

絶滅危惧種と生息国事象間の意味的關係解釈を実現する コンテキスト・ベース知識獲得システム

井上 紗里[†] 佐々木 史織[‡] 清木 康^{†‡}

[†] 慶應義塾大学環境情報学部 〒252-0882 神奈川県藤沢市遠藤 5322

[‡] 慶應義塾大学大学院政策・メディア研究科 〒252-0882 神奈川県藤沢市遠藤 5322

E-mail: [†] t14094si@sfc.keio.ac.jp, [‡] shiori.sasaki@gmail.com ^{†‡} kiyoki@sfc.keio.ac.jp

あらまし 絶滅危惧種への影響を様々なコンテキストによって分析し、可視化するシステムは、環境問題、特に絶滅危惧種の大局的かつ局所的状況を把握し、新たな対策の実行を進めることに貢献する。本論文で提案する知識獲得システムは、ある絶滅危惧種の生息に影響を与える様々な因子をコンテキスト（例えば生息地の消滅）とし、その種の生息国の事象（ニュース記事）とコンテキストの関連性を計量し、その結果をコンテキスト指向図及び 5D World Map 上で可視化する。記事の伝える感情、すなわち、絶滅危惧種に関する肯定的（絶滅危惧種の状況が良くなることを伝える）・否定的（絶滅危惧種の状況が悪くなることを伝える）感情を、記事とコンテキストの関連度計算に含めることで、関連性計量精度の向上を実現し、また可視化の手法として、コンテキスト指向図と 5D World Map を組み合わせることで、包括的な情報（生息国間やコンテキスト間の比較など）の可視化を実現させた。このシステムは、専門知識のない一般の人々が、絶滅危惧種に纏わる環境問題の包括的且つ深層的な知識を直感的に得ることを支援する。

キーワード multi-database, text processing, news articles, visualization, environmental science, context-oriented diagram, 5D World Map, semantic analysis, endangered species, 意味的連想検索, 空間・時空間データベース, テキストデータ処理

1. はじめに

地球上の自然と社会との接点において、自然環境の持続性を高めることは急務である。絶滅危惧種は、自然環境の実情を示すバロメーターとなっており、絶滅危惧種の現状を具に把握することは、地球全体で起きている環境問題を解釈するひとつの有力な方法であると考えられる。2015年10月25日には、国連総会で『我々の世界を変革する：持続可能な開発のための2030アジェンダ』と題する具体的行動指針が採択された[1]。世界的に、環境問題への危機感が高まりつつあることが窺える。しかしながら、環境問題はあまりにも対象が大きく、様々な因子が複雑に絡み合っているという性質から、一般人には捉えがたい問題である。国際的な組織がどれほど声高に問題を喧伝しても、環境問題の解決には、各個人が問題の本質を把握して危機意識を持ち、解決のために行動を起こしていくことが不可欠だと考える。

我々の提案するシステムは、環境問題、特に絶滅危惧種の問題を各個人が全容把握することを支援し、この社会の持続可能性を高める一助となる可能性を持っている。本論文で提案するコンテキストをベースとした知識獲得システムは、絶滅危惧種に焦点を当て、絶

滅危惧種の現状解釈を促すことを目的としている。

本論文では本システムの実現と発展の可能性を、絶滅危惧種にマウンテンゴリラを設定した実験によって検証する。

2. 基本方式

本システムは、ドキュメントの収集、ドキュメントの内容分析、可視化の3つのプロセスによって構成される。

2.1. ドキュメント(コンテキスト・ニュース記事)の収集

本システムは、絶滅危惧種の生息に影響を与える人為的な因子、例えば密猟などの影響因子についての説明文を“コンテキスト”として収集対象とする。ここで、本システムにおける“コンテキスト”とは、絶滅危惧種の状況を解釈する対象となる事象の表現である。コンテキストはベクトルとして定義されるデータ構造で、各事象は影響因子（コンテキスト）の特徴量によって表現される。

コンテキストにはポジティブなものとネガティブなもの2種類があり、絶滅危惧種の生息数を増やすものをポジティブなコンテキスト、減らすものをネガティブなコンテキストとして集めている。本システム

は、社会的な事象や自然の事象が記されたニュース記事を対象とし、様々なニュースサイトを情報源としている。例えば、絶滅危惧種の生息国での、絶滅危惧種の名前が含まれる記事を連結し、分析対象とする。

2.2. ドキュメントの内容分析

本システムは、コンテキストとニュース記事の内容分析において、意味的連想検索方式[2][3][4]の概念を適用している。コンテキストとニュース記事の関連度を求める際に、それぞれの情報に含まれる意味に着目することで、人間の感覚に近い分析を実現している。

2.2.1. 文書の特徴抽出

本ステップでは、収集された絶滅危惧種の影響因子の説明文とニュース記事、それぞれの文書の傾向・特徴を、単語の出現頻度として抽出し、ベクトル化する。ここでは TF-IDF を適用する。TF-IDF は文書の特徴を算出するための計算方法で、単語出現頻度 (Term Frequency, TF) と、逆文書頻度 (Inverse Document Frequency, IDF) を掛け合わせるものである。

$$tfidf_{i,j} = tf_{i,j} \cdot idf_i$$

単語出現頻度 (TF) は、文書 d_j における単語 t_i の出現回数 $n_{i,j}$ を、文書 d_j における全単語 (k 個) の出現回数の和で割ったものである。

