

連想ゲームによるコモンセンス知識の獲得

大谷直樹[†] 河原大輔[†] 黒橋禎夫[†] 鍛冶伸裕[‡] 颯々野学[‡]

[†] 京都大学大学院 情報学研究科 [‡] ヤフー株式会社

[†]otani.naoki.65v@st.kyoto-u.ac.jp, {dk,kuro}@i.kyoto-u.ac.jp

[‡]{nkaji,msassano}@yahoo-corp.jp

概要

コンピュータが知的な処理を行うためには人間が持つ常識 (コモンセンス知識) を収集することが必要である。我々は低コストかつ大量にコモンセンス知識を獲得するために、知識獲得プロセスを含んだ連想ゲームを開発し、音声対話システム上で公開した。このような獲得プロセスをゲーム化する枠組みは GWAP (Game With A Purpose) と呼ばれる。本研究では連想ゲームから得られた知識を評価し、プレイヤーの信頼度やヒントの情報を使った品質推定の有効性について検証する。

1 はじめに

高度な言語処理を実現するために、人間が持つ常識 (コモンセンス知識; Commonsense Knowledge) を整理する試みが長年行われている [5]。Open Mind Common Sense (OMCS) プロジェクトはインターネット上で一般の人々の協力を募り、データベースを構築してきた。そのデータベースは **ConceptNet** [4] と呼ばれ、(携帯電話, *IsA*, 家電) のようなファクト (2つの概念とその関係のトリプル) と「携帯電話は家電の一種である」のような言語表現を登録している。日本語のファクトも存在するが、量が少なく、拡充が求められている [7]。

一般的に人手による知識獲得は、品質が良い一方でコストが高いため、大規模な実行が難しい。この問題に対して、近年はインターネットを通して不特定多数のワーカーに仕事を依頼するクラウドソーシングの利用に注目が集まっている。その一分野として、データ獲得プロセスをゲームに落とし込み、ゲームで遊んでもらいながら知識獲得をする **GWAP** (Game With A Purpose) が盛んに研究されている。

町田らは GWAP を用いて自動獲得された関連語クラスタの評価を行った [8]。ゲームは Yahoo! JAPAN が

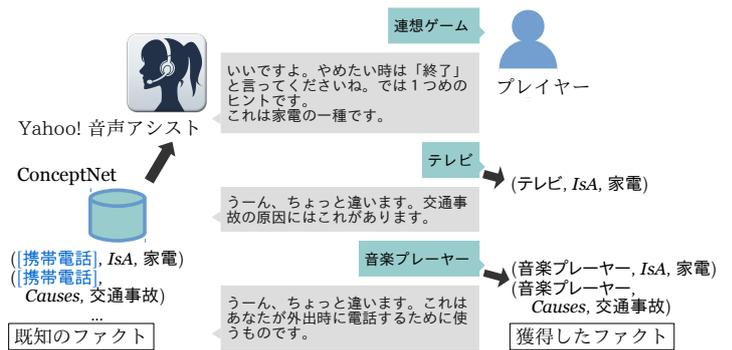


図 1: 連想ゲームの進行例。ConceptNet に登録されている既知のファクトからキーワードとヒントを生成し、プレイヤーが解答を行う。プレイヤーの解答から新しいファクトを獲得できる。

提供す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**¹ (以下では音声アシストと呼ぶ) の雑談対話機能の一つとして公開された。音声アシストは 2016 年 1 月現在 175 万ダウンロードの実績がある。これを利用する多数のユーザーにゲームに参加してもらうことにより、低コストでデータを収集することに成功した。町田らはプレイヤーの正答に主眼を置いていたが、誤答の中から自動獲得できなかった関連語が新たに得られる可能性があると報告している。

そこで本研究は、プレイヤーの誤答からコモンセンス知識を獲得する連想ゲームを開発し、町田らと同じく音声アシスト上に公開した。このゲームは、GWAP による知識獲得の先行研究 [2, 7] と同様に、既知の知識の一部をプレイヤーに推測してもらうことで新しい知識を獲得する (図 1)。連想ゲームでは、システムがある語 (キーワードと呼ぶ) についてのヒントを与え、プレイヤーがそのキーワードを推測する。例えば、システムがキーワード「携帯電話」から「これは家電の一種です」というヒントを提示し、ユーザーが「テレビ」と答えたとする。この解答はゲームでは不正解と判定されるが、我々は「テレビは家電の一種である」

¹<http://v-assist.yahoo.co.jp>

という知識を得ることができる。

一方、こうして集められた知識は品質に問題があることが多い。そこで、既存手法は複数のプレイヤーが答えた知識を妥当とするという考えに基づき多数決で獲得知識をフィルタリングしている [2, 1, 7]。

