

カルマンフィルタによるストーリー生成手法の提案

Story Generation with Kalman Filter

佐藤 真*¹ 赤石 美奈*¹ 堀 浩一*²
 Makoto Sato Mina Akaishi Koichi Hori

*¹法政大学 Hosei University
 *²東京大学 University of Tokyo

We propose a framework that interfaces with the story generation process. Moreover, based on the framework, we propose a new model for story generation with the Kalman filter and smoother can be used. Then, we show examples of the application.

1. はじめに

知識の伝達のために物語性は重要である。また知識創造活動支援のために、知識の分節技術が求められている。表現された知識のダイナミクスを取り扱うモデルは存在するが、動的な再構成を扱うために用いられていることは少ない。またより複雑なダイナミクスを想定することもできるだろう。本稿では、知識の表現としてのストーリーの生成過程のモデルのフレームワークを提案する。さらに、そのフレームワークに基づき、カルマンフィルタを利用できる新たなストーリー生成のモデルを提案する。

ストーリーの自動生成システムについては数多く提案されている [Meehan 77, Bringsjord 99, Montfort 07]。またユーザとシステムにインタラクションを通じて関連するストーリーを提示する手法も数多く提案されている [Cavazza 02, Shahaf 10]。小方は多くの物語生成システムの共通構成要素を、語られるべき内容の生成、語りの方法、語りと受容の担い手と分けており、語りの方法という分類を以下のようにまとめている [小方 11]。

語りの方法 (1) 一般にプロット、構成のように呼ばれる抽象的な層。言説技法、ジャンルも含まれる。実現技術としては構造変換、文法的・修辭的知識構造など。

語りの方法 (2) 言語表現、映像表現、音楽表現といった表層的媒体と直接結びついた表現の層。実現技術としては、自然言語生成、映像メディア技術、自動作曲・変奏・編曲など。

ドキュメント群が蓄えられており、ここからいくつかのドキュメントを抽出し再構成して、新しいストーリーとして関連するドキュメント列を機械的、自動的に生成あるいは生成支援するというとき、ストーリーの語りの方法に応じて変化する文脈のモデル化は不可欠であろう。そこで、本稿では、ストーリーが複数のドキュメントの連鎖で構成されているとして、以下のようなストーリー生成過程のフレームワークを定義する。

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}, \mathbf{u}_t, \mathbf{w}_t) \quad (1)$$

$$\mathbf{y}_t = h(\mathbf{x}_t, \mathbf{v}_t) \quad (2)$$

上式は、ストーリーが分岐なしの T 個のドキュメント列で構成されているとき、語りによるプロット構成による文脈の変遷先: 佐藤真, 法政大学情報科学部, 東京都小金井市梶野町 3-7-2, 042-387-4544, makoto.satoh.47@hosei.ac.jp

化 $f(\cdot)$ と文脈のドキュメントとしての表現 $h(\cdot)$ を記述したものである。文脈は状態 \mathbf{x} 、ドキュメントは文脈の観測 \mathbf{y} で特徴付けられる。上式の \mathbf{u} は入力、 \mathbf{w}, \mathbf{v} は誤差とする。

このモデルでストーリー生成が説明可能でかつ関数 $f(\cdot)$ と $h(\cdot)$ が数式で記述可能ならば、推論問題の分野における種々の手法の知見を活かすことができるであろう。本論文ではこのモデルにのっとりストーリー生成過程の簡易なモデルを提案し、ストーリー生成をドキュメントの連鎖推定問題として捉え、カルマンフィルタを用いたストーリー生成手法を提案する。

2. カルマンフィルタを用いたストーリー生成

ストーリーがあるダイナミクスに従って文脈が変化するならば、生成されるドキュメントを観測とみなし、推定問題として文脈を予測できる。すなわち、ストーリー生成のモデルに対して状態方程式と観測方程式を与えることができれば、カルマンフィルタで推定することができる。

2.1 状態方程式

ここでは文脈の状態方程式が線形であるとする。

$$f(\mathbf{x}, \mathbf{u}, \mathbf{w}) = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{w} \quad (3)$$

