九州大学学術情報リポジトリ
Kyushu University Institutional Repository

[2014]九州大学情報統括本部年報 : 2014年度

https://doi.org/10.15017/1560528

出版情報：九州大学情報統括本部年報. 2014, pp.1-, 2015. 九州大学情報統括本部
バージョン：
権利関係：

第1章　学術情報研究部門

1.1　スタッフ一覧

職名	氏名	研究キーワード
教授	廣川 佐千男	検索エンジン, Web マイニング, 推論システム, 証明論
准教授	鈴木 孝彦	演繹データベース, 論理プログラミング, 機能論理プログラミング, CAD データベース, 認証, PKI, 属性認証
准教授	伊東 栄典	Web マイニング, Web サービス, 情報検索, 情報統合, XML, 分散システム , ネットワーク , 協調システム , ソフトウェア工学
助教	中藤 哲也	Web データーベース, Web データマイニング, マッシュアップ, 観光情報, 可視化, 検索エンジン

1.2　研究事例紹介

ベクトル類似性を利用した日本語 WordNet の誤り検出
学術情報研究部門　鈴木孝彦

1-1.　はじめに

日本語 WordNet [1] は, 人間にとってもコンピュータインターフェースからも使いやすい概念辞書である．しかしながら，現バージョンの日本語 WordNet には 5%の誤りが含まれている [2]．我々は，日本語 WordNet における誤りの自動的検手法を研究している．これまで，WordNet の構造を利用した誤り検出法が有効であることを確かめた [3]．2014 年度は，語のベクトル化を利用した誤り検出に取り組んだ．以下にその研究状況を紹介する．

1-2.　バックグラウンド

1-2-1.　日本語 WordNet とその誤り

日本語 WordNet は Princeton WordNet を基にした，日本語の概念辞書である．単語は，その単語が属する概念(Synset)の元に同義語(Synonym)としてまとめられている．また，概念同士は，上位概念，下位概念，反対概念などのリンクで相互参照されている．現行の日本語 WordNet(Ver. 1.1)は，57,238 概念（Synset 数），93,834 語（日本語）を網羅している [2]．　日本語 WordNet には，5% 程度の誤りが含まれている [2]．各種の誤りの中で，我々は，同義語の誤り，すなわち，「Synset の同義語中に意味の異なる語が含まれる」誤りに着目した．

1-2-2.　Word2Vec による語のベクトル化

Word2Vec [4,5]は, 大規模コーパスを利用して, 語をベクトル化(10^2～10^4　次元)する手法である．Word2vec でベクトル化された語については，さまざまなベクトル演算が意味を持つことが知られている．例えば，ベクトル化された語の対についてコサイン類似度を計算した場合，コサイン類似度の大きい語の対の意味は類似している．すなわち，コサイン類似度の高い語同士は，同義語である可能性が高い(**傾向 A**).

1-3.　実験

1-3-1.　コサイン類似度を利用した誤り検出の試み

我々は，日本語 WordNet 上の 7000 個の Synset について，あらかじめ人間の手でチェックし，161 個の同

義語の誤りリストを作成した. 次に, 7000 個の Synset それぞれに含まれる単語同士のコサイン類似度を計算し, Synset 内の最小コサイン類似度(minimum cosine similarity in Synset S: MCS(S))順に並べた.

図 1 は，横軸に MCS(S)順位，縦軸に，その順位までに含まれる同義語誤りの総数をプロットしたグラフである．なお，対象としたコーパス（Wikipedia 日本語記事全文）[7] に出現しない語を含む Synset については評価から除外した.

MCS(S)順位と，その順位までの誤り総数は比例している．これは，コサイン類似度の大小と同義語誤りに関連が無いことを意味する．すなわち，日本語 WordNet と word2vec のベクトル化においては，**(傾向 A)** の逆および裏,「同義語であるならばコサイン類似度が高い」と「コサイン類似度が低いならば同義語ではない」は成立しない.

