Conditional VAEに基づく多様性を考慮したイベント予測

清丸 寛一† 大村 和正‡ 村脇 有吾† 河原 大輔† 黒橋 禎夫†
†京都大学 大学院情報学研究科 ‡京都大学 工学部

{kiyomaru, omura, murawaki, dk, kuro}@nlp.ist.i.kyoto-u.ac.jp

1 はじめに

イベント予測とは，あるイベントに続いて起きるイベントを予測するタスクである．イベントとは「自転車を漕ぐ」や「パンクする」など一つの出来事や状態を表す意味単位であり，イベント予測では「自転車を漕ぐ → パンクする」のようなイベント系列のもっともらしさをモデル化する．

本研究では，イベントが自然言語文で記述され，イベント系列が2要素からなる場合のイベント予測に取り組む．従来研究はこの問題を sequence-to-sequence (seq2seq) として定式化し，前件のイベントから後件のイベントを生成するモデルを学習している．しかし，「自転車を漕ぐ」には「パンクする」の他にも「目的地に着く」など複数のイベントが続く可能性があり，そうした多様性は考慮されていなかった．

本研究では Conditional VAE (CVAE) に基づくイベント予測モデルを構築し，入力から出力への決定的な変換ではなく，確率的生成を学習することで後件のイベントの多様性を捉える．またモデル自身の出力から入力信号を復元する reconstructor を導入することで一般的な出力の生成を抑制し，多様かつ情報量に富んだイベント生成を促す．

実験では，従来のデータセットに加え，日本語イベントペアデータを新たに構築し，イベント予測を学習した．自動評価指標とクラウドソーシングによる評価を行い，提案手法が妥当性・多様性においてベースラインの手法を上回ることを示した．

2 関連研究

2.1 イベント系列のモデル化

大規模コーパスの自動解析に基づきイベント系列を抽出・モデル化する研究において，イベントの表現には述語項構造が広く用いられてきた [1]．ここでの述語項構造は，述語と項がそれぞれ句で表され，連体修飾節などを含まないことが一般的である．例えば，「学生が驚くべき研究成果を得る」という文からは (得る; ガ:学生，ヲ:研究成果) という述語項構造がイベントとして抽出され，連体修飾節の「驚くべき」は抜け落ちてしまうが，これはイベント予測における重要な手がかりとなり得る．より柔軟なイベントの表現として，近年は自然言語文でイベントを表す取り組みが増えつつある [2, 3]．

イベント系列のモデル化は，narrative cloze task やその派生タスクを通じて行われるのが一般的である [1]．narrative cloze task はイベント系列から1要素を取り出し，候補イベントの中からその要素を選択するタスクである．このタスクを通じて学習したモデルは入力に完全なイベント系列を要し，あるイベントの後続イベントを得るには候補の数だけ計算を繰り返す必要がある．深層学習技術の進展に伴い，seq2seq の枠組みでイベント系列をモデル化する手法が提案されている [2, 3]．この枠組みでは生成モデルを学習するため，新たなイベント系列を効率的に獲得できる．しかし，従来研究では入出力間の決定的な変換を学習している．あるイベントに後続し得るイベントは一般に複数存在するため，この仮定は妥当ではない．

2.2 CVAEによる出力の多様性のモデル化

Variational Autoencoder (VAE) は，ニューラルネットワークに基づく確率的生成モデルの1つである．VAE は，関連する Autoencoder (AE) と同様に，表層表現 y から潜在表現 z を経由し y を再構成するネットワークであるが，AE が決定的な変換を行うのに対し，VAE は確率的な変換を行う点が特徴的である．VAE において，z には適当な事前分布 (典型的には多変量標準ガウス分布) が仮定される．Conditional VAE (CVAE) は VAE を構成する各分布が共通の観測変数 x で条件付けられるような拡張である．

CVAE を seq2seq モデルで構成することで，入力文に対する出力文の多様性がガウス分布に反映されると期待される．これまでに対話や機械翻訳など，入力に対して出力が一意に定まらないタスクに CVAE に基づくモデルが適用されている [4, 5]．

