

ビジネスチャットアプリ上の応答間隔に対する潜在クラスモデル分析

1X16C0062-7 齊藤 芙佑
指導教員 後藤 正幸

1. 研究背景と目的

近年、多くの企業では、簡潔な文章で密なコミュニケーションが可能なビジネスチャットアプリが浸透しつつあり、これらのコミュニケーションデータを人事や組織構築に活用する動きがある。チャットは電子メールと異なり、受信から返信までの時間（以下、返信時間）に、ユーザのコミュニケーション特性が現れやすい。また、発言者が受信者と築きたいと望む関係性によっても大きく変化し、ユーザ間の関係性も影響を与える可能性が高い。特に、相手との情報共有の重要度が高いほど迅速な返信を行うなど、返信時間にはビジネス展開に大きな影響を与えるコミュニケーション領域での特性が表れているはずである。そのため、関係を築き得る最小単位である二者間のコミュニケーション特性について、ビジネスチャットアプリ上のデータを活用して返信時間の観点から分析ができれば、職場のチーム作りや組織マネジメントにも有用と考えられる。

複数のユーザが参加可能で返信時間の傾向がユーザごとに異なるチャットでは、様々なコミュニケーション特性を持つユーザが混在している。このような統計的性質の異なるグループの集合体で与えられるデータに対する有効な分析手法として、潜在クラスモデル [1],[2] が知られている。このモデルは、観測データ間の背後に潜在的な変数の存在を仮定して、データ間の関係性をモデル化することで、高次元データを潜在クラスという軸での解釈を可能にする。また、発言者と受信者が複数の異なる潜在クラスに所属する度合いを所属確率で表すことで、各ユーザの相手の影響による多面的なコミュニケーション特性を表現する。

以上より本研究では、発言者、受信者、返信時間の関係性を定量的に表現する潜在クラスモデルを提案し、返信時間の観点からコミュニケーション特性を分析する方法論を示す。提案モデルの有用性を示すため、提案モデルを実データに適用し、得られる結果について考察を与える。

2. 準備

2.1. ビジネスチャットアプリ

ビジネスチャットアプリは、ビジネス上での利用が想定されたコミュニケーションツールで、近年多くの企業で導入されている。場所と時間を限定せず同時に複数人と簡潔な文章を用いた手軽な情報共有が可能である。従来多用されてきた電子メールより双方向での効率的な情報共有が可能のため、コミュニケーション量が大幅に増え、ビジネスの早い展開が期待できる。

2.2. ピープルアナリティクス

従来の人事領域では、人事部門の社員の主観的な判断に頼った意思決定も多く行われており、それが原因で発生する問題も非常に多かった。これに対し、近年では、社員の労働やコミュニケーションのログを収集することが技術的に可能と

なっており、これらの収集した情報を分析し人事に活用しようとするピープルアナリティクスが注目されている。事実と科学に基づく分析により客観的な判断材料が得られ、公正な意思決定に繋がること期待されている。

2.3. 問題設定

近年、チャットデータでのピープルアナリティクスに基づく企業向けの分析サービスが提供されるようになっており、実際に様々な企業で意思決定に活用されている。しかし、その指標は会話数など単純な統計量であることが多く、チャットの特徴を活かした研究が望まれている。チャットの特徴の「手軽さ」を考慮すると、会話が迅速に展開可能な条件下で「返信に前向きであるか」、あるいは「緩慢であるか」は大きな差異であるため、返信時間には相手との関係性が特に現れやすいと考えられる。本研究ではこの点に注目し、返信時間という切り口からチャットの特徴を生かした分析モデルを提案する。

