

# 離散イベント時系列からの潜在的な競争の推定

## Estimation of Latent Competition from Discrete Event Time Series

宮澤脩一 \*1\*2 持橋大地 \*3

Shuichi Miyazawa Daichi Mochihashi

\*1総合研究大学院大学

The Graduate University for Advanced Studies

\*2AGC 株式会社

AGC Inc.

\*3統計数理研究所

The Institute of Statistical Mathematics

To investigate the industrial market, we analyze patent information to estimate the latent competitive relationships between industrial firms. In existing studies, text analysis methods are often used for patent analysis to detect correlations between technologies and companies. However, existing methods cannot capture the competitive relationships between companies that might provide deeper insights for technology marketing. To solve this problem, we propose a model named Ordinary Differential Equation integrated Log Gaussian Cox Process (ODE-LGCP) that combines the ODE systems, for modeling competition in this study, with generative processes of event data. In the experiment, we use the real patent data and show that our proposed model can estimate competitive relationships within companies.

### 1. はじめに

特許は産業界の企業間競争において企業が自社技術を保護する最も基本的な方法であり、特許データの解析は産業界の技術トレンドを捉えたり、重要な技術を保有する企業を発見するのに有用である。特許情報を技術マーケティングに活用するための解析は多くの文献があり、例えば、[Tang 12] は文書の潜在トピックを推定する確率的トピックモデルにより、産業界の技術トレンドや各技術テーマに関連する企業の分析を行っている。しかし、既存手法は技術-企業の相関関係は推定できるものの、産業界の構造に対するより深い洞察を与えるであろう企業間の競争関係を推定することができない。そのため本稿では、イベントデータの生成過程に常微分方程式 (ODE: Ordinary Differential Equation) によってモデル化された競争システムを仮定する ODE 対数ガウス Cox 過程 (ODE-LGCP) を提案する。提案モデルによって、ODE システムの逆問題、すなわち観測データから ODE システムパラメータを推定する問題を近似的に解くことができ、本稿においては、企業の特許出願イベント時系列データから、企業間の競争関係をモデル化する ODE システムのパラメータを定量的に推定する。

図 1 に提案モデルの概要を示す。ODE システムの逆問題は、通常は離散時系列上で連続的に観測される数値データを扱うが、本研究では、データは連続時系列上で離散的にしか観測されず、なおかつ数値のないイベントデータである。そこで、ガウス過程により補完を行い、観測のない時刻も含めて潜在的な関数を推定し、その関数を支配する ODE を求める。提案モデルは、対数ガウス Cox 過程を用いてイベントの生成過程をモデル化し、イベントの生成されやすさを表す強度関数に ODE に支配された潜在時系列を用いることで、離散観測イベントデータから ODE パラメータを推論することを可能にする。

ODE システムの逆問題の解法については、ODE パラメータを所与として数値積分でデータをシミュレートし、観測データを再現するようなパラメータを探索するのが基本的なアプローチであるが、このアプローチは計算量が多く、大規模問題へ適用できない。計算量の問題への対応策として、勾配マッチングと呼ばれるベイズ的手法が近年成功を取っており、本稿に

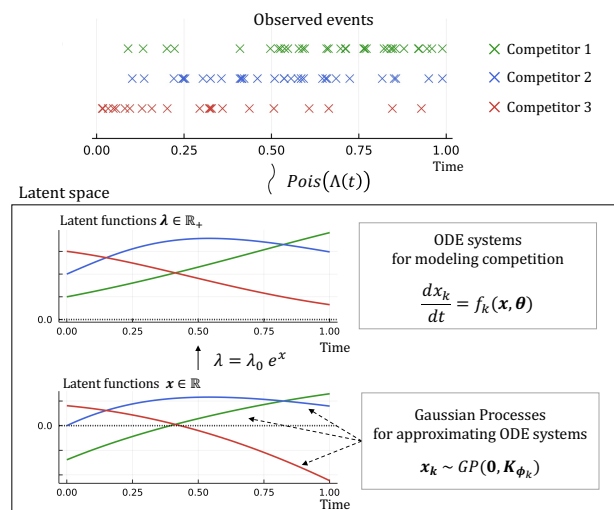


図 1: 提案モデルの概要. ODE ダイナミクスを近似する潜在時系列を用いた離散イベントの生成モデルであり、未知の ODE パラメータを離散イベントから推論することを可能とする。

