新アイデア創造のための日常会話解析の一手法

A Method for Analyzing Daily Conversation for New Idea Creation

山口広樹，大澤幸生，西原陽子

Hiroki YAMAGUCHI, Yukio OHSAWA, and Yoko NISHIHARA

東京大学大学院工学系研究科システム創成学専攻
Department of Systems Innovation,
School of Engineering, University of Tokyo.

Abstract: We develop a system for analyzing daily conversation. Utterances which are ignored in conversation are focused in this system. These utterances can involve speaker's passion and be buried away. The system extracts and displays these utterances. Thus users can reconsider speaker's passion from ignored utterances. In this paper, we adopt Peer Counseling view point and make online-chat experiment for evaluation. The experiment shows the significance of the method.

1. はじめに

情報技術の飛躍的な発達に伴い，我々人類が扱う情報亮は 2010 年までにはゼッタバイト（10 の 21 乗バイト）を超えると予測され[1]，現代は情報爆発の時代とされている．

こうした中，この膨大な情報の内の果たしてどの程度の情報を有効利用できているであろうか．ここでは精確な計算は省くが，この答えは天文学的に小さい値となると予想される．なぜなら現存する情報の中には，近年急速に普及してきた電子掲示板やブログも少なからず含まれているからである．これらはインターネットを利用する個々人が各々の想いを綴ったものであり，そこから社会に持続的成長をもたらすような新たな価値が生まれることは至極稀であると言える．

その一方で，「必要は発明の母」という言葉も存在するように，掲示板やブログ，オンラインチャットなどを介して行われるコミュニケーションから，革新的なアイデア・イノベーションを生むことができると考えられる．実際，セレンディピティ[2]と呼ばれる偶発的な出来事から偉大な発明が生まれたというエピソードも少なくない．

そこで本論では，日常の何気ない会話から新しいアイデアの種を掘り起こし，人の生活を向上させることのできるアイデア（本論ではこれを価値とする）を創造する過程を支援するシステムを開発する．本論では，テキスト処理技術を応用した会話解析によってこの実現を試みる．

本提案手法の一番の特徴は，会話内では特に注目されておらず，通常は重要でないと考えられている発言に注目している点である．本論ではこのような発言を被放置発言と呼ぶ．被放置発言は，“発言の背景に発言者独自の強い想いが存在するがために相手に伝わらなかった発言“や”卓抜な先見性の強さのために理解されずに放置されてしまった発言”も存在する．これら発言が注目されずに埋もれてしまっていることは，非常に“もったいない”ことである．本研究はこうした発言から新たな価値を見出すことを主目的とする．

また本論のもう一つの特徴は，何らかの共通点（同じような環境や悩み）を持つ（又は経験した）グループを対象として，対等な立場で同じ仲間として行われるカウンセリングであるピアカウンセリング[3,4]の視点を導入している点である．

ここでは大学院生を対象として，学生生活における悩みの共有とその解決方法をピア（仲間）同士で相談するという実験を行う．また本論では，ピアカウンセリングの起源である医療現場への適用可能性についても考察する．

2. 関連研究

関連研究として，まずテキストマイニング技術の従来研究を挙げる．テキストマイニングでは，文書内容をそこに含まれる単語群のベクトルで表すベクトル空間法[5]が基礎技術として広く利用されている．ベクトル空間法はシンプルかつ汎用性が高い手法であるという理由から，必ずしも正しい文法が使われておらず構造が無秩序である，日常会話を解析対象とする本研究では，これを採用する．

ベクトル空間法を基として，多くの研究が派生的になされてきた．それらの中には，ベクトルを拡張してより単語間の関連性を考慮したもの[6]や単語の意味を考慮できる行列に拡張したもの[7]，文書構造に注目して抄録を自動生成するもの[8]などが存在する．

また本研究は，社会学的側面や認知科学的側面も有している．特に人間の心理を扱うという点で，人

間の社会行動によって影響される社会的な過程における心理を対象とする社会心理学[9,10]との関連性が深いといえる．また認知科学的側面として，過去の会話のレビューによって自らの認知を認知することでメタ認知[11]を促すことも可能である．