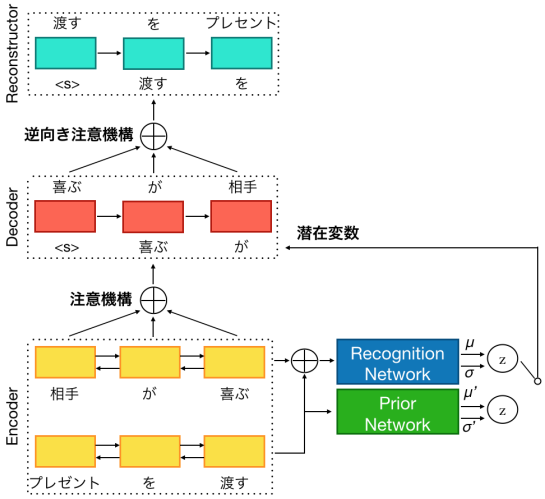

図 1: モデルの概要．この例では，前件イベントが「プレゼントを渡す」，後件イベントが「相手が喜ぶ」である．

3 提案手法

本研究では 2 要素からなるイベント系列を扱う．前件のイベントを x，後件のイベントを y とし，x と y の組をイベントペアと呼ぶ．x に対する y の多様性を捉えるため，提案手法は Conditional VAE (CVAE) に基づく seq2seq モデルを基本とする．さらに，情報量に富んだ後件イベントの生成を促す機構として，モデル自身の出力から入力信号を復元する reconstructor [6] を導入する．図 1 にモデルの概要を示す．

3.1 目的関数

x に対する y の多様性を捉えるため，y は x に加え，確率的潜在変数 z に依存して生成されると考える．すると，x に対する y の条件付き分布は式 1 となる．

$$p(y|x) = \int_z p_\theta(y, z|x)dz = \int_z p_\theta(y|z, x)p_\theta(z|x)dz \tag{1}$$

われわれの目的は，式 1 をモデルパラメータ θ について最適化することである．以降，$p_\theta(z|x)$ と $p_\theta(y|z, x)$ をそれぞれ prior network および decoder と呼ぶ．

式 1 の最適化を解く枠組みとして CVAE を用いる．CVAE の学習では $p(y|x)$ の対数の変分下限を最大化する．事後分布 $q(z|y, x)$ を近似する recognition network $q_\phi(z|y, x)$ (ϕ はモデルパラメータ) を導入し，式 2 の目的関数を得る．

$$\begin{aligned}\mathcal{L}'(\theta, \phi; y, x) &= -KL(q_\phi(z|y, x)||p_\theta(z|x)) \\ &\quad + \mathbb{E}_{q_\phi(z|y,x)}[\log p_\theta(y|z, x)] \qquad (2) \\ &\leq \log p(y|x) \qquad (3)\end{aligned}$$

モデルの出力は，入力イベントの情報を十分に反映していることが望ましい．そこで decoder の出力から x を予測する reconstructor $p_\theta(x|y)$ を導入する．最終的な目的関数を式 4 に示す．

$$\begin{aligned}\mathcal{L}(\theta, \phi; y, x) &= \mathcal{L}'(\theta, \phi; y, x) \\ &\quad + \mathbb{E}_{q_\phi(z|y,x)}[\log p_\theta(x|y)p_\theta(y|z, x)] \qquad (4)\end{aligned}$$

3.2 モデルの詳細

単層の双方向 LSTM を encoder とする．はじめに encoder で x と y を読み込み，各単語の表現を前向きと後ろ向きの LSTM の出力を結合して得る．

z は非対角成分が 0 の多変量ガウス分布に従うものとし，$q_\phi(z|y, x) \sim \mathcal{N}(\mu, \sigma^2\boldsymbol{I})$，$p_\theta(z|x) \sim \mathcal{N}(\mu', \sigma'^2\boldsymbol{I})$ とする．文表現を前向き·後ろ向きの LSTM の最終状態を結合したものとし，μ と σ^2 は x と y の文表現を線形変換，μ' と σ'^2 は x の文表現を多層パーセプトロンで変換して得る．z のサンプリングには reparametrization trick を用いる． 学習時は recognition network，推論時は prior network の出力から z を得る．

単層 LSTM を decoder とする．i 番目の出力単語 y_i を得る際の decoder の入力は，y_{i-1} の単語埋め込みベクトル，潜在変数 z，文脈ベクトルからなる．文脈ベクトルは decoder の $i-1$ 時点の隠れ状態に基づく encoder の各時点の出力の重み付き和であり，注意機構で計算する．

単層 LSTM を reconstructor とする．i 番目の出力単語 x_i を得る際の reconstructor の入力は，x_{i-1} の単語埋め込みベクトルと文脈ベクトルからなる．文脈ベクトルは注意機構を用いて reconstructor の $i-1$ 時点の出力に基づく decoder の各時点の出力の重み付き和として得る．

