

Acceleration Innovation Through Analogy Mining

Tom Hope, Joel Chan, Aniket Kittur, Defna Shahaf

https://scholar.google.com/scholar?as_q=&num=10&btnG=Search+Scholar&as_epq=Accelerating+Innovation+Through+Analogy+Mining&as_oq=&as_eq=&as_occt=any&as_sauthors=Hope

KDD/2017, Best Paper Award

Seeking in the vast Internet World

Summary

By Yohei Kawakami

2019/01/29



Acceleration Innovation Through Analogy Mining

Tom Hope, Joel Chan, Aniket Kittur, Defna Shahaf

https://scholar.google.com/scholar?as_q=&num=10&btnG=Search+Scholar&as_epq=Accelerating+Innovation+Through+Analogy+Mining&as_oq=&as_eq=&as_occt=any&as_sauthors=Hope

- 1, What is Analogy ...??
- 2, Summary
- 3, Experiments, conclusion, and discussion

Seeking in the vast Internet World

Summary

By Yohei Kawakami

2019/01/29





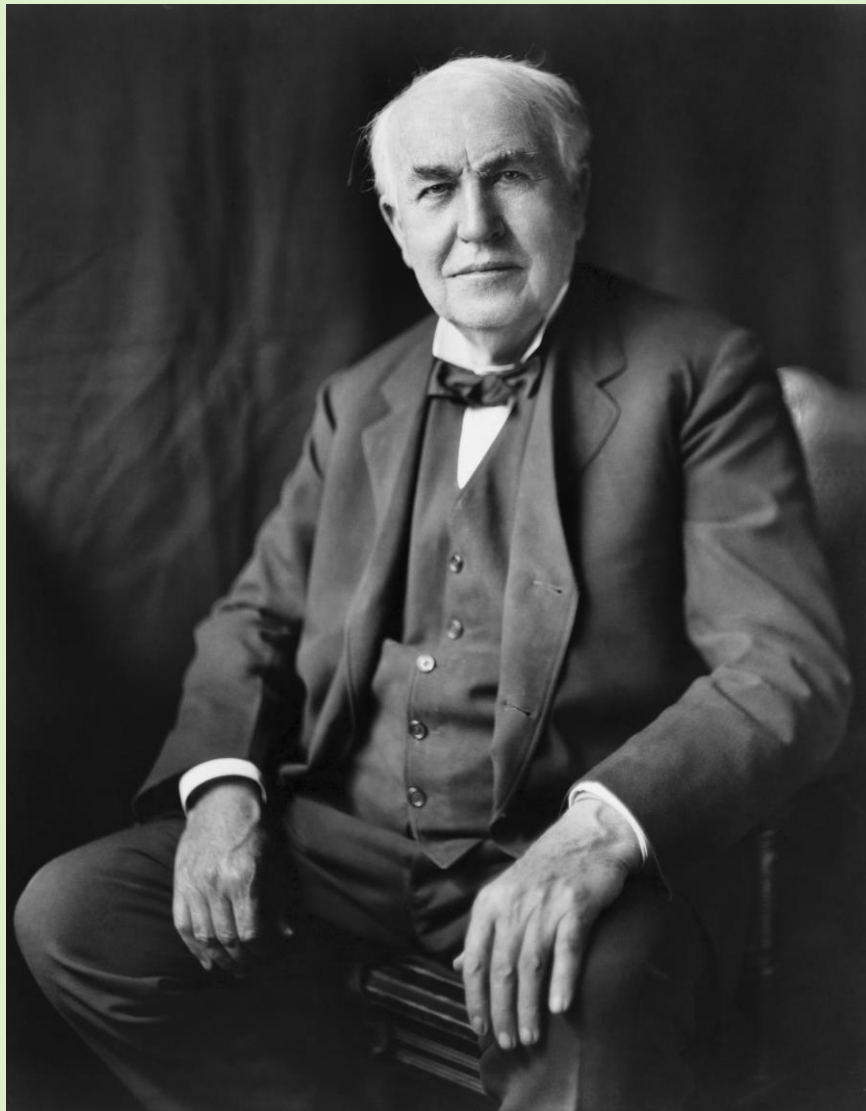
What is Analogy...???