3. 提案手法

3.1. 提案モデルへの着想

上述の通り、ビジネスチャットアプリ上の返信時間には、二者間の関係性が表れやすい。例えば、双方の返信時間が均等なユーザの組ほど関係構築への意識が一致している関係であると解釈できる。さらに、迅速に返信を行うユーザほど会話の展開を求めていると解釈できる。そこで、会話の返信時間を分析し二者間の関係性について評価する。なお、ユーザ A からユーザ B に発言があったもとの、その直後の B から A への発言を、「B の A への返信」と見なし、返信時間は「A から B への発言」から「B の A への返信」までの所要秒数とする。データ数は膨大なものとなり、発言者・受信者・返信時間の関係性を直接解釈することは困難である。そこで、それらを潜在クラスを通して紐付け、共起関係を潜在クラスにより低次元データで表現することで、相互に影響しあう発言者と受信者の関係性を解釈するモデルを提案する。

3.2. 提案モデル

提案モデルでは、発言者、受信者、返信時間の共起関係を表現し、提案モデルは「任意の発言者が任意の受信者に対してどのような返信を行うか」に関する分析を可能にする。

いま、発言者集合を $\mathcal{T} = \{t_l : 1 \leq l \leq Q_t\}$ 、受信者集合を $\mathcal{R} = \{r_m : 1 \leq m \leq Q_r\}$ 、返信時間を s 、潜在クラス集合を $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$ とする。このとき、確率モデルは以下の式 (1) で定義される。

$$p(t_l, r_m, s) = \sum_{k=1}^K p(z_k) p(t_l | z_k) p(r_m | z_k) p(s | z_k) \quad (1)$$

このモデルの生起事象 (t_l, r_m, s) は、発言者 t_l から受信者 r_m に返信時間 s の返信がされたことを意味する。ここで、 $p(t_l | z_k), p(r_m | z_k)$ には多項分布を、返信時間 s には対数正

規分布を仮定する。ここで、 μ_k, σ_k^2 をそれぞれ k 番目の潜在クラスにおける平均と分散とする。

3.3. 学習アルゴリズム

N 件の発言者・受信者・返信時間データのうち、 n 番目の発言者を $a_n \in \mathcal{T}$ 、その時の受信者を $b_n \in \mathcal{R}$ 、返信時間を c_n とすると、 n 番目の発言者・受信者・返信時間データはこれらの共起 (a_n, b_n, c_n) で表現できる。このとき、各パラメータは対数尤度関数を最大化するように EM アルゴリズムを用いて推定する。

4. 実データ分析

4.1. 実験条件

本章では、Laboratik 社より提供された某企業の職場における Slack の発言者・受信者・返信時間のデータに対して提案モデルを適用することにより、モデルの有用性を検証する。分析対象期間は 2018 年 11 月 2 日から 2019 年 9 月 24 日である。データ数は $N = 97,536$ 、発言者数 $Q_t = 54$ 、受信者数 $Q_r = 54$ である。なお、 $l = m$ ならば t_l と r_m は同一人物である。潜在クラス数 K については、解釈性の高いクラス数を探索的に求め $K = 10$ とした。また返信時間の計算は、営業時間外に起因する返信時間の遅延の影響を除くため、1 日単位に区切り行う。

4.2. 潜在クラスの解釈

各潜在クラスについて、推定されたパラメータ μ_k を秒換算した $\exp(\mu_k)$ を横軸、 σ_k^2 を縦軸に示したグラフを図 1 に示す。なお、円の大きさは推定確率 $p(z_k)$ の値に比例する。

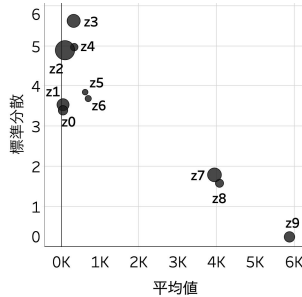


図 1: 各潜在クラスに対する推定結果

μ_k が小さいほど迅速な、 σ_k^2 が小さいほど安定した返信を行う潜在クラスと解釈でき、 $p(z_k)$ が大きいほどその特徴を持つ組が多く存在する。図 1 より、返信の速さと安定性は負の相関を持つ。円同士の距離は潜在クラスの類似度を現しており、迅速かつ不安定な返信を行う組が多いことがわかる。