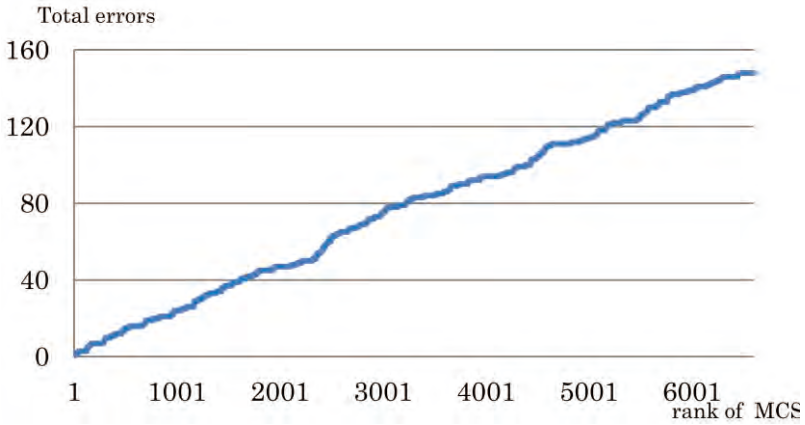

図 1 :最小コサイン類似度順位と同義語誤り数

1-3-2.　ベクトル類似性の有用性

コサイン類似度が有用でない原因は，いくつか考えられるが，我々は次の 2 つの仮説を立てた.
仮説 1 ：word2vec ベクトルの中に，同義語判定のために有用なベクトル要素が存在する.
仮説 2 ：有用でないベクトル要素はランダムな値を取り，コサイン類似度の計算に影響を及ぼす.
この 2 つの仮説が成り立てば，広い意味でのベクトル類似性を同義語判定に用いることができる.

1-3-3.　仮説 1 の検証:決定木の構築

まず，ある Synset について，その Synset の関連語集合（ Synset の下位語集合 ∪ Synset の下位語の下位語の集合) を作った. 次に, 関連語集合と, 関係の無い語の集合を判別するため, CART アルゴリズム[8]を使って決定木[8]を求めた.

図 2 は, Synset:07891726-n の関連語集合={赤ワイン,モーゼル，シャンパン,シャンペン, ボルドー,ベルモット,ポートワイン,シェリー, マスカット ,ビショップ}10 語と，無関係な 10 語の集合を識別するための決定木である．2 つのベクトル要素 X[706]と X[104]によって，完全な識別が可能である.

他の Synset についても，少数のベクトル要素を用いて, 精度の良い関連語の識別が可能であった．決定木に使われたベクトル要素は, 同義語判定のためにも有用と推測できる. なお, 決定木構築において, 同義語ではなく関連語集合を用いた理由は, 一般に同義語の語数が少なく, 少数の同義語だけでは決定木に基づく検証が困難なためである.

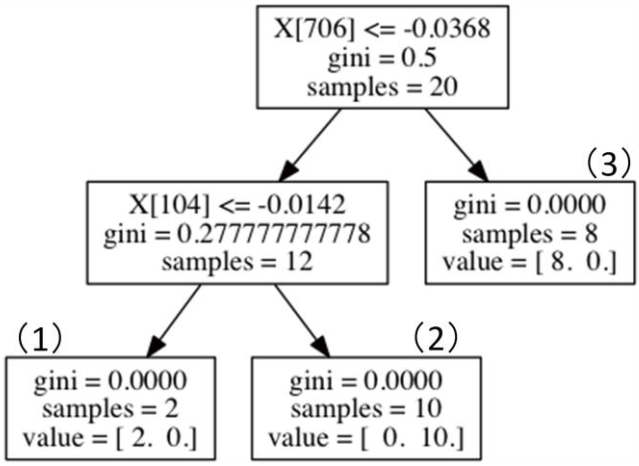

図 2 ：07891726-n [概念：葡萄酒，ワイン]の関連語(10 語)を識別するための決定木

1-3-4.　仮説 2 の検証: ベクトル要素の変動係数(C.V.)頻度分布

仮説 1 の検証と同様に，ある Synset の関連語集合を作り，その集合における各ベクトル要素の変動係数を求めた．図 3 は，Synset 00021265-n （概念：養分)の関連語 108 語と，無関係な 108 語の集合について，ベク

トル各要素の変動係数(coefficient of variation: C.V.)の絶対値のヒストグラムを取ったものである(ベクトルは 800 次元). 関連語集合において|C.V.|≧3 以上を取るベクトル要素が 全体の 58%を占め，これらの要素はコサイン類似度の計算に強く影響する.