3. アプローチ

本論で提案するシステムは，コンピュータによるテキスト処理と人間による感性処理とから成る．図 1 にテキスト処理部分のフローチャートを示す．

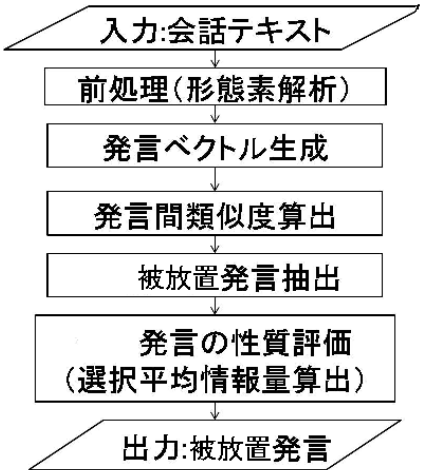

図 1 テキスト処理部フローチャート

4.1. 入力テキスト

発言者名と発言内容を一対として一発言とした会話ログを入力テキストとする．

4.2. 前処理

入力テキストに対して，前処理として，形態素解析[12]を行い，各発言からそこに含まれる単語群を抽出する．

次の前処理として，同義語処理を行う．これによって，同義語は同じ単語であると認識して計算することができる．

最後の前処理として，発言をベクトルする際の基準となるユニークタームベクトルを生成する．ユニークタームベクトルは，解析対象の文書に現れる単語全てを過不足なく列挙してベクトル化したものである．例として，10 種類の単語（形態素）が含まれる場合，ユニークタームベクトルは 10 次元となる．

4.3. 発言ベクトル生成

形態素解析によって得られた単語群を，ベクトル空間法を用いてベクトル化する（(1)式），この時，ベクトルを正規化しておく．

$$\mathbf{U}_k = (t_0, t_1, \Lambda\ , t_n), \left|\mathbf{U}_k\right| = 1 \qquad (1)$$

但し，t_k は k 番目の単語の出現頻度，　は総単語種類数である．また，後の処理で扱いやすくするために正規化を行っている．

4.4. 発言間類似度算出

発言ベクトルを基に各発言とその前後の発言群（コンテクスト）との類似度を(2)式に示すように定義する．ここでは，before 類似度(2.a)と after 類似度(2.b)の二つの類似度を定義する．

$$S_{before \cdot k} = \mathbf{W}^T_{before \cdot k} \cdot \mathbf{U}_k \qquad (2.a)$$

$$S_{after \cdot k} = \mathbf{U}_k \cdot \mathbf{W}^T_{after \cdot k} \qquad (2.b)$$

$$\text{但し,} \left(\begin{array}{l} \mathbf{W}_{before \cdot k} = \sum_{i<n} \mathbf{U}_i, \left|\mathbf{W}_{before \cdot k}\right| = 1 \\ \mathbf{W}_{after \cdot k} = \sum_{i>k} \mathbf{U}_i, \left|\mathbf{W}_{after \cdot k}\right| = 1 \end{array} \right)$$

4.5. 被放置発言抽出

二つの類似度に閾値を設けることによって，被放置発言を抽出する．被放置発言は周囲から排除されている発言であるため，閾値よりも小さい類似度を持ち，前後の会話と内容が食い違っていると考えられる発言を被放置発言として抽出する（(3)式）．なお，会話毎に妥当な閾値を定められるように閾値はユーザーが決定できるように設計している．

$$S_{before \cdot k} < \alpha \quad \mathrm{I} \quad S_{after \cdot k} < \beta \qquad (3)$$

（α，β は 0 から 1 の範囲の値に設定）

4.6. 被放置発言の性質評価

被放置発言には潜在価値を持っているものとそうでないものが混在する．なぜなら，放置されるべくして放置された発言と，発言者の強い想いを含んでいる本来注目されるべき発言が存在するからである．そこで本研究では被放置発言の性質評価によって，これらの発言の見極めを試みる．

これと同時に，性質評価によって，抽出手法の問題点の一つを解決することができる．抽出時には，注目発言とその前後の発言群との類似度のみを考慮しているため，含まれる単語が少ない発言は，言外に含まれる想いの強さに関係なく被放置発言として抽出されやすくなってしまう．発言の資質評価によってこうした発言の重要度を下げて，この問題をある程度解消することができる．