What is Analogy...??:

類推

特定の事物に基づく情報を、他の特定の事物へ、それらの間の何らかの類似に基づいて適用する認知過程である。

reference



エジソンの言葉：『よい発明家の条件は...?』

①Plenty money to throw away

②Imagination

③Logistic mind that sees **analogies**

What is Analogy...??:

類推

特定の事物に基づく情報を、他の特定の事物へ、それらの間の何らかの類似に基づいて適用する認知過程である。

reference



本研究における類推とは？：『携帯の充電器からの類推』

【①問題提起】携帯電話を充電器できるケースを再設計すること

- ・携帯電話
- ・充電器
- ・ケース

【②類推】①から類推されるものをデータベースから引き出す

- ・フラッシュチャージカラビナ
- ・バックアップバッテリー付属USB
- ・人的エネルギー発電機

【③人間の発想】②を受けて、人間が新商品を連想する

- ・ステップを追跡し、人の動きを使って充電できるケース
- ・複数のワイヤレス電源ハブがあり、どこにでもワイヤレスで携帯電話を接続して充電できるケース
- ・ワイヤレスキーボードで使用されているテクノロジーのように近くの光源から充電できるケース

Figure 3: Overview and excerpts of the ideation experiment. Top: Seed product. Workers were asked to solve the same problem in a different way. Middle: Top 3 inspirations for each of the conditions. Note that the TF-IDF baseline returns results from the same domain, while our method returns a broader range of products. Bottom: Ideas generated by users exposed to the different conditions.

Conclusion:

本研究モデルの類推は、人々が創造的なアイデアを生み出す可能性を著しく高めた。

Problem>>Farmer Work>>

問題) データベース上に大量、かつ乱雑なデータが蓄積されており、人間にとっても、機械にとっても類推することは難しい。
初期は、人間がリンクを手作業で作成していた。

先行研究) 教師ラベルを整えたデータを使い、ランダムサンプリングにより予測モデルを生成し、類推をする。
大規模な自然言語データセットに拡張できるが、構造的類似性を説明することが難しい。

本研究) 製品の説明から目的とメカニズムのベクトル表現を抽出するために、クラウドソーシングとリカレントニューラルネットワークを組み合わせたもので、製品の目的とその目的を達成するためのメカニズムを指定する、より単純な構造表現、具体的には「problem schemas」を学習することによって、より高い精度での類推を可能にした。

What is this thesis for?

大規模で、雑然とした現実の知識レポジトリ内の類推検索に関して、自動的に類推(アナロジー)をみつけることができるプロダクトの開発。

Where is an important point compared to previous researches?

教師学習ラベルなしで、自動的に探索して、類推をみつけることができる。

Where are the key points of technology and method?

- *Same purpose, different mechanism.* Given the corpus of all products \mathcal{P} , a product i with (normalized) purpose and mechanism vectors $\mathbf{p}_i, \mathbf{m}_i$, and distance metrics $d_p(\cdot, \cdot), d_m(\cdot, \cdot)$ between purpose and mechanism vectors (respectively), solve:

$$\begin{aligned} \operatorname{argmin}_{\hat{i} \in \mathcal{P}} d_p(\mathbf{p}_i, \mathbf{p}_{\hat{i}}) \\ \text{s.t. } d_m(\mathbf{m}_i, \mathbf{m}_{\hat{i}}) \geq \text{threshold}, \end{aligned} \quad (1)$$

- *Same Mechanism, different purpose.* Solve:

$$\begin{aligned} \operatorname{argmin}_{\hat{i} \in \mathcal{P}} d_m(\mathbf{m}_i, \mathbf{m}_{\hat{i}}) \\ \text{s.t. } d_p(\mathbf{p}_i, \mathbf{p}_{\hat{i}}) \geq \text{threshold}, \end{aligned} \quad (2)$$

How to verified whether it is valid?

Quirky.comの製品説明のデータを対象に、製品は何をするのか? 製品は、どこが良いのか? 製品はどのように機能するか? をラベル付けしたものと当研究モデルで比較した。

【質的研究】

→ 本研究モデル、先行研究モデル、ランダム検索モデルにより生成された言葉から影響を受けた案をアンケート調査し、案を比較する。

【量的検証】

→ 出た結果に対して、分散分析を行い、検定する。

Is there discussions?

