

中間発表

07t4072f

岡崎 駿

研究背景(1)

- 近年、日本語の意味の多様化(多義語の増加)が進んできている。
- 問題点
 - 1.文書中の誤字の増加
 - 2.文書の意図を汲みとることが困難

研究背景(2)

- 具体例

- あう【合う・会う・遭う・遇う・逢う】

1. 寄り集まって一つになる。「三筋の流れが一つになって本流となる」
2. 互いに同じ動作をする。「話しー」「なぐりー」
3. 費やしたものに対し、損をしない結果が出る。引き合う。「千円に見切ってもー」
4. 対面する。会見する。「応接間で客とー」
5. 物事に出会う。「雨にー」「ひどい目にー」 等

【岩波国語辞典より】

研究目的

「多義語の意味を判別するシステム」
の開発



- 単語誤用の修正提案
- 辞書の項目にない新用法の発見

従来モデル(1)

多義語を含む一文を形態素解析(学習データ)



着目する単語の周辺の単語を抽出



周辺単語の頻度計算よりベクトル生成



【これを50文に対して行う】



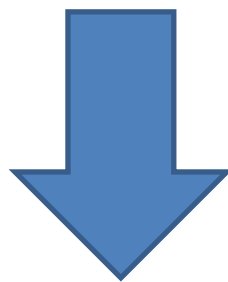
機械学習によるクラス間の識別面作成(LIBSVM)



テストデータを上記モデルに適応し、多義語の意味を識別

従来モデルの問題点

- クラス間分離性が低く、良好な識別面を得ることができない



クラス間分離性を最大化させ、良好な識別面を求めたい

提案手法(1)

一文を形態素解析



着目する単語の周辺の名詞を抽出



頻度計算よりベクトル生成



ベクトル空間を回転させ、同じクラス同士を近づけ、異なるクラスを離す(KLFDA)



機械学習によるクラス間の識別面作成(LIBSVM)

提案手法(2)

