

医学系研究課題の網羅的分析に向けての分類支援

No.1

木村泰知 (小樽商科大学), 渋谷英潔 (北海学園大学),

一瀬信敏 (札幌医科大学附属産学・地域連携センター, 文部科学省産学官連携コーディネーター)

背景

◆医学系研究

- 医学関係者以外に対しても理解しやすい表現が望まれている
- 従来
 - 個別案件の一覧表形式
 - 特徴を把握することが困難

◆自然言語処理技術の向上

- 分類技術
- 可視化技術

◆現状

- 人手による分類・可視化
- 対象とする病気, 人体の部位, 解決手法の観点から分類
- パブルチャート

◆問題点

- 人手作業にかかる労力が大きい
- 分類基準がぶれる



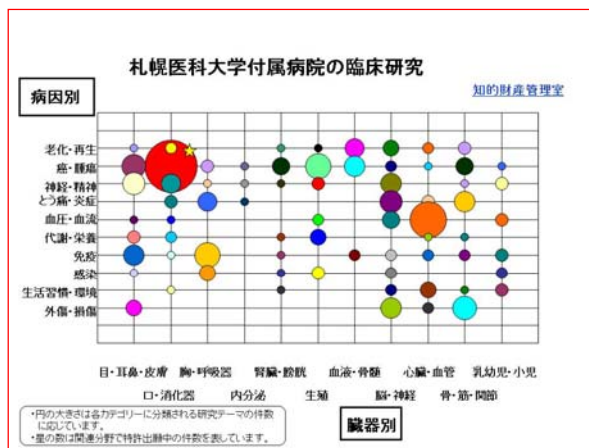
対象 課題 解決手法 の観点で分類

目的

- 研究活動の概要を客観的かつ定量的に提示
- 分類支援による人手作業の軽減

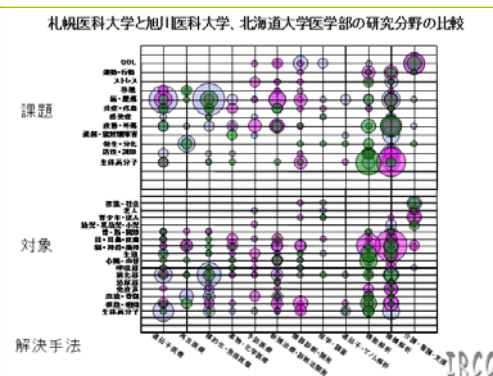
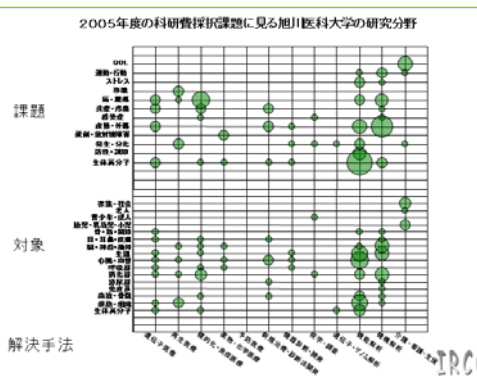
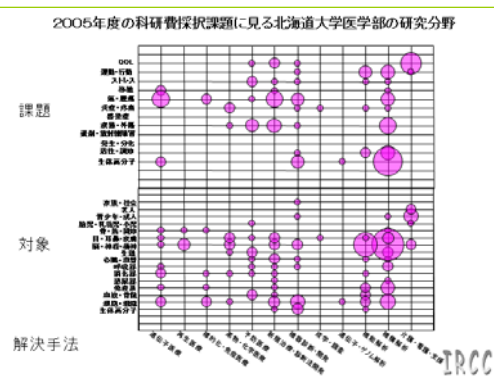
分類の仮定