これに対し我々は重み付き多数決によって獲得知識の品質を推定する。単純な多数決とは違いプレイヤー個別の信頼性を考慮するため、より正確な品質推定ができる。手法として Web ページの重要性を推定するために提案された **HITS (Hypertext Induced Topic Selection)** アルゴリズムを用いる。

連想ゲームは 2015 年 12 月に公開され、ログデータを収集している。本研究では得られた知識の質を評価し、品質推定方法の有効性についても検証する。

2 関連研究

本研究の対象であるコモンセンス知識は、コンピュータが知的な処理をするために必要な資源として重要視され収集が試みられてきた。OMCS プロジェクトは、十年以上にわたって人手でコモンセンス知識を収集しており、そのデータベース ConceptNet [4] はインターネット上に公開されている²。ConceptNet のデータは 2 つの概念とその間の 1 つの関係というファクトの形で登録されている。概念は語や短い句で表され、関係には、同位関係を表す *IsA*、全体部分関係を表す *PartOf* や因果関係を表す *Causes* など、予め定義された約 30 種類のタイプ (以下で関係タイプと呼ぶ) が使われている。

近年は本研究のように GWAP の枠組みによってコモンセンス知識を獲得する研究も活発化している。中国語では Kuo らが、日本語では中原らがゲームによりコモンセンス知識の大規模獲得に取り組んでおり、収集された知識は ConceptNet に登録されている [2, 7, 6]。本研究は彼らの研究と同じくプレイヤーの誤答から知識を獲得する方法を取っている。しかし彼らは獲得知識の品質を単純な多数決によって見積もっており、同頻度の知識間の品質差を見分けられないという欠点があった。その解決方法として、本研究はプレイヤーごとの信頼性やヒントと解答間の距離に着目する。

大規模文書集合から知識を自動的に獲得する手法と GWAP を組み合わせた方法も試みられている。町田らは自動で獲得した関連語クラスターを連想ゲームで評価する枠組みを提案した [8]。また、Herdağdelen らは自動的に収集した知識をゲームでフィルタリングす

²<http://conceptnet5.media.mit.edu>

る方法を提案した [1]。本研究は知識の自動獲得手法を取り入れていないが、今後自動獲得手法と GWAP を融合していくことも検討している。

3 連想ゲームによる知識獲得

連想ゲームは音声アシストの雑談機能の一つとして提供される。プレイヤーが「連想ゲームしよう」などと発声するとゲームが開始する。

ゲームが始まると、まず知識ベースからランダムにキーワードが選択される。そしてキーワードに紐付いている ConceptNet 中のファクトからヒントが選ばれ提示される。ヒントは関係タイプごとに用意したテンプレート文に当てはめることで自動生成される。図 1 の例ではキーワード「携帯電話」が選択され、最初のヒントとして「これは家電の一種です」が生成されている。このヒントはランダムに選択したファクト (携帯電話, *IsA*, 家電) をテンプレート文に当てはめ、キーワードを「これ」に置き換えることにより生成されている。提示されたヒントに対してプレイヤーが解答を行う。解答がキーワードと一致すれば、正解と判定されゲームが終了する。一致しなければ次のヒントが提示される。

ヒントに対するプレイヤーの解答から、中原らの方法に従いファクトを獲得する [7]。図の例ではプレイヤーは最初のヒントに対して「テレビ」、次に出たヒント「これは交通事故の原因の一つです」に対して「音楽プレーヤー」と答えている。これらは不正解と判定されるが、我々はこの誤答から新たに 3 つのファクト (テレビ, *IsA*, 家電)、(音楽プレーヤー, *Causes*, 交通事故)、(音楽プレーヤー, *IsA*, 家電) を得ることができる。以下ではこれらを獲得ファクトと呼ぶ。

プレイヤーがそれまでのヒントすべてに適合する解答を出していれば、3 回ヒントが提示されたときに獲得できるファクトは 6 件である。ただし実際はプレイヤーは直前のヒントを重視し、前のヒントほど適合度は低くなると考えられるので、解答とヒントとの距離 (ヒント距離と呼ぶ) を獲得ファクトごとに考慮する。上の例の (音楽プレーヤー, *Causes*, 交通事故) と (音楽プレーヤー, *IsA*, 家電) のヒント距離はそれぞれ 1 と 2 である。

4 重み付き多数決による品質推定

2 章で述べた通り、既存手法は多数決により獲得知識の品質を推定していた。しかし実際には解答が正確

でなかったり、真剣に解答しないプレイヤーが存在する。そこで本研究では、プレイヤーごとの正確さを考慮し、それに重み付けされた多数決により獲得ファクトの品質を推定する。このように作業者の信頼性を仮定した方法は、クラウドソーシングの品質管理の分野でよく利用されており、本研究は情報の信憑性を重み付け多数決で推定した Pasternack と Roth の研究 [3] と同様に、HITS アルゴリズムを応用する³。