過去の文脈 \mathbf{x} が変換行列 \mathbf{A} によって線形変換される。文脈を表すベクトルの要素が語の重要度を表すとしたら、 \mathbf{A} の要素はある語が他の語の重要度にどれほど影響を及ぼすかということを示す。

外部からの入力 \mathbf{u} があれば、変換行列 \mathbf{B} により線形変換される。ユーザが次に繋げたいドキュメントの方向性や趣向を入力すると、このテーマに応じて文脈を示すベクトルの要素にどれほど影響を与えるかということを示す。たとえばニュースであるならば、政治、経済、スポーツ、エンターテインメントといったようなテーマを入力すると、テーマに応じたベクトルの要素が影響を受ける。

\mathbf{w} はどれくらいランダムにベクトルの要素の値が大きくなるかあるいは小さくなるかということを表す。

2.2 観測方程式

つぎに観測方程式について考える。観測方程式も線形とする。

$$h(\mathbf{x}, \mathbf{v}) = \mathbf{C}\mathbf{x} + \mathbf{v} \quad (4)$$

文脈のベクトルに線形変換を施す行列 \mathbf{C} は変換行列の要素はある語がある重要度を持つときどれほど他の語が出現するかと

いうことを表す指標を示す。ただし、ベクトル \mathbf{x} は潜在的な変数で表される文脈が語の重要度を表し、ベクトル \mathbf{y} は観測変数であるドキュメントが語の出現を特徴付けるを表す。

\mathbf{v} は \mathbf{w} 同様にどれほどランダムにその語が観測されるかを表す指標となる。

2.3 ストーリー生成

ある時点までの誤差のあるデータセットに対して、データのパターンを定義する関数を定義し、過去、現在、未来の状態を推定することは、それぞれ、平滑化 (smoothing), 濾波 (filtering), 予測 (prediction) と呼ばれる。カルマンフィルタ (Kalman Filter; KF) は、ひとつの状態推定手法であり、誤差のある離散的な過去の観測から時間に応じて変化する現在の状態を推定する。カルマンフィルタはセンサーデータを用いて現在状態量を推定するような問題などに広く用いられている。たとえば、カーナビでは加速度計や GPS によって得られたデータから現在位置の推定する。

上に挙げた状態方程式と観測方程式が線形であるとして、カルマンフィルタを適用するために以下のようにまとめる。

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}u_t + \mathbf{w}_{t+1} \quad (5)$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}_t \quad (6)$$

誤差がそれぞれ平均 0, 共分散行列 \mathbf{Q}, \mathbf{R} のガウス分布に従い、

$$\mathbf{w}_t \sim N(0, \mathbf{Q}), \mathbf{v}_t \sim N(0, \mathbf{R}) \quad (7)$$

推定分布 $p(\mathbf{x}_t | \mathbf{y}_{0:t})$ も常にガウス分布と仮定するとき、

$$\mathbf{x}_t | \mathbf{y}_{0:t} \sim N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t) \quad (8)$$

カルマンフィルタでは $t+1$ における推定分布 $p(\mathbf{x}_{t+1} | \mathbf{y}_{0:t+1}) = N(\mathbf{x}_{t+1}; \boldsymbol{\mu}_{t+1}, \boldsymbol{\Sigma}_{t+1})$ を t における推定分布 $p(\mathbf{x}_t | \mathbf{y}_{0:t}) = N(\mathbf{x}_t; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ からガウス分布の性質を利用して以下のように予測 (Prediction) と更新 (Update) を行うことで求めることができる。入力は $\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t, \mathbf{y}_{t+1}$ であり、出力は $\boldsymbol{\mu}_{t+1}, \boldsymbol{\Sigma}_{t+1}$ である。

Prediction

$$\boldsymbol{\mu}'_{t+1} \leftarrow \mathbf{A}\boldsymbol{\mu}_t + \mathbf{B}u_t \quad (9)$$

$$\boldsymbol{\Sigma}'_{t+1} \leftarrow \mathbf{Q} + \mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}^T \quad (10)$$