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

逆文書頻度 (IDF) は、全文書数 D における単語 t_i を含む文書数 d の逆数の対数をとったものである。

$$idf_i = \log \frac{|D|}{|\{d : d \ni t_i\}|}$$

2.2.2. 文書ベクトルの類似度計算

本ステップは、TF-IDF によって算出された、コンテキスト (絶滅危惧種の影響因子) のベクトルデータとニュース記事のベクトルデータの類似度を計量する。これは、各コンテキストと各ニュース記事の関連性を推定するための計算である。ここではコサイン類似度計算を適用する。コンテキストのベクトルデータ \vec{c} とニュース記事のベクトルデータ \vec{a} の角度において、ベクトル同士の角度が近いほど値は 1 に近く、遠いほど 0 に近づく。コンテキストとニュース記事の内容が近いほど値は 1 に近くなる。

$$\cos(\vec{c}, \vec{a}) = \frac{\vec{c} \cdot \vec{a}}{|\vec{c}| |\vec{a}|}$$

2.2.3. ニュース記事の感情分析

本ステップは、コンテキストとニュース記事間の関連度計算の精度を上げるために、ニュース記事の感情分析を行う。これは、文書に含まれる肯定的・否定的な感情を数値化する計算である。単語感情極性 (polarity) という -1 から 1 までの連続した数値として各文書に含まれる感情が表され、単語感情極性が 1 に近いほどその文書は肯定的な内容に、-1 に近いほど否定的な内容に偏っているとされる。

2.2.4. コンテキストとニュース記事間の複合相関量計算

本ステップは、コンテキストとニュース記事間の複合相関量計算として、コサイン類似度と感情分析結果を組み合わせる。コンテキスト c とニュース記事 a の最終的な相関量 $V_{c,a}$ は、それぞれのベクトル \vec{c} と \vec{a} のコサイン類似度とニュース記事の単語感情極性 p_a を掛け合わせ、それに重み w_c を掛けて算出される。重み w_c は、コンテキスト c がポジティブなものであれば 1、ネガティブであれば -1 に設定される。この重み付けによって、ポジティブなコンテキストと肯定的感情に偏った記事、ネガティブなコンテキストと否定的感情に偏った記事の相関量がより大きくなる。

$$V_{c,a} = \cos(\vec{c}, \vec{a}) * p_a * w_c, \quad w_c = \begin{cases} 1 & \text{if } c \text{ is positive} \\ -1 & \text{if } c \text{ is negative} \end{cases}$$

そして、この複合相関量 V_i は、その範囲を最小値 V_{\min} -1、最大値 V_{\max} 1 から、最小値 v_{\min} 0、最大値 v_{\max} 1 の値 v_i に変換される。

$$v_i = (v_{\max} - v_{\min}) * \frac{V_i - V_{\min}}{V_{\max} - V_{\min}} + v_{\min}$$

2.3. 可視化 (5D World Map、コンテキスト指向図)

本システムは、複合相関量計算の結果を 5D World Map[5][6][7][8]とコンテキスト指向図として視覚化する。

5D World Map では、絶滅危惧種とそのコンテキスト、そしてニュース記事が 5 次元的に可視化される。5D World Map は、マルチメディア・データを空間的、時間的、そして意味的に分析し、その結果を 5 次元の動的な歴史地図として統合する知識表出システムである。その構成要素は、空間次元 (3 次元)、時間次元 (4 次元目)、そして意味次元 (5 次元目) である。

コンテキスト指向図では、絶滅危惧種とコンテキスト、そしてニュース記事の関係が視覚化される。コンテキスト指向図では、コンテキスト間の関係や生息国間の数値的な比較など 5D World Map では表現するのが困難な情報を可視化できる。これらの情報は、時空

間的なコンテキストから独立しているため、5D World Map 上では表せない。5D World Map とコンテキスト指向図を組み合わせることによって、現実世界のより大局的、かつ、深く直感的な知識を得ることができる。

3. 実現方式

本システムは、図 1 に示す通り、4つの機構によって構成される。

3.1. Document Crawler (ドキュメント収集機構)

ドキュメント収集機構には、コンテキスト収集機構とニュース記事収集機構が含まれる。

3.1.1. Context Crawler (コンテキスト収集機構)

コンテキスト、つまり絶滅危惧種の影響因子はWWF (世界自然保護基金) のウェブサイト[9]から見つけられる。WWFは、100カ国以上で活動し、約500万人の会員によって支えられている国際的な自然保護組織である[10]。WWFは、絶滅危惧種の生物学的特徴や生息地、脅威や保護活動など様々な情報を、ウェブサイト等を通して提供している。影響因子のテキストデータはWWFのウェブサイトから得られ、コンテキスト・データベースに蓄積される。以下に、コンテキスト・データベースの構成要素を示す。

Context DB	
column name	note
id	context id
title	title of context
content	content of context
p_n	positive(=1) or negative(=-1)

3.1.2. Article Crawler (ニュース記事収集機構)

ニュース記事は、コンテキストに応じて検索され、Google News[11]のウェブサイトから収集される。Google Newsは、幅広い種類のニュースサイトを含んでおり、様々なジャンルの事象がニュース記事・データベースに集められる。

News Article DB	
column name	note
id	article id
title	title of article
content	content of article
date	yyyy-mm-dd
country	country covered by article
url	URL of article

3.2. Context-based Similarity Calculator (コンテキスト・ベース関連度計算機構)

コンテキストとニュース記事間の関連度合いの計算は、ベクトル化機構、コサイン類似度計算機構、そして感情分析機構、最終計算機構が組み合わさってい

る。

3.2.1. Vectorizer (ベクトル化機構)

コンテキスト・データベースとニュース記事・データベースに集められたデータに、TF-IDFを行う。TF-IDFの実施には、Pythonの機械学習ライブラリのScikit-learnが使われる[12]。

3.2.2. Cosine Similarity Calculator (コサイン類似度計算機構)

コンテキストとニュース記事間のコサイン類似度計算の前に、ニュース記事のベクトルデータはコンテキストのベクトルデータと互換性を持つよう調整される。その後、各コンテキストと各ニュース記事のベクトルデータのコサイン類似度が計算される。コサイン類似度計算には、Scikit-learnが使われる[13]。計算結果は、コサイン類似度・データベースに格納される。

