

ブレインストーミング事例の再利用のための 関連情報推薦機構の試作

Developing a Recommendation Mechanism to Reuse Brainstorming Cases

伊藤 栄俊^{*1} 鈴木 智也^{*1} 丹羽 佑輔^{*1} 大園 忠親^{*1} 新谷 虎松^{*1}
Eishun Ito Tomoya Suzuki Yusuke Niwa Tadatika Ozono Toramatsu Shintani

^{*1}名古屋工業大学大学院工学研究科情報工学専攻

Department of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology

We are developing a recommendation mechanism to present relevant information in brainstorming using Mahoca, which is a reality-based e-sticky notes system. In the brainstorming with Mahoca, ideas are expressed on card objects. We aim to realize effective reuse of cards. However, it is hard to reuse cards because it requires a context-aware recommendation system. We implemented and evaluate four recommendation methods that have different strategies in terms of randomness and word models. Our evaluation showed that the methods have their unique characteristics. We expect the methods to promote more creativity.

1. はじめに

会議などにおける発想法の一つに、ブレインストーミング (BS) が存在する。BS では、集団でアイデアを出し合い続けることで、互いにアイデアを刺激し合い、一層の独創的なアイデアを生み出すことを目指す。BS を行う際に、付箋の紙を用いられることがある。アイデアを付箋に書き出し、ホワイトボードなどに貼り付ける。付箋は自由に記述、配置ができ、チーム内でアイデアを共有しやすいという利点がある。過去に行われた BS における、紙の付箋の内容だけでなく、配置方法や付箋のグループといった情報は、現在の BS の参考にするとときに有益である。しかし、再利用を考えた場合、電子化された付箋の方が、過去の状態を保存、管理できる点で優れている。我々は実世界型電子付箋システム Mahoca [Ozono 16] を開発した。Mahoca は、物理付箋を取り込み、電子付箋に変換したり、直接電子付箋を作成したりするシステムである。Mahoca を用いた、過去の BS 時の情報を BS 事例と呼ぶ。本研究では、Mahoca 上の電子付箋をもつ、テキストやタグ情報から、関連情報を提示するシステムの開発が目的である。

本研究における課題は以下の 3 点である。1) 付箋に関連する情報の推薦。2) 現在の BS 状況の推定。3) 過去の BS 事例の再利用。本研究では、3 つ目の課題を解決するために、1 つ目と 2 つ目の課題に対して取り組んだ。1 つ目の課題に対して、2 種類のモデルと 4 種類の選択手法を用いて問題解決を図った。

2. 実世界型電子付箋システム Mahoca

実世界型電子付箋システム Mahoca は、アイデアの再利用のために電子付箋にアイデアを入力し、会議参加者間で共有するためのシステムである。Mahoca では、会議参加者が「スクリーン」と「クライアント」の二つのアプリケーションを利用することで、電子付箋を利用した会議を行う。図 1 に Mahoca を使用している様子を示す。スクリーンは、公開された電子付箋を画面に表示し、会議参加者に共有するアプリケーションである。スクリーンでは、会議に対応する「ベースボード」と呼

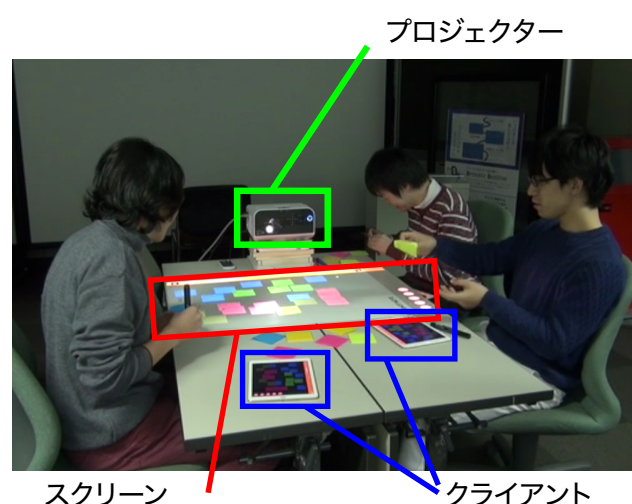


図 1: Mahoca を使用している様子

ばれる画面を全面に表示する。ベースボード上には会議参加者が公開した電子付箋が表示される。

クライアントは、電子付箋を作成、または編集してベースボードに公開するために利用されるアプリケーションである。クライアントは、各会議参加者が持つスマートフォンやタブレット端末上で動作する。電子付箋には、付箋に入力したテキストや手書きメモ他に、メタ情報としてテキストのメモと任意の URL を入力して付加できる。

