

抽象化した大規模グラフの描画法

古田 悠人[†] 阿部 昇[†]

[†]大阪電気通信大学 情報通信工学部

1. はじめに

グラフとは頂点と辺からなる数学的構造である。グラフ化することで、数値だけでは読み取れない情報を視覚化し理解しやすくすることができる。

本研究では、複数の小説作品の類似度をグラフ化することを考える。これを単にグラフ化しただけでは、そのグラフは完全グラフとなり、構造の理解が難しい。よって、本研究では k-means 法を用いて頂点の抽象化を行い、クラスカル法を基にした手法を用いることで重要な辺を抽出する。その後、川西法[1]を適用することで理解しやすいグラフを描画することを目指す。

2. k-means 法

本研究では、始めに k-means 法に基づいて頂点の抽象化を行う。この章では本研究における k-means 法について説明をする。

まず、k 個の代表点を選び、他の頂点をいずれかの代表点に属させる。このようにして得られる頂点の集合のことをクラスと呼ぶ。本研究では、k は 10 から 20 までの任意の数とする。より重要な頂点を代表点とするため、代表点は、それぞれの頂点から出ている辺の重みの平均値の大きい順に選ぶ。他の頂点は類似度を基準にいずれかの代表点に属させる。

3. グラフ描画に用いたデータ

本研究で使用するデータは、青空文庫の小説作品である。小説作品をそれぞれの頂点とし、各小説作品を結ぶ辺を作成する。

辺の重みは類似度であり、doc2vec という自然言語処理分野で広く用いられているツールを用いて測定した。

4. 提案手法

本研究では、2 つの手法について評価を行う。この章では、その 2 つの手法について説明する。

手法 1 は、k-means 法に基づいた手法で頂点の抽象化を行い、その後クラスカル法を用いて 1 つ目のスパニングツリーを作成する。次に、1 つ目のスパニングツリーで採用されなかった辺のみに対してクラスカル

法を用いて 2 つ目のスパニングツリーを作成する。そして両者を結合したグラフを作成した後、川西法を用いて再描画する。

手法 2 は、手法 1 で 1 つ目のスパニングツリーを作成した後、その中から重みが最小の辺 s を探す。1 つ目のスパニングツリーで、採用されなかった辺のみに対して、辺 s より大きい重みの辺をスパニングツリーに加えていく。その後、川西法を用いて描画する。

5. 実験結果

手法 1 の描画を図 1、手法 2 の描画を図 2 に示す。共にクラス数 k は 10 である。また、表 1 に描画の良さを示す評価値としてよく用いられる項目についての比較を示す。

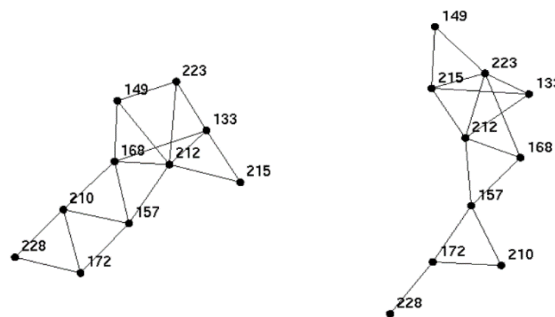


図 1. 手法 1 描画結果

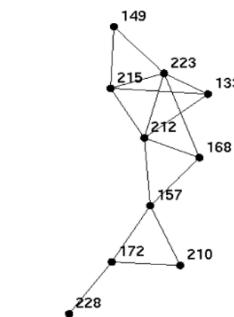


図 2. 手法 2 描画結果

表 1. 手法 1 と手法 2 の比較

	辺数	辺 交差数	辺長 分散	頂点と辺 の近接
手法 1	18	2	0.12	0
手法 2	16	3	0.11	0

表 1 から、手法 1, 2 に大きな差はないが、手法 2 の方が辺数は少なく、辺の重みも大きい。どちらの手法も頂点の抽象化、辺の抽出を行い良い描画が得られたと考えられる。

6. 参考文献

[1]川西・増田・山口:”2 種類の理想距離による Eades のグラフ描画法の改良” 電子情報通信学会論文誌 vol. J83-A, No. 9pp. 1117-1121, 2000.