Cosine Similarity DB	
column name	note
id	cosine similarity value id
c_id	context id
a_id	article id
ca_cs_value	cosine similarity value

3.2.3. Sentiment Analyzer (感情分析機構)

コンテキストとニュース記事間の関連度計算の精度を上げるために、感情分析が行われる。これは、文書に含まれる肯定的・否定的な感情を数値化する計算である。Pythonの自然言語処理ライブラリのVaderを使用して、感情分析がなされる[14]。Polarity (単語感情極性) という-1から1までの連続した数値として各文書に含まれる感情が表されており、polarityが1に近いほどその文書は肯定的な内容に、-1に近いほど否定的な内容に偏っているとされる。計算結果は、記事極性・データベースに格納される。

Article Polarity DB	
column name	note
a_id	article id
polarity	polarity of article
pos	positive value of article
neg	negative value of article

3.2.4. Final Calculator (最終計算機構)

可視化のために、コンテキストとニュース記事間の最終的な関連度を、コサイン類似度と感情分析結果を組み合わせで算出する。複合相関量計算結果は、最終値・データベースに蓄積される。

Final Value DB	
table name: cs_vader_p	
column name	note
id	final value id
c_id	context id
a_id	article id
cs_vader_p	final calculated value

3.3. Visualizer (可視化機構)

コサイン類似度と polarity を掛け合わせて計算された複合相関量は、5D World Map とコンテキスト指向図として可視化される。計算結果の時空間データとイメージデータは 5D World Map に投影される。コンテキスト指向図においては、コンテキスト間や生息国間などの視覚的且つ数値的な比較がなされる。

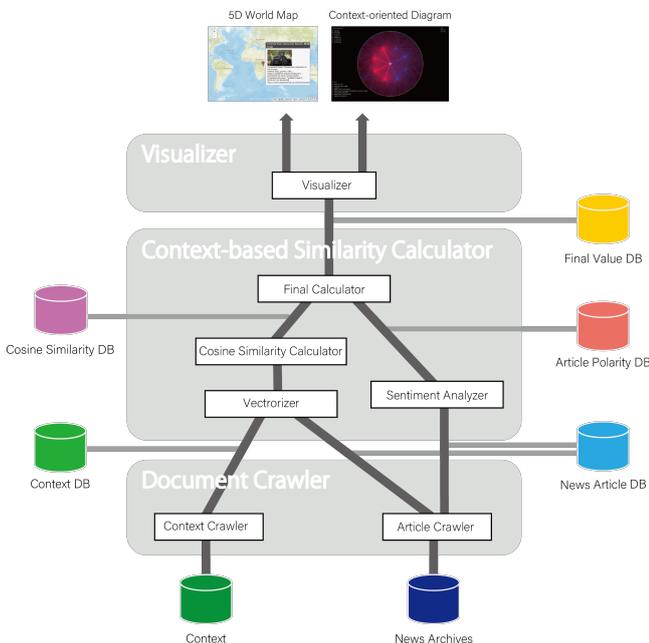


図 1: システム構成図

4. 実験

このシステムの実現可能性を示すために、実験を行った。

4.1. 実験の設定

4.1.1. 絶滅危惧種

この例では、マウンテンゴリラを絶滅危惧種として設定している。マウンテンゴリラはゴリラの4つの亜種のうちのひとつである。国際的な自然保護団体である IUCN は、野生生物をその生息状況を 10 のカテゴリーに分けており、マウンテンゴリラを「絶滅」から4つ目に近いカテゴリーである「絶滅危惧」と認定している[15][16]。

4.1.2. コンテキスト

絶滅危惧種のコンテキスト (影響因子) は 2 種類に分けられる。一つ目はネガティブなコンテキストであ

る。これは、絶滅危惧種の生息数を減少させる因子、つまり脅威のことである。マウンテンゴリラの場合、「生息地の消滅」(context id: 0)、「狩猟と取引」(context id: 1)、「病気」(context id: 2)、「法の不整備」(context id: 3)がこれに該当する。二つ目は、ポジティブなコンテキストである。これは絶滅危惧種の生息数を増加させる因子、つまり保護活動のことである。マウンテンゴリラの場合、「密猟と野生生物違法取引の防止」(context id: 4)、「観光の発展」(context id: 5)、「生息数の監視」(context id: 6)、「生息地の保全」(context id: 7)がこれに該当する。

4.1.3. 対象とするニュース記事

ニュース記事は、絶滅危惧種の生息国に関する記事のみとしており、この例では、野生のマウンテンゴリラの生息国であるコンゴ民主共和国、ルワンダ、ウガンダの計3国での記事を収集した。ニュース記事の出版時期は、2018年1月1日から10月1日と設定した。記事数は、コンゴ民主共和国のものが66件、ルワンダのものが264件、ウガンダのものが184件で計514件であった。

4.2. 計算結果

4.2.1. TF-IDF によってつくられた、文書のベクトルデータ

先述の通り WWF によると、マウンテンゴリラの状況を解釈するコンテキストは、ポジティブなもの4つとネガティブなもの4つの計8つある。以下に、8つのコンテキストのうちの1つ、「生息地の消滅」のベクトルデータを例として示す。

habitat loss (生息地の消滅) :

```
{agricultural: 0.226994955308,
companies: 0.226994955308,
cut: 0.226994955308,
destruction: 0.226994955308,
fields: 0.226994955308,...}
```

コンテキストとそのベクトルデータに多く含まれる単語との関係を示す。マウンテンゴリラの生息地の消滅(habitat loss)の主因は、大規模な企業(companies)が農業(agricultural)目的などで森林を切り倒して(cut)破壊(destruction)し、野(fields)を切り拓いていることである。このように、各コンテキストのベクトルデータはそれぞれのコンテキストの内容を正確に反映していると言える。