においても、勾配マッチングの一種である適応的勾配マッチング (AGM: Adaptive Gradient Matching) [Dondelinger 13] を用いて ODE パラメータの推定を行う。

実験では、米国特許庁 (USPTO) の公開する特許データを用いて、産業界の特定の技術分野における企業間競争関係を推定する。特許を分類するため国際的に統一された分類である国際特許分類 (IPC) を用い、特許母集団を作成する。さらに、発明の要約テキストに対してトピックモデルを用い、母集団内の技術的なトピックを推定後、特定のトピックに関連する出願イベント時系列データに対し提案モデルを適用することで、細分化された技術テーマごとに企業間競争関係を推定する。推定された ODE パラメータによって、競争の観点で注目すべき企業間関係が発見されることを示す。

### 2. 予備知識

競争をモデル化するための ODE である Lotka-Volterra 競争方程式並びに提案モデルの前提となる適応的勾配マッチング

連絡先: 宮澤脩一, 総合研究大学院大学, miyazawa@ism.ac.jp

(AGM) および対数ガウス Cox 過程 (LGCP) を紹介する.

## 2.1 Lotka-Volterra 競争方程式

ODE はシステムの振る舞いをモデル化する有用なツールで幅広い学問分野で応用されている. 本稿では, 競争をモデル化する ODE として, 数理生物学において代表的な, 生物種間の競争関係を記述する Lotka-Volterra 競争方程式 [Smale 76] を用いる. Lotka-Volterra 競争方程式は,  $K$  種の生物種の競争下における  $i$  種の個体数  $p_i$  の時間変化を下式で表す.

$$\frac{dp_i}{dt} = r_i p_i \left( 1 - \frac{\sum_{j=1}^K a_{i,j} p_j}{\eta_i} \right). \quad (1)$$

ここで,  $r_i, \eta_i$  は種  $i$  の内的自然増加率, 環境収容力と呼ばれる正の定数,  $a_{i,j}$  は, 種  $j$  の個体数が種  $i$  の個体数増加率に与える影響の大きさを示す競争係数と呼ばれる正の定数である. 特に,  $a_{i,i} = 1 \forall i = 1, \dots, K$  であり,  $a_{i,j}$  は種  $i$  の内的競争の影響の大きさに対する種  $j$  の影響の大きさの比を示す.

$p_i$  は種  $i$  の個体数を表す非負の変数であるが, 本稿では  $x_i \in \mathbb{R}$  とし,  $p_i = e^{x_i}$  と書き換えて, ODE を実数変数を用いて定義しなおす. すると,  $x_i$  についての導関数は,  $\theta = (r, \eta, \alpha)$  と ODE パラメータをまとめて以下の通り表せる.

$$\dot{x}_i = f_i(\mathbf{x}, \theta) = r_i \left( 1 - \frac{\sum_{j=1}^K a_{i,j} e^{x_j}}{\eta_i} \right). \quad (2)$$

## 2.2 適応的勾配マッチング

勾配マッチングは, ODE システムの逆問題のベイズ的アプローチによる近似解法であり, 計算量の多い数値積分を行うことなく ODE パラメータ  $\theta$  を推定する. 時系列観測データ  $\mathbf{Y}$  に対し, 潜在的な回帰曲線  $\mathbf{X}$  をフィットすることで ODE ダイナミクスを近似し, 回帰曲線の勾配および ODE の導関数の値が近い値を取るような確率モデルを定義する. [Dondelinger 13] は, 回帰にガウス過程を用いる適応的勾配マッチング (AGM: Adaptive Gradient Matching) を提案した.

