

自動獲得と集合知の併用による関連語知識の高度化と評価

町田 雄一郎[†] 河原 大輔[†] 黒橋 禎夫[†] 颯々野 学[‡]

[†] 京都大学大学院情報学研究所 [‡] ヤフー株式会社

[†] {machida, dk, kuro}@nlp.ist.i.kyoto-u.ac.jp [‡] msassano@yahoo-corp.jp

1 はじめに

近年、語と語の関係を扱った知識が様々な言語処理タスクに利用されており、質の高い関連語知識を獲得・整理することが重要である。単語間の関連を整理する手法はいくつか提案されているが、次のような問題が存在する。人手によって獲得・整理する場合は、品質は高いが、カバレッジが低く、時間と費用によるコストが高くなってしまふことが問題である。自動的に獲得する場合は、大規模に行うことができる反面、人手による手法よりも精度が低いことが問題である。そこで本研究では自動獲得と集合知を組み合わせ、低コスト・大規模で高品質な関連語知識を獲得する手法を提案する。具体的には、自動獲得した関連語知識を利用してスマートフォン上で遊ぶことができる連想ゲームを作成し、ログを利用して関連語知識の評価を行う。例えば、「ガラス」に対して「透明、窓、割れる、薄い」などの関連語を自動獲得し、これらを連想ゲームの問題として提示した場合にプレイヤーが正しく「ガラス」と回答できるなら、この関連語は妥当なものであると考える。ゲーミフィケーションを利用することで金銭的成本を抑え、かつ付随的な情報から関連語知識の評価だけではなく新たな知識を獲得する。

2 関連研究

関連語知識の獲得に関しては以下の研究が行われている。岡本ら [7] は実際に被験者に連想させることで、連想概念辞書を構築している。少数の刺激概念を被験者に見せ、そこから連想される単語を連想概念として獲得している。奥村ら [5] は関連の強さをモデル化し、国語辞典と新聞記事から概念ベースとよばれる関連語知識のデータベースを構築している。進ら [6] はこれを多義語にも対応できるように拡張し、Web コーパスと Wikipedia の 2 種類の知識源を用いて名詞関連語を獲得している。

言語処理に集合知を利用する研究には以下のものがある。Snow ら [3] は様々な自然言語処理のタスクをクラウドソーシングで実施し、専門家によるものと同程度の質が得られることを示した。クラウドソーシングとは、Web サービスを通じて、不特定多数のユーザーに作業を依頼する仕組みのことである。また、プレイヤーがゲームを楽しむその行為自体が、何か別の作業となっている、ゲーミフィケーションを利用する研究も行われている。プレイヤーは楽しみを享受することを目的としてゲームを行うため、ゲーミフィケーションは低コストで作業を実施できるという利点がある。自然言語処理の分野においては、ESP ゲームを行うことで Semantic Network を作成することや [1]、関連語のアノテーションをビデオゲームの形で行うといった研究 [4] が行われている。

3 関連語知識の自動獲得

本節では関連語知識を自動獲得する手法について説明する。獲得された関連語知識は 4.1 節で示す連想ゲームを用いて評価する。まず、本研究で扱う関連語知識に関して 3.1 節で定義し、名詞・動詞・形容詞の関連語知識を獲得する。更に 3.2 節では、連想ゲームでの効率的な出題を目的とした関連語知識のクラスタリングを行う。

3.1 関連語の定義と自動獲得

単語間の関連を整理する方法は、次のように大別できる。

1. シソーラスなど、あらかじめ上位・下位などの関連の種類を定義しそれに合わせて分類するもの
2. 分布類似度の利用などで、特に同義関係に着目しコーパスから同義語を抽出するもの
3. 関連の種類によらず、関連の度合いを定義し、分類するもの

本研究で扱う関連語知識は、3 の方法であり、進ら [6] が提案する関連語のモデルを利用して関連語知識を自

動獲得する。進らは単語間の関連の種類に関わらず、2つの名詞 w_1, w_2 間の関連の強さを次のように定義されたスコアで測っている。

$$\begin{aligned} CoScore(w_1, w_2) &= f(w_1, w_2) * PMI(w_1, w_2) \\ &= f(w_1, w_2) * \log \frac{\frac{f(w_1, w_2)}{f(ALL)}}{\frac{f(w_1)}{f(ALL)} * \frac{f(w_2)}{f(ALL)}} \end{aligned}$$