今回の実験においては、コンテキストのベクトルデータの次元数は 159、ニュース記事のベクトルデータの次元数は 16817 であった。相関量計算に移る前に、ニュース記事の次元数をコンテキストの次元数に合わ

せる処理を行った。

4.2.2. 各コンテキストと各ニュース記事間のコサイン類似度計算結果

以下に、コンテキストのベクトルデータとニュース記事のベクトルデータのサイン類似度計算の一例を示す。この例でのコンテキストは「生息数の監視」(context id: 6)である。以下に示しているのが、最もサイン類似度が大きい記事 20 本である。列左端からコンテキストの ID、ニュース記事の ID、ニュース記事のタイトル、コサイン類似度の値である。

c_id	a_id	title	cs_cs_value
6	79	11 Primal Facts About Dian Fossey	0.4096922884
6	259	11 Primal Facts About Dian Fossey	0.4096922884
6	211	Dian Fossey (1932)	0.3739208560
6	454	Dian Fossey (1932)	0.3739208560
6	318	Why is Dian Fossey a key icon for gorilla tourism and conservation?	0.3721196632
6	359	The Tragic Exploits Of Dian Fossey And Her "Gorillas In The Mist"	0.3697695213
6	327	DR Congo: Why do tourists go to Virunga National Park?	0.3371976244
6	151	Good news for mountain gorillas as researchers reveal there are now over 1,000 in the wild	0.335534758
6	378	Good news for mountain gorillas as researchers reveal there are now over 1,000 in the wild	0.335534758
6	0	"We know we may be killed": the rangers risking their lives for Virunga's gorillas	0.3274011998
6	67	"We know we may be killed": the rangers risking their lives for Virunga's gorillas	0.3274011998
6	253	"We know we may be killed": the rangers risking their lives for Virunga's gorillas	0.3274011998
6	155	Violence shuts Africa's Virunga gorilla park till 2019	0.3227739813
6	379	Violence shuts Africa's Virunga gorilla park till 2019	0.3227739813
6	88	Congo authorities continue search for kidnapped British tourists	0.3209942779
6	282	Congo authorities continue search for kidnapped British tourists	0.3209942779
6	24	Mountain Gorilla Population Climbs Above 1,000	0.3158915794
6	85	Mountain Gorilla Population Climbs Above 1,000	0.3158915794
6	278	Mountain Gorilla Population Climbs Above 1,000	0.3158915794
6	266	Understanding group relationships in gorillas	0.3141507263

最もサイン類似度が小さい記事 20 本は以下の通りである。

c_id	a_id	title	cs_cs_value
6	53	12 animals that are in danger of disappearing forever	0.0000000000
6	72	Arsenal Unveil 'Visit Rwanda' as New Sleeve Sponsor Supporting Country's Booming Tourism	0.0000000000
6	94	Competition: win the trip of a lifetime	0.0000000000
6	196	12 animals that are in danger of disappearing forever	0.0000000000
6	222	Gorilla Trekking in Uganda	0.0000000000
6	255	Arsenal Unveil 'Visit Rwanda' as New Sleeve Sponsor Supporting Country's Booming Tourism	0.0000000000
6	285	Ellen DeGeneres Cries After Receiving the 'Best' Birthday Gift From Portia de Rossi - El Online	0.0000000000
6	292	Watch Portia de Rossi Surprise Ellen and Make Her Burst into Tears	0.0000000000
6	331	Ellen DeGeneres' Birthday Party Had More Celebs Than an Awards Show	0.0000000000
6	333	Portia de Rossi Gives Ellen DeGeneres The Most Memorable 60th Birthday Gift	0.0000000000
6	354	Adorable photos show baby mountain gorilla playing around before tumbling into nanny's tummy	0.0000000000
6	446	Ellen DeGeneres weeps as Portia de Rossi reveals 60th gift	0.0000000000
6	459	Arsenal's new kit is not only pinkish - It has Visit Rwanda on it	0.0000000000
6	462	Ellen DeGeneres Cries After Receiving the 'Best' Birthday Gift From Portia de Rossi	0.0000000000
6	473	Ellen DeGeneres and wife Portia de Rossi spend time at Giraffe Manor in Nairobi, Kenya	0.0000000000
6	481	Gorilla Trekking in Uganda	0.0000000000
6	581	Portia de Rossi and Jamie Foxx, Sofia Vergara And More Celebrate Ellen's 60th Birthday	0.0000000000
6	589	Channel24.co.za Ellen DeGeneres' massive 60th birthday celebration - including Portia's heartfelt gift	0.0000000000
6	135	10 reasons to visit "shit-hole" countries	0.0149917334
6	390	10 reasons to visit "shit-hole" countries	0.0149917334

表中の、コンテキストと直接的に関係する記事をオレンジ色にマーカーをつけた。以上の例では、「生息数の監視」というコンテキストに意味的に非常に近い記事である、研究者によるマウンテンゴリラの生息数増加の発表に関する記事等が色付けされている。サイン類似度が小さい記事群には意味的に近い記事はなかった。以上より、サイン類似度計算によってコンテキストとニュース記事の意味的距離がある程度の精度をもって計量できることがわかる。

4.2.3. 各ニュース記事の感情分析計算結果

以下に、ニュース記事の感情分析の結果の例を示す。この例でのコンテキストは 4.2.2. と同様「生息数の監視」(context id: 6)である。列左端からニュース記事の ID、ニュース記事のタイトル、polarity、ポジティブ値、ネガティブ値が並んでいる。

