Vol.2022-HCI-197 No.29
2022/3/15

ISeedS：単語ベクトルの変化に基づく意外な関係のある単語推薦システム

渡邉 一樹[†1,a)]　濱川 礼[†2]

概要：本稿は，ベクトル空間上にある単語の変化を用いた, 意外な関係のある単語の推薦と, 単語同士の関係を文章生成技術により説明することでアイデア創造の支援を行う ISeedS について述べる. 創作活動や研究テーマ決めなど, 創造的なアイデアが必要な機会は数多く存在する. しかし, それらを漠然と考えても創造的なアイデアが出ることはない. そのため, 集団で思考の発散を行うブレインストーミングなどの方法が提案されてきた. しかし, ブレインストーミングは「集団であるがゆえに時間の無駄となる」と結論付ける研究が数多く報告されている. そのため, 集団での相乗効果によらない思考の発散方法が必要である. そこで, 我々は意外性のある単語をユーザに推薦し, それら単語が使われた説明文の生成を行い利用者に提示することで, 集団による思考の発散に代わる発想支援システム ISeedS を提案する.

1. 背景

創作活動や研究開発, 研究テーマ決めであったりコンペへの応募作品立案など, 創造的なアイデアが必要な機会は数多く存在する. しかし, それらを漠然と考えても創造的なアイデアを出すことは困難であろう. そのため, 多くの方が個別具体的だったアイデア創造の過程を, 様々なアプローチからモデル化してきた.

1940 年に出版された, ジェームス・W・ヤング著のロングセラー本である「A Technique for Producing Ideas（邦題: アイデアのつくりかた)」[1] では, アイデアが生まれるまでの過程を五段階で分割し, モデル化した. これらの手順を踏むことで容易にアイデア創造することが可能となる.

■第一段階：資料集め

当面の課題の為の資料と, 一般的知識の貯蔵を絶えず豊富にすることから生まれる資料を集める段階.

■第二段階：関係性・組み立て

あなたの心の中でこれらの資料に手を加え, 関連性, 組み合わせを考える段階.

■第三段階：孵化（ふか）

意識の外で何かが自分で組み合わせの仕事をやるのにまかせ,「ぼーっ」と, 映画を見たり, とにかく, 少しの時間, 他のことをする段階.

■第四段階：アイデアの誕生！

アイデアの実際上の誕生（ユーレカ！わかった！見つけた！）と言う段階.

■第五段階：アイデアの具体化・現実化

現実の有用性に合致させるために最終的にアイデアを具体化し, 展開させる段階.

心理学者であるジョイ・P・ギルフォードが 1956 年に提唱した「知能構造論」[2] によると, アイデアの思考法には人間の思考にみられる 2 つの側面,「拡散的思考」と「収束的思考」によって構築されているとした. なかでも, 新しいアイデアや解決策を発想する考え方である「拡散的思考」の方が創造的思考であると重視されている.

■発散的思考

情報を元に様々な方向に考えを巡らせ, まだ存在しない新しいアイディアを生み出す思考方法

■収束的思考

すでにある情報を元に,1 つの正しい答えへ辿り着く思考方法

それらアイデア創造の過程・思考法を, 実問題に使えるようフレームワーク化したものとして, 収束的思考を支援する DEMATEL 法や ISM 法, 発散的思考を支援するマンダラート [3] やマインドマップ [4] など, アイデア発想法が数多く提案されてきた. その中でも特に有名な発想法として, ブレインストーミング [5] というものがある.

ブレインストーミングとは,1953 年にアレックス・F・オズボーンによって考案されたアイデア発想法であり, 集団発想法とも言われる. ブレインストーミングの特徴として,

†1 現在，中京大学工学研究科情報工学専攻
Presently with Chukyo University
†2 現在，中京大学工学部情報工学科
Presently with Chukyo University
a) kazuki.wata2012@gmail.com

© 2022 Information Processing Society of Japan

集団でアイデアを出し合うという点がある. これによって相互交錯の連鎖反応や発想の誘発に期待出来る. 人数に制限はなく,5 名から 7 名, 場合によっては 10 名程度が好ましいというやり方もある.

しかしながら, 近年の研究で「ブレインストーミングは時間の無駄である」と結論付ける研究が数多く挙げられてきた. Furnham ら [6] や,Paul ら [7] によると, 個人でアイデア出しを行った場合の方が, グループで行った場合より,30〜40 %多くのアイデアを出し, クオリティ面でも個人で行った場合の方が高くなった. さらに,5 人,7 人,9 人といった様々なグループ規模を変えた実験を行ったが, グループ人数が増えるほど, アイデアの生産性が落ちていった. これらの原因として, グループでアイデア出しを行うと, 誰かが話している間その人の話を傾聴するため, 聞き手は思考が妨げられることになる, 思いついたことを忘れてしまうといったことや, 突飛な意見が受け入れられづらいといったグループ環境が原因ではないかと考察されている.