ただし、 $f(w_1, w_2)$ は w_1 と w_2 の共起頻度を、 $f(w_i)$ は w_i の共起頻度を、 $f(ALL)$ は全ての名詞表現の共起頻度合計をそれぞれ表す。 PMI は w_1 と w_2 の自己相互情報量である。 $CoScore$ は、単語同士が相対的によく共起し、かつコーパス全体にわたる出現頻度もある程度大きい時に高くなる。そのため、ある単語から簡単に連想できるような単語の関連度が高くなるようなモデルとなっていると解釈できる。ここで、ある単語に対して $CoScore$ が高い順に 25 個を関連語として獲得し、最も高いスコアが 0.9 となるように正規化したものを関連度とする。進らは Web テキストと Wikipedia の第一段落の 2 種類の知識源を用いて $CoScore$ を計算し、 $CoScore$ が高いものから 25 個を関連語として獲得している。

本研究ではこの手法を拡張し、名詞・動詞・形容詞に対しても関連度を計算することで、あるキーワードに対する関連名詞・関連動詞・関連形容詞を獲得した。例えば「ガラス」の関連形容詞として、「透明だ、美しい、モダンだ、特殊だ」などが獲得された。コーパスは 4.2 億文の Web テキストと Wikipedia の第一段落を使用した。本論文では、関連語を獲得する元になる単語をキーワード、獲得された単語集合を関連語集合と呼ぶ。

3.2 関連語のクラスタリング

本研究では、このようにして作成した関連語集合の妥当性をゲーミフィケーションを利用して評価するが、各キーワードに対して獲得される関連語集合の数は各品詞ごとに 20 語以上あり、プレイヤーに単語全てを一度に提示し妥当性を判定させることは難しい。そこで、あるキーワードに対する関連語知識を意味のまとまりでクラスタリングし、そのクラスタを利用して出題することで効率的な判定を行う。例えば、「ガラス」の関連形容詞には「透明だ、クリアだ、繊細だ、明るい」というような「透明感」に関するものや「厚い、薄い、脆い」のような「強度」に関する語がある。妥当な関連語が獲得されていれば、この意味のまとまりも妥当であると考えられ、クラスタ内の数語を出題するだけでも十分に関連語の妥当性を判断できると考えた。

今回は関連語間の関係を利用したグラフを作成し、それを利用してクラスタリングを行った。グラフ構造を使用する理由は、Newman アルゴリズム [2] を使用することで、関連語辞書の情報のみでクラスタリングできる、事前に分割数を決めずにクラスタリングできる、という点からである。あるキーワードの関連語集合と、更にその関連語集合までを利用し、次の手順でグラフを構築する。

1. 単語をノードとし、それぞれ関連語の関係にあればエッジを張る。エッジの重みとして関連度を用いる。
2. 次数が 1 のノードを削除する。
3. キーワードの直接の関連語間の重みに 1 を加える。

このようにして構築されたグラフは、互いに関連のある単語間にエッジが密に張られている構造を持っており、この構造を利用してクラスタリングを行う。次数 1 のノードを削除し、キーワードの直接の関連語の重みを増やしているのは、このような操作を行うことで、キーワードの直接の関連語が効果的にクラスタリングされることが期待されるからである。

Newman アルゴリズムでは、ネットワーク中のエッジの媒介中心性に着目し、分割結果の良さの指標である Modularity が最も高くなる分割を出力とする。具体的なアルゴリズムを次に示す。

1. 各エッジの媒介中心性を計算する。
2. 媒介中心性が最大のエッジを除去する。
3. Modularity Q を計算する。

以上のステップを、ネットワーク内のエッジの数だけ実行する。このとき、各段階で計算された Modularity Q が最大となる分割を最終的な出力とする。エッジ e の媒介中心性 $g(e)$ は以下のように定義される。

$$g(e) = \sum_{s \neq t} \frac{\sigma_{st}(e)}{\sigma_{st}}$$

ここで $\sigma_{st}(e)$ はノード s, t 間の最短経路中、エッジ e を通るもの数であり、 σ_{st} はノード s, t 間の最短経路の総数である。Modularity Q は以下のように定義される。

$$\begin{aligned} Q &= \sum_i (e_{ii} - a_i^2) \\ e_{ij} &= \sum_{i=1}^N \sum_{j=1}^N \frac{A_{ij}}{2M}, \quad a_i = \sum_j e_{ij} \end{aligned}$$