Polarity の値が最も大きい記事 20 件が以下の通りである。Polarity の値が大きいというのは、記事に含まれる感情が肯定的であるということである。全ての記事がポジティブな内容であることが記事のタイトルから推測できる。

a_id	title	polarity	pos	neg
6	Humble in the jungle: Meeting the mountain gorillas of war-torn Congo	1.0000000000	0.1250000000	0.0530000000
23	Cheap Flights From Sweden to Uganda or Rwanda From only €348!	1.0000000000	0.2150000000	0.0530000000
24	Cheap Turkish Airlines Flights From Oslo to several African destinations from only €361!	1.0000000000	0.1700000000	0.0530000000
27	How This Coffee Entrepreneur Beat Incredible Odds to Give Back to War-Torn Rwanda	1.0000000000	0.1520000000	0.0530000000
31	The best places for hiking and hiking in Rwanda	1.0000000000	0.0520000000	0.0530000000
32	The 50 greatest wildlife holidays on Earth	1.0000000000	0.0530000000	0.0530000000
43	How to see Rwanda's thriving mountain gorillas	1.0000000000	0.1520000000	0.0530000000
56	Evaluating recovery potential of the northern white rhinos from cryopreserved somatic cells [RESEARCH]	1.0000000000	0.0510000000	0.0530000000
61	Destination Flyfreak: Uganda	1.0000000000	0.1520000000	0.0530000000
63	Cheap Flights From France to Uganda or Rwanda From only €344!	1.0000000000	0.1040000000	0.0530000000
65	How Rwanda Made an Astonishing Recovery From Genocide	1.0000000000	0.0770000000	0.0530000000
94	Competition: win the trip of a lifetime	1.0000000000	0.2520000000	0.0530000000
105	If the Boston Marathon Is too Easy for You, Try These Insane Races	1.0000000000	0.0750000000	0.0530000000
109	America's Scenic Scenic Experience On Traveling World Helping Exotic Animals	1.0000000000	0.2120000000	0.0530000000
110	Top 5 Tourism activities for Your Uganda Safari Vacation	1.0000000000	0.0070000000	0.0530000000
113	10 Hotels to Go Off The Map You Need a Private Plane Or Helicopter Just To Get There	1.0000000000	0.1500000000	0.0530000000
117	Cheap Flights From Sweden to Uganda or Rwanda from only €348!	1.0000000000	0.1510000000	0.0530000000
118	Obsessed with monkeys? Go here.	1.0000000000	0.1130000000	0.0530000000
132	How This Coffee Entrepreneur Beat Incredible Odds to Give Back to War-Torn Rwanda	1.0000000000	0.1230000000	0.0530000000

以下に示すのが、Polarity の値が最も小さい記事 20 件である。Polarity の値が小さいというのは、記事に含まれる感情が否定的であるということである。多くの記事がネガティブな内容についての記事であることがわかる。人質が保護された、などポジティブな内容の記事であっても、その根底にある紛争などの社会問題が分析に影響していると考えられる。

a_id	title	polarity	pos	neg
0	"We know we may be killed": the rangers risking their lives for Virunga's gorillas	-1.0000000000	0.0650000000	0.1220000000
2	These Congo park rangers are risking their lives to protect endangered gorillas	-1.0000000000	0.0420000000	0.2240000000
5	Mountain gorilla population rises above 1,000	-1.0000000000	0.0620000000	0.1600000000
34	Two Britons who were kidnapped and held hostage in the Congo have been released	-1.0000000000	0.0460000000	0.1060000000
55	Female park ranger who tried to save two British tourists who were kidnapped and shot dead in Congo	-1.0000000000	0.0450000000	0.1160000000
68	500 Years Is Long Enough! Human Depravity in the Congo	-1.0000000000	0.0860000000	0.1560000000
67	"We know we may be killed": the rangers risking their lives for Virunga's gorillas	-1.0000000000	0.0660000000	0.1220000000
80	Mountain gorilla population rises above 1,000	-1.0000000000	0.0620000000	0.1600000000
169	Okhona, 65, fatally stung by wasps while tracking gorillas in Uganda with her husband	-1.0000000000	0.0400000000	0.1520000000
173	Atacama mummy mystery solved: Alien-like skeleton was a human	-1.0000000000	0.0450000000	0.1160000000
211	Dian Fossey (1932)	-1.0000000000	0.0710000000	0.0970000000
263	Kigali Genocide Memorial: How Rwanda overcame its violent past	-1.0000000000	0.0750000000	0.1000000000
158	500 Years Is Long Enough! Human Depravity in the Congo	-1.0000000000	0.0620000000	0.1260000000
253	"We know we may be killed": the rangers risking their lives for Virunga's gorillas	-1.0000000000	0.0660000000	0.1220000000
269	Mountain gorilla population rises above 1,000	-1.0000000000	0.0620000000	0.1600000000
359	The Tragic Exploits Of Dian Fossey And Her "Gorillas In The Mist"	-1.0000000000	0.0400000000	0.1020000000
489	Atacama mummy mystery solved: Alien-like skeleton was a human	-1.0000000000	0.0450000000	0.0770000000
173	Okhona, 65, fatally stung by wasps while tracking gorillas in Uganda with her husband	-1.0000000000	0.0400000000	0.1520000000
454	Dian Fossey (1932)	-1.0000000000	0.0710000000	0.0970000000
500	Kigali Genocide Memorial: How Rwanda overcame its violent past	-1.0000000000	0.0750000000	0.1000000000

