

## LSA を用いた要約文自動評価の試み

牛 娜\* 村木英治\*\*

\* 東北大学大学院教育情報学教育部

\*\* 東北大学大学院教育情報学研究部

**要旨：**木村宗男 (1982)『日本語教授法』には、読解指導や聴解指導における学習者の理解の是非を確認する手段としての要約作文の有効性が随所に説かれている。しかし、採点には、客観的な評価の基準が明確にされていない、公平な基準の保証がないのが現状である。本研究の目的は、Dumains (1998) による“Latent Semantic Analysis 手法 (潜在意味分析)”に着目し、要約文の評価への適用可能性について検討し、その有用性を検証することである。本稿は、以下のような3部構成になっている。まず、本研究の背景と目的を説明した。次に、LSAを用いた日本語要約文の評価方法を説明した。最後に、LSAによる評価結果は日本語教師による評価結果と比較し、LSAの有用性を検証する。

**キーワード：**自然言語処理、評価基準、潜在意味分析、特異値分解、

### 1. はじめに

「日本語教授法」には、読解指導や聴解指導における学習者の理解の是非を確認する手段としての要約文の有効性が説かれている。「要約文を書かしてみれば、学習者の日本語の理解力と表現力がわかる」という指摘もなされている。しかし、要約文の評価には、「採点者」の主観が入りやすく、全体的に統一した信頼性の高い評価を行うのが難しい。

一方、“潜在意味分析”(Latent Semantic Analysis 以下 LSA) という名前であり、Deerwester, Dumais, Furnas, Landauer, & Harshman (1990) によって開発され、複数の文章間、単語間および文章-単語間の意味的な類似度を検出する手法がある。潜在意味分析を用いてテキストの一貫性と内容の理解を予測することができる。潜在意味分析を用いてテキストの主観評定(小論文の成績等)を予測することができる。

要約文評価には、主観に拠らない信頼性の高い評価を行うにはどうすべきか。この問題を解消する一つ的手段として、潜在意味分析(LSA)という数学方法を活用できるのではないかと考えられる。本研究の目的は、潜在意味分析を用いた日本語要約文の自動評価を試み、その有用性を検証することである。

### 2. LSA を用いた要約文自動評価方法

#### 2.1 要約文の評価基準

佐久間(1997)より、要約文の評価には、原文の内容理解と要約文における表現技法という二つの観点が考えられる。例えば、内容理解面では、要約文に原文の重要な内容が書かれているか、表現面では、要約文に語彙・文法・文字表記の誤りはないか、などを見る。しかし、LSAの特徴により、すべての評価基準をLSAで再現できるわけではない。本研究の研究方法としては、LSAで再現可能な評価基準からアプローチしていきたいと思う。

#### 2.2 LSA を用いた要約文自動評価方法

##### 2.2.1 実験

- ①材料：「要約力を磨く説明文の指導」という本の説明文1編と要約文の解答例
- ②使用した文章：恐竜を探る 分析に使用した品詞：名詞のうち一般名詞、サ変接続名詞、副詞可能、未知語の計4種類を用いた。未知語は茶筌に登録されていないが、その説明文の内容に特徴的な単語である可能性がある。
- ③手続き：原文と要約文にLSAを行い、原文における重要な内容がどれくらい書かれているかを測る。実験手続きを詳しく説明する。