#Update

$$\mathbf{K}_{t+1} \leftarrow \boldsymbol{\Sigma}'_{t+1}\mathbf{C}^T(\mathbf{C}\boldsymbol{\Sigma}'_{t+1}\mathbf{C}^T + \mathbf{R})^{-1} \quad (11)$$

$$\boldsymbol{\mu}_{t+1} \leftarrow \boldsymbol{\mu}'_{t+1} + \mathbf{K}_{t+1}(\mathbf{y}_{t+1} - \mathbf{C}\boldsymbol{\mu}'_{t+1}) \quad (12)$$

$$\boldsymbol{\Sigma}_{t+1} \leftarrow \boldsymbol{\Sigma}'_{t+1} - \mathbf{K}_{t+1}\mathbf{C}\boldsymbol{\Sigma}'_{t+1} \quad (13)$$

さらにカルマンフィルタによって時刻 T まで推定分布 $p(\mathbf{x}_t | \mathbf{y}_{0:t}) = N(\mathbf{x}_t; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ と予測分布 $p(\mathbf{x}_{t+1} | \mathbf{y}_{0:t}) = N(\mathbf{x}_{t+1}; \boldsymbol{\mu}'_{t+1}, \boldsymbol{\Sigma}'_{t+1})$ が既に求まっているとき平滑分布 $p(\mathbf{x}_t | \mathbf{y}_{0:T}) = N(\mathbf{x}_t; \boldsymbol{\mu}''_t, \boldsymbol{\Sigma}''_t)$ を求める手法はカルマン smoother (Kalman Smoother) と呼ばれ、以下のように求められる。

$$\boldsymbol{\mu}''_t = \boldsymbol{\mu}_t + \mathbf{J}_t \cdot (\boldsymbol{\mu}''_{t+1} - \boldsymbol{\mu}'_{t+1}) \quad (14)$$

$$\boldsymbol{\Sigma}''_t = \boldsymbol{\Sigma}_t + \mathbf{J}_t \cdot (\boldsymbol{\Sigma}''_{t+1} - \boldsymbol{\Sigma}'_{t+1})\mathbf{J}_t^T \quad (15)$$

ただし $\mathbf{J}_t = \boldsymbol{\Sigma}_t\mathbf{A}(\boldsymbol{\Sigma}'_{t+1})^{-1}$ とする。

カルマンフィルタとカルマン smoother を用いて以下の 2 通りのストーリー生成手法を提案する。

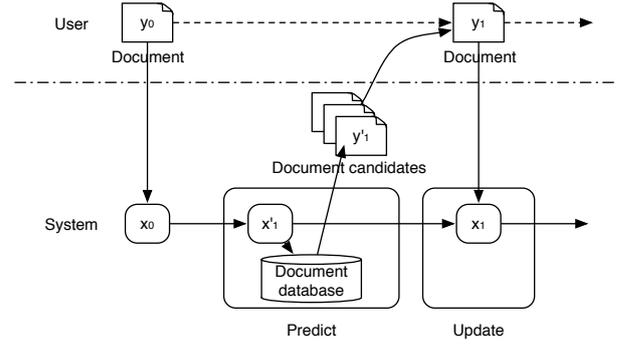


図 1: カルマンフィルタを用いた手法

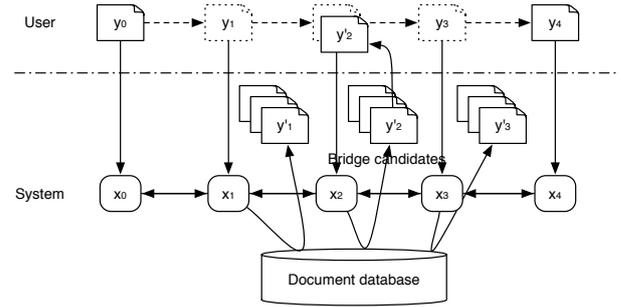


図 2: カルマン smoother を用いた手法

2.3.1 カルマンフィルタを用いた手法

まずカルマンフィルタを用いた手法を提案する。図 1 に概要を示す。以下の手順でユーザとシステムがインタラクションを通じてストーリーを生成する。外部入力 u_t はないものとして、 t 番目のドキュメントとしてユーザが選択したドキュメントの特徴量を観測 \mathbf{y}_t とする。