学習時には KL cost annealing [7] を行う．また，日本語のイベントペアデータを学習する際は文末から順に系列を生成する．語順を反転すると意味の中核である述語を出力した後に項などの補足情報を出力することになり，予備実験において，定性的・定量的な品質の向上が確認できた．

	EventGraph			Wikihow			Descript		
	BLEU	Dist-1	Dist-2	BLEU	Dist-1	Dist-2	BLEU	Dist-1	Dist-2
S2S	1.87	0.61	2.08	1.63	3.10	12.72	3.84	3.75	14.14
S2S+ATT	**2.08**	**0.70**	**2.50**	2.04	3.41	14.86	3.95	4.29	18.27
S2S+RC	1.75	0.46	1.43	1.47	3.30	12.50	3.45	5.18	19.64
S2S+ATT+RC	1.44	0.46	1.43	1.67	3.60	14.53	4.52	5.57	22.33
CVAE	0.33	0.37	1.47	1.64	3.48	13.24	4.64	8.89	39.35
CVAE+ATT	0.36	0.51	2.04	1.96	3.69	14.65	**5.26**	8.61	38.22
CVAE+RC	1.60	0.61	1.43	1.63	3.56	13.51	4.66	8.75	40.26
CVAE+ATT+RC	1.15	0.28	0.65	**2.24**	**4.02**	**16.31**	4.69	**9.79**	**46.81**

表 1: 各データセットにおける BLEU および Distinct-N.

4 実験

4.1 データセット

12 億文の日本語ウェブテキストからイベントペアを抽出する．抽出には齋藤らのイベントグラフを用いた [8]．イベントグラフは述語項構造を拡張した自然言語文に近い単位でイベントを表し，イベント間の係り受け関係と談話関係を整理する．イベントグラフを適用し，約 7 億件のイベントペアを得た．

時系列に沿ったイベントペアを得るため,「原因・理由」「条件」「時間経過:同時」「時間経過:後」の談話関係を持つものを取り出した．また，ウェブテキストは顔文字など自動解析の困難な表現を多く含むため，特殊記号を含むイベントペアを除いた．さらに，ウェブテキストには特定のフォーマットに基づき自動生成された文も多く，こうした文から抽出されたイベントペアを除くため，系列長の長いイベントを含みかつ高頻度なイベントペアを除いた．開発データ・テストデータとして 10 万件ずつイベントペアを取り出した．それらと重複のない残りのイベントペアから無作為に 500 万件を取り出し学習データとした．これを **EventGraph** と呼ぶ.

先行研究 [2] で使用されたイベントペアデータである **Wikihow** と **Descript** においても実験を行う．Wikihow はタスクの実施手順を共有するウェブサイトから自動構築されている．Descript はいくつかのシナリオについて典型的なイベント系列をクラウドソーシングで収集したものである．これらはイベントを英語文で記述している．表 2 に各データセットの統計を示す.

	学習	開発	テスト
EventGraph	5,000,000	100,000	100,000
Wikihow	1,287,360	26,820	26,820
Descript	23,320	2,915	2,915

表 2: データセットの統計.

4.2 評価

出力された後続イベントの妥当性を **BLEU** で評価する．しかし，前件イベントに対する妥当な後件イベントは複数存在し得るため，BLEU は必ずしも適切でない [2]．そこで，Li らの流暢さ，関連性，論理的一貫性の観点に基づく評価 [9] をもとに，次の 5 段階評価を行う．

0 ：不自然な文が含まれる
+1 ：関連が全くない
+2 ：関連が薄い
+3 ：関連はあるが論理的に一貫していない
+4 ：関連があり論理的に一貫している

はじめに，テストデータの前件イベントから学習データ・開発データの前件イベントに存在しないものを 50 件取り出す．それぞれのイベントについて各モデルで後件イベントを 100 回サンプリングし，高頻度の 5 件を得る．イベントペアはそれぞれ 5 名のクラウドワーカが評価する．この評価は EventGraph でのみ実施する.

また，出力の多様性を Distinct-N (**Dist-N**) で評価する．Distinct-N は出力系列全体の n-gram のうちユニークなものの割合である．Distinct-N が高いほど，モデルは多様な単語や単語系列を出力をしていることが分かる．本研究では $N = 1, 2$ の場合を評価する.