AGM では,  $K$  個の式から ODE システムが構成されるとすると, モデルの同時分布を回帰曲線の勾配  $\dot{\mathbf{X}}$  を周辺化したものとして以下のように定義する.

$$p(\mathbf{Y}, \mathbf{X}, \theta, \phi, \sigma, \gamma) = p(\mathbf{Y}|\mathbf{X}, \sigma) p(\sigma) p(\mathbf{X}, \theta, \phi, \gamma), \quad (3)$$

$$p(\mathbf{X}, \theta, \phi, \gamma) = \int p(\dot{\mathbf{X}}, \mathbf{X}, \theta, \phi, \gamma) d\dot{\mathbf{X}}, \quad (4)$$

$$p(\dot{\mathbf{X}}, \mathbf{X}, \theta, \phi, \gamma) = p(\mathbf{X}|\phi) p(\theta) p(\phi) p(\gamma) \times \prod_{k=1}^K p(\dot{\mathbf{x}}_k | \mathbf{X}, \theta, \phi, \gamma). \quad (5)$$

ここで,  $\phi$  はガウス過程のカーネル関数  $k(t, t')$  のパラメータ,  $\gamma$  は ODE 勾配の確率モデルのノイズの分散を表す. さらに, (5) 式に含まれる勾配の条件付き確率分布を, ガウス過程回帰曲線の勾配の確率モデルと ODE に従う勾配の確率モデルを Product-of-Experts [Hinton 02] の枠組みで統合した確率モデルとして以下のように定義する.

$$p(\dot{\mathbf{x}}_k | \mathbf{X}, \theta, \phi, \gamma) \propto p(\dot{\mathbf{x}}_k | \mathbf{x}_k, \phi) p(\dot{\mathbf{x}}_k | \mathbf{X}, \theta, \gamma) = \mathcal{N}(\dot{\mathbf{x}}_k | \mathbf{g}_k, \Psi_k) \mathcal{N}(\dot{\mathbf{x}}_k | \mathbf{f}_k(\mathbf{X}, \theta), \gamma \mathbf{I}). \quad (6)$$

ここで, 右辺一つ目の確率項はガウス過程回帰曲線の勾配  $\dot{\mathbf{x}}_{\text{GP}}$

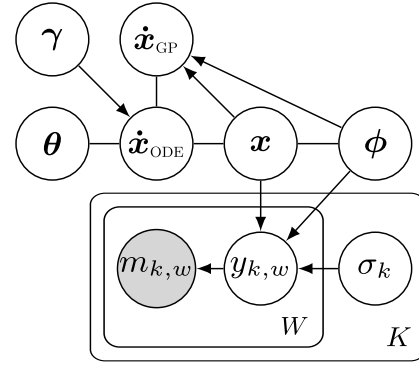


図 2: ODE 対数ガウス Cox 過程推論モデルのグラフィカルモデル.

の確率モデルを, 二つ目の確率項は ODE の勾配  $\dot{\mathbf{x}}_{\text{ODE}}$  の確率モデルを表しており, 勾配を評価する時刻において, 回帰曲線が ODE の導関数  $f(\mathbf{X}, \theta)$  の制約に従った勾配を持つよう  $\mathbf{X}, \theta$  が学習される. なお,  $\mathbf{g}_k, \Psi_k$  は  $k$  個目のガウス過程回帰曲線の勾配の平均および共分散であり, 下式により得られる.

$$\mathbf{g}_k = \mathbf{K}'_{\phi_k} \mathbf{K}_{\phi_k}^{-1} \mathbf{x}_k, \quad \Psi_k = \mathbf{K}''_{\phi_k} - \mathbf{K}'_{\phi_k} \mathbf{K}_{\phi_k}^{-1} \mathbf{K}'_{\phi_k}. \quad (7)$$

ここで,  $\mathbf{K}_{\phi_k}$  は  $\mathbf{x}_k$  の共分散,  $\mathbf{K}'_{\phi_k}$  は  $\mathbf{x}_k$  およびその導関数の交差共分散,  $\mathbf{K}''_{\phi_k}$  は  $\mathbf{x}_k$  の導関数の自己共分散を示す.