原文と要約文から作成された行列は表1に示







表4 センテンス間の相関係数

	01-1	01-2	01-3	02-1	02-2	03-1	03-2	03-3	03-4	04-1	04-2	04-3	04-4	05-1	05-2	06-1	06-2	06-3	06-4	06-5	07-1	07-2	08-1	08-2	08-3	09-1	09-2	09-3	S1	S2		
01-2	0.972																															
01-3	0.989	0.996																														
02-1	0.76	0.588	0.658																													
02-2	0.626	0.426	0.506	0.983																												
03-1	0.165	-0.07	0.02	0.766	0.872																											
03-2	-0.12	-0.35	-0.27	0.55	0.696	0.958																										
03-3	-0.62	-0.79	-0.73	0.039	0.224	0.672	0.856																									
03-4	0.193	-0.04	0.048	0.784	0.886	1	0.95	0.651																								
04-1	-0.17	-0.39	-0.31	0.514	0.664	0.945	0.999	0.877	0.935																							
04-2	-0.17	-0.39	-0.31	0.515	0.665	0.945	0.999	0.877	0.936	1																						
04-3	-0.16	-0.39	-0.3	0.52	0.67	0.947	0.999	0.874	0.938	1	1																					
04-4	0.642	0.444	0.524	0.986	1	0.862	0.681	0.204	0.876	0.649	0.65	0.654																				
05-1	0.772	0.601	0.671	1	0.979	0.754	0.535	0.021	0.773	0.498	0.499	0.504	0.983																			
05-2	0.184	-0.05	0.04	0.779	0.882	1	0.952	0.657	1	0.938	0.939	0.941	0.872	0.767																		
06-1	0.719	0.535	0.61	0.998	0.992	0.805	0.601	0.101	0.821	0.566	0.567	0.572	0.995	0.997	0.816																	
06-2	-0.44	-0.56	0.251	0.427	0.814	0.946	0.977	0.798	0.959	0.959	0.957	0.408	0.233	0.803	0.31																	
06-3	-0.04	-0.27	-0.18	0.62	0.755	0.979	0.996	0.809	0.973	0.992	0.992	0.993	0.741	0.605	0.975	0.667	0.915															
06-4	-0.62	-0.79	-0.73	0.033	0.218	0.667	0.853	1	0.646	0.874	0.873	0.87	0.197	0.014	0.652	0.094	0.976	0.806														
06-5	-0.74	-0.88	-0.83	-0.13	0.055	0.536	0.755	0.985	0.512	0.783	0.782	0.778	0.034	-0.15	0.519	-0.07	0.926	0.696	0.986													
07-1	0.389	0.162	0.251	0.895	0.962	0.973	0.865	0.482	0.979	0.843	0.844	0.847	0.956	0.886	0.977	0.92	0.657	0.905	0.476	0.325												
07-2	0.97	1	0.995	0.578	0.417	-0.08	-0.36	-0.79	-0.05	-0.4	-0.4	-0.4	0.436	0.593	-0.06	0.527	-0.64	-0.28	-0.8	-0.88	0.153											
08-1	0.919	0.8	0.851	0.955	0.883	0.542	0.278	-0.26	0.565	0.237	0.238	0.244	0.893	0.96	0.558	0.935	-0.05	0.359	-0.27	-0.42	0.722	0.794										
08-2	-0.66	-0.82	-0.76	-0.01	0.176	0.635	0.83	0.999	0.613	0.853	0.852	0.849	0.156	-0.03	0.62	0.062	0.965	0.779	0.999	0.993	0.438	-0.82	-0.31									
08-3	0.409	0.184	0.272	0.904	0.968	0.968	0.854	0.463	0.974	0.831	0.832	0.835	0.962	0.896	0.972	0.929	0.641	0.896	0.457	0.305	1	0.174	0.737	0.419								
09-1	0.994	0.991	0.999	0.687	0.54	0.06	-0.23	-0.7	0.088	-0.27	-0.27	-0.26	0.557	0.7	0.079	0.641	-0.53	-0.14	-0.7	-0.81	0.29	0.99	0.871	-0.73	0.31							
09-2	0.628	0.428	0.509	0.983	1	0.871	0.694	0.222	0.884	0.662	0.663	0.668	1	0.98	0.88	0.993	0.424	0.753	0.215	0.062	0.961	0.419	0.885	0.173	0.967	0.542						
09-3	0.999	0.963	0.983	0.784	0.655	0.202	-0.09	-0.59	0.229	-0.13	-0.13	-0.12	0.67	0.795	0.221	0.744	-0.4	-0	-0.6	-0.72	0.423	0.96	0.933	-0.63	0.443	0.99	0.657					
S1	0.989	0.996	1	0.658	0.506	0.02	-0.27	-0.73	0.048	-0.31	-0.31	-0.3	0.524	0.671	0.04	0.61	-0.56	-0.18	-0.73	-0.83	0.251	0.995	0.851	-0.76	0.272	0.999	0.509	0.983				
S2	0.237	0.002	0.093	0.811	0.906	0.997	0.935	0.616	0.999	0.919	0.919	0.921	0.897	0.8	0.999	0.846	0.77	0.962	0.611	0.473	0.987	-0.01	0.601	0.577	0.983	0.132	0.906	0.273	0.09			
S3	0.859	0.715	0.776	0.985	0.937	0.646	0.4	-0.13	0.657	0.361	0.362	0.368	0.944	0.988	0.661	0.973	0.083	0.477	-0.14	-0.3	0.806	0.709	0.992	-0.18	0.818	0.8	0.938	0.878	0.78	0.7		

