な異分野連携サーバを「ユーザの要求を満たすSTCを多く持ち、それらの残り有効時間が長いサーバ」と定義する。つまり、ユーザの要求を満たすSTCを多く持つだけでなく、古いSTCを多くもつサーバよりも新しいSTCを多く持つサーバのほうがユーザにとって有益である。そこで、メッシュ内に複数台設置されるDFサーバの中から最適なDFサーバを決定する方法について述べる。「ユーザ要求を満たすSTC数」と「STCの有効時間がより長い」という観点からそれぞれの項目で各DFサーバに順位をつけ、その順位の和が値が最も小さいサーバが最適なDFサーバとする。順位の和が同じサーバがあった場合はユーザ要求を満たすSTC数が多いDFサーバを最適なDFサーバとする。このように定義することでSTC数は多いものの古いSTCばかりのサーバや、一つのSTCの有効時間がすごく長いものの、STC数が少ないサーバは選択されず、STC数が多く、STC全体の有効時間の長いSTCを保持するサーバを最適サーバとして決定する。

4.4 STC数が多いサーバの選択手法

ユーザ要求を満たすSTCを多く持つDFサーバをマッチングするために、「ユーザ要求の重要度」と「各サーバの使用トピックのSubscriptionに占める割合」を用いて最適なDFサーバの推定を行う。ユーザ要求の各トピックの重要度と各トピックのSubscriptionに占める割合に近いほど、ユーザとDFサーバがそのトピックを重要視している割合が同じだと考えられる。そのため本研究では、ユーザの重要度と各サーバの使用トピックのSubscriptionに占める割合の値がどれだけ近いのかを確認するためにコサイン類似度を用いる。コサイン類似度の算出、及び活用について、次節で説明する。

4.4.1 コサイン類似度の算出方法

4.1節で説明したように、DSサーバは各DFサーバのSubscriptionのトピックの組み合わせ、最終送信時刻、送信間隔、Subscriptionを受け取った回数を保持している。そこでDSサーバにおいて、「DFサーバから受信したSubscriptionに基づく各トピックの使用割合」と「ユーザ要求に含まれるトピック毎の重要度」のコサイン類似度を算出し、類似度が高いサーバの中が最適なDFサーバであると考え、その精度を確認する。

DSサーバは、各DFサーバから受信したSubscribe要求から、以下の情報を保持可能であるものとする。

- メッシュ内に存在するサーバの台数: L
- DFサーバのリクエストしたトピックの組み合わせ集合:

$C = \{c_1, c_2, \dots, c_m, \dots, c_M\}$, M は DF サーバのリクエストした組み合わせ数.

- 組み合わせ c_m の最終送信時刻: t_{c_m}
- 組み合わせ c_m の送信間隔: i_{c_m}
- 組み合わせ c_m の送信回数: n_{c_m}
- ユーザ要求を満たす STC の総 Subscription 回数: n_{sum}

次に, 組み合わせ c_m が 1 台の DF サーバの Subscription に占める割合を求める. ユーザの指定したあいまい度をしきい値 α とし, 検索パケットに指定されたトピックの重要度の総和が α を超える組み合わせのみを使用する. ユーザが Topic a,b,c を指定しており, $(a,b,c) = (80,15,5)$ と指定していた場合, しきい値 α を 80 とすると, 組み合わせ $\{a, \{a,b\}, \{a,c\}, \{a,b,c\}\}$ が該当する. α を満たす組み合わせ c_m の STC の重みを $w_{c_m} = n_{c_m}/n_{sum}$, C の中でトピック j を含む c_m の集合 C'_j と定義する. トピック j の重み w_j を C'_j に含まれる要素の w_{c_m} の総和と定義する. 各トピックの重みを正規化したベクトルを DF サーバの重みベクトル W とする.