4.1 HITS アルゴリズム

HITS アルゴリズムは Web ページとページ間のリンクが与えられたときに、各 Web ページの重要性を求める方法である。各 Web ページはオーソリティスコアとハブスコアを持つ。オーソリティスコアはそのページの価値を表し、高いオーソリティスコアを持つページは、高いハブスコアを持つページからリンクされている。ハブスコアは重要なページにリンクしている度合いを表し、高いハブスコアを持つページは、高いオーソリティスコアを持つページにリンクしている。

このアイデアのもとで、オーソリティスコアはリンク元ページのハブスコアの和、ハブスコアはリンク先ページのオーソリティスコアの和として計算される。スコアを交互に更新すると、やがてある値に収束する。これが求めるスコアである。

4.2 重み付き多数決

HITS アルゴリズムを獲得ファクトの品質推定に応用する。獲得ファクトの品質がオーソリティスコア、プレイヤーの正確さがハブスコアに対応する。正確さの高いプレイヤーに答えられたファクトほど品質が高く、品質の高いファクトを答えたプレイヤーほど正確さが高いと考える。

ファクトとプレイヤーをそれぞれ i と j で表す。HITS アルゴリズムではそれぞれのページが 2 種類のスコアを持っていたが、本研究ではファクトのみがオーソリティスコアを持ち、プレイヤーのみがハブスコアを持つ。ファクトのスコア f_i 、プレイヤーのスコア p_j を以下の式により更新する。

$$f_i = \sum_{j \in \mathcal{I}_i} p_j \quad (1)$$

$$p_j = S \left(\frac{1}{|\mathcal{I}_j|} \sum_{i \in \mathcal{I}_j} f_i \right) \quad (2)$$

³Pasternack と Roth は HITS アルゴリズムを Hubs and Authorities と呼んでいる

ゲーム	発話	ユーザー	獲得ファクト
1605	5528	984	13072 (6932)

表 1: 集められたログデータ。獲得ファクトはフィルタリング前の値。ユーザーと獲得ファクトの括弧内の数字はユニーク数である。

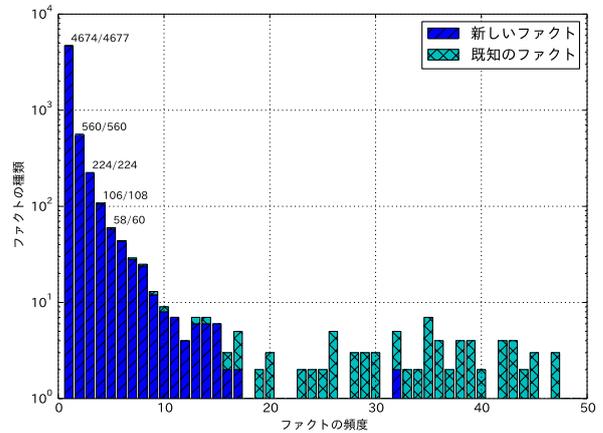


図 2: 獲得ファクトの分布。横軸はファクトの頻度で、縦軸は種類数 (対数スケール) である。グラフ上に頻度 1 から 5 までの全獲得ファクト数とそのうちヒントで用いなかった新しいファクトの数を示す。

ここで \mathcal{I}_i はファクト i を解答したプレイヤーの集合、 \mathcal{I}_j はプレイヤー j から獲得したファクトの集合で、 $|\mathcal{I}_j|$ はそのファクト数である。値の発散を防ぐため、プレイヤーのスコアが平均 1 分散 1 の正規分布に従うと仮定してスケーリング S を行っている。

3 章で述べた通り、一般的にファクトのヒント距離が大きくなるほど品質は下がっていきと考えられるので、ヒント距離を更新式に取り入れることを考える。プレイヤー j が解答したファクト i のヒント距離が $d_{ij} (\geq 1)$ のとき、重みを $w_{ij} = 1/d_{ij}$ で定義する。式 (1),(2) に対応する更新式は次のようになる。

$$f_i = \sum_{j \in \mathcal{I}_i} w_{ij} p_j \quad (3)$$

$$p_j = S \left(\frac{1}{\sum_{i \in \mathcal{I}_j} w_{ij}} \sum_{i \in \mathcal{I}_j} w_{ij} f_i \right) \quad (4)$$

5 実験

2015 年 12 月から 2016 年 1 月にかけて連想ゲームのログを収集した。984 人のプレイヤー⁴ から 5,528

⁴ここでいうプレイヤー数は、厳密には連想ゲームを起動した端末数のことを指す。

ファクト	キーワード	頻度
1) (目玉焼き, <i>MadeOf</i> , 卵)	ケーキ	2
2) (犬, <i>IsA</i> , 動物)	キリン	3
3) (雑巾, <i>UsedFor</i> , 鼻をかむこと)	ティッシュペーパー	2
4) (富士山, <i>RelatedTo</i> , 絵の具)	スケッチ	3