相関から文脈同士の意味的な結びつきの強弱の程度、つまり、センテンス間の意味的な類似度を調べるためにセンテンス間の相関係数を算出する。

要約文の各センテンスと原文の重要なセンテンスの相関係数を陰影の数値で示す。この方法を使って、原文中で重要な内容がどれぐらい書かれるかを算出できる。

## 2.2.2 実験の結果

ここで、陰影の数値の平均値は0.74である。つまり、要約文と原文中で重要な内容の相関係数が0.74である。強い相関関係があると判断できる。原文中で重要な内容の54%が書かれているということが分かった。

## 3. 日本語教師による要約文の評価結果

### 3.1 日本語教師による評価方法

本研究では、30名の実験協力者（日本人25名、留学生5名）が書いた要約文を、5人の日本語教師に評価してもらった。評価欄に①内容、②つながり、③構成、④具体例、⑤文法、⑥語句、⑦文字、⑧句読点、⑨字数という9つの評価項目を入れた。実験協力者が書いた要約文に対して、5段階（5優れている、4やや優れている、3まあまあ、2やや劣る、1劣る）を評価欄（各評価項目）に記入してもらった。

### 3.2 日本語教師による評価結果

一人の実験協力者が書いた要約文を、5人の日本語教師に評価してもらい、評価項目ごとの平均値と標準偏差を算出した。内容面に関する30人の実験協力者の得点の平均値と標準偏差を表5に示す。

表5 内容に関する30人の実験協力者の得点の平均値と標準偏差

平均値	標準偏差
2.59	0.35

図1は日本語教師による内容に関する評価結果を表すヒストグラムである。グラフの縦軸は人数を表し、横軸は得点区分を表す。

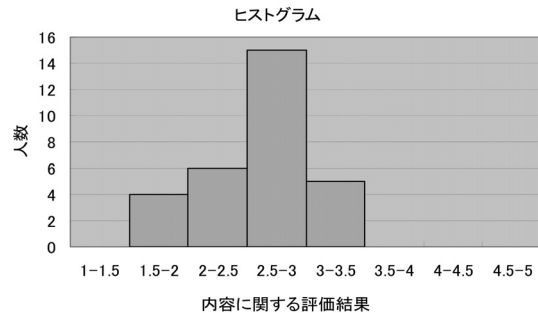


図1 日本語教師による内容に関する評価結果

図1からわかるように、実験協力者の得点は2.5-3の区間に集中している。得点が3.5以上であった実験協力者はいない。全体的に得点が低い。日本人

表6 内容に関する日本人25人と留学生5人の得点

日本語教師による評価結果	日本人(25人)	留学生(5人)	全体(30人)
内容に関する得点の平均値	2.6	2.4	2.59

25人と留学生5人の得点は表6のとおりである。

#### 4. LSAの有用性の検証

LSAによる評価結果を日本語教師による評価結果と比較し、その評価方法の有用性を検証する。LSAによる評価結果と日本語教師による内容に関する評価結果の相関係数の散布図を図2にあげた。

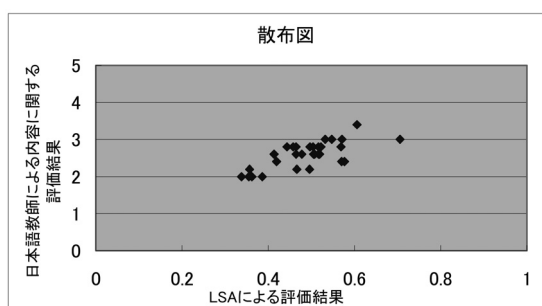


図2 日本語教師による内容に関する評価結果とLSAによる評価結果の相関係数散布図

LSAによる評価結果と日本語教師による内容に関する評価結果の相関係数は0.72であり、かなり強い相関関係がある。

#### 5. 考察とまとめ

##### 5.1 考察

日本語教師による評価結果と比較し、LSAの有用性を検証した。平均値、標準偏差、相関係数などの統計方法を使用し、日本語教師による評価結果と比較した。比較結果に基づき、以下の考察を行う。

