

機械学習を用いた集団における影響予測
Influence prediction in the group using machine learning

槻瀬 誠
Makoto TSUKISE

1. はじめに

農村地域では農業用施設の維持管理や農村住民と農村地域外の住民との交流、水利慣行などの共同活動が行われており、近年その重要性が注目されている。これらの共同活動を行う集団内における対話が個人の心理的状況に及ぼす影響は、社会科学や心理学などの分野において研究が行われており、定量的な影響予測については確率による影響モデルが提案されている。一方、近年、ノードとエッジから構成される人間関係や分子構造などのグラフ構造を対象とした機械学習手法が複数提案されており、グラフ構造自体の予測やグラフ内のノード又はエッジの状態を予測することなどが可能であるとされている。そこで、本実験では集団の例として、漫画「みゆき」（あだち充著、小学館刊）（以下、物語）における主要登場人物について、確率による影響モデルを基に機械学習を行い、各人物の状態の予測を試みた。

2. 方法

(1) 実験対象グラフの選定 物語における主要登場人物の男性 6 名、女性 2 名から構成される集団におけるグラフを採用した。各人物をノードとし、ノードの特徴を物語において各人物が好意を寄せる人物とした。また、他の人物への発言をエッジと定め、エッジを発言元、発言先及び発言回数を持つ重み付き有向エッジとした。物語全 1 2 巻中の人物の発言を集計し(Table.1)、ノードとエッジを整理してグラフ構造を作成した。(Fig.1)

発言先 (発言回数 (回))

登場人物名	若松真人	若松みゆき	鹿島みゆき	間崎 竜一	中田 虎夫	鹿島 安次郎	香坂 健二	沢田 優一	村木 好夫	特徴 (物語中で好意を寄せる人物)
若松真人	0	605	323	69	2	19	25	0	99	鹿島みゆき
若松みゆき	657	0	37	107	60	38	0	0	8	若松真人
鹿島みゆき	335	42	0	4	1	4	18	0	1	若松真人
間崎 竜一	106	146	5	0	22	0	0	0	2	若松みゆき
中田 虎夫	9	90	1	29	0	4	3	0	0	若松みゆき
鹿島 安次郎	25	58	8	0	0	0	0	0	0	若松みゆき
香坂 健二	37	0	36	0	1	0	0	0	0	鹿島みゆき
沢田 優一	0	0	0	0	0	0	0	0	0	若松みゆき
村木 好夫	142	8	5	2	0	0	1	0	0	村木 好夫

発言元

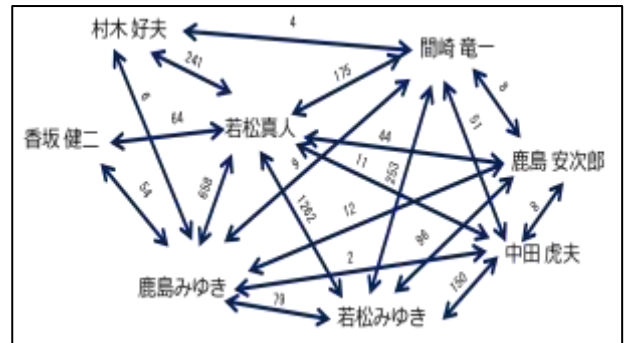


Table.1 主要登場人物間発言回数表
Statement between main characters

Fig.1 主要登場人物グラフ(第 1 0 巻まで)
Graph of main characters in the story

(2) 説明変数及び目的変数の選定 説明変数を各ノードにおけるエッジとし、目的変数をノードの特徴とした。全 1 2 巻のうち、第 1 巻から第 1 0 巻までの説明変数及び目的変数（学習データ）で機械学習を行い、第 1 1 巻及び第 1 2 巻における説明変数を用いて第 1 1 巻から登場する人物におけるノードをテストデータとして特徴を予測すると共に各ノードの特徴の変化を確認した。