同義語を 7 個持つ Synset 05562038-n に対し、同義語のベクトルについてヒストグラムを取った(図 4)ところ，|C.V.| ≧3 の要素が 42% に登った．実際に Synset 05562038-n 内の同義語についてコサイン類似度を計算したところ、4 組について負の値を取り、コサイン類似度による同義語の判定に問題があることがはっきりした.

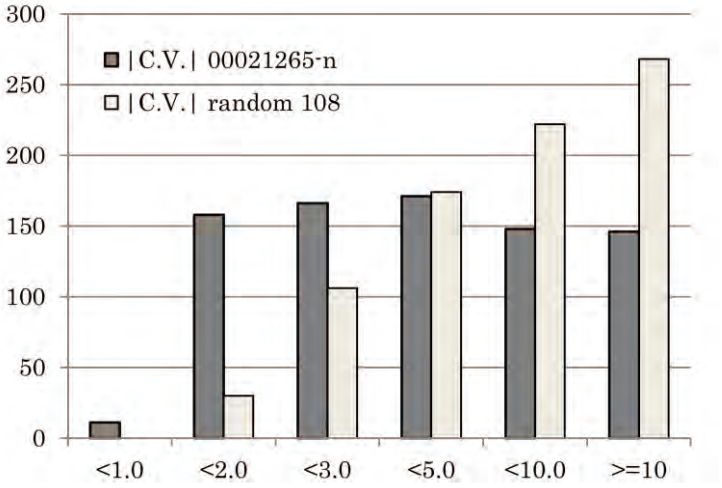

図 3：Synset　00021265-n 関連語　108 語における
ベクトル各要素の|C.V.|ヒストグラム
(ランダムな 108 語と比較)

1-4.　まとめ

ベクトル化した語のベクトル類似性を用いて日本語 WordNet の同義語誤りを検出する研究の現状を紹介した．現在までに,

・コサイン類似度はそのままでは同義語の誤り判定に利用できない.

・意味を表す有用なベクトル要素が存在する.

・同義語であっても，ランダムな値を取るベクトル要素があり，コサイン類似度の値に強く影響する.

との知見を得た．これらの性質を利用して，同義語誤りの検出を行うことが次の課題である.

上記の事実は，日本語 WordNet の誤り検出だけでなく，新しい同義語の発見や，感情分析における特徴語の発見のために有用である.

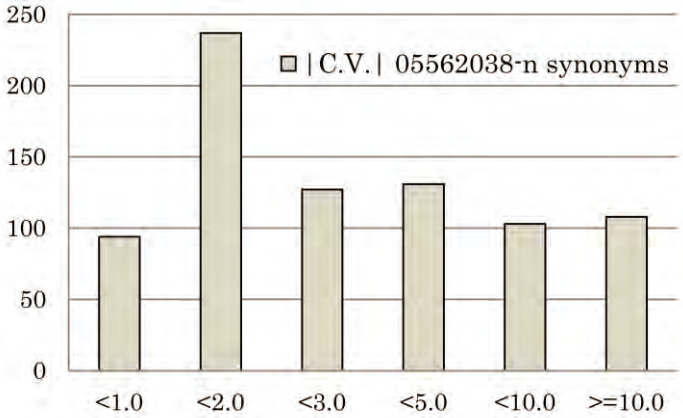

図 3：Synset　05562038-n 同義語 7 語における
ベクトル各要素の|C.V.|ヒストグラム

この研究および関連研究で得た同義語誤りのリストは，日本語 WordNet の管理者の一人である Francis Bond [1]に提供されている.