このように、ニュース記事の感情分析では、ニュース記事に含まれる感情が精度良く推定できる。

4.2.4. 各コンテキストと各ニュース記事間の複相関量計算結果

コンテキストとニュース記事間のサイン類似度と感情分析を掛け合わせた、意味的距離複合計算結果の例を示す。以下の表では、コンテキスト「生息数の監視」と相関量が最も大きい記事 20 本を表示している。列左端からコンテキストの ID、ニュース記事の ID、ニュース記事のタイトル、そして複相関量の値である。

c_id	a_id	title	cs_vder_p
6	79	11 Primal Facts About Dian Fossey	0.7827976228
6	259	11 Primal Facts About Dian Fossey	0.7827976228
6	151	Good news for mountain gorillas as researchers reveal there are now over 1,000 in the wild	0.6641897208
6	378	Good news for mountain gorillas as researchers reveal there are now over 1,000 in the wild	0.6641897208
6	266	Understanding group relationships in gorillas	0.6391385039
6	178	Rwanda-Kagame Commends Conservationists For Gorilla Population Growth	0.6524672954
6	481	Rwanda-Kagame Commends Conservationists For Gorilla Population Growth	0.6524672954
6	450	To Spot Gorillas on a Budget, Try This Hike and Hike	0.6355769893
6	257	Ellen DeGeneres Tears Up After Receiving the BEST Birthday Present From Portia de Rossi - Plus, More AMAZING Celeb Gifts!	0.6381451483
6	120	Number of wild mountain gorillas exceeds 1,000	0.6498179881
6	341	Number of wild mountain gorillas exceeds 1,000	0.6498179881
6	263	Jennifer Lopez's New Song 'u' - REACTING LIVE!	0.6493324751
6	264	The Cost of Celebrity Big Brother America Is So BAD That...	0.6493324751
6	255	Listen To This: Post-Post-Traumatic!	0.6493324751
6	266	Iggy Azalea's New Song, 'Savior' - REACTING LIVE!	0.6493324751
6	172	Number of wild mountain gorillas exceeds 1,000!	0.6468745818
6	256	Number of wild mountain gorillas exceeds 1,000	0.6468745818
6	15	Hope for mountain gorillas: new census results reveal the population is increasing	0.6448538594
6	32	Hope for mountain gorillas: new census results reveal the population is increasing	0.6448538594
6	289	Hope for mountain gorillas: new census results reveal the population is increasing	0.6448538594

コンテキスト「生息数の監視」と相関量が最も小さい記事 20 本は以下の通りである。

c_id	a_id	title	cs_vder_p
6	211	Dian Fossey (1932)	0.3136095720
6	454	Dian Fossey (1932)	0.3136095720
6	359	The Tragic Exploits Of Dian Fossey And Her "Gorillas In The Mist"	0.3151152394
6	124	Why is Dian Fossey a key icon for gorilla tourism and conservation?	0.3362994041
6	0	"We know we may be killed": the rangers risking their lives for Virunga's gorillas	0.3362994041
6	253	"We know we may be killed": the rangers risking their lives for Virunga's gorillas	0.3362994041
6	327	DR Congo: Why do tourists go to Virunga National Park?	0.3381451483
6	155	Violence shuts Africa's Virunga gorilla park till 2019	0.3402768793
6	379	Violence shuts Africa's Virunga gorilla park till 2019	0.3402768793
6	281	Violence shuts Africa's Virunga gorilla park till 2019	0.3581781897
6	461	Violence shuts Africa's Virunga gorilla park till 2019	0.3581781897
6	42	First picture of female park ranger, 25, killed trying to save two British tourists from being kidnapped in Congo national park	0.3517871897
6	158	6 park rangers killed in DR's Virunga National Park	0.3535043938
6	464	6 park rangers killed in DR's Virunga National Park	0.3535043938
6	78	How East Africa is winning the battle to save its mountain gorillas - and where to see the great apes	0.3557176668
6	17	How East Africa is winning the battle to save its mountain gorillas - and where to see the great apes	0.3557176668
6	261	How East Africa is winning the battle to save its mountain gorillas - and where to see the great apes	0.3557176668
6	37	Virunga park rangers in DR killed in ambush near border checkpoint	0.3598484885
6	156	Virunga park rangers in DR killed in ambush near border checkpoint	0.3598484885

4.2.2 と同様に、コンテキストと直接的に関係する記事を色付けしている。4.2.2. と比較すると、関係の深い

記事がより多く上位に入っているのが分かる。複合相関量が小さい記事群には意味的に近い記事はなかった。以上より、コサイン類似度と感情分析を組み合わせた複合相関量計算によって、コンテキストとニュース記事の意味的距離計量の精度が高まることがわかる。

4.3. 可視化

ここまでの計算結果は、5D World Map とコンテキスト指向図で可視化される。

4.3.1. 5D World Map

5D World Map System 上での時空間画像と情報の可視化を図2から図5に示す。

ニュース記事の、画像、位置情報、記事内容が 5D World Map に表示されている。図3は、コンテキスト「生息地の消滅」と深い関連がある事象はコンゴ民主共和国、ルワンダ、ウガンダ3国で起きていることを表している。図4からは、コンテキスト「病気」と深い関連がある事象はコンゴ民主共和国のみで起きていることがわかる。

5D World Map は、統計情報を表現することもできる。図5では 1972 年から 2010 年までのコンゴ民主共和国にあるヴィルンガ国立公園でのマウンテンゴリラの頭数時系列変化を示している[17]。図5中のカラーマーカーの明暗度は、マウンテンゴリラの頭数の大小を表現していて、頭数が多いほどマーカーの色は明るく表示される。

これらの結果は、5D World Map System のヴィジュアルデータマイニングのツールとしての、時空間マッピングと可視化機能の有効性を示している。

図2から図5より、5D World Map System は、時空間的そして意味的コンテキストを組み合わせたり切り替えたりなどして、自然事象と社会事象の時間的推移と空間的移動を俯瞰的に捉えることを可能にすることがわかる。



図 2: 5D World Map System での情報可視化 (記事の画像と内容)



図 3: 5D World Map System での情報可視化 (コンテキスト=「生息地の消滅」)



図 4: 5D World Map System での情報可視化 (コンテキスト=「病気」)