Mahoca では、会議参加者は、クライアントを利用して自分の意見を書いた電子付箋を作成する。作成した電子付箋をスクリーンで表示したベースボード上に公開することで、他の会議参加者に意見を共有する。ベースボード上の電子付箋で他者の意見を確認し、議論をすることで会議を進行できる。

本研究では、Mahoca における関連情報推薦機構を試作した。関連情報推薦機構では、ユーザの連想のヒントを提示することで、BS におけるアイデア出しを支援する。図 2 に関連情報推薦機構のシステム構成図を示す。関連情報推薦機構は、主に 4 つのモジュールからなる。関連情報推薦機構は Mahoca 上のインタフェースから、検索のもとにする情報を取得し、クエリ生成モジュールで検索クエリを作成する。生成された検索

連絡先: 伊藤 栄俊, 名古屋工業大学工学研究科情報工学専攻, 愛知県名古屋市昭和区御器所町, 052(735)5584, ashun@toralab.org

クエリから、単語モデルと、推薦語選択手法に基づいて関連語検索モジュールで単語を検索する。得られた単語の中から、情報提示モジュールにおいて提示する単語を選択する。選出された単語をもとに、提示方法決定モジュールでクライアントやスクリーンに関連情報を提示する。検索結果は付箋作業管理機構にフィードバックされる。

3. 関連情報推薦機構

BS 事例の再利用のためには、BS 状況の定量的な推定が必要である。システムが BS 中におけるコンテキストを取得することは難しい。そこで、本研究では、BS 状況を推定する方法として、関連情報を提示した際に、ユーザが選択した情報に注目し、ユーザが選択した関連情報をもとに、BS 状況を評価し、付箋状況と BS 状況から BS 事例として記録する。

ユーザが求める関連情報の傾向は BS 中のコンテキストに応じて変化する。例えば、ある話題に対して、議論を深めるため類似する事例を列挙していく場合、意味的距離が近い情報が求められる。一方で、議論が停滞し、話題を多方面に展開したい時、セレンディピティが高い情報が求められる。

関連情報として、画像や単語など様々な形式があるが、本研究において、提示する関連情報はテキスト情報に限定して述べる。電子付箋のテキスト情報から、単語を取り出す。単語の取り出しには、MeCab を用いて品詞が名詞である単語のみを抽出する。自然言語処理における関連する単語の取得では、TF-IDF など単語の共起頻度に基づく \cos 類似度の高い単語を推薦する手法が用いられる。しかし、BS などにおける発想支援を想定した場合、セレンディピティのような、意外性のある単語の推薦が好ましい。そこで、本研究では、類似度の高い単語を推薦するだけでなく、発想を広げるような意外性のある単語の推薦システムの開発を行った。

関連情報を検索するためには検索クエリを生成する必要がある。クエリ生成では、Mahoca における電子付箋やそのグループの情報を利用する。利用者が最適な関連情報の提示を可能にするために、以下の項目に対する選択肢を与える。

- アンカー 関連情報を検索する際のクエリに含めるキーワード
- 検索の対象 関連情報を検索する対象のドメイン
- 提示方法 検索して抽出した関連情報を提示する方法

アンカーは、関連情報を検索する際のクエリに含めるキーワードである。例えば、Mahoca 上における電子付箋や、そのグループの情報である。検索の対象は、関連情報を検索する対象のドメインである。関連情報として提示する検索エンジンや Web ページ、使用する文書の種類などを指定する。提示方法は、検索して抽出した関連情報を提示する方法である。関連情報として提示する情報の数や、画像や単語など提示する情報の種類の指定を行う。