## 2.3 対数ガウス Cox 過程

時系列上のイベントの発生時刻について, イベントの発生率が時刻によって変動する場合の確率過程として Cox 過程がよく知られている. 時系列上の区間  $A$  におけるイベント発生数は, 時刻  $t$  における発生率を  $\lambda(t)$  として,  $\Lambda(A) = \int_A \lambda(t) dt$  をパラメータとするポアソン分布  $Po(\Lambda(A))$  に従うと考える. ここで, 時間によって値が変動する未知の強度関数  $\lambda(t)$  をデータから推定するために関数生成器としてガウス過程を仮定した確率過程を対数ガウス Cox 過程 [Møller 98] という. 非負の強度関数を実数空間上で定義される関数  $f(t)$  によって表すため, 指数変換を用いて  $\lambda(t) = \lambda_0 \cdot e^{f(t)} (\geq 0)$  とする. ここで,  $\lambda_0$  は区間内の平均的な発生率とする.

## 3. ODE 対数ガウス Cox 過程

### 3.1 生成モデル

ODE 対数ガウス Cox 過程は, ODE ダイナミクスを近似する潜在時系列関数を用いた対数ガウス Cox 過程, すなわち ODE に従った時系列上のイベントデータの生成モデルである.  $\theta$  をパラメータとする  $K$  個の式から構成される ODE システムの下で,  $k$  個目の導関数に従った時系列関数を  $F_k(t)$ , 区間内平均発生率  $\lambda_0^{(k)}$  として, 強度関数を  $\lambda_k(t) = \lambda_0^{(k)} F_k(t)$  と定義する Cox 過程から  $k$  種目のイベントデータを生成する.

### 3.2 推論モデル

図 2 に提案手法の推論モデルに関するグラフィカルモデルを示す.  $\phi_k$  をパラメータとするガウス過程から生成する潜在時系列  $\mathbf{x}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_{\phi_k})$  を用いて  $F_k(t) \approx e^{\mathbf{x}_k}$  と近似することを考える. そのために, (4)–(6) 式の AGM の枠組みを用い,  $\mathbf{x}_k$  の回帰曲線の勾配  $\dot{\mathbf{x}}_k$  が, (2) 式のように実数変数の導関数に変換した ODE の勾配  $f_k(\mathbf{x}, \theta)$  とマッチするような尤度を導入し, 潜在時系列  $e^{\mathbf{x}_k}$  が ODE に従うように誘導する.

通常の AGM は観測時系列データ  $\mathbf{y}$  に対し, ガウス過程で回帰関数をフィッティングするが, 提案モデルでは観測時系列では

なく潜在時系列  $\mathbf{y}$  に対するフィッティングを行う。  $\mathbf{x}_k, \mathbf{y}_k$  は、それぞれ有限の時点数  $T, W$  の時刻  $\boldsymbol{\tau} = \{\tau_t\}_{t=1}^T, \boldsymbol{\omega} = \{\omega_w\}_{w=1}^W$  においてのみ評価するのものとし、潜在時系列  $\mathbf{y}_k$  は潜在時系列  $\mathbf{x}_k$  のガウス過程の予測分布に標準偏差  $\sigma_k$  の出力ノイズが加えて以下のように生成する。

$$\begin{aligned} \mathbf{y}_k &\sim \mathcal{N}(\hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k), \\ \hat{\boldsymbol{\mu}}_k &= \mathbf{K}_{xy}^\top \mathbf{K}_{xx}^{-1} \mathbf{x}_k \\ \hat{\boldsymbol{\Sigma}}_k &= \mathbf{K}_{yy} - \mathbf{K}_{xy}^\top \mathbf{K}_{xx}^{-1} \mathbf{K}_{xy} + \sigma_k^2 \mathbf{I} \end{aligned} \quad (8)$$