(3) 機械学習の手法 この研究における機械学習の手法は、グラフ構造を対象とした機械学

農林水産省近畿農政局, Ministry of Agriculture, Forestry and Fisheries Kinki Regional Agricultural Administration Office

Keyword 機械学習, グラフ, 予測

習手法である R-GCN の手法を採用した。各ノードの特徴予測については、Influence Model を参考に条件付き確率モデルを設定した。機械学習モデルの構成は、入力層を各ノードのエッジとし、隠れ層を1層、出力層をノードの特徴を表す5種類のカテゴリにおける確率分布とした。エポック（学習回数）は1巻あたり100回とした。学習データにおける第1巻から第7巻までを訓練データとし、各巻毎に訓練を行った。第8巻から第10巻を検証データとし、学習後のモデルを用いて第11巻及び第12巻のデータによる各ノードの特徴を予測した。

3. 結果および考察

実験結果を Table.2 および Table.3 に示す。第1巻から第7巻まで、学習を進めるに伴って正解率が上昇し、学習後の訓練データ及び検証データにおける正解率が共に1.0となった。第11巻および第12巻のデータを用いた予測では、新たな登場人物の特徴を正しく予測することができ、他のノードにおいても特徴の変化が生じた。また、ノードの特徴はある特定の値に集中する傾向が見られた。これは本実験に用いたグラフが比較的小規模なもので、学習ノードが少数であったこと、エッジの重みが特定のノードに偏る傾向があったことが原因と考えられる。

	巻数	平均正解率 (100ステップ当たり)
訓練データ	1	0.69
	2~5	0.86
	6	0.89
	7	0.97
検証データ	8~10	1.0

Table.2 各巻における平均正解率の変化

Change of average accuracy rate in each volume

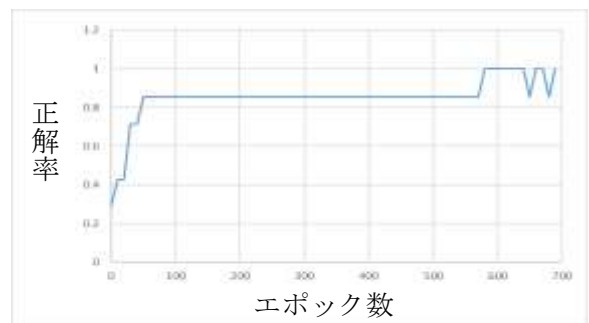


Fig.2 訓練における正解率の変化

Change of the values of accuracy in training

ノード (登場人物名)	特徴		備考
	予測前 (第1巻から第10巻までの 物語中設定)	予測後 (第11巻及び第12巻データ による予測)	
若松真人	鹿島みゆき	若松みゆき	
若松みゆき	若松真人	若松真人	
鹿島みゆき	若松真人	若松真人	
間崎 竜一	若松みゆき	若松みゆき	
中田 虎夫	若松みゆき	若松みゆき	
鹿島 安次郎	若松みゆき	若松みゆき	
香坂 健二	鹿島みゆき	鹿島みゆき	
沢田 優一	—	若松みゆき	※1
村木 好夫	村木 好夫	若松真人	※2

Table.3 予測前及び予測後における各ノードの特徴

Features of each node in before and after prediction

- ※1：第11巻から登場のため予測前の特徴なし
- ※2：物語中で好意を寄せる人物が設定されていないため予測前特徴を自己に設定

4. おわりに

本実験によって、集団内の人間関係をグラフ構造として整理し、確率モデルを基にした機械学習により客観的データを基に集団における各個人間の影響予測を行う可能性を示した。本手法を応用することで農村における様々な共同活動についても定量的又は定性的な予測が可能と考える。また、今後は水利システムなど、集団以外のグラフ構造への応用も期待される。今回の実験では静的なグラフ構造を対象としたが、今後は動的に変化するグラフ構造を扱う手法の開発も必要と考える。