3.1 関連語検索手法

検索クエリを定式化したものを示す。文書集合 D は、News コーパスや Wikipedia、Web ページなどの検索もとにする文書全体である。関連情報推薦機構では、文書 d に含まれる単語群 $t_1 t_2 \dots t_k \dots$ から選出する。検索クエリは、 $Q = \{q_A, q_1, q_2, \dots, q_m\}$ とする。 q_A をアンカーとする。 q_1, q_2, \dots, q_m はキーワードとし、例えば、アンカーをグループ名とした場合、グループに含まれるカード中の単語が相当する。検索の対

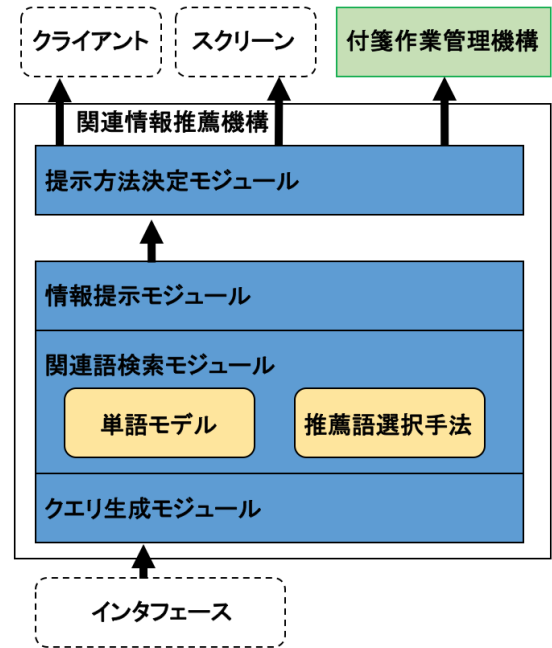


図 2: 関連情報推薦機構

象は文書集合 D を決める。提示方法は提示方法決定モジュールに送信する。

$$\text{文書集合 } D = \{d_1, d_2, \dots, d_n\}$$

$$d = t_1 t_2 \dots t_k \dots$$

$$\text{検索語集合 } Q = \{q_A, q_1, q_2, \dots, q_m\}$$

q_A : アンカー (グループ名)

$q_1 \sim q_m$: キーワード (グループ内のカード)

文書集合 D に対して、アンカーとキーワードをもとに検索する関数を $\text{search}(D, Q) = \text{search}(Q) = \{d_i | t_k \in Q, t_k \in d_i, d_i \in D\}$ と定義する。関連情報として選択する単語は $\text{search}(Q)$ に含まれる。類似度が高い単語を取得するための $\text{rank}(d, \text{key}, L)$ を定義する。 rank は key をもとに sort で降順に並び替えた単語から、 L 個を高スコアな単語として返す。

$$\text{search}(Q) = \text{search}(D, Q) = \{d_i | t_k \in Q, t_k \in d_i, d_i \in D\}$$

$$\text{rank}(d, \text{key}, L) = \text{limit}(\text{sort}(\{t | t \in d\}, \text{key}), L)$$

本研究における関連情報推薦機構では、2 種類の単語モデルと、4 種類の推薦語選択について検討し、試作した。単語モデルには TF-IDF モデル、日本語 Wikipedia エンティティベクトルを採用した。推薦語選択には、ランダム型選択、近傍型選択、連想型選択 (TF-IDF)、連想型選択 (EV) を提案する。

3.2 単語モデル

TF-IDF モデル

文書頻度 df を用いる手法である。まず検索クエリから関連する文書集合を取り出す。検索結果の文書集合の文章から名詞を抽出する。抽出した名詞の特定の文書集合内に対する逆文書頻度 idf と、アンカーに指定した単語との共起文献数 codf を利用して関連度 relation を算出する。計算に用いる文書集合として、2014 年 2 月から 2015 年 2 月までに発行された 651,508 件の Web ニュースを利用した。関連度の高い単語からランダムに単語を選択し、関連情報とする。

日本語 Wikipedia エンティティベクトル

学習済みコーパスとして、日本語 Wikipedia エンティティベクトル [Suzuki 16] を利用した。日本語 Wikipedia エンティティベクトル (以降 EV) は、Wikipedia をテキストデータに学習したモデルである。エンティティを生成する際に、単語の固有表現抽出に注目している。ある固有表現において、Wikipedia の記事上でハイパーリンクが貼られている単語を置き換えることで、固有表現と他の単語の区別を行う。