また, アイデア出しに伴う社会的問題も存在する. 株式会社 TIS による, 企画業務に携わる男女計 1,032 名に対しての 2019 年度のアンケート調査結果では, 年々求められるアイデアの量・質ともに高くなっているという実態がある [8]. また, 働き方改革による労働時間制限により, 業務時間外に一人でアイデア出しを実施している担当者が多く存在するという問題も報告されている.

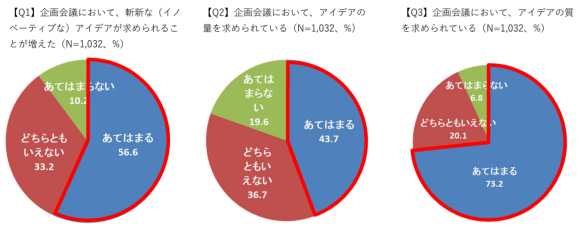


図 1 株式会社 TIS による調査結果 [8]

これらのことから, 集団での相乗効果によらない発散的思考を刺激するような仕組みや, 効率的なアイデア出しが行える発想法が必要であるといえる.

そこで, 我々はランダムな単語による刺激に着目した. ランダムな単語による刺激とは, 文字通りランダムに単語を推薦し, 利用者は推薦された単語と核となるテーマとを強制的に関連付け思考することで, 発散的思考を刺激する方法である. この, ランダムな単語を推薦する部分が, 一人では思いもつかないような概念を与えてくれるため, 集団での相乗効果によらない発散的思考の刺激を行うことが出来る. ランダムな単語を提供するシステムは数多く存在し,WEB アプリやスマホアプリのような形態で提供されている [9][10][11]. しかし, 提供される単語はランダム性がゆえに, 利用者は核となるテーマと単語との関連性を見つけることは難しく, ほとんどが新しいアイデアに昇華する事無く, 切り捨てられてしまう. そのため, 我々はこのランダム単語推薦に, 利用者が刺激を受けるような意外性のある単語かつ, 利用者が核となるテーマと単語が結びつけやすいような何かしらの部分で関連性がある単語を推薦し推薦するシステムと, テーマと推薦された単語を用いた文章を生成することで, 利用者の関連性理解への補助を行うシステムからなる ISeedS を開発した.ISeedS とは「Idea Seed Spreader」の略である.

2. 関連研究・関連システム

2.1 アイデア創造支援

アイデア創造に関する研究として, 発想方法に工夫を入れた研究がある.

植野らは [12],「Y（喩辞）のような X（主題）」という比喩表現を用いた支援の方法を提案した. 主題を先に決めてから喩辞を埋めるとアイデアの意外性が向上した. また, 喩辞を先に決めてから主題を埋めると, アイデア生成の容易さが向上した.

梅村らは [13], アイデア連想のつながりを記録することで, 他者のアイデア連想を参考にしたアイデア発想や, アイデア連想の振り返りにより支援を行う方法を提案した.

小野寺らは [14] 時間経過によるアイデアの質や量が低下する問題に対し, アイデア生成の対象となる主テーマを, 関連するいくつかのテーマに分割し, 短時間で切り替えながらアイデア生成を行う TKTS 法を提案した.

趙らは [15], 子供は大人が思いつかないようなアイデアを思いつくと仮定し, 子供の発想を専門知識を有する大人に提供することで, これを参照しながら実用的なアイデア生成を行う手法を提案した.

北山らは [16], 過去のエピソードを思い出すことによる刺激という観点に着目し, アイデア出しのテーマに沿うような過去のエピソードを,「たねでぃ」と呼ばれる音声認識デバイスを用いて,4W1H 形式（「いつ行ったのか」「どこで行ったのか」「誰と行ったのか」「なぜ行ったのか」「どのように行ったのか」）で聞き出し整理することにより, アイデア創造の支援を行う手法を提案した.

川上は [17], 不便さならではの効用,「不便益」という観点に着目し, アイデアを「便利軸」「益軸」の 2 軸 4 象限に分類することで支援を行った.

小山らは [18], ブレインストーミングの補助を目的に, 多人数で同時書き込みができる共有可能なキャンパスと word2vec を用いた類似単語や文章の推薦により, アイデアの発散を促すシステムを提案した.

ISeedS は, 単語を自動で生成し提示することができるため, 従来手法に比べ短時間で多くのアイデア発想の支援を行うことができる. また, 多くの研究がアイデアと単語との関連性に着目しているが,ISeedS は意外性に着目してい

© 2022 Information Processing Society of Japan

るという点で異なる.

2.2 意外性を考慮したアイデア創造支援

伊藤らは [19], 関連性はあるが共起度が低い単語がアイデア創造の刺激になると仮定し, 設定されたテーマに対してあらかじめ文章収集とクラスタリングを行い, テーマと単語の共起度が低いものが一番多く含まれているクラスタから単語を得ることで, アイデア創造の支援を行う手法を提案した.

Kai らは [20], 文章間のリンク構造に着目し,wikipedia のページから 2 回リンクをたどったページが意外性のある概念だと仮定し, そこに含まれている単語や文章を提供することで, アイデア発想の支援を行うシステムを構築した.