ここで、  $\mathbf{K}_{xx}, \mathbf{K}_{xy}, \mathbf{K}_{yy}$  は、それぞれカーネル関数値  $k(\tau_t, \tau_{t'})$ ,  $k(\tau_t, \omega_w)$ ,  $k(\omega_w, \omega_{w'})$  を要素とする行列である。こうして生成された実数空間上の潜在時系列  $\mathbf{y}_k$  を指数変換すると、  $F_k(t) \approx e^{\mathbf{y}_k}$  と近似でき、生成モデルで用いる対数ガウス Cox 過程の強度関数が得られる。

対数ガウス Cox 過程の学習のため、連続時間を時間間隔  $\Delta$  の  $W$  個の時間窓に離散化し、それぞれの窓でポアソン分布の尤度を評価する。  $w$  番目の時間窓の中間時刻  $\omega_w$  における  $k$  種の潜在状態を  $y_{k,w}$ 、同窓内における  $k$  種のイベント頻度を  $m_{k,w}$  とし、各時間窓における平均イベント頻度  $\lambda_0^{(k)} = \frac{\sum_w m_{k,w}}{\Delta}$  を用いて  $\text{Pois}(m_{k,w}; \lambda_0^{(k)} e^{y_{k,w}})$  の尤度を評価する。

まとめると、提案モデルの同時分布は下式の通り表される。

$$\begin{aligned} p(\mathbf{m}, \mathbf{y}, \mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma}) &= p(\mathbf{m}|\mathbf{y})p(\mathbf{y}|\boldsymbol{\sigma}, \boldsymbol{\phi})p(\boldsymbol{\sigma}) \\ &\quad \times \int p(\dot{\mathbf{x}}, \mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma}) d\dot{\mathbf{x}}, \\ p(\mathbf{m}|\mathbf{y}) &= \prod_k \prod_w \text{Pois}(m_{k,w}; \lambda_0^{(k)} e^{y_{k,w}}), \\ p(\mathbf{y}|\boldsymbol{\sigma}, \boldsymbol{\phi}) &= \prod_k \mathcal{N}(\mathbf{y}_k; \hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k), \\ \int p(\dot{\mathbf{x}}, \mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma}) d\dot{\mathbf{x}} &\propto p(\boldsymbol{\theta})p(\boldsymbol{\phi})p(\boldsymbol{\gamma}) \prod_k \mathcal{N}(\mathbf{x}_k; \mathbf{0}, \mathbf{K}_{xx}) \\ &\quad \times e^{-\frac{1}{2}(\mathbf{f}_k - \mathbf{g}_k)^\top (\boldsymbol{\Psi}_k + \boldsymbol{\gamma} \mathbf{I})^{-1} (\mathbf{f}_k - \mathbf{g}_k)}. \end{aligned} \quad (9)$$

### 3.3 MCMC 法による推論

(9) 式は解析的に計算できないため、MCMC による近似推論を行う。ODE パラメータ  $\boldsymbol{\theta}$  に対してはロジット変換、その他のパラメータ  $\boldsymbol{\sigma}, \boldsymbol{\phi}, \boldsymbol{\gamma}$  に対しては対数変換を行って得られる実数変数について、それぞれ事前分布を正規分布とすることで、ハミルトニアンモンテカルロ法 (HMC)[Neal 11] 等の実数空間上の潜在状態に対する推論アルゴリズムが利用可能になる。また、[Dondelinger 13] で指摘されている通り、AGM の推論において、非線形な ODE システムの影響で、事後分布はごつごつとした多峰形状となる可能性があるため、レプリカ交換モンテカルロ法 [Hukushima 96] によりサンプラーが局所的な溝に陥ることを防ぐ。

## 4. 実験

### 4.1 実験データの準備

提案手法の有効性を検証するために、米国特許庁 (USPTO) が公開する現実の特許データに対して、特定の技術テーマに関する企業間競争関係を推定する。Handling natural language data という国際特許分類 (IPC) グループを含む 2005 年から 2020 年の間に出版され特許の内、第一出願人に組織名が含まれる特許に限定し、計 17,422 の特許を解析対象として選定した。