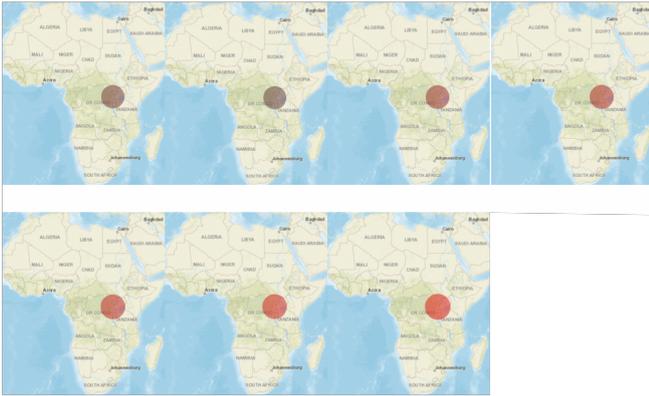


図 5: 5D World Map System での情報可視化 (ヴィルンガ国立公園のマウンテンゴリラの頭数時系列変化、図左端から 1972, 1977, 1981, 1986, 1989, 2000, 2003, 2010 年の頭数データ)

4.3.2. コンテキスト指向図

コンテキスト指向図は、図 6 から図 9 に示されている。絶滅危惧種、この実験例ではマウンテンゴリラが図の中心に位置している。8つのコンテキストが絶滅危惧種の周りに円状に配置されている。絶滅危惧種の点とコンテキストの点をつなぐ線の色は、コンテキストがポジティブなもの(赤)かネガティブなもの(青)かを表している。絶滅危惧種の生息国のニュース記事はコンテキストの点8つの周りに一回り大きく円状に並べられている。ニュース記事の点は書かれた国によって色分けされている。今回の例では、コンゴ民主共和国が緑、ルワンダがオレンジ、ウガンダが水色で表示されている。コンテキストの点とニュース記事の点の間の線の太さは、それぞれの間の複合相関量計算結果の値の大きさと比例する。つまり、相関量が高ければ、線は太くなる。絶滅危惧種の点とコンテキストの点の間の線の太さは、各コンテキストとニュース記事間の関連度の合計の大きさと比例する。

図 6 はマウンテンゴリラ生息国 3 国すべてのニュース記事、図 7 はコンゴ民主共和国のニュース記事のみ、図 8 はルワンダのニュース記事のみ、図 9 はウガンダのニュース記事のみが含まれたコンテキスト指向図である。

図 6 からは、図全体的に赤の色味が強いことから、ポジティブなコンテキストの方がネガティブなものよりマウンテンゴリラの生息との相関が高い、つまりマウンテンゴリラの全体的な状況は良い、と解釈できる。また、コンゴ民主共和国が他の生息国 2 国と比べて極端に記事数が少ないことから、コンゴ民主共和国は他国に比べてマウンテンゴリラへの関心が低いことが推測できる。図 7、図 8、図 9 からは、共通してコンテキスト「観光の発展」(context id: 5) が最も存在が大きい

いが、それ以外のコンテキストの傾向はそれぞれの国によって異なる。どの生息国においても、マウンテンゴリラの生息と観光は大きく相関しているが、マウンテンゴリラの生息に関する課題や強みは国によって異なることが推察できる。

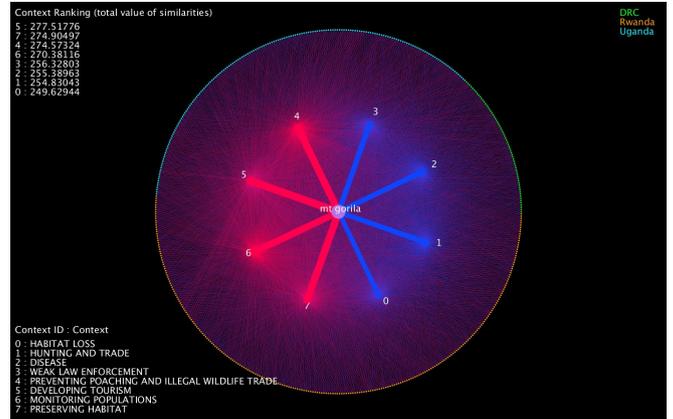


図 6: コンテキスト指向図 (コンゴ民主共和国、ルワンダ、ウガンダのニュース記事を含む)

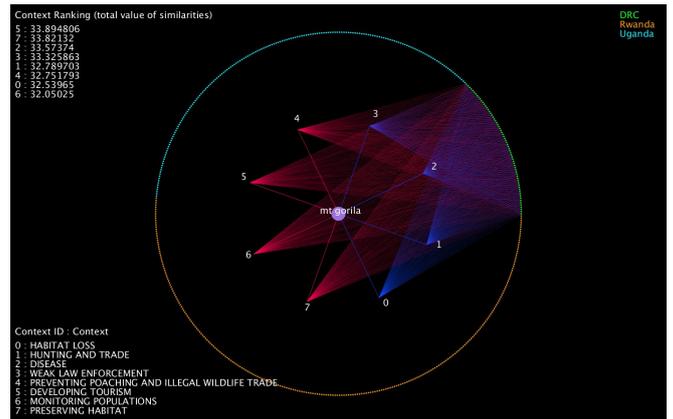


図 7: コンテキスト指向図 (コンゴ民主共和国のニュース記事のみを含む)

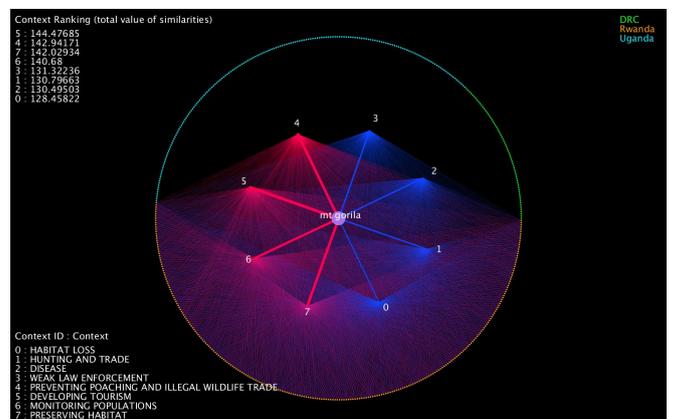


図 8: コンテキスト指向図 (ルワンダのニュース記事のみを含む)