ここで A_{ij} は隣接行列、 M はグラフ中のエッジの総本数である。この手法で「ガラス」の関連形容詞をクラスタリングした結果を表 1 に示す。

表1 「ガラス」の関連形容詞のクラスタリング例

cluster1	明るい, 綺麗だ, 透明だ, 鮮やかだ, 白い, 美しい, 繊細だ, クリアだ, 色取り取りだ
cluster2	無機質だ, シンプルだ, モダンだ, おしゃれだ, 高級だ, スタイリッシュだ, 洒落だ, オリジナルだ
cluster3	厚い, 強度だ, 粉々だ, 脆い, 薄い, 微細だ, 特殊だ, 丸い
cluster4	涼しい

表2 半月分の連想ゲームログの統計情報

端末 id(プレイヤー数に相当)	5291 台	出題数	7493 問
<i>Exact</i>	2367 問	<i>Near</i>	1470 問
<i>Bad</i>	3656 問	回答時間最頻値	20.0 秒
端末 id 平均プレイ回数	1.4 回	最大プレイ回数	23 回

ている音声対話スマートフォンアプリである「Yahoo! 音声アシスト」の追加機能として動作させる。^{*1} ユーザーは音声対話システムと雑談したり暇つぶしを目的として話かけることがあり、ゲームを促進する要素を備えていると考えられる。Yahoo!音声アシストはAndroidとiOSで動作する。

4.2 ログの統計情報

連想ゲームは実際に2014年12月中旬から本番環境で動作させており、今回は、86語のキーワードに対する関連語を連想ゲームとして出題した。本稿では現時点で取得できている約半月分のログを元に分析を行った。ログから得られた情報を表2にまとめた。ここで、プレイヤーの回答がキーワードと完全に一致したものを*Exact*、一致していないが類似していると判定されたものを*Near*、それ以外を*Bad*とした。約半月ではあるが、プレイヤー数は約5000人、出題数は約7000問でありゲームの規模は十分大きいと言える。回答時間を10秒刻みに分割したところ、最頻値は約20秒で少し考えて答えている程度であった。また、繰り返し何度も遊んでいるプレイヤーの存在が確認できる点でも、ゲームとして十分成り立っていると考えている。

自動獲得された関連語集合の評価には、各キーワード毎のプレイヤーの回答で、頻度の高い上位5つの回答を利用した。これは音声対話を利用しているために発生する、意図しない発話や、音声認識のミス等を除外するためである。これらに対して以下のように定義した正答率を計算した。

$$\text{正答率} = \frac{\text{Exact} + \text{Near}}{\text{Exact} + \text{Near} + \text{Bad}}$$

この正答率に着目してログを分析することで、自動獲得結果の妥当性を評価した。プレイヤーの回答の例を表3にまとめた。出題した問題例は一例であり、実際にはクラスタの組み合わせにより複数の問題セットが出題されている。

4.3 正答率に基づく分析

出題した問題セットに関わらず正答率が低かったものは、関連語の自動獲得結果が妥当でなかった可能性がある。例えば表3のキーワード「直線」は正答率が低

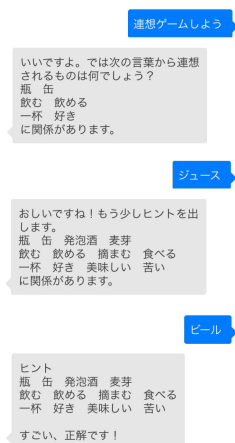


図1 実際の連想ゲーム (青はプレイヤーの発話)

4 連想ゲームによる実験

このようにして自動獲得された関連語集合の妥当性についての人間の判断を、ゲーミフィケーションを通して低コストかつ大規模に獲得する。このような判断は一般的な知識を持った人間であれば誰でもできるということと、現在では多くの人がスマートフォンを利用していることから、スマートフォン上でのゲームを用いて実現する。

4.1 連想ゲームの枠組み

基本的なアイデアは、人間であれば、関連語の自動獲得結果が妥当ならば関連語集合を見るだけで、キーワード、もしくはそれに近いものを連想できるだろうというものである。そこで、プレイヤーに自動獲得した関連語集合をヒントとして提示し、キーワードを当ててもらおうという連想ゲームを実施する。プレイヤーの回答とキーワードを比較し、一致している、もしくは類似しているものが多ければ、自動獲得結果は妥当であると判断できる。回答とキーワードの比較には、分布類似度と word2vec によって類似度を計算し、閾値以上であれば類似していると判定した。回答のチャンスは2回とし、1回目が不正解、もしくは類似した回答であれば、ヒントを追加する。図1に連想ゲームのやりとりを提示する。