3.3 推薦語選択手法

本システムでは、BS における状況変化に対応した推薦を実現するために、ランダム型選択、近傍型選択、連想型選択 (TF-IDF)、連想型選択 (EV) の 4 つの推薦語選択手法を使い分けることを考えた。ランダム型探索は、アンカーをもとにランダムに単語を選択する手法である。近傍型探索は、アンカーとキーワードから近傍となる文書集合を取得し、取得した近傍文書集合内での出現頻度が重要度となる。連想型選択では、重要度が高い単語をもとにさらに単語を取得していき、近傍には存在していない意外性のある単語を見つける手法である。TF-IDF と EV で手法が異なる。

ランダム型選択

意外性のある単語を選択する方法として、単語の無作為抽出法がある。search(q_A) から、無作為に単語 $t \in search(Q)$ を N 個選出する。ランダムに抽出することにより、単純に類似度の高い単語を選出するよりも、意外性のある単語が期待される。

近傍型選択

近傍型選択について述べる。近傍型選択は、アイデアの漏れに対処するための選択手法である。search(q_A, q) を q_A, q の近傍と定義する。単語 t が多くの近傍に含まれている場合、アンカー q_A に対して関連度が高いと考えることができる。そこで、単語 w を含む近傍の数を典型度と定義する。典型度が高い単語を選択することで、ランダム型選択よりも、アンカーに関連する単語を選択することが期待できる。手順について述べる。検索クエリ $Q = \{q_A, q_1, q_2 \dots q_m\}$ から近傍 $d_i = search(q_A, q_i), q_i \in \{q_1, q_2 \dots q_m\}$ を求める。全近傍集合を D_{near} とする。キーワードに対応する各近傍 d_i から TF-IDF 値が高い単語を L 個取り出す。取りだされた全単語集合を \mathbb{T} と定義する。

$$D_{near} = \{d_i | d_i \in search(\{q_A, q_i\}), q_i \in \{q_1, q_2 \dots q_m\}\}$$

$$\mathbb{T} = \{T_i | T_i = \{t_k | t_k \in rank(d_i, TF-IDF, L)\}, d_i \in D_{near}\}$$

取りだされた単語 w に対して、典型度 score を求める。典型度 score は各近傍 $d_i \in D_{near}$ において出現した回数の総和となる。degree(t, d) は、文書 d 中に単語 t が登場した時に 1 を返し、登場しない場合は 0 を返す関数である。典型度をもとに高スコアの単語を関連情報として取り出す。

$$score(t) = \sum_{d_i \in D_{near}} degree(t, d_i)$$

$$C = rank(\mathbb{T}, score, num)$$

連想型選択 (TF-IDF)

連想型選択 (TF-IDF) とは、近傍 search($\{q_A, q_0\}$) を探索した際に取得した単語 $t_1 \in search(\{q_A, q_1\})$ から、search($\{q_A, t_1\}$) を探索する。すると、近傍の周辺から単語 $t'_1 \in search(q_A, t_1)$ が得られる。近傍 search($\{q_A, q_1\}$) から取得する単語 t_1 に、近傍周辺 search(q_A, t_1) から単語 t'_1 を候補として加えることで、近傍の探索だけで見つからない自明でない単語も探索することが可能となる。

連想型選択 (EV)

Word2vec を使った連想手法の研究に、稲垣ら [Inagaki 16] の研究がある。稲垣らの手法では、連想のもとになる刺激語に対して、直接関係する語だけでなく、連想された単語からさらに単語を連想することで、人間における発散的思考の模倣を試みた。

本研究においても、探索的に単語から単語を連想する手法である連想型選択を採用した。TF-IDF と EV の違いは、EV では、アンカーとキーワードを加算した結果に対して近傍を求め、連想時は、類似度の高い未発見の語だけを選ぶ点である。連想時は、アンカーやキーワードから意味的に少し離れた位置にある単語を取得したい。そのため、未知の単語のみを連想していくことで、できるだけ、意味的に離れた部分の探索を試みる。