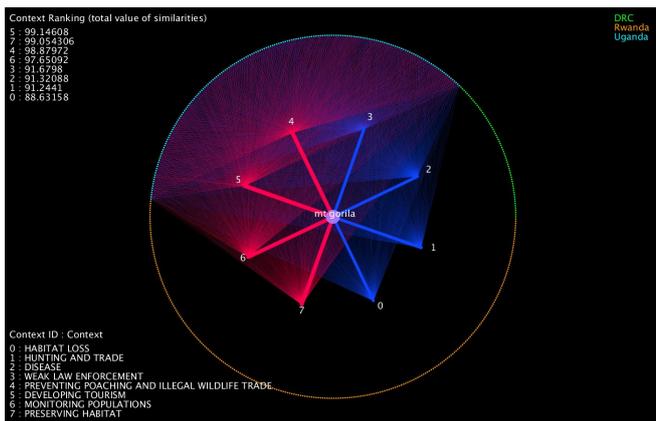


図 9: コンテキスト指向図 (ウガンダのニュース記事のみを含む)

5. 結論

本論文では、我々の提案するシステムの実現可能性を検証した。我々のシステムは、環境問題とニュース記事として表出する日常の事象との意味的関係のデータを蓄積、分析、可視化する新しい知識獲得システムである。その実現可能性は、絶滅危惧種にマウンテンゴリラを設定し、日常の事象としてマウンテンゴリラの生息国3国でのニュース記事を設定して検証した。コンテキストをベースにニュース記事を分析し、その計算結果は 5D World Map とコンテキスト指向図として可視化した。ニュース記事の画像、地理情報、内容の可視化は 5D World Map で実現できた。コンテキスト間の比較や数値的關係の可視化はコンテキスト指向図で実現した。以上より、このシステムの実現可能性と発展可能性を示した。

6. 今後の展望

今回は、コンテキストの文書とニュース記事の文書間の意味的距離計量の精度を上げるために、コサイン類似度と感情分析の結果の数値を組み合わせる計算し、実験を行なった。今後も、計算精度の更なる向上のために様々な計算方法を試していく。また、より感覚的に情報を伝えられるよう、可視化の強化も行う。さらに、広く情報発信できるよう、ウェブアプリケーション化も目指す。

参 考 文 献

- [1] United Nations. Website: <https://sustainabledevelopment.un.org/sdgs>
- [2] Kitagawa, T. and Kiyoki, Y., "The mathematical model of meaning and its application to multidatabase systems," Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering : Interoperability in Multidatabase Systems, pp.130-135, April 1993.
- [3] Kiyoki, Y., Kitagawa, T., and Hayama, T. "A metadatabase system for semantic image search by a mathematical model of meaning," ACM SIGMOD Record, Vol. 23, No. 4, pp.34-41, 1994.
- [4] Kiyoki, Y., Chen, X., Sasaki, S., and Koopipat, C., "Multi-Dimensional Semantic Computing with Spatial-Temporal and Semantic Axes for Multi-spectrum Images in Environment Analysis," Information Modelling and Knowledge Bases, Vol. XXVIII, IOS Press, pp.14-30, May 2016.
- [5] Sasaki, S., Takahashi, Y., and Kiyoki, Y., "The 4D World Map System with Semantic and Spatio-temporal Analyzers," Information Modelling and Knowledge Bases, Vol. XXI, IOS Press, pp.1-18, May 2010.
- [6] Sasaki, S. and Kiyoki, Y., "Real-time Sensing, Processing and Actuation Functions of 5D World Map System: A Collaborative Knowledge Sharing System for Environmental Analysis," Information Modelling and Knowledge Bases, Vol. XXVIII, IOS Press, pp.220-239, May 2016.
- [7] Mori, M., Sasaki, S., and Kiyoki, Y., "A Bird-World Image-Map System with Spatiotemporal and Color Analysis Functions." 10th International Conference on IEEE ICT and Knowledge Engineering (ICT-KE 2012), Siam University, Bangkok, Thailand, pp.104 - 111, 21 Nov - 23 Nov 2012.
- [8] Sasaki, S., Kiyoki, Y., Fujioka, H., Watanabe, T., Otsuka, K., Ishii, M., "Application of 5D World Map System to Large New-Article Database for Realizing Context-Diversity-Responsive Semantic Associative Search," Information Modelling and Knowledge Bases, Vol. XXX, IOS Press, accepted 20 pages, May 2019 (In press).
- [9] World Wildlife Fund. Website: <https://www.worldwildlife.org/species>
- [10] World Wildlife Fund. Website: <https://www.worldwildlife.org/about>
- [11] Google News. Website: <https://news.google.com/?hl=en-US&gl=US&ceid=US:en>
- [12] Scikit-learn. Website: http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html
- [13] Scikit-learn. Website: http://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.cosine_similarity.html
- [14] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.
- [15] World Wildlife Fund. Website: http://wwf.panda.org/knowledge_hub/endangered_species/great_apes/gorillas/
- [16] IUCN Red List. Website: <http://www.iucnredlist.org/>
- [17] Gray, M., Roy, J., Vigilant, L., Fawcett, K., Basabose, A., et al., "Genetic census reveals increased but uneven growth of a critically endangered mountain gorilla population," Biological Conservation, Vol. 158, pp.230-238, 2013.