本研究では、このゲームを Yahoo!JAPAN が提供し

^{*1} <http://v-assist.yahoo.co.jp/>

表3 プレイヤーの回答例

キーワード	問題の一例	プレイヤーの回答	正答率
メッセージ	カード 応援 一言 コメント 祝う 添える 伝える 込める 力強い 強烈 オリジナル 最後	メール:41, 手紙:10, 言葉:9, メッセージ:8, 誕生日:4	0.94
ネクタイ	シャツ スーツ ブランド 柄 結ぶ 織る 締める 外す 恰好 だらしがない 黒い 白い	ネクタイ:10, ベルト:8, ファッション:5, サラリーマン:4, ふく:4	0.71
映画	主演 監督 シーン 音楽 描く 演じる 見る 泣ける 切無い 最後 有名 ポップ	ラブストーリー:3, 高倉健:3, ミュージカル:3, 映画:2, 野球:2	0.39
距離	直線 馬 移動 位置 撃つ 詰める 縮める 縮まる 短い 長い 高速 速い	競馬:19, ピストル:12, メジャー:12, 車:10, 距離:10	0.16
直線	曲線 脚 距離 馬 抜け出す 差す 伸びる 結ぶ 長い 速い 垂直 ジグザグ	競馬:35, 競馬場:4, 料理:3, 位置:3, 馬:3	0.0

表4 人手ラベルとの評価

		人手評価	
		妥当である	妥当でない
ログ	妥当である	65	6
	妥当でない	9	6

い。多くのプレイヤーは「競馬」関連の単語を回答している。これは「直線」の関連語としては不適切であり、自動獲得結果が妥当ではなかったからであると考えられる。このように正答率が低かったものに着目することで、妥当性の低い関連語集合を特定することが可能である。

一方、正答率が高い問題セットに着目すると、プレイヤーの回答集合から新たな関連語を獲得することができる。例えば表3のキーワード「ネクタイ」では、最多回答が「ネクタイ」となっているが、「ネクタイ」以外の回答にある「ベルト」や「サラリーマン」は「ネクタイ」の関連語でもある。事前に「ベルト」や「スーツ」といった服装に関する単語は自動獲得されていたが、「サラリーマン」の様な新しい関連語知識も獲得することができた。ゲームログ全体でも、正答率が高かった問題に関しては、プレイヤーの回答集合自体がキーワードの関連語となっている場合が多いことがわかっている。

また、ログを集約して得られた判定が、人手で付けたラベルとどれだけ一致するかを測定するため、出題したキーワード86語の関連語集合に対して人手で「妥当である」、「妥当でない」の2値のラベルを付けたものを正解とし、Precision、Recall、F値を計算した。認識ミスなど除外しきれなかったノイズの存在を考え、正答率が0.3以下のものを「妥当でない」としてログを集約

した。その結果 Precision:0.91、Recall:0.87、F値:0.89であり、連想ゲームの実施で十分な評価が可能であると考えられる。

5 まとめと今後の課題

ゲーミフィケーションを用いることによって自動獲得された関連語知識の人手評価を低コストに行うための手法を提案し、実際に関連語知識の妥当性を判定できることを確認した。今後は評価に十分なログが集まる期間を推定し、問題セットを更新することで、より多くの関連語知識の評価を行う。また、出題した関連語クラスタと正答率を比較することで、関連語クラスタの関連度を計算することが可能であると考えており、ログの量が増えた段階でそのような分析を行いたい。

参考文献

- [1] Mathieu Lafourcade. Making people play for lexical acquisition. In *Proc. of ENLP2007*, pp. 13–15.
- [2] M E Newman. Modularity and community structure in networks. *PNAS2006*, Vol. 103, No. 23, pp. 8577–8582, 2006.
- [3] Rion Snow, Brendan O'Connor, Daniel Jurafsky, and Andrew Y. Ng. Cheap and fast—but is it good?: Evaluating non-expert annotations for natural language tasks. In *Proc. of EMNLP2008*, pp. 254–263.
- [4] Daniele Vannella, David Jurgens, Daniele Scarfini, Domenico Toscani, and Roberto Navigli. Validating and extending semantic knowledge bases using video games with a purpose. In *Proc. of ACL2014*, pp. 1294–1304.
- [5] 奥村紀之, 土屋誠司, 渡部広一, 河岡司. 概念間の関連度計算のための大規模概念ベースの構築. *自然言語処理*, Vol. 14, No. 5, pp. 41–64, 2007.
- [6] 進義治, 黒橋禎夫. 名詞関連語知識に基づく文章のグラフ表現とその応用. *言語処理学会 第20回年次大会*, pp. 1007–1010, 2014.
- [7] 岡本潤, 石崎俊. 概念間距離の定式化と既存電子化辞書との比較. *自然言語処理*, Vol. 8, No. 4, pp. 37–54, 2001.