参考文献

[1] F. Bond, H. Isahara, S. Fujita, K. Uchimoto,T. Kuribayashi, Enhancing the Japanese WordNet, *ALR7 Proc. the 7th Workshop on Asian Language Resources*, pp. 1-8 ,*Association for Computational Linguistics*. pp. 1-8, 2009

[2] NICT Information Analysis Laboratory, National Institute of Information and Communications Technology, *Japanese WordNet*, http://nlpwww.nict.go.jp/wn-ja/index.en.html

[3] T. Hirao, T.Suzuki, K.Miyata, S.Hirokawa, Detection Methods for Misplacement of Synonyms in the Japanese WordNet, International Journal of Computer & Information Science , vol. 15, no.2, pp.26-35, 2014

[4] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient Estimation of Word Representations in Vector Space. In Proceedings of Workshop at ICLR, 2013.

[5] T. Mikolov et al. ,Distributed representations ofwords and phrases and their compositionality Proc. 27th Annual Conference on Neural Information Processing Systems, 2013

[6] 平尾，鈴木，宮田，廣川，日本語 WordNet 類義語の誤り検出　ーコーパス利用の試みー，信学技報，vol. 114, no. 339, AI2014-18, pp. 13-18, 2014

[7] http://ja.wikipedia.org/wiki/

[8] L. Breiman, J. Friedman, C. J. Stone, R.A. Olshen, Classification and Regression Trees, Wadsworth & Brooks, 198

1.3　研究内容紹介

1.3.1　廣川 佐千男

研究内容

- **WWW 空間の解析, WWW からの知識発見と情報統合**

  インターネット上のホームページ群は世界最大の知識の書物といえます. 人類はかってこれだけの知識の素を共有したことはありません. しかも我々はそこから膨大な量の情報を瞬時に集めることができます. インターネットのページには, 作った人や組織の意図に従って, 特定のパターンに従って構成された役に立つリストや, 他のページへの明示的な関連づけ（リンク）があります. WWW におけるこのリンクによる関連はどのようになっているのでしょうか？増え続ける WWW 空間からどのようにすれば効率良く知識を獲得できるのでしょうか？沢山ある情報をどのように組み合わせればより価値ある情報を作れるでしょうか？私は, ホームページを点, ホームページ間のリンクを枝とするグラフを可視化し, WWW 空間を解析する研究を行なっています. また、ホームページを表現する HTML ファイルに現れるパターンに着目し, 情報の抽出と統合を行なっています.

- **計算論理学**

  インターネットの他に, 計算論理学の研究も行なっています. 人間の論理的思考を証明図という図形で表現し, 各種論理体系における推論の構造を分析します.

所属学会名

人工知能学会, 情報処理学会, ソフトウェア科学会, 電子情報通信学会,
Association for Symbolic Logic. , 日本数学会

主な研究テーマ

- データマップ法と概念グラフによる次世代検索エンジンの研究開発
  キーワード：検索エンジン, マトリックス検索, 概念グラフ, 大学発ベンチャー, 2006.08～.
- リンク情報と Web データの半構造性を融合した高品質コンテンツ・マイニング
  キーワード：Web マイニング, データマイニング, WWW, XML, 文書検索, 検索エンジン, 特許検索, データベース, 2001.04～.
- 専門検索サイトの動的統合による次世代検索システムの研究開発
  キーワード：検索エンジン, メタサーチ, 1993.04～.
- 構成的論理の推論構造の研究
  キーワード：証明論, 型理論, ラムダ計算, 線形論理, 適切さの論理, 直観主義論理, 1997.01～.

研究業績

- 原著論文

  1. Sachio Hirokawa, Brendan Franagan, CHENGJIU YIN, Hiroto Nakae,Vizualization of Relation and Generality of Words in Search Result,Proceedings of the Third Asian Conference on Information Systems,90-95,2014.12.