- 文中単語には推測に不要な単語が存在する。
単語毎の重みを仮定し学習する
- 同じ単語でも出現位置によりカテゴリーが異なる。
文構造の重みを仮定し学習する



可視化

題目	対象	課題	解決手法
0-15標識水による冠動脈硬化病変の早期診断と治療効果判定に関する研究	心臓・血管	疾患・外傷	新規治療・診断法開発
3次元CTIによる気腫・気道病変の評価と慢性閉塞性肺疾患病態解析への応用	呼吸器	疾患・外傷	機器診断・開発
3次元空間での追跡眼球運動信号の形成における小脳片葉の役割	脳・神経・精神	運動・行動	機能解析
FDG-PETIによる炎症病変診断法の開発:膜輸送遺伝子、病理組織学的検討	生体高分子	炎症・疼痛	機器診断・開発



医学系研究課題の網羅的分析に向けての分類支援

No.2

木村泰知 (小樽商科大学), 渋木英潔 (北海学園大学),

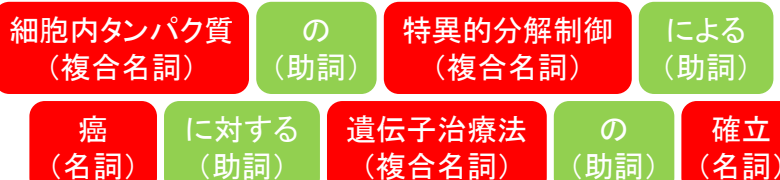
一瀬信敏 (札幌医科大学附属産学・地域連携センター, 文部科学省産学官連携コーディネーター)

処理過程

入力文解析

細胞内タンパク質の特異的分解制御による癌に対する
遺伝子治療法の確立

分類処理



選択処理

対象

家族・社会
血液・骨髄
生殖
胎児・乳幼児・小児
免疫系
消化器
呼吸器
目・耳鼻・皮膚
青少年・成人
泌尿器
老人
脳・神経・精神
細胞・組織
骨・筋・関節
生体高分子
心臓・血管

課題

感染症
運動・行動
疾患・外傷
移植
薬剤・放射線障害
癌・腫瘍
QOL
認知・運動・行動
発生・分化
活性・調節
生体高分子
ストレス
炎症・疼痛

解決手法

機器診断・開発
疫学・調査
遺伝子医療
機能解析
標的化・免疫医療
再生医療
機構解析
遺伝子・ゲノム解析
看護・介護・支援
薬物・化学医療
新規治療・診断法開発
予防医療

分類処理

単語 の重みを学習する
文構造の重みを学習する

スロットとは、名詞、複合名詞
助詞などの名詞以外の品詞で区切られる。



各スロットには複数の単語(形態素) M1, M2...
単語ごとに重み MW1, MW2...

・スロット単位でのスロット尤度の和

$$\text{本手法(頻度)} \quad SP = \sum (MW_n \cdot Fq(M_n))$$

・尤度最大のカテゴリを推測

$$P = \sum (SW_i \cdot SP_i)$$

◎学習

・推測誤りを重みを変化

・以下の条件で変化させる

[1]スロット尤度で正解カテゴリを推測成功

スロットの重みを変化

[2]スロット尤度で正解カテゴリを推測失敗

単語の重みを変化

◎重みの変化量

正解を推測している場合 +1

不正解を推測している場合 -1

◎終了条件

推測精度が悪化する直前まで繰り返し

※10回繰り返し精度が変化しない場合も終了

評価実験

■ 実験目的

■ 対象、課題、解決手法から選択

■ 実験データ

■ 札幌医科大学
■ 北海道大学医学部
■ 旭川医科大学

■ 比較実験

■ Baseline(キーワードマッチング)
■ SVM
■ 提案手法A(頻度による重み)
■ 提案手法B(TFIDFによる重み)
※10分割交差検定による精度

実験結果の例

■ 3次元CTによる気腫・気道病変の評価と慢性閉塞性肺疾患病態解析への応用

■ Baseline(キーワードマッチング)
該当なし
■ SVM
生体高分子
■ 提案手法
運動・行動

課題

感染症
運動・行動
疾患・外傷
移植
薬剤・放射線障害
癌・腫瘍
QOL
認知・運動・行動
発生・分化
活性・調節
生体高分子
ストレス
炎症・疼痛

今後の課題

- 学習方法の改善
- 少ない学習データ → 高い精度
- 確信度の提示

手法	Baseline	SVM	本手法 (単語頻度)
精度	19.7%	54.4%	37.1%