製品の説明以外に他の分野に一般化するには拡張が必要である。
比較的単純な製品説明から科学論文のような、より複雑なデータに移行する場合。高次元の拡張を今後検討していくべきである。

Which reserches should I read next?**【先行研究】**

JoelChan, TomHope, DafnaShahaf, and AniketKittur. 2016.
ScalingupAnalogy with Crowdsourcing and Machine Learning.
In Workshop on Computational Analogy at ICCBR.



Two Experiments, Conclusion, Discussion

Quirky.comの商品データを対象に調査をする

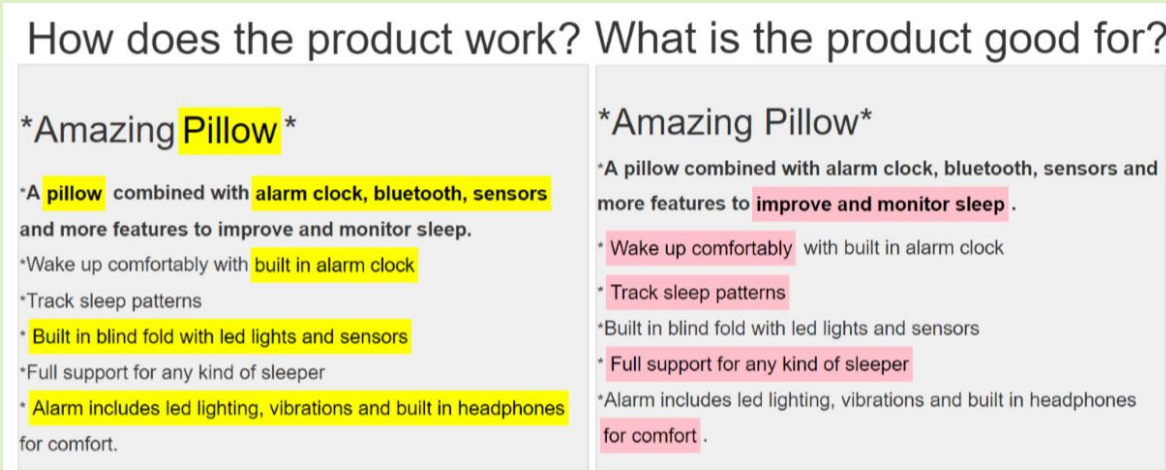


Figure 1: Collecting purpose and mechanism annotations from the crowd.

対象のデータセット

- Quirky.com: crowdsourced innovation site
- ~10,000 product ideas
- Many categories
- Natural language

Thirsty too
Pet water bowl/dispenser for your vehicle cup holder.
Over spill lip to catch water
Has optional sleeve for larger cup holders
Optional floor base
One way valve so water cant over flow from bottle
Small reservoir
Reservoir acts as backsplash
Water bottle attachment
Holds water in your vehicle cupholder for pet

Table 1: Purpose and Mechanism vector interpretation examples. Descriptions shortened. Sparse coding shows only words with $|\alpha| \geq 0.1$

Product	Purpose words	Mechanism words
A small yogurt maker machine for concentrating yogurt under heat and vacuum. Has a round base in drum with customized scooper, washable stainless steel drum parts. Reduce time and energy used.	Top similar: concentrate, enough, food, even, much, especially, reduce, produce, whole Sparse coding: making, energy, yogurt, drum, concentrate, vacuum, heavy, foods, aches, service	Top similar: liquid, heat, cooling, pump, steel, machine, water, heating, electric Sparse coding: vacuum, cooled, drum, heavy, ingredients, design, renewable, stainless, vending
A cover placed on a car truck to protect from hail. Elastic perimeter to prevent wind from blowing under cover. Snap or velcro slits to open door without removing cover. Strong attachment so it wonfit blow away. Inflatable baffles that cover the top, front windshield, side.	Top similar: storm, hail, rain, roofs, doors, wind, front, winds, walls Sparse coding: roof, hail, padded, obstructing, defenses, diesel, windshield, wets	Top similar: roof, cover, lining, zipper, bottom, hood, plastic, flap, rubber Sparse coding: front, cover, insulation, hail, buckle, sling, watertight, cutter, blowing
A leash accessory with removable compartments for phone, cards cash, keys, poop bags, treats, bowl. Walk your dog and carry your essentials without pockets or a purse bag.	Top similar: bags, purse, wallet, carry, leash, backpack, pocket, dog, luggage Sparse coding: bag, leash, compartments, pets, phone, eats, practical, hand-ing, pull	Top similar: leash, pouch, purse, pocket, bags, pockets, strap, compartment, backpack Sparse coding: leash, bag, compartments, hand, holders