2. Toshiro Minami, Kensuke Baba, Sachio Hirokawa,Should University Library Collect New Books or Old Books?,Proc. ISAAC2014 & ICACT2014, AACL03,34-37,2014.12.
3. Toshiro Minami, Kensuke Baba, Sachio Hirokawa, Eriko Amano,A Trichotomic Approach to Approximate Representation of Concepts, Applied Computing and Information Technology,Studies in Computational Intelligence,61-75,2014.12.
4. Yuichi Ono, Manabu Ishihara, Sachio Hirokawa,Real-time Feedback Systems in a Foreign Language Teaching: A Case of Presentation Course,Proc. ICCE2014,779-784,2014.11.
5. Brendan Franagan, CHENGJIU YIN, Sachio Hirokawa,Learning by Search & Log,Proc. ICCE2014,391-393,2014.11.
6. Yuichi Ono, Manabu Ishihara, Sachio Hirokawa, Mitsuo Mamashiro,Real time text-based feedback systems – From frequency-based feedback to mindmap feedback in foreign language teaching,Proceedings of IEEE International Conference on Systems, Man and Cybernetics,2150-2153,2014.10.
7. Shaymaa Sorour, Kazumasa Goda, Sachio Hirokawa, Tsunenori Mine,Predicting Students' grades based on free style Comments Data by Artificial Neural Network,Proc. FIE2014,2014.10.
8. Takanori Yamashita, Yoshifumi Wakata, Satoshi Hamai, Yasuharu Nakashima, Yukihide Iwamoto, Brendan Flanagan, Naoki Nakashim, Sachio Hirokawa,Extraction of Key Factors from Operation Records by Support Vector Machine and Feature Selection,8,70-71,2014.10.
9. Yuichi Ono, Sachio Hirokawa, Manabu Ishihara, Mitsuo Yamashiro,Implementation and evaluation of real time qualitative feedback systems in a foreign language presentation course,Proc. AAI2014,372-376,2014.08.
10. Sachio Hirokawa, Tetsuya Nakatoh, Hiroto Nakae, Takahiko Suzuki,Discovery of Implicit Feature Words of Place Name,Proc. AAI2014,561-566,2014.08.
11. Takuya Hirao, Koki Miyata, Takahiko Suzuki, Sachio Hirokawa,Detection of Misplacement of Synonyms in the Japanese WordNet,Proc. AAI2014,31-36,2014.08.
12. Takanori Yamashita, Yoshifumi Wakata, Naoki Nakashim, Satoshi Hamai, Yasuharu Nakashima, Yukihide Iwamoto, Brendan Flanagan, Sachio Hirokawa,Construction of Dominant Factor Presumption Model for Postoperative Hospital Days from Operation,Proc. AAI2014,19-24,2014.08.
13. Shaymaa Sorour, Tsunenori Mine, Kazumasa Goda, Sachio Hirokawa,Comments data mining for evaluating student's performance,Proc. AAI2014,25-30,2014.08.
14. Brendan Franagan, CHENGJIU YIN, Takahiko Suzuki, Sachio Hirokawa,Classification and clustering English Writing Errors based on Native language,Proc. AAI2014,318-324,2014.08.
15. Jun Zeng, Min Gao, Junhao Wen, Sachio Hirokawa,A Hybrid Trust Degree Model in Social Network,Proc. AAI2014,37-41,2014.08.
16. Sachio Hirokawa, Emi Ishita,Non-Topical Classification of Healthcare Information on the Web,Frontiers in Artificial Intelligence and Applications,383-393,2014.07.
17. Sachio Hirokawa, Brendan Flanagan, Takahiko Suzuki, CHENGJIU YIN,Learning Winespeak from Mind Map of Wine Blogs,Proc. HIMI 2014, Part II, LNCS 8522,383-393,2014.07.
18. Takanori Yamashita, Yoshifumi Wakata, Naoki Nakashim, Sachio Hirokawa, Satoshi Hamai, Yasuharu Nakashima, Yukihide Iwamoto,Extraction of Determinants of Postoperative Length of Stay from Operation Records,Proceedings of 2014 IEEE Workshop on Electronics, Computer and Applications,8,822-827,2014.05.

19. Shaymaa Sorour, Tsunenori Mine, Kazumasa Goda, Sachio Hirokawa,Efficiency of LSA and K-means in perdicting studnt's academic performance based on comments data,Proc. CSEDU2014,43-52,2014.04.