1. ユーザがストーリーの最初のドキュメントを指定する。
2. システムがユーザが指定したドキュメントを観測として、次のカルマンフィルタによって内部状態を予測する。この内部状態を観測したとしたときの値と類似したドキュメントをあらかじめ蓄えておいたドキュメントデータベースから探し出し、次に繋げるべきドキュメントとして推薦する
3. ユーザがシステムの推薦したドキュメントを選ぶ。あるいはドキュメントを自分で書く。
4. ユーザが満足するまで 2 と 3 を繰り返す。

T 個のドキュメントをつなげれば、最終的には最初に指定したドキュメントを含めて $T+1$ 個のドキュメントの連鎖からなるストーリーが生成される。

推定された現在の状態 $p(\mathbf{x}_t | \mathbf{y}_{0:t}) = N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ に対して、次のステップにおけるドキュメントの状態の予測 $p(\mathbf{y}_{t+1} | \mathbf{y}_{0:t})$ は

$$\begin{aligned} p(\mathbf{y}_{t+1} | \mathbf{y}_{0:t}) &= p(\mathbf{x}_{t+1} | \mathbf{y}_{0:t})p(\mathbf{y}_{t+1} | \mathbf{x}_{t+1}) \\ &= N(\mathbf{C}(\mathbf{A}\boldsymbol{\mu}_t + \mathbf{B}u_t), \mathbf{R} + \mathbf{C}(\mathbf{Q} + \mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}^T)\mathbf{C}^T) \end{aligned} \quad (16)$$

と求められる。

表 1: タイトル一覧

No.	Title
0	堀 浩一 教授 東京大学
1	堀 浩一
2	Koichi Hori
3	堀 浩一 詳細履歴
4	創造活動支援システムの研究
5	文化国家としての技術立国 — 猪瀬博教授の思い出
6	堀の自作ソフト KNC(Knowledge Nebula Crystallizer)
7	共同通信連載「仕事のデッサン」 創造性をめぐる話題
8	IBM ワトソン
9	工学と理学の違い
10	ことばと発想のテクノロジー
11	知識の姿 — 人工知能研究者の立場から
12	「評価」から「解説」へ
13	事業仕分けをきっかけに思う
14	Rails + MySQL
15	文脈依存システムの気持ち悪さ
16	堀洋一さんと堀浩一は別人です
17	原子力発電所事故現場に何故ロボットを ロボットを送り込めないのか
18	KNC(Knowledge Nebula Crystallizer) の原理
19	堀 浩一 の ひとりごと
20	このサイトでの人気ストーリー
21	文理融合研究
22	CREST プロジェクト 「情報があふれる社会から表現が編みあがる社会へ」
23	堀研究室所属を希望する皆さんへ
24	「科学技術」か「科学・技術」か
25	活躍する教え子たち
26	志高き産学連携?
27	僕が欲しい携帯デバイス
28	人工知能学会創立 25 周年にあたって

2.3.2 カルマンスムーザを用いた手法

次にカルマンスムーザを用いた手法を提案する。図 2 に概要を示す。ユーザがスタートとゴールを指定して、システムがスタートとゴールを繋げるドキュメントの連鎖を推薦する手法(私たちはこれをトピックブリッジと呼んでいる)を提案する。なおスタートとゴールをつなげるドキュメントの一つ一つをブリッジと呼ぶ。

1. ユーザがスタートのドキュメントとゴールのドキュメントとブリッジの数を指定する。
2. システムがスタートとゴールをつなぐようなドキュメント列をカルマンスムーザによって推定し、推定された各ドキュメントの状態変数と類似するドキュメントをあらかじめ蓄えておいたドキュメントデータベースから推薦する。初期状態ではすべての観測について仮想的なものを用いる。
3. ユーザは推薦されたドキュメント列のうち適当であると評価したドキュメントを部分的に選択し、選択されたド