Figure1) 4名による商品に対するラベル付けを行う

Table1)商品から類推される言葉を抽出する

本研究手法における抽出方法の提示

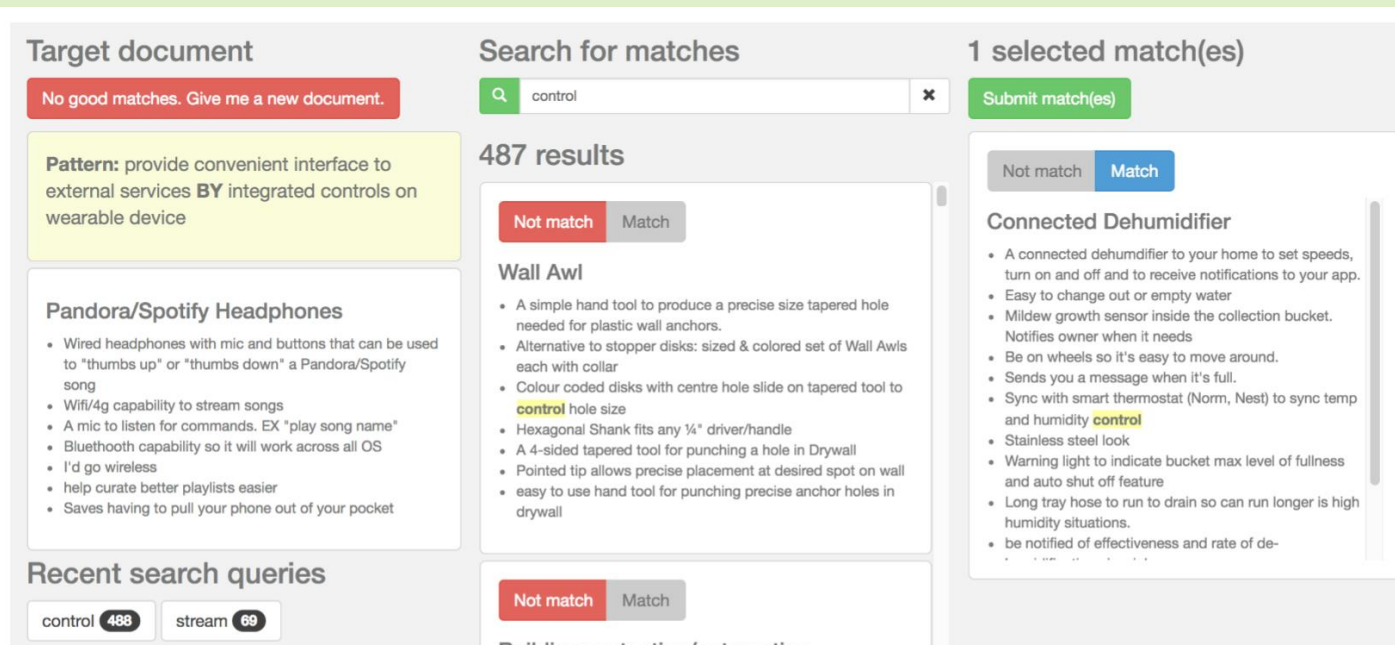


Figure 2: Screenshot of analogy search interface

Figure 2) 本研究手法インターフェース

類推される言葉が抽出される

[下図]