表 1: トピックの例とそれを代表する確率上位の単語。

トピック	上位単語
1	content item file page web_page display
2	word text sentence candidate model term
3	message query entity natural_language node question
4	user interface template provide field form
5	text display image unit input character

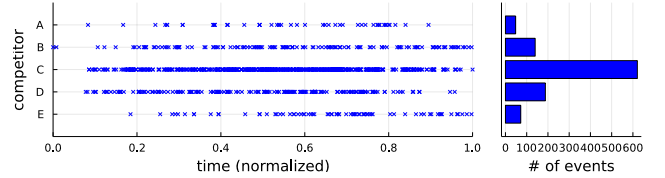


図 3: トピック 2 に関連する特許出願イベントの時系列。

分類の粒度が粗い国際特許分類を用いて集めた特許集合に対し、より細分化された技術的なトピック毎に分析を行うため、対象特許集合の要約テキストに潜在的に含まれるトピックのモデリングを行った。LDA[Blei 03] に代表されるトピックモデルは、文書をトピックの多項分布とし、さらにトピックを単語の多項分布とする階層モデルである。特に、Correlated Topic Models (CTM)[Blei 06] は文書のトピック事前分布を多変量正規分布とするモデルであり、グローバルなトピックの出現確率やトピック間の相関関係を推定できる。本稿では、対数尤度を指標 [Griffiths 04] として用いたトピック数を 9 と設定して CTM によるトピックの推論を行った。表 1 に一部のトピックのトピック-単語分布における確率上位 6 単語を示した。これらの各トピックについて、各文書の文書-トピック分布における当該トピックの確率が経験的に設定した閾値 15% を超える特許のみを各トピックの関連特許として抽出した。本稿では、図 3 で示されるトピック 2 に関する特許出願件数上位 5 企業 A, B, C, D, E のイベントデータのみを実験対象を絞り、5 社間の競争モデルを仮定した実験結果を示す。

### 4.2 実験設定

本稿では、  $\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}$  のみを推論対象としてサンプリングを行い、  $\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\phi}$  については推論対象とはせずハイパーパラメータとして扱った。ハイパーパラメータは経験的に  $\sigma = 0.02$ ,  $\gamma = 0.1$  と設定した。また、ガウス過程のカーネル関数に RBF カーネル  $k(\mathbf{x}, \mathbf{x}') = \phi_1^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\phi_2^2}\right)$  を用い、  $\phi_1 = 5.0, \phi_2 = 0.2$  で固定した。また、ODE パラメータの値域は、自然増加率  $r$  について [1, 10]、環境収容力  $\eta$  について [5, 15]、競争係数  $a$  について [0, 10] を用いた。潜在変数  $\mathbf{x}$  および  $\mathbf{y}$  を評価する時刻については、  $T = 21, W = 100$  で、0 から 1 に正規化した時刻において、  $\boldsymbol{\tau} = [0.00, 0.05, 0.10, \dots, 0.95, 1.00]$ ,  $\boldsymbol{\omega} = [0.005, 0.015, 0.025, \dots, 0.985, 0.995]$  とした。

MCMC シミュレーションは、異なる 10 段階の温度  $1.1^0, 1.1^1, \dots, 1.1^9$  でチェーンのレプリカを作成し、各温度で HMC により事後分布からのサンプルを得、チェーン間のスワップ試行を毎イテレーション行う設定とした。HMC のパラメータは、リープフロッグステップ数は  $L = 30$  で固定し、ステップサイズは事前のシミュレーションで調整した値を用いた。2,000 回のバーンインシミュレーションを行い、その後の 10,000 回分のシミュレーションで得られたサンプルを 10 サンプル毎に間引いて収集した 1,000 個のサンプルを用いて事後パラメータ推論を行った。