キュメントを観測列の部分列と置換して、指定された数のブリッジが置換されるまで 2 に戻る。

最終的には、すべての仮想的な観測が置換され、スタートからゴールへ、指定された数だけのブリッジのドキュメントを媒介してつながるドキュメントの連鎖のストーリーが生成される。

3. 適用例

堀浩一の Web サイト^{*1}を対象にした適用例を示す。カルマンフィルタを用いた手法については、あるドキュメントを指定したときに、つぎにどのようなドキュメントが推薦されたかを示す。カルマンスムーザを用いた手法については、あるスタートとゴールを指定したときに、推薦された上位のドキュメント列の集合を提示する。

サイトは、構成されるページにおける訪問者の履歴によって、異なる関連ページが推薦される仕組みを備えている。2012 年 4 月 1 日の時点でトップページから辿れるページは全部で 29 ページあり、すべて HTML で記述されていた。言語は 28 ページが日本語、1 ページが英語であった。

解析にあたりまず HTML タグの除去を行い自然言語で記述されたドキュメントを抽出した。ドキュメントに対して MeCab^{*2}を用いて形態素解析を行い名詞のみを抽出、語として扱った。なお一般的な語(「これ」「あれ」など)は除去した。

状態方程式と観測方程式の係数についてはさらに簡単のために以下のように設定した。

$$A = I_N, B = 0, C = I_N, Q = R = N(0, I_N) \quad (18)$$

ただし I_N は N 次元の単位行列を表す。これらは、状態変数は 1 ステップ後にガウス分布に従って遷移し、観測変数は状態変数がガウス分布に従う誤差を含む形でそのまま観測されるというモデルである。

観測として得られるドキュメントの特徴量は、全 29 ページで出現する全 N 個の語に対して、 N 次元のベクトル y で表し、 y の要素 $y[i]$ は語 $t_i (1 \leq i \leq N)$ に関する tf-idf 値とした。

$$y[i] = tfidf(t_i) \quad (19)$$

なお tf-idf 値は

$$idf = \log \frac{|D|}{|\{d : t \in d\}|} \quad (20)$$

として

$$tfidf(t) = tf(t, d) \times idf(t) \quad (21)$$

と求めた。 $|D|$ は全ドキュメント数 29, $|\{d : t \in d\}|$ は語 t が含まれるドキュメント d の数, $tf(t, d)$ はドキュメント d に出現する語 t の回数として計算した。

フィルタリング・スムージングによって得られた文脈の特徴量からドキュメントの特徴量を推定し連鎖させるべき候補のドキュメントを探し出すために、ドキュメントの類似度に従ってスコアを付け、スコアの高い順にユーザに提示した。予測したドキュメントの平均 $\hat{\mu}$ と候補のドキュメントの特徴量 y_c の類似度はコサインを用いて測った。

$$score(\hat{\mu}, y_c) = \frac{\langle \hat{\mu}, y_c \rangle}{|\hat{\mu}| |y_c|} \quad (22)$$

今回対象としたページのタイトルを表 1 に示す。

*1 <http://www.ailab.t.u-tokyo.ac.jp/horiKNC/knc/index.jp>

*2 <http://mecab.sourceforge.net/>

3.1 カルマンフィルタを用いた手法

ページ(1)「堀 浩一」に対して推薦されたドキュメントの順位とスコアは表2に示すとおりである。候補順位1位のドキュメントはページ(3)「堀 浩一 詳細履歴」でありスコアは0.081であり、候補順位2位のドキュメントはページ(28)「人工知能学会創立25周年にあたって」でありスコアは0.065であった。

次にページ(1)「堀 浩一」に対して推薦されたドキュメントに対して上位2件をそれぞれ選択し、次に推薦されるページの違いを見る。ページ(1)「堀 浩一」、ページ(3)「堀 浩一 詳細履歴」を選んだとき推薦されるドキュメントは表3に示すとおりである。1位の推薦を辿ると「堀 浩一」「堀 浩一 詳細履歴」「評価」から「解説」へ」というストーリーが生成される。