受賞

1. 科学技術振興機構データサイエンス・アドベンチャー杯言語部門最優秀賞, 科学技術振興機構,2015.03.

研究資金

- 科学研究費補助金

  1. 2012 年度～2015 年度, 基盤研究 (C), 代表, 手掛語と内容語の双対ブートストラップ・マイニング.

教育活動

- 教育活動概要

  大学院システム情報科学府情報知能工学専攻専攻の協力講座として広域分散データ特論・演習を担当し, インターネット検索エンジンなどの基礎となる技術を教えています. 大学院統合新領域ライブラリーサイエンス専攻博士課程も担当しています.　また, 学部の講義では工学部電子情報工学科において, プログラミング論 I を担当しています. 同学科の学部学生についての卒業研究と, 大学院情報科学府の修士, 博士課程の指導も行っています.

- 担当授業科目

  1. 2014 年度・後期,PTL1.
  2. 2014 年度・後期, 分散システム特論.
  3. 2014 年度・前期, プログラミング論Ｉ.

大学運営

- 学内運営に関わる各種委員・役職等

  1. 2009.04～2016.03, 国際交流専門委員会

1.3.2　鈴木 孝彦

研究内容

- 化学プラントデータベース上での知識利用に関する研究
- 九州大学教務事務システムの作成と運用支援
- 九州大学学務情報システムの計画支援
- 日本語 WordNet およびテキストマイニング

所属学会名

情報処理学会

主な研究テーマ

- 日本語 WordNet における意味構造の研究
  キーワード：ワードネット，オントロジー，テキストマイニング，2012.05～2016.05.
- 混合道路交通における低速車両の移動ロケーションデータの収集と分析方法の研究
  キーワード：ロケーションデータ，Web マイニング，低速車両，　2012.04～2017.04.

研究業績

- 原著論文

  1. Takahiko Suzuki, Sachio Hirokawa, 平尾　拓也, 宮田　光樹,Detection Method for Misplacement of Synonyms in the Japanese WordNet,15,2,26-35,2014.12.
  2. Takahiko Suzuki, Sachio Hirokawa, 平尾　拓也, 宮田　光樹,Detection of Misplacement of Synonyms in the Japanese WordNet,Proc. of International Conference on Advanced Applied Informatics (AAI 2014), Kitakyushu, Japan, 2014.8.31-9.4. ,2014.09.
  3. Takahiko Suzuki, Sachio Hirokawa, Tetsuya Nakatoh, 中江　博人,Discovery of Implicit Feature Words of Place,Proceedings of International Conference on Advanced Applied Informatics (AAI 2014), Kitakyushu, Japan, 2014.8.31-9.4. ,2014.09.

- 学会発表

  1. 鈴木 孝彦, 廣川 佐千男, 平尾　拓也, 宮田　光樹, 日本語 WordNet 類義語の誤り検出　-コーパス利用の試み- , 電子情報通信学会,2014.11.27.
  2. 鈴木 孝彦, 宮田　光樹, 廣川 佐千男,WSD における単語の曖昧性と難易度との関係, 電子情報通信学会 ,2015.02.05.

教育活動

- 担当授業科目

  1. 2014 年度・後期, プログラミング言語特論.
  2. 2014 年度・後期, 情報処理概論.

社会貢献・国際連携等

1. 社会連携活動

   共同研究 General Product Model 上の論理検索（株）エーアイイー研究社との共同研究

大学運営

- 学内運営に関わる各種委員・役職等

  1. 2012.04～2015.03, 広報委員会委員.