ある DF サーバが Topic a~e を均等に使用している時, DF サーバの重みベクトルは, $W = (20, 20, 20, 20, 20)$ となる, この際, ユーザが Topic a,b,c を指定しており, 重要度を $(a,b,c) = (35,35,30)$ のように指定していた場合はユーザが指定していないトピックの重要度は 0 と扱い, $W_U = (35, 35, 30, 0, 0)$ と表される, これらのベクトルを用いてコサイン類似度を式 1 のように計算する. このうち, コサイン類似度が最大となる DF サーバを選択する.

$$CS_l = \text{CosSim}(W, W_U) \quad (1)$$

4.4.2 コサイン類似度の活用方法

本調査で想定するシミュレーション環境を図 3 に示す. メッシュ内に存在する DF サーバは 10 台存在し各 DF サーバは特定のトピックを多く Subscribe 要求するように偏りを持たせる. DF サーバごとに Subscribe 要求時に特定のトピックを取得する確率を 50%(Subscription bias 0.5:以下 SB=0.5 とする), 100%(Subscription bias 0.5:以下, SB=1 とする)と設定し, DF サーバ毎に異なるトピックを多く要求する環境を想定する. 生成コンテンツを構成するトピック数はランダムに設定する. ユーザ要求はユーザが所望する STC が明示的でトピックに割り当てる重要度を, 特定のトピックに重要度を大きく指定した明示的な検索の場合と, ユーザの所望する STC が曖昧で各トピックの重要度の差がほぼ見られない曖昧な場合で検索を行う. シミュレーション内では 10 分間 STC を生成し, 次の時刻に検索パケットを送信して評価を行う. また本シミュレーションでは, 表 1 のようにパラメータを設定する. ユーザは明示的な検索パケット (Explicit Request, 以下 ER とする), もしくは曖昧な検索パケット (Unclear Request: 以下 UR とする) を DS サーバに送信する場合の性能をそれぞれ 1000 回試行する,

コサイン類似度の値が大きいほど, 最適な DF サーバであると考えられるため, 式 1 で求められた CS_l が最大の DF サーバを最適な DF サーバであると推定する. 図 4 に最適な異分野連

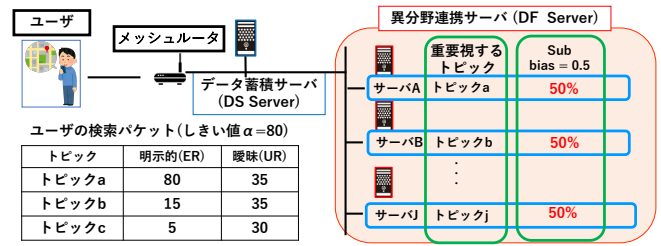


図 3 シミュレーショントポロジ

表 1 シミュレーションパラメータ

サーバ数	10 台
単位時間あたりの生成コンテンツ数	100[個/台]
トピックの種類	10[種類]
連携トピック数	2~5
STC の平均有効時間	10[分]
しきい値 α	80
平均有効時間 λ	10[分]

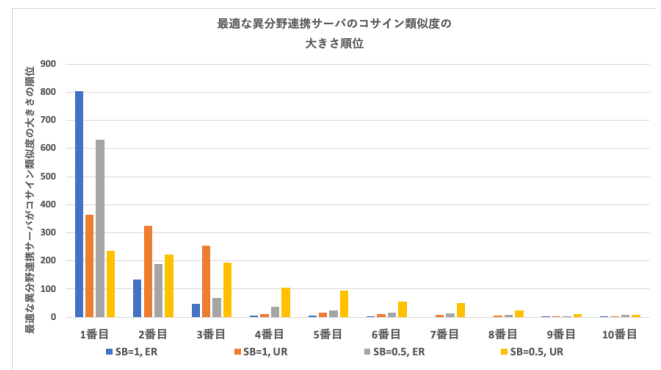


図 4 コサイン類似度による順位 vs. 最適な DF サーバの選択回数

表 2 最適な DF サーバのコサイン類似度の範囲

	最小値	最大値
ER	0.63	0.82
UR	0.82	0.97

携を推定したシミュレーション結果を示す. 図 4 の「1 番目」の部分にそれぞれの環境で 1000 回のシミュレーション中, 4.3 節で定義した最適な DF サーバを推定できた回数を示す. この結果からコサイン類似度 1 位のサーバが最適な DF サーバとは必ずしも限らないことが明らかになった.

そこで, 最適な DF サーバをコサイン類似度の大きさの順番で並べたときの分布と, その際に取り得るコサイン類似度の値の範囲を調査した. 式 1 で求められた CS_l を降順に並べた際に最適な DF サーバはコサイン類似度が何番目になるか, を調査する. 調査結果を図 4 に示す. このときのコサイン類似度のとりうる値の範囲を表 2 に示す. これらの結果から最適な異分野サーバはコサイン類似度が大きなサーバ 5 台の中に約 90% 含まれていることから, コサイン類似度が上位 5 台以内かつ, 表 2 のしきい値を満たすサーバ郡を検索結果の候補として扱う.