Han らは [21], 収集したアイデアに対し, 関連するような単語で繋げ合わせることと, アイデアに関係のある画像どうしを合成や結合を行い, ユーザに提示することにより発想支援を行うシステムを提案した.

これらの研究では, あらかじめデータの収集や分析が必要であるため, 特定のドメインに特化してしまっている. また, 単語同士の組み合わせ方に対し必要な支援がされていない.ISeedS では, テーマごとにデータの収集は必要なく, 特定のドメインだけでなくあらゆるテーマでアイデア支援を行うことができる. また, 推薦された単語が使われた文章を生成することで, 推薦された単語とテーマとの組合わせ方に対しても支援を行うことが可能である.

2.3 提案システムの位置づけ

西本ら [22] によると, アイデア創造において発散的思考過程で, いかに思考の幅を広げ, 意外な情報や知識を獲得できるかがが重要だと述べている. また, 高橋 [23] によると, 発散的思考の中でも, 流暢性（数の多さ）, 柔軟性（多様さ）, 独自性（ユニークさ）の三基準が重要になってくるとしている. そこで, 本節では発散的思考で重要な役割を持つ流暢性, 柔軟性, 独自性という観点と, 単語やアイデア同士の関係性理解の支援について, 意外性を考慮した関連研究との比較を行う. 以下の表 1 に, 関連研究との位置づけを示す.

	流暢性	柔軟性	独自性	関係性理解の支援
伊藤ら [19]	△	△	○	×
Kai ら [20]	△	△	○	×
Han ら [21]	△	△	×	△
ISeedS	○	○	○	○

表 1 ISeedS との位置づけ

3. システム概要

ISeedS は, 以下の図 2 のシステム構成にあるように, 意外性のある単語推薦部と, 推薦された単語を用いた文章生成部からなる. ユーザはアイデア出しのテーマに沿うような単語を ISeedS に入力することにより, 意外性のある単語と, それら二つの単語を用いた文章が生成される. それらを基にし, ユーザはアイデア出しを行う. 以下では, 意外性のある単語推薦部と推薦された単語を用いた文章生成部につい述べる.

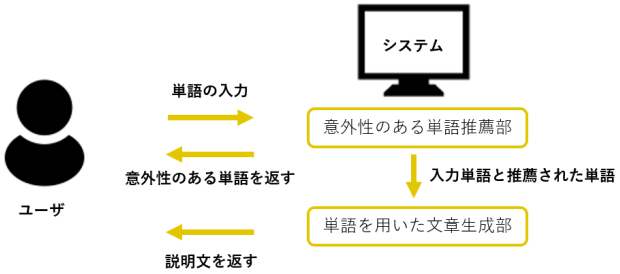


図 2 ISeedS のシステム構成

3.1 意外性のある単語推薦部

3.1.1 手法

先行研究の手法にもあるように, 単語同士の共起度を下げる方法であったり, Wikipedia のような, 一般的な概念がまとめ上げられた百科事典のみから取得する方法では, 関連性がないようなノイズが数多く混ざってしまうため独自性が薄くなってしまったり, 柔軟性に必要なニッチな使われ方の単語は得ることが出来ないと考えられる.

そこで,ISeedS では SNS の投稿やブログなどの個人が各々執筆した文章に着目した. Wikipedia の文章と比べ,SNS の投稿やブログなどの文章には型にはまらない自由な形式なため, 型にはまらないような単語が含まれていたり, 分野に対しての詳しい知識も存在する. そのため,SNS の投稿やブログなどの尖った文章から,Wikipedia のような一般的な概念を無くすことで, 意外性のあるかつ, 何かしらで関連性を持つ単語を検出することができるのではないかと考えた. 以下の図 3 を用いて具体例の説明を行う. 左は Wikipedia で仮想現実 (VR) について説明した文章の一部であり, 右は Twitter から「VR」でツイートを検索した結果である. それらは同じ概念について述べられているが, 出現する単語に違いがみられる.Wikipedia 側では「感覚」「刺激」「テレビ」など,「VR」の特徴や概念を表すような単語が多いのに対し,Twitter の文章は,「パイロット体験」「睡眠」「歯磨き」など,「VR」と関連するが, 初学者がすぐには思いつかないような単語が多く含まれている. それらの文章中に出現する単語の違いや, 使われ方の特徴を活用することで, 関連性は薄いが無関係ではない意外な単語を獲得する.