1.3.3　伊東 栄典

研究内容

- **情報検索・情報統合・情報連携**

  Web 上や電子データとして蓄積されている膨大なデータから，意味のある知識抽出や，抽出した知識を統合する研究を行っている．具体的には，Web データからの情報抽出，利用者コメントからの知識発見，コメントやリンク構造を利用したコンテンツ推薦などを行っている．ソーシャルブックマークからの新規情報の発見や,「クチコミ」と呼ばれる情報サービスからの知識抽出について研究している．

- **集合知を利用した高品質コンテンツ検索**

  情報検索および知識発見に関する研究の具体的な対象として，ネット上のコンテンツを対象とした研究を行なっている。近年，ユーザ投稿型のコンテンツサービスが普及しており，動画・写真・静止画・小説などが増大している。これらのサービスでは，投稿機能だけでなく，視聴者からのコメント・タグ付け・リンクなどのフィードバック機能もある。視聴者からのフィードバックは，Folksonomy や集合知と呼ばれるもので，これを活用したコンテンツ検索を行なっている。

- **電子認証基盤および認証フェデレーション構築**

  特定メンバーへの情報提供サービスや，オンラインでの商取引および申請のように，利用者を正しく認証する電子認証が求められている．本研究では電子認証機構および認可機構の構築と，情報サービス連携のための認証連携機構の構築を目的としている．多様な組織が柔軟にサービスを提供しあうための，電子認証基盤の構築についての研究開発を行う．その実現のために，同じポリシー・標準規格での認証基盤となるフェデレーションについての研究開発を行う．電子認証証基盤として，認証のためのデータベース構築，アプリケーションとなる情報サービスでの認証・認可機構の実現などが課題となる．組織内で使う，柔軟な利用者認証・認可システムの実現を目指す.

所属学会名

電子情報通信学会, 人工知能学会

主な研究テーマ

- 利用者投稿型コンテンツの多様性動向分析
  キーワード：コンテンツ, 利用者投稿型, 多様性, エントロピー, コンテンツ選択モデル, 2014.04〜2018.03.

- 大規模データ解析
  キーワード：大規模データ, 分散処理, 統計解析, クラウド・コンピューティング, 2012.06〜

- 集合知に基づくコンテンツの高品質検索
  キーワード：集合知, 利用者投稿型, オンラインコンテンツ, タグ, コメント, 2011.04〜2015.03.

- 情報検索・情報統合・情報連携
  キーワード：Information retrieval / extraction / integration / discovery, Data mining, Web mining, 2000.04〜

研究プロジェクト

- 集合知に基づく高品質コンテンツ検索
  2011.04～2015.03, 代表者：伊東栄典, 日本学術振興会 科学研究費助成事業
  近年 YouTube 等の利用者投稿型コンテンツサービスが人気である. 動画以外でも，小説・漫画・画像の投稿型サービスが普及されつつある. これらのサービスでは，毎日多数のコンテンツが投稿・蓄積されており，その数は膨大である. 利用者が求めるコンテンツを探すためには検索システムが必要である. コンテンツを高い品質で検索するためには. コンテンツの品質（面白さ）評価尺度と，コンテンツのカテゴリ分類機能が必要である. 本研究では，閲覧者がコンテンツに付与するタグやコメントなどを集合知として用いる検索を考える.

研究業績

- 原著論文

  1. 川谷 卓也, 伊東 栄典, 笠原 義晃, 藤村 直美, 適切なクォータ値設定のためのディスク使用量分析, 情報処理学会,2015-IOT-28,32,1-7,2015.02.
  2. Yoshiaki Kasahara, Eisuke Ito, Naomi Fujimura,Introduction of New Kyushu University Primary Mail Service for Staff Members and Students,SIGUCCS ’14 (Proceedings of the 2014 ACM SIGUCCS Annual Conference on User Services Conference),103-106,2014.11.
  3. Zechen Wu, Eisuke Ito,Correlation Analysis between User’s Emotional Comments and Popularity Measures,Proceedings of IIAI-AAI 2014/ESKM 2014,280-283,2014.09.
  4. 藤村 直美, 笠原 義晃, 伊東 栄典, 尾花昌浩, 井上 仁, 学生番号と異なる学内情報サービス専用 ID 付与, 情報処理学会,2014-IOT-26,1,1-6,2014.06.
  5. Xiao Lin, Eisuke Ito, Sachio Hirokawa,Chinese Tag analysis for foreign movie contents,13th IEEE/ACIS International Conference on Computer and Information Science (ICIS2014),163-166,2014.06.

受賞

  1. 科学技術振興機構データサイエンス・アドベンチャー杯言語部門最優秀賞, 科学技術振興機構,2015.03.