(4.a),(4.b)式によって定義する情報量を，選択平均情報量とし，抽出した例外発言の評価として利用した．

$$E_{\mathbf{W}_{before \cdot k}\mathbf{U}_k} = \sum_{t_i \in \mathbf{U}_k} p_{\mathbf{U}_k t_i} I_{\mathbf{W}_{before \cdot k} t_i} \qquad (4.a)$$

$$E_{\mathbf{W}_{after \cdot k}\mathbf{U}_k} = \sum_{t_i \in \mathbf{U}_k} p_{\mathbf{U}_k t_i} I_{\mathbf{W}_{before \cdot k} t_i} \qquad (4.b)$$

$$\text{但し,} \left(\begin{array}{l} \mathbf{W}_{before \cdot k} = \sum_{i \le n} \mathbf{U}_i \\ \mathbf{W}_{after \cdot k} = \sum_{i \ge k} \mathbf{U}_i \end{array} \right)$$

上式で定義した情報量を各発言に対して算出すると，二つの情報量は表 1 に示す性質を持つ．

表 1　選択平均情報量の性質

Before 値	大	新しい話題を挿入
	小	以前の話題と似通った内容
After 値	大	以後の会話とは異なる内容
	小	以後の会話と似通った内容

また，定義した情報量を使って例外発言の性質を評価するため，(5)式のように評価関数を定義する．

$$\overline{E} = E_{\mathbf{W}_{before \cdot k}\mathbf{U}_k} - E_{\mathbf{W}_{after \cdot k}\mathbf{U}_k} \quad (5)$$

(5)式の評価関数によって得られた評価値は，表 2 に示す特徴を持つ．

表 2　評価値の特徴

評価値 大 (before 大且 after 小)	新たな話題を導入している発言
評価値 小 (before 小且 after 大)	会話に特に影響を与えていない発言

表 2 の特徴から，評価値が大きい発言ほど，前後の会話コンテクストと食い違っている例外発言でありながら会話に影響を与えた発言であるといえる．こうした発言には，発言の裏側に何らかの意図や知識（暗黙知や経験知）が潜んでいる発言であると考えられる．したがって，評価値の高い発言から今まで共有されていない新たな価値を創造できる可能性が高いといえる．

4.7. 出力：被放置発言

抽出した被放置発言を評価値に従ってランク付けをした後，ランク順に出力する．出力項目を表 3 に示す．

表 3　出力項目

Rank	発言 No	発言者	発言内容	Score

ユーザーは上記の出力結果を確認しながらインタラクティブにパラメータの調整でき，より妥当な結果を追究することができる．

4. オンラインチャット実験

評価実験として，大学という共通の環境を持つ被験者で，日常会話に近いオンラインチャットでの会話を解析した．

4.1. 実験概要

大学という共通の環境を持つ小集団として，被験者は大学教員 1 名と大学院生 6 名で行った．日常会話でしばしば話題となる話題をチャットのテーマとして与え，テーマに沿って自由に会話を行い，そのログデータを取得した．ここでのテーマは「大学生のお財布事情」として，各自の節約術などについての会話データが得られた．チャット実験は 30 分程度行われた．

4.2. パラメータ設定

本実験用いたパラメータを表 4 に示す．パラメータの設定は，解析対象データに合わせてインタラクティブに設定する．計算範囲窓 w は話題展開の速さに合わせて設定した．二つの閾値は，目的とする被放置発言は発言後の会話に多少の影響を与えていると考えられるため $\alpha < \beta$ とした．また，抽出された被放置発言の数が 15~30 個程度になるように値を調整した．

表 4　パラメータ設定

計算範囲窓 w	before 閾値 α	after 閾値 β
20	0.2	0.25

4.3. 解析結果

表 5 に解析結果の一部（評価ランキング上位 5 発言）を示す．合計 24 個の被放置発言が抽出された．

表 5　解析結果

Rank	発言 No	発言者	発言内容	Score
1	111	GU	K はハイボール飲め	0.88
2	72	GO	地デジに対応していません	0.55
3	104	KO	おかわり系のところは，大抵無理をして後悔するのである．	0.36
4	55	KO	うーん，ひっかからない www	0.35
5	62	KA	家でのエアコンはドライのみ　省エネ　たぶん	0.29

