

極大 2 部クリーク列挙アルゴリズムの実装とビジネスキャリア形成支援への応用

Implementation of Algorithm for Enumerating Cliques in Biparite Graphs and application of support formation of business career by using its

谷 研究室 重田 琴美
Kotomi Shigeta

概要

ある条件を満たす仕事や就きたい職種に就く為には、どのようなキャリアアップを計った方がいいのかいいのだろうか。本研究では、実際の求人データを極大 2 部クリーク列挙アルゴリズムにかけて、データマイニングし、就きたい仕事に就く為に必要となる条件や経験等のデータが得られないか、検討した。

1 はじめに

近年、一つの会社に定年までいるという事自体が珍しいものとなった。人生において何度も転職するのが当たり前の時代だ。転職サイトを見てみると、転職をする理由として「キャリアアップを計る」、「年収や勤務形態への不満」が大部分を占めているようだ。ここからほとんどの人が、今の仕事よりも条件の良い仕事に就く為に転職活動をしている事が分かる。本研究では、ある職種に求められるキャリアや資格、年齢、年収等の「条件」が入った求人データを集め、そのデータから、同じ「条件」を必要とする職種を引き抜き、その得られた「条件」と、職種等を比べる事によって、自分の就きたい職種に就く為にはどのような「条件」をクリアすれば就けるか、といった情報が手に入らないだろうか、検討してみる。もし、自分の就きたい職種に就く為の「条件」の目安等が分かれば、転職活動に多いに役立つだろう。今回のデータマイニングに使うアルゴリズムは極大 2 部クリーク列挙アルゴリズムである。今回参考にしたのは、国立情報学研究所の宇野教授のアルゴリズムである。アルゴリズム理解の際必要となってくるグラフ理論の定義と、基本となる定義を 2 章で紹介する。それから、3 章で極大 2 部クリーク列挙アルゴリズムの説明を行い、4 章で実際の集めた求人データをどのように 2 部グラフに適用するかを説明し、5 章で実験結果を記し、6 章で考察・まとめとする。

2 定義

2.1 グラフ理論

- 誘導部分グラフ：グラフ G の部分グラフで、 $V(G)$ のある部分集合 S と、両端点 u, v が S 中にあるような G の辺全体から得られるグラフをいう。

- クリーク：グラフ $G = (V, E)$ の頂点集合 S で任意の 2 つの頂点間に枝がある S を G のクリークという。
- 2 部グラフ：グラフ中の頂点 V を V_1 と V_2 に分割し、任意の V_1 の頂点の組、および任意の V_2 の頂点の組の間に枝無いグラフ $G = (V_1 \cup V_2, E)$ をいう。
- 2 部クリーク：2 部グラフ $G = (V_1 \cup V_2, E)$ の頂点部分集合 $S_1 \subseteq V_1, S_2 \subseteq V_2$ に対して S_1 の任意の頂点と S_2 の任意の頂点の間に枝があるとき、頂点部分集合 $S = S_1 \cup S_2$ をいう。また、 $S_1 = \emptyset, S_2 \subseteq V_2$ である頂点集合もまた 2 部クリークであり、またその逆もいえる。
また、 $S_1 = S \cap V_1, S_2 = S \cap V_2$ である。
- 極大 2 部クリーク：ある 2 部クリークが他の 2 部クリークに含まれないとき、その 2 部クリークの事をいう。

2.2 ここで使われる記号

- G_i ： G に対して頂点集合 $\{v_1, \dots, v_i\}$ によって誘導されるグラフ
- i 極大クリーク： G_i の極大クリーク
- $V_1 = \{v_1, \dots, v_{|V_1|}\}, V_2 = \{v_{|V_1|+1}, \dots, v_{|V_1|+|V_2|}\}$
- 自明なクリーク：2 部クリーク S が $S \subseteq V_1$ または $S \subseteq V_2$ のとき、 S を自明なクリークという。本研究で用いるアルゴリズムでは、自明なクリークは列挙しない。