教育活動

- 担当授業科目

  1. 2014 年度・後期, プログラミング演習 1.
  2. 2014 年度・前期, 暗号と情報セキュリティ特論.

大学運営

- 学内運営に関わる各種委員・役職等

  1. 2012.06～, 情報統括本部・キャンパスクラウド事業室 室長

2. 2012.06～, 情報統括本部 認証基盤事業室
3. 2011.04～, 情報統括本部・全学基本メール事業室

1.3.4　中藤 哲也

研究内容

研究活動：

- **近似文字列照合アルゴリズムとその応用**

  文章などの長い文字列中から, 与えた文字列を見つけ出す処理を文字列照合処理と言います. 与えた文字列に似た文字列を見つけ出す処理を特に近似文字列照合と呼び, 曖昧な検索や似た機能の遺伝子を探す処理など, 多くの応用があります. この近似文字列照合を高速に行うアルゴリズムの研究をしています. また, 近似文字列照合アルゴリズムを用いたパターン発見や盗作検出などの研究も行っております.

- **WWW 上からの情報抽出, 分析, 可視化**

  WWW 上の膨大な情報から必要とされる情報の抽出, 分析, 可視化に関する研究を行っております. 特に近似文字列照合アルゴリズムを用いたパターン発見による Web データ・マイニング, 視点や構造を与えることによる文書群の分析・可視化の研究を進めています.

- **観光情報処理**

  近年, 我が国において観光立国を推進する立場から, WWW 上の観光情報をもっと適切に利用しようと言う動きがあります. WWW 上からの情報抽出アルゴリズムを用いるだけではなく, 観光情報に特化した手法も含めて, 観光情報処理の研究を進めています.

業務：
学生情報処理に関するサポート, 職場の安全・衛生に関するサポートを行っております.

所属学会名

電気学会, 情報処理学会, 人工知能学会, 言語処理学会, 日本データベース学会, ACM SIGMOD 日本支部

主な研究テーマ

- 研究動向分析、研究調査手法に関する研究
  キーワード：研究動向分析、計量書誌学, 2014.04～.
- 近似文字列照合アルゴリズムを用いたウェブマイニング手法の研究
  キーワード：ウェブマイニング, パターン発見, 近似文字列照合, 2004.04～.
- Web データベースの機能化
  キーワード：Web データーベース，Web データマイニング，マッシュアップ, 2001.04～.
- 観光情報マイニング
  キーワード：観光情報, データマイニング, ウェブマイニング, 情報抽出, 可視化, 検索エンジン, 2010.04～.

研究業績

- 原著論文

  1. Sachio Hirokawa, Tetsuya Nakatoh, Hiroto Nakae, Takahiko Suzuki,Discovery of Implicit Feature Words of Place Name,Proc. of IIAI International Conference on Advanced Applied Informatics,2014.08.

- 学会発表

  1. 中藤 哲也, 廣川 佐千男,HPC によるテキストマイニング, ワークショップ「数値シミュレーションだけではないスーパーコンピュータ活用」,2015.01.14.
  2. 中藤 哲也, 山田 泰寛, 馬場 謙介, 池田 大輔, 廣川 佐千男, 近似文字列照合を用いた剽窃検出手法の評価, 平成 25 年度電気関係学会九州支部連合大会（第 66 回連合大会）,2014.09.18.

研究資金

- 科学研究費補助金

  1. 2012 年度〜2014 年度, 基盤研究 (C), 分担, 手掛語と内容語の双対ブートストラップ・マイニング.
  2. 2012 年度〜2014 年度, 基盤研究 (B), 分担, 大規模テキストデータ中の部分構造と稀少な複合パタンの階層的な発見.

教育活動

- 担当授業科目

  1. 2014 年度・前期, 情報処理概論.
  2. 2014 年度・後期, 分散システム特論.

大学運営

- 学内運営に関わる各種委員・役職等

  1. 2014.04〜, 情報統括本部 広報事業室 室長.
  2. 1994.05〜, 入学試験機械処理委員会委員.