3.1.2 実装

3.1.2.1 データ収集

前節では,SNS の投稿やブログとしていた. 本論文では Twitter からアカウントごとに取得したツイートで仮説の

© 2022 Information Processing Society of Japan

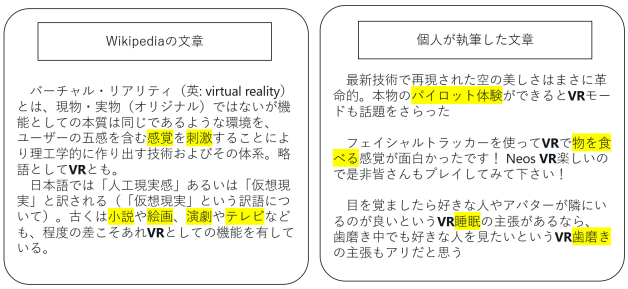


図 3 Wikipedia と Twitter で出現する単語の違い

検証を行う．これは,Twitter はブログ等の文章と違い字数の制限が設けられているため, 要点が要約されているため, 後述する手法と相性が良いためである．また, 取得するツイートが様々な分野を網羅させるため,WordNet と呼ばれる単語同士の概念関係をを木構造で表した辞書を用い, 様々な単語で検索を行った．本論文では WordNet からランダムに抽出した単語 1000 個を使用し, アカウントの検索を行った．また, ツイート取得の条件として, ナノインフルエンサーと呼ばれるフォロワー 1000 人以上のユーザかつ, 平均いいね数が 30 以上のアカウントからリツイートを除いた全ツイートの取得を行った．これは, あまりツイートしないユーザや Bot を排除するためである．以下の表 2 に, 取得したアカウントの数とツイート数を載せる．

アカウント数	3,245
総ツイート数	2,465,341

表 2 取得したアカウント数と総ツイート数

また, 一般的な概念や単語獲得には, 日本語 Wikipedia 全文を利用した．

3.1.2.2 前処理

Twitter の文章には,URL や絵文字, 顔文字など, 不要なものが多く含まれているため, それら該当する部分の削除を行った．また, 収集した文章には, 単語の関係を構築するうえで, 接続詞や助動詞など不要なものが存在する．そこで, 品詞分類や文章の分かち書きに Sudachi[24] を用いた．

Sudachi とは, 徳島人工知能 NLP 研究所が開発した OSS 形態素解析器であり, 日本語の分かち書きと品詞分類を行うことができる．また, 未知語や打ち間違えられた単語に対しての正則化も行うことが出来る．さらに,A〜C の辞書の種類によって分かち書きの大きさを変えられるため, タスクに応じて適切な分割を選ぶことが可能となる．本論文では, 様々な単語で学習を行うため, 固有名詞などでも分かち書きを行うことが出来る C モードを用いた. さらに, 学習データには「名詞」「動詞」「形容詞」を使うことにする．

3.1.2.3 単語関係学習手法

単語同士の関係を学習する方法として単語分散表現を用いた．単語の分散表現とは,1 つの単語を数 100 次元程度の固定長のベクトルで表す方法であり, 単語の意味を捉えているかのような性質をベクトル空間上で表現することができる．本論文では単語ベクトルの獲得方法に word2vec[25] を用いた．word2vec には,CBOW と skip-gram の二つの学習アルゴリズムがある. 本論文では skip-gram を用いる.skip-gram は, コンテクストの数だけ損失を求める必要があるため学習コストが大きく時間がかかるといった欠点を持つが, コーパスが大規模になるにつれ, 頻出の単語や類推問題の性能の点において優れた結果を得ることができる．

以下に,skip-gram モデルのイメージ図 4 を載せる．skip-gram は, 中央の単語から前後の複数のコンテクストを予測するモデルであり, 入力層はひとつで出力層はコンテクストの数だけ存在する. そして, それぞれの出力層で個別に損失を求め, それらを足し合わせたものを最終的な損失とする. 本論文では, パラメータの設定として, 次元数を 300, ウィンドウサイズを 5, 単語の最小出現回数を 2 とした．

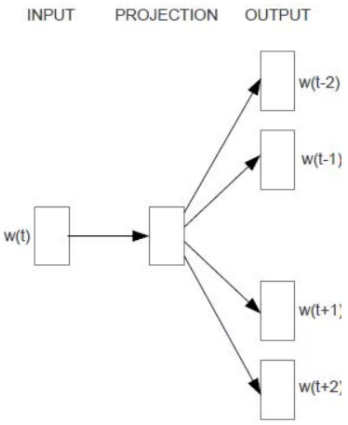


図 4 skip-gram モデルイメージ

最後に,word2vec の学習について述べる．今回は,Wikipedia 文章と Twitter 文章で単語分散表現の変化度合いを測るため,2 つのモデルの学習を行う．1 つ目のモデルは,Wikipedia の文章を word2vec で学習させたものである．2 つ目のモデルは,1 つ目のモデルを Twitter の文章で転移学習したものである．転移学習した理由として, 最初から Twitter 文章で学習してしまうと, 文章の学習順番や乱数の影響で word2vec によって得られる分散表現が 1 つ目のモデルとかなり異なってしまい, 単語間の変化の度合いを正確に測れない可能性があったためである．

3.1.2.4 意外性のある単語の算出

前節までに述べた方法で学習した二つのモデルそれぞれで, 単語類似度算出を利用して単語同士の類似度を求める．その後, 二つのモデルから得られた単語同士の類似度の変化量を求め,1 つ目のモデルと比べ 2 つ目のモデルでの類似

© 2022 Information Processing Society of Japan

度が上昇したものを意外性のある単語とする.