4.5 STC の有効時間が長いサーバの推定手法

4.4 節の調査から「ユーザの要求を満たす STC 数」と「STC

の残り有効時間」の2つのパラメータのうち、ユーザの要求を満たすSTCを多く持つサーバを探すことが可能な条件を示した。次にコサイン類似度の条件を満たしたサーバの中から有効時間推定による最適な異分野サーバ推定方法について説明する。

STCの有効時間が長いサーバを、DFサーバのもつユーザ要求を満たすSTCの中で、残り有効時間の中央値が最も大きなサーバとする。このように定義することで「STC数は多いが作成された時刻が古いサーバ」や「他のSTCよりも残りの有効時間が非常に大きなSTCを持つサーバ」が選択されることを排除できる。また各STCの有効時間をDSサーバは知ることができないため、平均有効時間 λ のポアソン分布の値 p_{c_m} を用いて推定する。具体的には、組み合わせ c_m の残り有効時間を e_{c_m} を式2で計算する。ユーザの検索パケットの到着時刻を t_{now} と置く。

$$e_{c_m} = p_{c_m} - (t_{now} - t_{c_m}) \quad (2)$$

DFサーバ S_l の e_{c_m} 集合の中央値を U_{S_l} とする。メッシュ内のすべてのDFサーバのうち U_{S_l} が最大となるサーバは最適なDFサーバである確率が最も高いため、DSサーバはユーザの検索パケットを転送し、検索を行う。

DFサーバは検索パケットに指定された重要度の合計がしきい値 α を超えるトピックの組み合わせを含むコンテンツを検索した上で、該当する全てのSTCをユーザに返信する。

5. 性能評価

5.1 実験環境

また本シミュレーションでは、図3のトポロジを使用する。また、表1の平均有効時間を10, 30, 60分としたパラメータを用いる。1回のシミュレーションではある環境に検索パケットを送信して評価を行う。これを環境を変えて1000回試行する。検索パケットはERとURをそれぞれ用いて評価する。

5.2 評価指標

本研究では4つの評価指標を用いる。1つ目は各手法のDFサーバの推定が何番目に最適なDFサーバを選択したのかという推定精度を用いる。これによってDSサーバは、ユーザの検索条件に適しているDFサーバを推定することが可能かどうかを評価する。2つめは推定されたサーバからユーザは所望するSTCの取得数を用いる。ユーザ要求に完全にマッチするSTCのみ取得した場合をcp, 及びユーザ要求以外に一つのトピックを含むSTCを取得するcp+1, 二つ含むcp+2と表記し、それぞれ比較する。ユーザの取得できるSTC数は多いほうが良いため、比較手法と提案手法の精度を確認する。3つめはユーザが指定していないトピックがどれだけ含まれているかを表す気づきの割合(Unconscious Contents Ratio: UCR)を用いる。UCRは式3を用いる。 C_{total} はユーザが取得することができるSTCの総数、 $C_{complete_match}$ はユーザの要求トピックとSTCの構成トピックが完全一致したSTCの数を表す。

$$UCR = \frac{C_{total} - C_{complete_match}}{C_{total}} \times 100 \quad (3)$$

最後の評価指標にはユーザが取得するSTCの残り有効時間の分布を用いる。式2で求めたSTCの有効時間からユーザが取得するSTCの残り有効時間を評価する。

5.3 比較手法

本シミュレーションでは2つの比較手法を用いて提案手法の有効性を評価する。比較手法1は参考文献[3]で提案された手法(期待値ベース)を用いる。この手法では、ユーザの要求を満たすSTC数と各トピックのSubscribe回数を用いてユーザ要求に完全一致するSTCを持つ期待値が最も大きなサーバを選択する。DFサーバ i における期待値 E_i を式7に示す。ユーザ要求を満たすトピックの組み合わせを c ($c \in C$)、DFサーバ i のトピック j の要求確率を $P_{i,j}$ 、DFサーバ i のDFサーバが組み合わせ c を持つ確率を $G_i(c)$ (式6)とする。

$$G_i(c) = \prod P_{i,j} \quad (4)$$

$$E_i = \sum_{c \in C} (G_i(c) \times N_i) \quad (5)$$

比較手法2はSTC有効時間の特性の考慮の有無での変化を確認するために、STC数が最も多いと思われるサーバを選択するコサイン類似度ベース手法を用いる。提案手法で導かれた CS_l の値が最も大きなサーバが最適なサーバを選択する。