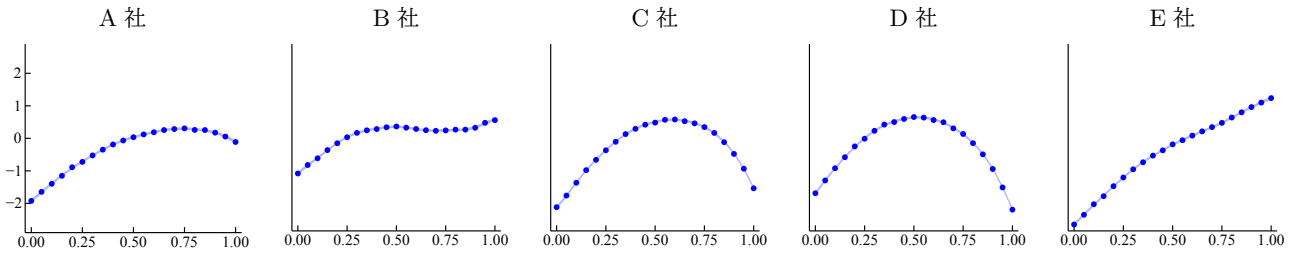


図 4: Lotka-Volterra 競争方程式に従うダイナミクスを近似する各社の潜在関数の事後サンプル. X 軸: 時刻, Y 軸:  $\alpha_k$  の値.

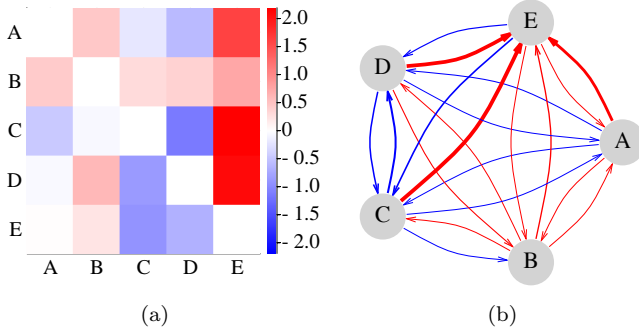


図 5: 競争関係の推定結果. (a)  $i$  行  $j$  列の要素が対数競争係数  $\log a_{ij}$  の事後平均を示す隣接行列. 赤が影響の強い関係, 青が影響の弱い関係を示す. (b) 競争関係の有効グラフ. エッジの太さは対数競争係数の絶対値に対応し, 影響の度合いは, 赤く太いエッジ, 赤く細いエッジ, 内的競争, 青く細いエッジ, 青く太いエッジの順に強い.

### 4.3 結果と考察

図 4, 5 にトピック 2 に関連する特許出願データからの提案手法によるパラメータの推定結果を示した. 図 4 は, 各社の潜在関数  $\alpha_k$  の事後分布からのサンプルであり, 関数の浮き沈みは 3 のイベントの出現傾向を反映している. また, 図 5 に競争関係の推定結果を可視化した. (a) は,  $i$  行  $j$  列の要素を対数競争係数  $\log a_{ij}$  とする隣接行列である. 対角成分は各企業の内的な対数競争係数  $\log a_{ii} = 0$  であり, これを基準として白色で示し,  $\log a_{ij} > 0$  を赤,  $\log a_{ij} < 0$  を青で示している. (b) は (a) の隣接行列をネットワーク状に可視化した結果である. 競争係数の値は, 関数  $\alpha$  に反映されており,  $\log a_{ij}$  が大きい場合,  $i$  の関数の下降と  $j$  の関数の上昇が同時期に起こり,  $\log a_{ij}$  が小さい場合,  $j$  の関数の  $i$  の関数への影響が弱いことを示す.

推定された対数競争係数からは, A, C, D が E から強い影響を受けることが示唆されており, A, C, D の潜在関数の下降と E の潜在関数の上昇が同時期に起きている結果に対応する. 上記の結果は, 競争係数が推定できなくとも, 例えば対数ガウス Cox 過程で関数  $x$  の形状さえ可視化できれば, 目視で発見可能かもしれない. しかしながら, B のように上昇, 下降, 再上昇とより複雑な動きをする関数については目視で企業間の関係性を捉えることは難しい. また, 競争モデルで取り扱う企業数  $K$  を増やした場合, 企業間の 1 対 1 の関係の数は  $\mathcal{O}(K^2)$  で増加するため, 目視での発見は困難になるため, 競争係数行列を図 5(a) のように可視化できることは実用上意味がある.