3.2 単語を用いた文章生成部

3.2.1 手法

深層学習を用いた文章生成手法の有名なものとして,LSTM[26] や GPT-2[27],Transformer[28] といったものがある. それらの特徴として, 入力された文章の次に続くような単語を予測し, それを逐次的に行うことで文章の生成を行っている. しかしながら, それらの手法では二つの単語のみを使って, 文章を生成することは難しい. 例えば, 単語 A を入力とし文章を生成すると, 単語 B が出現するまで単語予測を何度もやり直しを行い強引に文章を生成することになる. また,「単語 A と単語 B の関係性は,」や「単語 A・単語 B について」など, 単語 A と単語 B それぞれをつなげ, 文章として入力した場合, あらかじめテンプレートを用意しなければならないため, 生成される文章の自由度が下がる.

そこで, 本論文では RoBERTa[29] と呼ばれるモデルを基に, 文章生成用に転移学習を行うことによりそれらの問題の解決を図る. RoBERTa では, 与えられた文章に対し, ランダムに MASK と呼ばれるトークンに置き換え, それらを求めるような学習を行う. それにより, 文脈の意味や文章構造といったものを学習できるため, 文章生成の安定性や向上が期待できる.

3.2.2 実装

3.2.2.1 学習データの成型

二つの単語から文章を生成出来るようにするために, 学習データを文章生成用に成型する必要がある. 本論文では, 日本語 Wikipedia の文章から学習用のデータセットを構築した. データセットの構築方法として,Wikipedia の文章から, 一文ごとに抜き出し,Sudachi を用い品詞分類を行い,「名詞」「動詞」「形容詞」に該当する単語を抽出する. 抽出した単語から二つ選んだものを入力単語, 抽出元の文を出力文としてテーブルに格納しデータセットの構築を行った. 以下の表 3 に構築したデータセットのテーブル例を示す.

入力単語	出力文
単語 A, 単語 B	Wikipedia 文 1
単語 A, 単語 C	Wikipedia 文 1
… , …	…
単語 X, 単語 Y	Wikipedia 文 2

表 3 データセットのテーブル形式

3.2.2.2 転移学習

前節で構築したデータセットを用い,RoBERTa で文章生成を行えるように転移学習を行う. 本論文では,RoBERTa の日本語モデルとして,rinna 社の「japanese-roberta-base」モデル [30] を利用し, 転移学習を行った. 層への入力は 32 トークン, 出力は 128 トークンになるよう設計し, トークン長に満たないデータはゼロパディングで穴埋めを行った.

4. アプリケーション

4.1 モデルの軽量化

前章で述べた方法により学習を行った word2vec モデルは, 一つ当たり 2.74Gbyte とかなり大容量になってしまう. また, そのモデルを DRAM 上に展開すると, 一つ当たり 8Gbyte ほど使用してしまうため, 実用的ではない. さらに, モデル上に存在しないような単語を入力してしまうとエラーになってしまう.

そこで, モデルの軽量化と単語の正則化に Magnitude 化を行った. Magnitude とは [31],Ajay らにより提案された手法で,SQLite を用いることにより初期の呼び出し時には等速なものの, 二回目の呼び出し以降の高速化を実現した. また,SQLite が用いられているため,Magnitude ファイル自体は word2vec のときと比べて二倍になるものの,DRAM 上にモデルを展開する必要がなくなる. さらに,Magnitude には未知語への対応として,「未知語の文字 n-gram」と「文字列として似ている既知語のベクトル」をもとに, その場で作成されるため, どのような単語が入力されても値を返すことが出来る.

4.2 ワードクラウド化

ISeedS では, ワードクラウドを用いて単語同士の関連性の可視化を行った. ワードクラウド [32] とは, 頻出語を頻度に比例する大きさで雲のように並べることが出来るツールである. ISeedS では, 単語は一度しか出現しないので, 頻出度の代わりとして類似度からワードクラウドの生成を行った. ワードクラウドにより可視化したものを以下の図 5 に載せる.ISeedS へ入力する単語は「VR」とした.

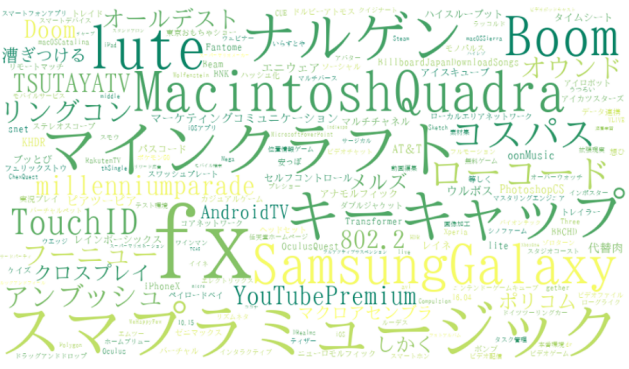


図 5 ISeedS での結果をワードクラウド化したもの

また,Wikipedia モデルと Twitter モデルで生成した結果も以下の図 6, 図 7 に載せる.