ネットワークの説明と数式

Learn a similarity metric reflecting analogy

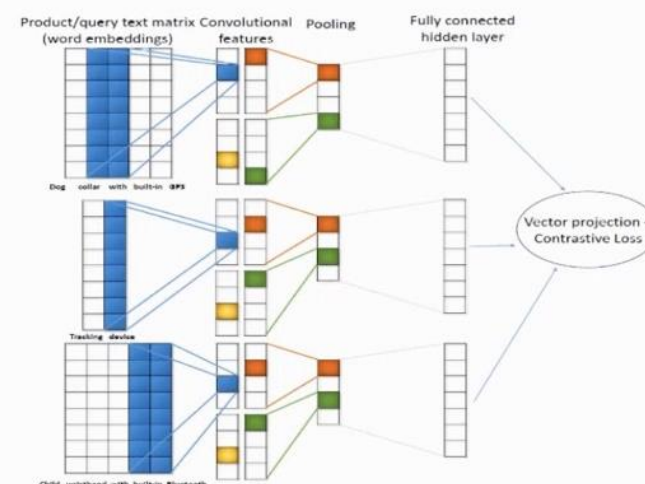
— Siamese Network

— Objective: Contrastive Loss

Incorporate user queries

— Compute product similarity relative to query

— Helps model focus on non-surface features



- *Same purpose, different mechanism.* Given the corpus of all products \mathcal{P} , a product i with (normalized) purpose and mechanism vectors $\mathbf{p}_i, \mathbf{m}_i$, and distance metrics $d_p(\cdot, \cdot), d_m(\cdot, \cdot)$ between purpose and mechanism vectors (respectively), solve:

$$\begin{aligned} \operatorname{argmin}_{\hat{i} \in \mathcal{P}} d_p(\mathbf{p}_i, \mathbf{p}_{\hat{i}}) \\ \text{s.t. } d_m(\mathbf{m}_i, \mathbf{m}_{\hat{i}}) \geq \text{threshold}, \end{aligned} \quad (1)$$

- *Same Mechanism, different purpose.* Solve:

$$\begin{aligned} \operatorname{argmin}_{\hat{i} \in \mathcal{P}} d_m(\mathbf{m}_i, \mathbf{m}_{\hat{i}}) \\ \text{s.t. } d_p(\mathbf{p}_i, \mathbf{p}_{\hat{i}}) \geq \text{threshold}, \end{aligned} \quad (2)$$

観念実験

本研究の類推モデルを使った方が、より

reference



観念実験の概要と抜粋：『携帯の充電器からの類推』

【①問題提起】携帯電話を充電器できるケースを再設計すること

- ・携帯電話
- ・充電器
- ・ケース

【②類推】①から類推されるものをデータベースから引き出す

- ・フラッシュチャージカラビナ
- ・バックアップバッテリー付属USB
- ・人的エネルギー発電機

【③人間の発想】②を受けて、人間が新商品を連想する

- ・ステップを追跡し、人の動きを使って充電できるケース
- ・複数のワイヤレス電源ハブがあり、どこにでもワイヤレスで携帯電話を接続して充電できるケース
- ・ワイヤレスキーボードで使用されているテクノロジーのように近くの光源から充電できるケース

Figure 3: Overview and excerpts of the ideation experiment. Top: Seed product. Workers were asked to solve the same problem in a different way. Middle: Top 3 inspirations for each of the conditions. Note that the TF-IDF baseline returns results from the same domain, while our method returns a broader range of products. Bottom: Ideas generated by users exposed to the different conditions.

量的検定結果： χ^2 検定

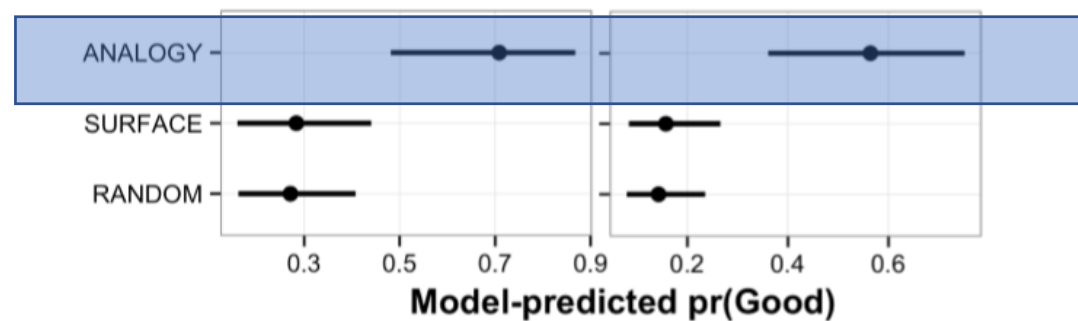


Figure 4: Showing proportion estimates by our random-effect logistic regression, for $k = 2$ (left) and $k = 3$ (right). Participants are significantly more likely to generate good ideas for the redesign ideation task when given inspirations from our analogy approach compared to baseline-surface and baseline-random approaches