4.3. 二者間の関係性の解釈

まず、発言者 t_l から受信者 r_m への一方向で行う返信の特徴を解釈する。 t_l が r_m に対して行う返信の特徴が潜在クラス z_k である確率は、 t_l が z_k の特徴を持ち、かつ r_m も z_k の特徴を持つ確率である。 t_l と r_m の潜在クラスに対する同時所属確率は式 (2) で求められる。

$$p(z_k | t_l, r_m) = \frac{p(t_l, r_m | z_k) p(z_k)}{p(t, r)} \quad (2)$$

また、同時所属確率 99% 以上の組の一部抜粋を表 1 に示す。

表 1 より、 t_l が r_m に行う返信の特徴解釈が可能となる。例えば、 t_{26} と r_{12} の同時所属確率が最大となる潜在クラス

表 1: 同時所属確率 99% 以上の組と潜在クラス (一部抜粋)

(t_l, r_m)	z_k	(t_l, r_m)	z_k
(t_0, r_{26})	z_2	(t_{26}, r_0)	z_2
(t_{47}, r_6)	z_3	(t_5, r_{47})	z_3
(t_{28}, r_5)	z_8	(t_{26}, r_{12})	z_2

は z_2 であることから、 t_{26} は r_{12} に対して z_2 の特徴である不安定だが迅速な返信をする傾向があると解釈できる。このような返信の特徴解釈を、 t_l と r_m を入れ替えて双方向で行うことにより、以下に示すようにその組の関係性の把握することが可能となる。

(1) 双方向の関係性の特徴に差がないユーザの組

表 1 で太字で表記された組のように、同時所属確率が最大となる潜在クラスが双方向で共通している組は、高い確率で共通する潜在クラスの特徴を持つ均衡な返信を、お互いが行う関係を築いている組である。例えば、 t_0, r_{26} と t_{26}, r_0 の同時所属確率が最大となる潜在クラスは z_2 であることから、この組は双方が比較的安定でありつつも迅速な返信を行っていると解釈できる。

(2) 双方向の関係性の特徴に差があるユーザの組

同時所属確率が双方向で最大となる潜在クラスが異なる組は、返信の特徴に差があり、会話の展開への貢献に対する意識に差がある組である。例えば、 $(t_5, r_{51}), (t_{51}, r_5)$ の同時所属確率が最大となる潜在クラスと同時所属確率を表 2 に示す。図 1 より、 μ_k, σ_k^2 共に $z_k = z_1$ のときより $z_k = z_2$ の方が大きいことから、ユーザ 51 のほうがユーザ 5 よりも安定的に迅速な返信をしていると解釈できる。

表 2: ユーザ 5 とユーザ 51 の同時所属確率

(t_l, r_m)	z_k	確率	(t_l, r_m)	z_k	確率
(t_5, r_{51})	z_1	0.67	(t_{51}, r_5)	z_2	0.79

5. 考察

4.2 節に示したように、提案モデルにより企業全体における返信の速さと安定性の傾向と、各特徴を持つ返信を行う組の割合を定量的に評価することが可能となった。また、4.3 節のように、社員の組としての特徴抽出が可能となり、これは組織改善の施策の有用な情報となると考えられる。例えば、企業にコミュニケーション不足が問題視されているチームが存在する場合、安定性のある迅速な返信を行うメンバーを配置することで、会話の展開を促し、職場チームの雰囲気改善を行うことが期待できる。

6. まとめと今後の課題

本研究では、発言者・受信者・返信時間の共起関係を分析するモデルを提案した。また、実データに提案モデルを適用し、各潜在クラスへの所属確率を用いた二者間の関係性の分析を行い、考察を行うことで提案モデルの有用性を示した。

今後の課題としては、言語データなど他の条件も考慮したモデルの修正や、実際には会話が行われていない二者間についての関係性予測などが挙げられる。

参考文献

- [1] Bishop, C.: *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 2nd printing, Springer (2010)
- [2] Hohmann, T.: Probabilistic Latent Semantic Analysis, *Proc. UAI '99*, pp.289-296 (1999).