図 5 と, 図 6, 図 7 からも分かるように, 類似度順に表示したものでは「BeatSaber」「ビデオアプリ」,「NUKESTUDIO」といった,「VR」に関連がある単語が表示されてい

© 2022 Information Processing Society of Japan

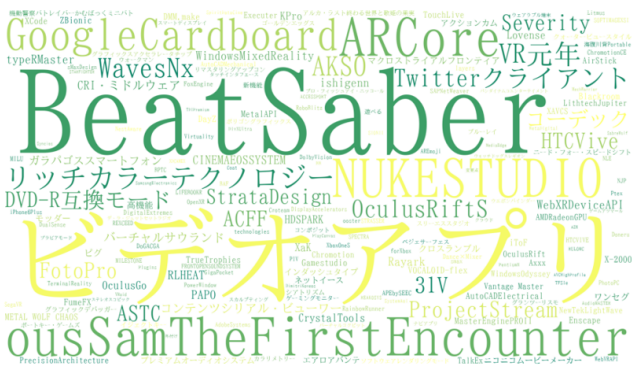


図 6 Wikipedia モデルをワードクラウド化したもの



図 7 Twitter モデルをワードクラウド化したもの

るのに対し,ISeedS では「マインクラフト」,「キーキャップ」,「fx」といった VR には直接の関係はないが, 何かしらの概念で関係を持つ単語が表示されている. 例えば,「VR」と「マインクラフト」では, ゲームといった概念での関連が思いつくだろう. また,「VR」と「キーキャップ」では, デバイスといった概念で関係を持っている.

5. 評価・考察

ISeedS の有効性を検証するため,「意外性のあるものなのかどうか」と生成された文章は「提案された単語ペアの関連性を説明できているか」という点について評価を行う. 以下では, 評価実験の内容と結果および考察について述べる.

5.1 評価手法

6 名の被験者に, それぞれ発想したいテーマを決めてもらい,ISeedS を実際に使用し意外性のある単語やアイデア文の生成を行ってもらった. その後, 精度面や, アイデア発想に参考になるかについて, 五段階でアンケートに回答してもらった.

また, 今回は本手法との比較用に, 入力された単語に対してランダムな単語を返す「ランダムワード生成」と, 入力された単語に対してる意義単語を返す「類義語生成」の二つを実際に使ってもらい, 本手法との比較も行ってもらった. 実験の流れは以下のとおりである.

(1) 出したいアイデアに沿ったテーマを決める
(2) ランダムワード生成とアイデア文生成でアイデア創造を行う
(3) 類義語生成とアイデア文生成でアイデア創造を行う
(4) 本手法とアイデア文生成でアイデア創造を行う
(5) アンケートに回答してもらう

また, アンケートの評価項目は以下の通りである.

- ISeedS は意外性のある単語が出てきたか
- ISeedS はアイデア出しの役に立つと思うか
- アイデア文生成は刺激になったか
- ランダムワード生成と ISeedS を比べ, どちらが刺激を受けたか
- 類義語生成と ISeedS を比べ, どちらが刺激を受けたか

被験者 6 名が決めたテーマを以下の表 4 に示す.

	テーマ
被験者 A	料理
被験者 B	就職活動
被験者 C	車
被験者 D	ゲーム
被験者 E	漫画
被験者 F	漫画

表 4 被験者それぞれが決めたアイデア発想テーマ

5.2 評価結果と考察

本節では, 被験者 6 名へのアンケート結果について述べる.

5.2.1 ISeedS の評価

ISeedS はアイデア創造の刺激となったか, 被験者 6 名のアンケート結果を表 5 に示す.

(1) ISeedS は意外性のある単語が出てきたか
(2) ISeedS はアイデア出しの役に立つと思うか
(3) アイデア文生成は刺激になったか

	平均	中央値	最大値	最小値
1	3.50	3.50	5.00	2.00
2	3.17	3.00	4.00	2.00
3	2.00	2.00	3.00	1.00

表 5 アンケート結果 1

評価項目 1

項目 1 については, 平均 3.5 と部分的に評価を得ることが出来た. 意見としては,「知らない単語がたくさん表示されて刺激を受けた」や「単語同士を組み合わせて考えるのが楽しかった」などの肯定的な意見が多かったが,「世界の料理がいろいろ表示されるだけで刺激を受けなかった」といった, 否定的な意見も寄せられた. これは, 被験者が欲しがっているアイデアと, システムによって提供された単語にギャップが存在したことが

© 2022 Information Processing Society of Japan

原因と考えられる. 現在は, 単語一語を利用者が入力しなければならない状態なので, 利用者が何を求めているのかをシステム側で補って行くような仕組みを導入する必要がある.