表 2: 獲得ファクトの例。1), 2) は妥当なファクト、3), 4) は妥当でないファクトである。

HITS	HDIST-HITS	MV	HDIST
0.746	0.796	0.670	0.784

表 3: 頻度 2 以上の 100 件のファクトに対する ROC AUC スコア。

件の発話があり、約 13,000 件 (重複含む) のファクトを得た (表 1)。獲得ファクトの頻度の分布を図 2 に示す。明らかにゲームと関係のない発話や、音声認識誤りと思われる発話はフィルタリングしたところ、獲得ファクトは全体で約 12,000 件となった。品質推定の実験のために、頻度 2 以上の獲得ファクトからランダムに 100 件を選び、人手で妥当性を二値で判定した。そのうち妥当であるファクトは 76 件であり、先行研究 [7] の報告と同程度の傾向が見られた。

全獲得ファクトに対して 4 つの方法で品質を推定する。

HITS HITS アルゴリズムに基づく重み付き多数決

HDIST-HITS ヒント距離を加味した HITS

MV 既存研究で用いられた単純な多数決

HDIST 4 章で定義したヒント距離の重み w の和を品質の推定値とする方法

ラベル付きの 100 件のファクトについての ROC AUC スコアを表 3 に示す。AUC スコアは 0 から 1 までの値を取り、高い値ほど妥当なファクトと妥当でないファクトを正しく区別していることを示す。ランダムに分類した場合の AUC スコアは 0.5 である。スコアは HDIST-HITS がもっとも高く、HDIST、HITS が続き、MV が最も低かった。本稿で議論したように、ヒント距離とプレイヤーごとの信頼性が品質推定に寄与することが示唆された。表 2 で示した頻度が同じファクトのペアについても HDIST-HITS と HDIST は妥当であるファクトの品質を他方より高く推定できた。ただし HITS だけは、(2) と (4) の優劣関係がわずかに逆転した。この原因はヒント距離を考慮しなかったことにあると考えられる。

6 おわりに

我々は人間から低コストかつ大量にコモンセンス知識を獲得することを目的とし、スマートフォンの音声対話アプリ上で動く連想ゲームを開発した。公開後に収集したログデータから実際に知識が得られることを確認した。さらに本研究ではその品質についても検討し、プレイヤーの信頼性とヒント距離が寄与することを示した。実際には、連想ゲームではキーワードやヒントの難しさや、ヒントの提示順なども得られる知識に影響すると考えられる。それらの検討は今後の課題としたい。

参考文献

- [1] Amaç Herdağdelen and Marco Baroni. Bootstrapping a Game with a Purpose for Commonsense Collection. *ACM Transactions on Intelligent Systems and Technology*, Vol. 3, No. 4, pp. 1–24, 2012.
- [2] Yen-ling Kuo, Jong-Chuan Lee, Kai-yang Chiang, Rex Wang, Edward Shen, Cheng-wei Chan, and Jane Yung-jen Hsu. Community-Based Game Design: Experiments on Social Games for Commonsense Data Collection. In *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP)*, pp. 15–22, Paris, France, 2009.
- [3] Jeff Pasternack and Dai Roth. Judging the Veracity of Claims and Reliability of Sources With Fact-Finders. *Computational Trust Models and Machine Learning*, pp. 39–72. Chapman and Hall/CRC, 2014.
- [4] Robert Speer and Catherine Havasi. Representing General Relational Knowledge in ConceptNet 5. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)*, pp. 3679–3686, Istanbul, Turkey, 2012.
- [5] Liangjun Zang, Cong Cao, Yanan Cao, Yuming Wu, and Cungen Cao. A Survey of Commonsense Knowledge Acquisition. *Journal of Computer Science and Technology*, Vol. 28, No. 4, pp. 689–719, 2013.
- [6] 中原和洋. コモンセンス知識獲得を目的としたソーシャルゲーム”日本人検定”. *Unisys 技報: Unisys technology review*, Vol. 32, No. 4, pp. 389–401, 2013.
- [7] 中原和洋, 山田茂雄. 日本でのコモンセンス知識獲得を目的とした Web ゲームの開発と評価. *Unisys 技報: Unisys technology review*, Vol. 30, No. 4, pp. 295–305, 2011.
- [8] 町田雄一郎, 河原大輔, 黒橋禎夫, 颯々野学. 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. *言語処理学会 第 21 回年次大会 発表論文集*, pp. 1060–1063, 2015.