$\log a_{ij} < 0$  である青で示される関係性に注目すると, 企業  $j$  の企業  $i$  への影響の小ささもまた評価できる. C, D の関数は似た形状で時系列としては相関するため, 一見すると関係性があるようにも見える. しかしながら, 競争係数からは互いに影響しあわない関係であるという推定結果が得られており, これは関数の目視確認では発見することが難しい関係であろう.

## 5. まとめと今後の展望

本稿では, 時系列イベントデータの生成過程に任意の ODE システムを仮定することを可能とする ODE 対数ガウスを提案した. 産業界の企業間の競争関係をモデル化するために, ODE として Lotka-Volterra 競争方程式を用い, 特定の技術に関連する特許出願イベント時系列から ODE パラメータの推定を行った. 推定された ODE パラメータの競争係数により, 企業間の 1 対 1 の競争関係における影響の度合いを定量化し, 注目すべき関係性を発見し, 解釈・考察を与えることを可能にした.

本稿では, 技術分野別にモデリングしたが, 企業活動が複数の異なる技術分野をまたがることも多いため, ある企業内における異なる技術分野間の資源の奪い合いもまた競争モデルで表現できる. 企業間の競争と企業内の競争を同時にモデリングすることで, より示唆に富む推定結果が得られると期待される.

また, 提案手法は特許テキストの内容に関しては考慮せず, 単にイベントデータとして扱うに限っている. 本稿では特許母集団を仕分けるために前段でテキスト解析を行い, その結果を提案モデルによるモデリングに用いているが, 確率的トピックモデルや文書分散表現といったテキストデータに対する手法と ODE システムを直接組み合わせるモデルも考えられる.

## 参考文献

- [Blei 03] Blei, D. M., Ng, A. Y., and Jordan, M. I.: Latent dirichlet allocation, *JMLR*, Vol. 3, No. Jan, pp. 993–1022 (2003)
- [Blei 06] Blei, D. and Lafferty, J.: Correlated topic models, *Advances in Neural Information Processing Systems*, Vol. 18, p. 147 (2006)
- [Dondelinger 13] Dondelinger, F., Husmeier, D., Rogers, S., and Filippone, M.: ODE parameter inference using adaptive gradient matching with Gaussian processes, in *AISTATS*, pp. 216–228PMLR (2013)
- [Griffiths 04] Griffiths, T. L. and Steyvers, M.: Finding scientific topics, *Proceedings of the National Academy of Sciences*, Vol. 101, No. suppl 1, pp. 5228–5235 (2004)
- [Hinton 02] Hinton, G. E.: Training products of experts by minimizing contrastive divergence, *Neural Computation*, Vol. 14, No. 8, pp. 1771–1800 (2002)
- [Hukushima 96] Hukushima, K. and Nemoto, K.: Exchange Monte Carlo Method and Application to Spin Glass Simulations, *Journal of the Physical Society of Japan*, Vol. 65, No. 6, p. 1604–1608 (1996)
- [Møller 98] Møller, J., Syversveen, A. R., and Waagepetersen, R. P.: Log Gaussian Cox Processes, *Scandinavian Journal of Statistics*, Vol. 25, No. 3, pp. 451–482 (1998)
- [Neal 11] Neal, R. M., et al.: MCMC using Hamiltonian dynamics, *Handbook of Markov Chain Monte Carlo*, Vol. 2, No. 11, p. 2 (2011)
- [Smale 76] Smale, S.: On the differential equations of species in competition, *Journal of Mathematical Biology*, Vol. 3, pp. 5–7 (1976)
- [Tang 12] Tang, J., Wang, B., Yang, Y., Hu, P., Zhao, Y., Yan, X., Gao, B., Huang, M., Xu, P., Li, W., et al.: Patentminer: topic-driven patent analysis and mining, in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1366–1374 (2012)