評価項目 2

項目 2 については, 平均 3.17 と高評価を得ることが出来なかった. 意見としては,「画像に大小いろいろな大きさでまとめられていてよかった」といった肯定意見を頂けたが,「画像の文字が小さくて読めないことがある」や「単語の意味がよく分からない」といった否定的な意見もあった. これは, 可視化手段が wordcloud 化と拡張性がないことが鯨飲だと考えられる. そのため, 利用者が自由に大きさなどを変えられるような GUI を実装する必要がある.

評価項目 3

項目 3 については, 平均 2 とかなり低い結果となり,「日本語としておかしい文ばかり表示される」や「名詞ばっかりでよくわかんない」など, 否定的な意見が多く寄せられた. これは, 文章生成部分がうまくいってないことが原因だと考えられる. 学習方法を変えたりなど, 根本的に見直す必要がある.

5.2.2 従来手法との比較評価

(1) ランダムワード生成と ISeedS を比べ, どちらが刺激を受けたか

(2) 類義語生成と ISeedS を比べ, どちらが刺激を受けたか

被験者 6 名のアンケート結果を表 6 に示す.

	刺激を受けた	刺激を受けなかった
4	4	2
5	6	0

表 6 アンケート結果 2

評価項目 4

項目 4 については,4 人の方に ISeedS の方が刺激を受けるとの回答を頂いた. ランダムワードを選んだ人の意見には,「たまに面白いのが出てくる」と, ランダムワードの方が意外性があるという認識を持たれてしまった. 意外性のある単語でも, 人によっては類語と判断されることもあるため, 利用者によって意外性の指標を変更できるような仕組みを整えたい.

評価項目 5

項目 5 については, すべての人に ISeedS の方が刺激を受けたとの回答を頂いた. 意見としては,「類義語システムは知っているようなものばかりで刺激を受けない」など,「類義語システムよりはましかな」, というニュアンスの意見がいくつかあった. これは, 項目 1 や項目 2 で考察したように, 利用者への伝え方の部分が問題となっている可能性がある.GUI にも力を入れていく.

6. 結論・今後の展望

我々は, ユーザが刺激を受けるような意外性のある単語かつ, テーマと単語が結びつけやすいような何かしらの部分で関連性がある単語を推薦し推薦するシステムと, テーマと推薦された単語を用いた文章を生成することで, 利用者の関連性理解への補助を行うシステムからなる ISeedS を開発した.

評価実験によって,「意外性のある単語推薦」部分では, 既存手法であるランダム単語推薦や類語推薦と比べ評価を得ることが出来たが. 「単語を用いた文章を生成」部分では, 生成される文章の正しさの部分で評価を得ることが出来なかった.

今後はよりアイデア創造の支援を向上させるため, 以下のような課題の解決を目指し,ISeedS を改良していく.

類似度算出単語数の見直し

ISeedS では,2 つの単語推薦モデルで上位 100,000 単語の類似度を取得し, 変化の度合いを求めた. しかし, 今回用意した word2vec モデルの語彙数は 140 万以上あるため, もっと深いところにはより大きな変化度合いを持つ単語があるかもしれない. しかし,100,000 単語でも 30 秒ほどかかるため, すべての単語を網羅するにはかなりの時間がかかると予測される. そのため, より早く比較できる手法を考えるか, すべての単語を予め計算しておくことにより, 解決できるのではないかと期待する.

単語のカテゴリー分類

現在, ワードクラウド上に表示される単語は分野問わず混ざってしまっている. しかし, 利用者によっては単語を特定のカテゴリーだけで表示してほしいと思うこともあるだろう. そこで, システムによって提案された単語群に対し, さらにカテゴリー分類をすることで利便性が向上するのではないかと考える. カテゴリー構築の手法としては,goo の辞書からカテゴリーを抽出することで一部の単語カテゴリーは直ぐにでも得られるだろう. しかし, 今回の word2vec の語彙数は 140 万を超えているため, すべての単語に対してカテゴリーがあるとは思えない. そのため, 深層学習によるカテゴリー分類を実装する必要がある.

文章生成による単語関係の説明の改善

アンケート結果より, 本手法による単語関係の説明では不十分であることが分かった. これは RoBERTa の転移学習ができていないことが原因であると考えられる. 損失関数の見直しや, 生成された文章に対し日本語校正を行うことで説明部分を改善できるのではないかと期待する.

文章生成以外による単語関係の説明

関連研究にもある通り, 画像生成による刺激も検討すべ

© 2022 Information Processing Society of Japan

き課題である. 従来手法では,Web で収集した写真に対しカテゴリー分類とタグ付けを行い, 入力された単語と近いタグがある写真を表示することで刺激を与えてた. しかしながら, 写真に写っているものに意外性を求める必要があるため, 収集した写真によってはよい刺激を与えることが出来ないこともある. そこで,CLIP などの, 文章から画像を新しく生成するモデルを応用することで, よりユーザに刺激を与えられるのではないかと期待する.