Algorithm 1 連想型選択 (EV)

Require: C : 提示する関連情報, Num : 提示する単語数, $Depth$: 探索する深さ

Ensure: 検索語集合 Q に関連する単語 t を探索的に見つける

```
1:  $d \leftarrow 0$ 
2:  $t \leftarrow t_{root}$ 
3: while  $n < Num$  do
4:    $\mathbb{T} \leftarrow most\_similar\_unknown\_words(\{q_A, t\})$ 
5:    $t \leftarrow random\_word(\mathbb{T})$ 
6:    $d \leftarrow d + 1$ 
7:   if  $d < Depth$  then
8:      $C.append(t)$ 
9:      $\mathbb{T}.clear()$ 
10:     $n \leftarrow n + 1$ 
11:    break
12:   end if
13: end while
```

4. BS 状況の推定と再利用

本研究では、BS 事例の再利用のために、ユーザの選択した関連情報に注目した。Algorithm 2 に関連情報をもとに、BS 中の状態を推定するアルゴリズムを示す。

状態集合 \mathbb{S} 、単語集合 \mathbb{T} とする。状態 s は、BS 中において、求められるアイデアが発散的である状況や、収束である状況に対応させる。関連情報としてユーザに提示される単語群 \mathbb{T}_s は、状態 s は状態評価値 V によって遷移先を決定する。ユーザの選択した単語 t から、行動評価関数 $Q(t|s)$ を求める。行動評価関数 $Q(t|s)$ は、ユーザの行動によって状態を推定するための評価関数である。ユーザの行動は、BS 中の状況に応じて変化するため、選択された単語の価値は状態 s によって変化する。行動評価関数 $Q(t|s)$ から状態評価値 V を更新する。状態評価値 V から状態遷移を行い、状態 s' を決定する。状態 s に応じて、行動評価関数 $Q(t|s)$ を定めることで、状態評価値 V の増減を制御することができ、議論の停滞や話題の転換に対応する。決定した状態 s' から行動評価関数 $Q(t|s)$ を更新する。

現在の状態 $s \in \mathbb{S}$ と、ユーザの選択した関連情報 $t \in \mathbb{T}$ に応じて、状態遷移先 $s' \in \mathbb{S}$ を決める。関連情報 t は評価関数 $V(t|s)$ によって重み付けされる。評価値をもとに遷移先 s を決定する。

過去の事例を再利用することで議論を促すことがある。事

Algorithm 2 BS 状況の推定

Require: \mathcal{S} : 状態集合, \mathcal{T} : 単語集合, V : 状態評価値,
 Q : 行動価値関数

Ensure: 関連情報をもとに状態を推定

- 1: 状態 s を元に関連する単語群 $\mathcal{T}_s \subset \mathcal{T}$ を得る
- 2: ユーザによって単語群 \mathcal{T}_s から単語 t が選択される
- 3: 状態評価値 V を更新する
 $V' \leftarrow V + Q(t|s)$
- 4: V' をもとに状態 s' に遷移する
- 5: 状態 s' に応じて Q を更新する

例数が少ない状況で過去事例の再利用を行うためには、現在の状況に類似した状況で、類似した情報が存在していることが重要である。現在や過去の BS 状況の推定は、ユーザからの関連情報の選択に基づいた評価値から推定する。現在と過去のそれぞれにおける、付箋中の単語同士の類似度を計算することで、似通った事例を検索することができる。

5. 評価と考察

ランダム型選択、近傍型選択、連想型選択 (TF-IDF)、連想型選択 (EV) の 4 種類を行った場合の結果について表 1 にまとめた。表 1 に示されている単語は、“原子力発電”というアンカーに対して検索した際に得られた関連情報である。また、各選択手法の特徴を表 2 にまとめた。

表 1: “原子力発電”に対して各選択手法で得られた単語一覧

ランダム型選択	近傍型選択	連想型選択 (TF-IDF)	連想型選択 (EV)
情報通信	東京電力	江戸時代	点火プラグ
沖縄振興	原子	誘導體	北海油田
代替エネルギー	エネルギー	エネ	生活科学部
中層企業	関西電力	説得	電力会社
ミサイル防衛	シナリオ	感動ポルノ	Web デザイン