4.4. 実験考察

表 5 の解析結果を見ると，抽出した発言のみでは解釈が不可能であることが分かる．発言前後の文脈を参考にして初めて，被放置発言が内包する発言者の物語を垣間見ることができる．本論では詳細は省き，簡単な説明をする．

ランキング 1 位の発言は，発言者と参加者の一部が共通の経験（物語）を持っており，敢えて深く言及されなかったものであることが，実験後のインタビュ

一で明らかになった.

ランキング 2,3,5 位の発言には，発言者が日頃感じている関心・問題意識が含まれる発言であるといえる.

表 5　チャット実験：解析結果

ランキング 4 位の発言は，実験後のインタビューによると，発言前の文脈に外れてはいないが，表現方法の違いで，他の参加者には意味が正確に伝わらなかった発言であるといえる.

このように，発言者の隠れた関心・問題意識，小グループ固有の物語が含まれている発言が被放置発言として抽出されていることが分かった.

また，これらの発言自体は新アイデアの種であり，被放置発言に再度注目するようフィードバックし会話を繰り返すことによって，新アイデアを創造していく過程が必要となる.

5. おわりに

本論では，ピアカウンセリングという視点を導入して，仲間同士での会話による問題意識の共有とその解決法の模索を支援することを目的の一つとしている. ピアカウンセリングの起源は障害者支援にあり，自立生活支援センター等において実践されているものである. ピアカウンセリングにおける「当事者こそ専門家」という観点を応用することで，本提案手法を医療現場だけでなく，システム開発等に適用が可能であると考えられる. 本研究の今後の課題は，これらの可能性を追究することと，本手法の方法論と評価手法を確立することである.

謝辞

著者の一人（山口）は，文部科学省 GCOE プログラム「機械システム・イノベーション国際拠点」による補助を受けた. また，被験者になっていただいた東京大学 大澤・西原研究室の皆様に感謝します.

参考文献

[1] How much information? 2003, University of California Berkeley *http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/*

[2] 三輪和久，飛躍を伴う発見における潜在的意識の関与，計測と制御，第 48 巻，第 1 号, pp.33-38, 2009.

[3] 安積遊歩，野上温子，ピアカウンセリングという名の戦略，青英舎，1999.

[4] 西山久子，山本力，ピアサポート-仲間支援活動の起源から現在まで-，岡山大学教育実践総合センター紀要，第 2 巻, pp.81-93, 2002.

[5] Salton, G, A. Wang, and C. S. Yang, “A Vector Space Model for Automatic Indexing”, Communication of the ACM, Vol.18, No.11, pp.613-620, 1975.

[6] Jing H., and Tzoukermann E., Information retrieval based on context distance and morphology, Proceedings of the 22nd Annual International ACM SGIR Conference on Research and Development in Information Retrieval (SIGIR ’99), pp.90-96, 1999.

[7] Geoffrey Z. Liu, Semantic Vector Space Model: Implementation and evaluation, Journal of the American Society for information Science, Vol.48, Issue.5, pp.395-417, 1997.

[8] 住田一男ほか，文書構造解析に基づく自動抄録生成と検索提示機能としての評価，電子情報通信学会論文誌，Vol.J78-D-Ⅱ, No.3, pp.511-519, 1995.

[9] 井上隆二，山下富美代，図解雑学社会心理学(図解雑学シリーズ)，ナツメ社, 2000.

[10] 松尾 太加志，コミュニケーションの心理学-認知心理学・社会心理学・認知工学からのアプローチ-，ナカニシヤ出版, 1999.

[11] 諏訪正樹，身体知獲得のツールとしてのメタ認知的言語化，人工知能学会誌, 20 巻, 5 号, pp.525-531, 2005.

[12] 形態素解析ソフト『茶筌 ver.2.3.3』 *http://chasen-legacy.sourceforge.jp*