参考文献

[1] 今井茂雄 Young. アイデアのつくりかた. 東京 : 阪急コミュニケーションズ, 1988.
[2] J. P Guilford. The structure of intellect. *Psychological Bulletin*, Vol. 53(4), , 1956.
[3] 加藤昌治. 考具. CCC メディアハウス, 2003.
[4] 神田 昌典トニー・ブザン. ザ・マインドマップ. ダイヤモンド社, 2005.
[5] A.F Osborn. *Applied imagination*. Scibners, 1953.
[6] Adrian Furnham. The brainstorming myth. *Business Strategy Review*, Vol. 11, pp. 21–28, 2000.
[7] Vincent R. Brown and Paul B. Paulus. Making group brainstorming more effective: Recommendations from an associative memory perspective. *Current Directions in Psychological Science*, Vol. 11, pp. 208 – 212, 2002.
[8] 企画職に聞く、アイデア出しに関する意識・実態調査. https://www.tis.co.jp/documents/jp/news/2019/tis_news/20190603_1_Attachment.pdf.
[9] 言葉をランダムに組み合わせ | gacha-n-co. https://gacha-n-co.com/.
[10] 日常系お題単語ガチャ. https://odaibako.net/gacha/3162.
[11] ランダム単語ガチャ. https://tango-gacha.com/.
[12] 西本一志植野涼. 比喩生成における喩辞先発想の効果とその発想支援への応用可能性. 情報処理学会研究報告. GN, pp. 1–5, 2020.
[13] 雄貴梅村, 隆也由井薗. アイデア連想のつながりを明示化するブレインライティング支援システムの提案. 第 80 回全国大会講演論文集, 第 2018 巻, pp. 53–54, mar 2018.
[14] 西本一志小野寺貴俊. Tkts 法：時間経過によるアイデア生産量の低下を生じない新たな発想技法の提案とその有効性検証. インタラクション 2019 論文集, pp. 403–408, 2019.
[15] 西本一志趙暁. 「子供の発想」を利用するアイデア生成技法の提案とその有効性の検証. 情報処理学会研究報告. GN, pp. 1–6, 2018.
[16] 市野昌宏 市川大佑北村章真. たねでぃ：抽象的なテーマを元に具体的なエピソードを想起させるアイデア創出支援ツール. エンタテインメントコンピューティングシンポジウム 2019 論文集, 第 2019 巻, pp. 379–382, sep 2019.
[17] 川上浩司. ユーザに不便の効用を与えるシステムの発想支援. 計測と制御 60 巻 12 号, pp. 840–843, 2021.
[18] 荻野 正小山 琢也. word2vec を用いたブレインストーミングシステムの実装. FIT2018（第 17 回情報科学技術フォーラム）, 2018.
[19] 淳子伊藤, 孝行東, 純宗森. 単語共起度の低い単語を提示する発想支援システムの提案と適用. 情報処理学会論文誌, Vol. 56, No. 6, pp. 1528–1540, jun 2015.
[20] Kai Wang and Jeffrey V. Nickerson. A wikipedia-based method to support creative idea generation: The role of stimulus relatedness. *Journal of Management Information Systems*, Vol. 36, No. 4, pp. 1284–1312, 2019.
[21] Ji Han, Feng Shi, Liuqing Chen, and Peter R. N. Childs. The combinator – a computer-based tool for creative idea generation based on a simulation approach. *Design Science*, Vol. 4, p. e11, 2018.
[22] 一志西本, 研二望月, 文郎岸野. 発想支援システムに関する一考察. 全国大会講演論文集, 第 46 回, システム, pp. 283–284, mar 1993.
[23] 高橋誠. 新編創造力事典: 日本人の創造力を開発する創造技法主要 88 技法を全網羅! 日科技連出版社, 2002.
[24] 久本空海 岡一馬 内田佳孝坂本美保. 形態素解析器『sudachi』のための大規模辞書開発. 言語資源活用ワークショップ 2018, pp. 118–129, 2018.
[25] Greg Corrado Jeffrey Dean Tomas Mikolov, Kai Chen. Efficient estimation of word representations in vector space. *Computation and Language*, 2013.
[26] 80290 München Germany Fakultät für Informatik, Technische Universität München. Long short-term memory. *Neural Computation*, 1997.
[27] Rewon Child David Luan Dario Amodei Ilya Sutskever Alec Radford, Jeffrey Wu. Language models are unsupervised multitask learners. 2019.
[28] Niki Parmar Jakob Uszkoreit Llion Jones Aidan N. Gomez Łukasz Kaiser Illia Polosukhin Ashish Vaswani, Noam Shazeer. Attention is all you need. *Neural Information Processing Systems*, 2017.
[29] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, Vol. abs/1907.11692, , 2019.
[30] rinna/japanese-roberta-base. https://huggingface.co/rinna/japanese-roberta-base.
[31] Ajay Patel, Alexander Sands, Chris Callison-Burch, and Marianna Apidianaki. Magnitude: A fast, efficient universal vector embedding utility package. *CoRR*, Vol. abs/1810.11190, , 2018.
[32] wordcloud. http://amueller.github.io/word$_c$*loud*/.

© 2022 Information Processing Society of Japan