表 2: 各選択手法に対する評価

	類似度	重要度	自明度
ランダム型選択	低	分散	低
近傍型選択	高	典型度	高
連想型選択 (TF-IDF)	中	近傍	中
連想型選択 (EV)	低	類似度	低

単語モデルの違いについて述べる。TF-IDF モデルの場合、News コーパスを用いたため、関係しない単語が関連情報として推薦される傾向が他のモデルよりも強かった。これは、News コーパス中での出現頻度のみを考慮したモデルであることが原因であると考えられる。

日本語エンティティベクトルの有効性について述べる。アンカーの単語に対して類似度の高い単語のみを選出した場合、

意味的に近い単語や、部分一致している単語が推薦された。こうした単語はアイデアを広げるには不向きである。しかし、ランダム型選択など、類似度の低い単語が推薦されるような選択手法を適用した時、自明ではないが関係性があるといえる単語が候補として現れる頻度が高い傾向があった。固有表現に重きを置いている日本語エンティティベクトルは、類似度の高い単語の中にノイズが含まれていることが少なかった。ノイズとは、人が判断した場合に、関係性を見いだせないと判断される単語を意味する。

次に、推薦語選択手法の違いについて考察する。ランダム型選択は、文書集合 $d = search(\{q_A, q\})$ の大きさに依存していた。文章集合の大きさが大きいほど、アンカーと関係性がなく、無意味な単語が選ばれる頻度が大きくなる傾向があった。逆に、文章集合の大きさが小さいほど、類似度の高い単語が選ばれやすくなり、近傍型選択の結果に近づいた。今回は、一様分布に従う乱数をもとに単語を選出した。他の手法として、類似度 *similarity* を確率変数として用いる方法があげられる。

近傍型選択と連想型選択 (TF-IDF) は組み合わせで使用することが可能である。近傍型選択は、典型度に基づき、アンカーに関連した単語が選ばれやすい。連想型選択 (TF-IDF) は、近傍中の単語からランダムに単語を選択し、探索を行うため、近傍には存在しない、自明でない単語を選ぶことができる。議論事例の再利用のために、提示する関連情報を制御する際に有効な手法である。

連想型選択 (EV) は、深く探索するほど、アンカーとの関連度が低くなるため、ランダム型選択のような結果に近づく。また、同一単語を再探索しないように、探索済み単語を記憶するため、探索コストは他の探索に比べて大きくなる。しかし、探索のために選択する単語の選び方をランダムではなく、意味のある語を選択するように改良することで、ランダムよりも高い頻度で意外性のある単語を見つけることが期待される。

6. おわりに

本研究では、Mahoca における関連情報推薦機構の試作を行った。ユーザの関連情報の選択から BS 状況の推定を行うことで、過去の BS 事例の再利用を試みた。今後は、Word2vec などで得られる単語ベクトルを操作することで、提示する情報の制御に関して取り組んでいく。例えば、単語ベクトルの成分に対して、遺伝的アルゴリズムにおける交叉処理を行ったり、ある単語に対する異なるモデル上でのベクトル表現を組み合わせたりすることで、発想支援につながる結果が得られるかについて、実験と考察を行っていく。

参考文献

- [Ozono 16] 大園忠親, 丹羽佑輔, 藤江翔太郎, 渡邊正人, 鈴木智也, 伊藤栄俊, 岩田知, 新谷虎松: Mahoca: 議論事例の再利用を可能にする実世界型電子付箋システム. 平成 27 年度 電気情報通信学会 人工知能と知識処理研究会, Vol.116, No.350, pp.33-38 (2016)
- [Suzuki 16] 鈴木正敏, 松田耕史, 関根聡, 岡崎直観, 乾健太郎: Wikipedia 記事に対する拡張固有表現ラベルの多重付与. 言語処理学会第 22 回年次大会 (NLP2016), 4p, (2016)
- [Inagaki 16] 稲垣健吾, 山本英子, 神崎享子, 井佐原均: 人間の連想を模擬するシステムの開発と分析. 言語処理学会第 22 回年次大会 (NLP2016), 4p (2016)