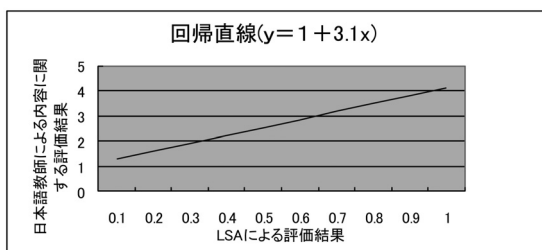


図3 回帰直線

① LSAによる評価結果と日本語教師による内容に関する評価結果との相関係数を算出した。その相関係数は0.72であり、かなり強い相関関係がある。実験

協力者30人のデータに基づき、回帰直線を作った。回帰直線を図3にあげた。

図3からわかるように、日本語教師による内容に関する評価結果が高い要約文に対して、LSAによる評価結果も高いことが分かった。つまり、この分析により、LSAを用いた要約文の評価方法は内容の評価に適用できる可能性があることが分かった。

② 25名の日本人大学生と5名の中国人留学生を対象として要約文調査を行った。それぞれの評価項目について、日本語教師による評価結果の平均値が低い5人の留学生に対して、LSAによる評価結果の平均値も低いことが分かった。

##### 5.2 まとめ

① 共起頻度を表す行列の作成には、LSAは作成した行列が大きければ大きいほど、計算結果の精度が高くなる。行列の再構成について、S(固有値)が大きいくつ選べばよいのかという問題がある。再検討の必要があると考えられる。

② LSAの検証の精度を増やすために、作文サンプル数を増やす必要があると考えられる。また、算出されたデータの処理についても、共分散構造分析などの多変量解析を加え、より多角的な分析を加えていきたいと考えている。

③ LSAによる内容に関する評価結果と日本語教師による内容に関する評価結果の重決定は0.52である。若干低いから、現在、日本語要約文評価に応用できないが、日本語の要約文評価へのLSA適用が可能であると明らかとなった。LSAを用いた要約文の内容に関する評価方法を改良した上で、日本語要約文評価に使えようと考えられる。

##### 参考文献と引用サイト

Landauer, T. k., Foltz, p. w., & Lahman, D (1998). *An Introduction to Latent Semantic Analysis, Discourse Processes*, 25, 259-284.

佐久間 まゆみ (1997) 要約文の表現類型と評価方法 -外国人留学生と日本人大学生の比較-

- 平成 6 年度～平成 8 年度科学研究費補助金 研究成果報告書
- Steinhart, D.(2001). *Summary Street: An intelligent tutoring system for improving student writing through the use of latent semantic analysis*  
Unpublished doctoral dissertation, Institute of Cognitive Science, University of Colorado, Boulder.
- 木村宗男 (1982) 『日本語教授法』 凡人社  
椿本弥生 (2002) LSA 手法解説  
吉本清久 (2005) 『要約力を磨く説明文の指導』  
明治図書  
Chasen:茶筌 (<http://chasen.aist-nara.ac.jp/chasen/distribution.html.ja>)  
<http://lsa.colorado.edu/>

## The try of evaluating summary automatically by Latent Semantic Analysis

Niu Na\*, Eiji MURAKI\*\*

\* *Graduate School of Educational Education Division, Tohoku University*

\*\* *Graduate School of Educational Informatics, Tohoku University*

The purpose of this study is to try of evaluating summary automatically by Latent semantic analysis and verifying that of it is useful. In this paper, contains three parts: First, the background and objectives of this paper. Then, a method that try to valuate the Japanese summary's level by LSA will be presented. Finally, verifying that if it is useful.

**Key words:** Natural language processing, Latent semantic analysis, Singular values decomposition (SVD)