ちなみにページ(1)「堀 浩一」から推薦順位2位のページ(28)「人工知能学会創立25周年にあたって」を選んだとき推薦されるドキュメントは表4に示すとおりである。ページ(3)「堀 浩一 詳細履歴」を選んだ時とは推薦されるドキュメントが異なっていることは確認できる。

表2: 1の連鎖ドキュメント候補
表3: 1-3の連鎖ドキュメント候補
表4: 1-28の連鎖ドキュメント候補

No.	Score	No.	Score	No.	Score
3	0.081	12	0.106	3	0.068
28	0.065	28	0.060	5	0.041
23	0.057	5	0.056	7	0.035
25	0.034	23	0.052	13	0.034
7	0.029	26	0.034	19	0.031

3.2 カルマンスムーザを用いた手法

スタートをページ(1)「堀 浩一」、ゴールをページ(28)「人工知能学会創立25周年にあたって」、ブリッジ数3とする。仮想的な観測を0ベクトルとしたときの推薦されるブリッジの候補は、表5に示すとおりである。

2番目のブリッジにページ(3)「堀 浩一 詳細履歴」を選択したとき、再び計算を行い新たに推薦されるドキュメント列は、表6に示すとおりである。そして2番目のブリッジにページ(19)「堀 浩一 の ひとりごと」を選択したとき、再び計算を行い新たに推薦されるドキュメント列は、表7に示すとおりである。

結果を見ると、ある選択を行うことで推薦されるドキュメントは変わるが、ブリッジはどこでも推薦されるものは同じである。たとえば表6において1番目のブリッジにも3番目のブリッジにも、スコアは異なるものの、ページ17,15,20,5,13の順で推薦されている。

3.3 考察

今後、閲覧者の推薦結果に対するリアクション、推薦なしでの閲覧履歴との比較を検討している。厳密な比較はしていないが、サイトの仕組みによる関連ページを辿ると「堀 浩一」「創造活動支援システムの研究」「知識の姿 - 人工知能研究者の立場から」となっている。このような他手法の推薦との比較も行いたい。

表5: スタートをページ(1)、ゴールをページ(28)としたときのブリッジの1,2,3番目の候補

#1	Score	#2	Score	#3	Score
3	0.096	3	0.075	3	0.061
23	0.056	5	0.042	5	0.040
7	0.040	7	0.037	7	0.032
13	0.037	13	0.036	13	0.031
19	0.037	19	0.033	19	0.029

表6: ページ(3)を2番目のブリッジとしたときの1,3番目の候補
表7: ページ(19)を2番目のブリッジとしたときの1,3番目の候補

#1	Score	#3	Score	#1	Score	#3	Score
17	0.261	17	0.243	12	0.107	12	0.107
15	0.229	15	0.214	5	0.058	5	0.064
20	0.077	20	0.075	23	0.056	23	0.053
5	0.047	5	0.055	26	0.036	26	0.038
13	0.046	13	0.050	7	0.033	7	0.038

4. おわりに

本稿ではカルマンフィルタを用いたストーリー生成支援手法を提案した。上で述べたモデルは閲覧者の興味がブラウン運動のように変化するモデルであった。今後、定性的な評価を通じて、状況に応じたモデルの設定を考察していく予定である。

参考文献

- [Bringsjord 99] Bringsjord, S. and Ferrucci, D., Artificial Intelligence and Literary Creativity: Inside the Mind of Brutus, a Storytelling Machine (1999)
- [Cavazza 02] Cavazza, M. and Charles, F. and Mead, S.J., Character-based interactive storytelling, IEEE Intelligent Systems, Vol. 17, No.4, pp. 17-24 (2002)
- [Shahaf 10] Shahaf, D. and Guestrin, C., Connecting the dots between news articles, Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 623-632 (2010)
- [Montfort 07] Montfort, N., Generating narrative variation in interactive fiction (2007)
- [Meehan 77] Meehan, J.R., Tale-spin, an interactive program that writes stories, Proceedings of the fifth international joint conference on artificial intelligence, pp. 91-98 (1977)
- [小方 11] 小方 孝, 「物語の情報学」の実践としての物語生成システム, 日本知能情報ファジィ学会誌, No.5, Vol. 23 (2011)