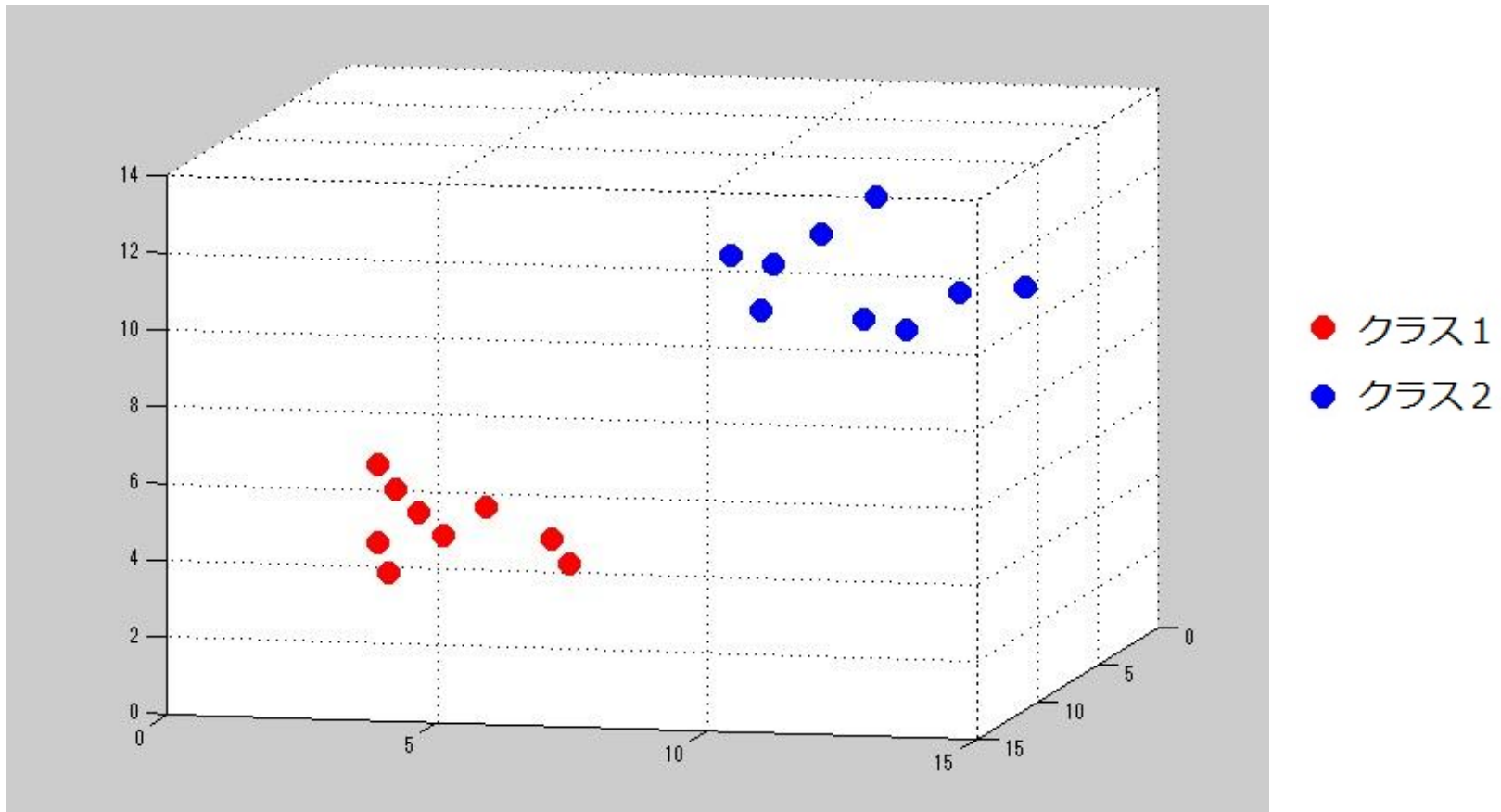


図2 提案手法イメージ

KLFDAの概要(1)

- 「局所クラス内散布行列」と「局所クラス間散布行列」を求めそれらの一般化固有値問題を解き、変換行列を作成する。
- 結果
 - 同じクラスの標本は近づける
 - 異なるクラスの標本は遠ざける
 - 行列の次元を削減することができる

KLFDAの概要(2)

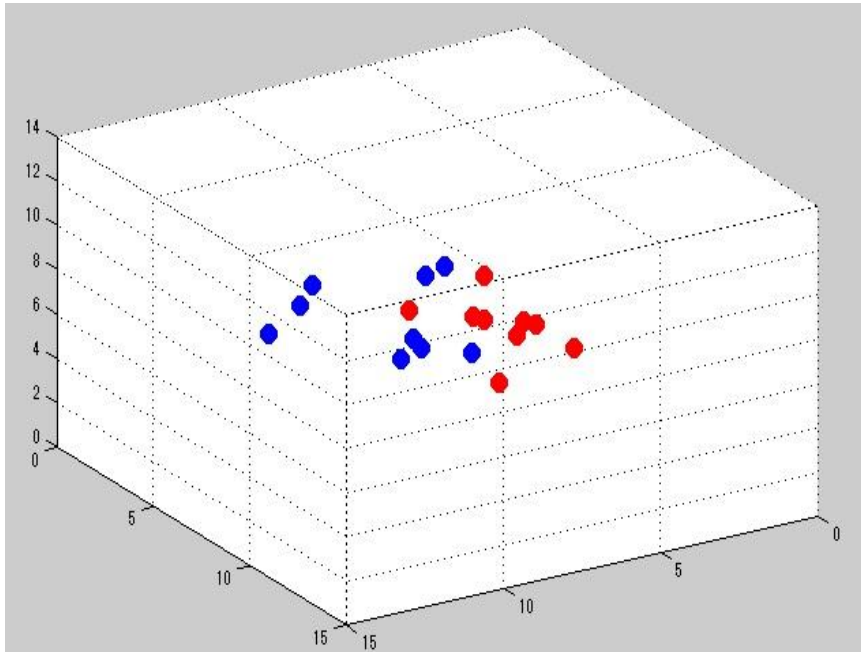


図1 従来モデルイメージ

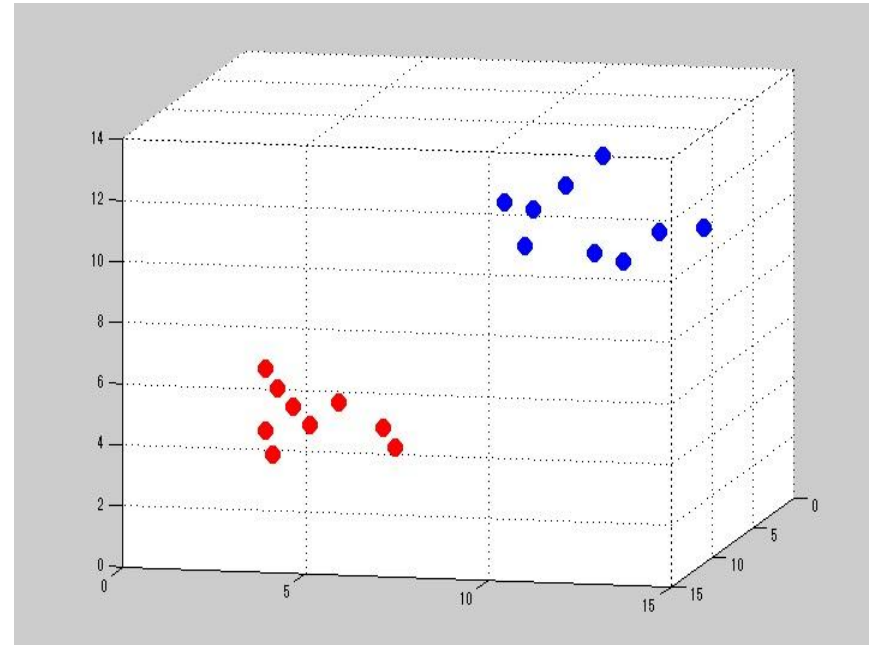


図2 提案手法イメージ

対称データ

- 多義語
 - 岩波国語辞典の分類項目
- 文章データ
 - SEMEVAL2010より提供された文章

進捗状況(1)