5.4 結果・考察

図5に提案手法、比較手法の推定精度を示す。ユーザリクエストER/UR, Subscription bias(SB) 1.0/0.5を合わせた4種類の環境でそれぞれ、推定した最適なサーバの順位の割合を表す。提案手法はユーザリクエストもSBも偏りが大きい環境では、92%の精度で順位が3番目までのDFサーバを推定できている。ユーザリクエストもSBも偏りが小さい環境では、81%の精度で順位が3番目までのDFサーバを推定できている。次に表3, 4に各SBにおけるユーザのSTC取得数の平均値とUCRを示す。ユーザの取得STCの残り有効時間のグラフを図6に示す。STCを多く収集している提案手法(Proposed)とコサイン類似度ベース(Cosine Similarity, CS)の結果を示している。

最適なDFサーバの推定精度は全ての環境で提案手法が最も精度が高く、最も検索が難しいSB=0.5, URでも47%の推定ができている。次にコサイン類似度ベースの精度が高く、SB=0.5, URでは20%の推定をしている。提案手法の推定精度はSubscribeの傾向と、ユーザリクエストの偏りが大きい環境ほど高く、小さいほど低い結果となった。Subscribe傾向の偏りが小さいほどサーバの重要視するトピックが含まれにくくなり、コサイン類似度の値が小さくなるためにサーバが選ばれにくくなるためである。

期待値ベースは全ての環境で推定精度が1%未満である。期待値ベースはユーザの要求と完全一致するSTCが最も多いサーバを選択するための手法であり、STCの有効時間を考慮していない。その結果、期待値ベースでは5番目までに最適なDFサーバを選ぶことができず、10台のDFサーバの中で最も悪いサーバを選択することが最も多い。コサイン類似度ベースでは期待値ベースよりも大幅に改善し、最も最適なDFサーバを選

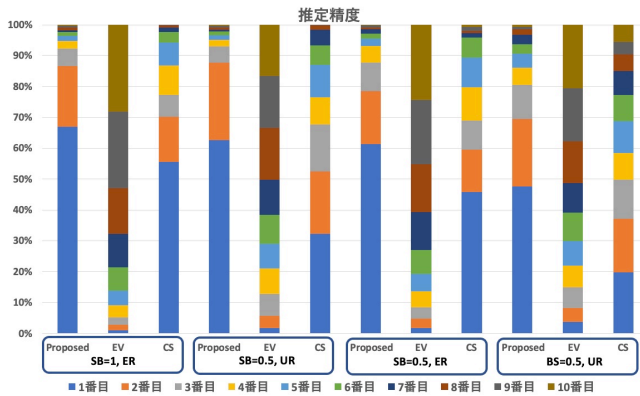


図5 推定精度 (提案手法)

表3 SB=1 の STC 取得数

	ER				UR			
	cp	cp+1	cp+2	UCR	cp	cp+1	cp+2	UCR
提案手法	4	14	28	95	4	15	29	93
期待値ベース	0	2	9	100	0	3	11	1
コサイン類似度ベース	5	18	34	95	8	20	32	87

表4 SB=0.5 の STC 取得数

	ER				UR			
	cp	cp+1	cp+2	UCR	cp	cp+1	cp+2	UCR
提案手法	3	8	15	96	3	8	14	87
期待値ベース	1	4	13	97	1	4	13	94
コサイン類似度ベース	3	9	15	95	5	10	12	82

択することが可能である。期待値ベースは cp+1 や cp+2 に関しては考慮していないため、cp+1 や cp+2 も含めて取得 STC が最大となる DF サーバの推定ができていない。時間の特性を考慮していないため、すべての環境で提案手法よりも最適な DF サーバを推定できた回数が少ない。