- $N(S)$: 頂点集合 S の全ての頂点に隣接する頂点集合
- $l(S)$: 頂点集合 S の最小の添え字の頂点. S が空なら架空の頂点 v_{-1}
- $X^*(i, S)$
 $= S \cup N(S_2) \cup N(S \cup N(S_2) \cap \{1, \dots, i\})$
 任意の 2 部クリーク S を部分集合とする i 極大 2 部クリークを求めるアルゴリズム.
- $\bar{X}(S, v_i)$:
 S_1 から v_i と隣接しない頂点を除き, v_i を加える操作.

3 極大 2 部クリーク列挙アルゴリズム

国立情報学研究所、宇野助教授らの「極大 2 部クリークの高速列挙法とデータマイニングへの応用」[1]「大規模 2 部グラフに対する極大クリーク列挙アルゴリズムの改良と実装」[2]「2 部クリークを用いた closed item set の効率的」[3] を参考にした. これは深さ優先探索法により極大 2 部クリークを列挙するアルゴリズムである.

1. $i = |V_1| + 1, \dots, |V_1| + |V_2|$ の各頂点 v_i について,
 $S := \{v_i\}$ として 2 に進む.
2. $S := X^*(i, S)$ を求め, S_2 が i 以上の頂点のみからなるとき, S を出力し 3 に進む.
3. S_1 のある頂点に隣接し, S_2 に属さない i 以上の頂点集合 J を求め, 4 に進む.
4. J の集合が空になるまで以下の手順を繰り返す.
 $v_k := l(S_2 \cup J)$;
 If $v_k \in S_2$ then
 $S_2 := S_2 \cup \{v_k\}$;
 End if
 If $v_k \in J$ then
 $J := J \cup \{v_k\}$;
 5 に進む.
 End if
5. $S' := \bar{X}(S, v_k)$;
 If $S = X^*(k-1, S' \cup \{v_k\})$ then
 $S' := X^*(|V|, S')$ を求めて, 出力;
 新たに $i := k$, $S := S'$ として 3 に進む;
 End if

4 求人データを 2 部グラフにする

今回の実験で使われる求人データはキャリアデザインセンター (type@[4] という転職サイトと type という雑誌) からデータを収集した.

求人募集内容の条件を, 募集している会社の業界と募集職種, 年齢, 学歴, 年収幅, 経験していると望ましい業界や職種, 資格とし, データ収集を行った.

このデータに対して, 2 部グラフの V_1 の要素を各求人情報とし, V_2 の要素を各条件とし, 各求人情報とその求人情報の持つ条件とで隣接関係を持たせて 2 部グラフを構築する.

5 実験結果

<実験データの URL>

<http://www.tani.cs.chs.nihon-u.ac.jp/g-2006/kotomi/program/INPUT/all.xls>

集めた求人データを極大 2 部クリークの列挙を使用し, データマイニングを行った.

<実験結果 URL>

<http://www.tani.cs.chs.nihon-u.ac.jp/g-2006/kotomi/program/MCBG.txt>

6 考察・まとめ

- 同じ業界・職種を経験していて, ある一定の年齢幅だと, 似た様な年収結果となった.
- 27, 28 才が 1 番条件で多い.
- 年齢・年収の組み合わせが多い
- 年収・年齢・経験職種が揃ったデータが転職者にとって, 最も価値があると思われる.

7 参考文献

[1]: 「極大 2 部クリークの高速列挙法とデータマイニングへの応用」

<http://research.nii.ac.jp/uno/papers/03071a.pdf>

[2]: 「大規模 2 部グラフに対する極大クリーク列挙アルゴリズムの改良と実装」

<http://research.nii.ac.jp/uno/papers/03071a.pdf>

[3]: 「2 部クリークを用いた closed item set の効率的」

<http://research.nii.ac.jp/uno/papers/03071a.pdf>

[4]: 転職サイト@type

<http://type.jp/>