- KLFDAによる処理を行ったあとの学習データモデル作成
 - 実行プロセス
 1. 学習データから頻度ベクトル生成
 2. そのベクトルをKLFDAにより、次元削減及びベクトル空間の変換
 3. LIBSVMよりモデルの作成

進捗状況(3)

	0	1	2	3	4	5	6	7	8
1	117002	1:695.527148	2:0.000000	3:0.000079	4:0.000001	5:0.007483	6:-0.008967	7:-0.004416	
2	117002	1:695.527148	2:0.000000	3:0.000079	4:0.000001	5:0.007483	6:-0.008967	7:-0.004416	
3	117002	1:695.535012	2:0.000000	3:0.000079	4:0.000001	5:0.007485	6:-0.008970	7:-0.004418	
4	117001	1:-3365.155729	2:-0.000075	3:-0.009415	4:-0.000079	5:-1.052698	6:-1.519207	7:-0.004416	
5	117001	1:-3365.155492	2:0.015542	3:-1.719589	4:-0.000542	5:0.671945	6:-0.088781	7:-0.004416	
6	117001	1:-3365.155509	2:-0.819611	3:0.573989	4:1.409220	5:0.649487	6:-0.087467	7:-0.004416	
7	117001	1:-3364.994569	2:-0.000094	3:-0.007915	4:-0.000097	5:-0.757086	6:0.901325	7:1.391325	
8	117001	1:-3365.084558	2:-0.000095	3:-0.007906	4:-0.000097	5:-0.758246	6:0.906357	7:-1.391325	
9	117001	1:-3365.155490	2:1.622308	3:0.596380	4:0.001225	5:0.649172	6:-0.087433	7:-0.004416	
10	117001	1:-3365.155509	2:-0.817974	3:0.575006	4:-1.409622	5:0.649771	6:-0.087523	7:-0.004416	
11	117002	1:695.527298	2:0.000000	3:0.000079	4:0.000001	5:0.007483	6:-0.008967	7:-0.004416	
12	117002	1:695.535999	2:0.000000	3:0.000079	4:0.000001	5:0.007485	6:-0.008970	7:-0.004418	
13	117002	1:695.529444	2:0.000000	3:0.000079	4:0.000001	5:0.007483	6:-0.008967	7:-0.004417	
14	117002	1:695.527439	2:0.000000	3:0.000079	4:0.000001	5:0.007483	6:-0.008967	7:-0.004416	
15	117002	1:695.530138	2:0.000000	3:0.000079	4:0.000001	5:0.007484	6:-0.008968	7:-0.004417	
16	117002	1:695.527446	2:0.000000	3:0.000079	4:0.000001	5:0.007483	6:-0.008967	7:-0.004416	
17	117002	1:695.618044	2:0.000000	3:0.000079	4:0.000001	5:0.007503	6:-0.008991	7:-0.004426	
18	117002	1:695.527149	2:0.000000	3:0.000079	4:0.000001	5:0.007483	6:-0.008967	7:-0.004416	
19	117002	1:695.527398	2:0.000000	3:0.000079	4:0.000001	5:0.007483	6:-0.008967	7:-0.004416	
20	117002	1:695.558477	2:0.000000	3:0.000079	4:0.000001	5:0.007495	6:-0.008981	7:-0.004422	
21	117002	1:695.587017	2:0.000000	3:0.000079	4:0.000001	5:0.007494	6:-0.008980	7:-0.004422	
22	117002	1:695.590254	2:0.000000	3:0.000079	4:0.000001	5:0.007504	6:-0.008992	7:-0.004426	

現状の課題

- KLFDAによって変換された学習データと、テストデータのベクトル空間をどう合わせるか。
- KLFDAでは周囲の4点のベクトルを参照するため、入力行列で4つ以下クラスがあった場合どのように対処するか

今後の予定

- 現状の課題の解決
- テストデータを作成されたモデルに適応する実験
 - 上記の実験より正解率を求める
- KLFDA以外の手法との比較

スケジュール

- 9月下旬: 中間発表、実験開始
- 10月上旬～12月中旬: 実験から改良及び修正
- 12月下旬: 論文作成
- 1月下旬: 論文完成
- 2月中旬: 卒研発表