提案手法とコサイン類似度ベースで取得できる STC の有効時間の中央値の差は、SB=1 や、ER の環境では 1 分以内だった。しかし、偏りの小さい SB=1, UR では約 4 分、SB=0.5, UR では約 5 分の差があった。提案手法は SB=0.5 や UR の環境で類似度手法と比べて有効時間の長い STC を提供できる。

すべての手法でユーザが未指定のトピックを含む STC を取得でき、ユーザに新しい気づきを与えることができる。また期待値ベースでは、順位の高いサーバを選べていないためユーザ要件に完全一致する STC の取得はできず、ユーザの未指定トピックを含む STC のみを取得できる。

6. まとめ

GCIP では時々刻々と生成される STC によってユーザは自身が所望する STC が存在するか、どのサーバに存在するのかわからない。そこで、本研究では GCIP の時々刻々と存在する STC が変化する環境で、「ユーザ要求の重要度」と「各サーバの使用トピックの Subscription に占める割合」の類似度と STC

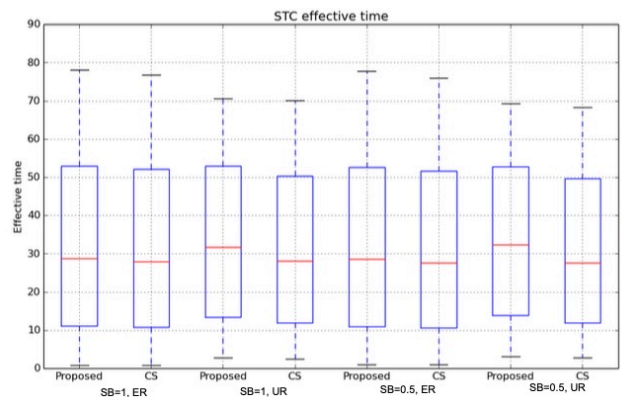


図6 取得 STC の残り有効時間

の有効時間に着目し、ユーザの要求を満たす STC を検索する方法を提案し、ユーザが STC を得ることができることを示した。今後は同一の Subscription の送信間隔からいくつの STC を生成しているのか推定を行い、検索精度を上げる手法を検討する。

謝辞

本研究成果は、JSPS 科研費 JP18H03234 及び国立研究開発法人情報通信研究機構の委託研究により得られたものです。

文献

- [1] A. Al-Fuqaha, et al. "Internet of Things: A Survey on Enabling Technologies, Protocols and Applications," IEEE Communications Surveys & Tutorials, vol. 17, no. 4, pp. 2347-2376, June 2015.
- [2] K. Tsukamoto, et al. "Geolocation-centric Information Platform for Resilient Spatio-temporal Content Management," IEICE Trans. Commun., Online ISSN 1745-1345, Print ISSN 0916-8516, Sep. 2020.
- [3] K. Nagashima, et al. "Matching based content discovery method on Geo-Centric Information Platform," INCOS 2020, vol 1263, pp 470 - 479, Sept 2020.
- [4] S. Pattar, et al., "Searching for the IoT resources: fundamentals, requirements, comprehensive review, and future directions," IEEE Commun. Surv. Tutorials vol.20, pp. 2101-2132 (2018).
- [5] S.Mayer, D. Guinard, V. Trifa: Searching in a web-based infrastructure for smartthings. In: 2012 3rd IEEE International Conference on the Internet of Things, pp.119-126, October 2012
- [6] Mayer, S., Guinard, D.: An extensible discovery service for smart things. In: WoT 2011: Second International Workshop on the Web of Things, June, pp. 1-6 (2011)
- [7] Pintus, A., Carboni, D., Piras, A.: Paraimpu: a platform for a social Web of Things. In: Proceedings 21st International Conference on Companion World Wide Web (WWW Companion), pp. 401-404, April 2012
- [8] G. Xylomenos, et al.: A survey of information-centric networking research. IEEE Commun. Surv. Tutorials vol 16, pp. 1024-1049 (2013)
- [9] 栗原理聡他, Twitter を利用した地域毎の要望抽出, 人工知能学会全国大会論文集, 2015, JSAI2015 巻, 第 29 回全国大会 (2015).
- [10] 金子 貴美他, 商品レビューからの購買意図の抽出と自動階層化, 人工知能学会全国大会論文集, 2016, JSAI2016 巻, 第 30 回全国大会 (2016), セッション ID 2P1-13in1, 公開日 2018/07/30