Figure 4) 本研究モデルと他二つのモデルには有意な差がある

左図は、変量効果ロジスティック回帰による推定値を示す。先行研究モデルと比較して、本研究アナロジーアプローチからのインスピレーションを与えられたとき、参加者は再設計思想課題に対して良いアイデアを生み出す可能性が、かなり高い

Conclusion:

本研究モデルの類推は、人々が創造的なアイデアを生み出す可能性を著しく高めた。

Problem>>Farmer Work>>

問題) データベース上に大量、かつ乱雑なデータが蓄積されており、人間にとっても、機械にとっても類推することは難しい。
初期は、人間がリンクを手作業で作成していた。

先行研究) 教師ラベルを整えたデータを使い、ランダムサンプリングにより予測モデルを生成し、類推をする。
大規模な自然言語データセットに拡張できるが、構造的類似性を説明することが難しい。

本研究) 製品の説明から目的とメカニズムのベクトル表現を抽出するために、クラウドソーシングとリカレントニューラルネットワークを組み合わせたもので、製品の目的とその目的を達成するためのメカニズムを指定する、より単純な構造表現、具体的には「problem schemas」を学習することによって、より高い精度での類推を可能にした。

What is this thesis for?

大規模で、雑然とした現実の知識レポジトリ内の類推検索に関して、自動的に類推(アナロジー)をみつけることができるプロダクトの開発。

Where is an important point compared to previous researches?

教師学習ラベルなしで、自動的に探索して、類推をみつけることができる。

Where are the key points of technology and method?

- *Same purpose, different mechanism.* Given the corpus of all products \mathcal{P} , a product i with (normalized) purpose and mechanism vectors $\mathbf{p}_i, \mathbf{m}_i$, and distance metrics $d_p(\cdot, \cdot), d_m(\cdot, \cdot)$ between purpose and mechanism vectors (respectively), solve:

$$\begin{aligned} \operatorname{argmin}_{\hat{i} \in \mathcal{P}} d_p(\mathbf{p}_i, \mathbf{p}_{\hat{i}}) \\ \text{s.t. } d_m(\mathbf{m}_i, \mathbf{m}_{\hat{i}}) \geq \text{threshold}, \end{aligned} \quad (1)$$

- *Same Mechanism, different purpose.* Solve:

$$\begin{aligned} \operatorname{argmin}_{\hat{i} \in \mathcal{P}} d_m(\mathbf{m}_i, \mathbf{m}_{\hat{i}}) \\ \text{s.t. } d_p(\mathbf{p}_i, \mathbf{p}_{\hat{i}}) \geq \text{threshold}, \end{aligned} \quad (2)$$

How to verified whether it is valid?

Quirky.comの製品説明のデータを対象に、製品は何をするのか? 製品は、どこが良いのか? 製品はどのように機能するか? をラベル付けしたものと当研究モデルで比較した。

【質的研究】

→ 本研究モデル、先行研究モデル、ランダム検索モデルにより生成された言葉から影響を受けた案をアンケート調査し、案を比較する。

【量的検証】

→ 出た結果に対して、分散分析を行い、検定する。

Is there discussions?

製品の説明以外に他の分野に一般化するには拡張が必要である。
比較的単純な製品説明から科学論文のような、より複雑なデータに移行する場合。高次元の拡張を今後検討していくべきである。

Which reserches should I read next?**【先行研究】**

JoelChan, TomHope, DafnaShahaf, and AniketKittur. 2016.
ScalingupAnalogy with Crowdsourcing and Machine Learning.
In Workshop on Computational Analogy at ICCBR.

Acceleration Innovation Through Analogy Mining

Tom Hope, Joel Chan, Aniket Kittur, Defna Shahaf

https://scholar.google.com/scholar?as_q=&num=10&btnG=Search+Scholar&as_epq=Accelerating+Innovation+Through+Analogy+Mining&as_oq=&as_eq=&as_occt=any&as_sauthors=Hope

KDD/2017, Best Paper Award

Seeking in the vast Internet World

Summary

By Yohei Kawakami

2019/01/29