比較手法として，一般的な seq2seq モデルに基づくイベント予測モデルを学習する (**S2S**)．また，CVAE，S2S それぞれについて注意機構 (**ATT**) と reconstructor (**RC**) を除いたモデルを学習する．BLEU と Distinct-N の計算時は，S2S は貪欲法でデコードし，CVAE は事前分布のモードから貪欲法でデコードする．後件イベントをサンプリングする際は，S2S は出力単語をカテゴリカル分布からサンプリングしながらデコードし，CVAE は事前分布から z をサンプリング

前件のイベント：ちょっと腹の具合が悪い			
S2S	S2S+ATT+RC	CVAE	CVAE+ATT+RC
1. 早朝から前述した車両を予約	1. ミルクを 1 日中心に摂取してます	1. 気にしない	1. 気にしている
2. その結果結構旨い餌が多い	2. 今回の増刊はあえなく断念	2. しない	2. 気になる
3. 使ってくれそう断る	3. ただ，エアコン効きとしては心配だ	3. 落ち着かない	3. 下痢をする
4. そのまま帰ってお帰る	4. いいのか検討してみる	4. でも，心配になる	4. 病院に行きたい
5. 満腹感はある	5. 余計に疲れる	5. 構わない	5. 諦めている

表 3: EventGraph を学習したイベント予測モデルの出力．

	0	+1	+2	+3	+4
S2S	.223	.278	.130	.111	.258
S2S+ATT	.183	.280	.194	.103	.240
S2S+RC	.244	.267	.132	.151	.206
S2S+ATT+RC	.205	.353	.115	.105	.222
CVAE	.094	.240	.176	.160	.330
CVAE+ATT	.077	.238	.230	.136	.319
CVAE+RC	.075	.185	.198	.170	.372
CVAE+ATT+RC	.074	.164	.202	.142	.418

表 4: クラウドソーシングによる 5 段階評価．

した後，貪欲法でデコードする．モデルは開発データの損失に基づき early stopping で選択する．

表 1 に BLEU と Distinct-N による自動評価の結果を示す．注意機構は BLEU・Distinct-N の両方について有効である一方，reconstructor は BLEU の評価値を損ねる場合があった．これは，reconstructor が多数の前件イベントに対応付く一般的な後件イベントの生成を抑制していることに起因すると考えられる．

表 4 にクラウドソーシングによる 5 段階評価の結果を示す．EventGraph を学習したモデルは，BLEU スコアにおいて CVAE は S2S に劣っていたが，人手による評価では CVAE の結果が勝る結果となり，BLEU による評価でモデルの品質を測ることは妥当でないことを確認した．

モデルの出力例を表 3 に示す．seq2seq に基づくモデルの出力は，非文が多く含まれ，前件イベントとの関係性も薄い傾向にあった．これは，出力単語をカテゴリカル分布からサンプリングして得ており，言語モデル的に不安定であることに起因すると考えられる．CVAE は簡素ではあるが文としては成立し，入力との関連性は比較的保たれていた．CVAE と CVAE+ATT+RC を比較すると，前者は後者より「〜ない」のような否定表現をより多く含む傾向にあった．

5 おわりに

本研究では，イベントが自然言語文で記述され，イベント系列が 2 要素からなる場合のイベント予測に取り組んだ．入力に対して出力が一意に定まらないという問題の特性に着目し，CVAE に基づく確率的生成モデルを学習した．また，入力イベントの情報を十分に反映した出力を促す機構としてモデル自身の出力から入力信号を復元する reconstructor を導入した．

実験では，従来のデータセットに加え，ウェブ文書から自動構築した日本語イベントペアデータを用いてイベント予測を学習した．自動評価指標とクラウドソーシングによる評価を行い，提案手法が妥当性・多様性においてベースラインを上回ることを示した．今後，構築した日本語イベントペアデータの大規模化・高品質化に取り組む他，述語項構造解析などへの応用を検討したい．

謝辞

本研究はヤフー株式会社の支援を受けた．ここに感謝の意を表する．

参考文献

[1] Karl Pichotta and Raymond Mooney. Statistical Script Learning with Multi-Argument Events. In *EACL*, pp. 220–229, 2014.

[2] Dai Quoc Nguyen, Dat Quoc Nguyen, Cuong Xuan Chu, Stefan Thater, and Manfred Pinkal. Sequence to Sequence Learning for Event Prediction. In *IJCNLP (2)*, pp. 37–42, 2017.

[3] Linmei Hu, Juanzi Li, Liqiang Nie, Xiao-Li Li, and Chao Shao. What Happens Next? Future Subevent Prediction Using Contextual Hierarchical LSTM. In *AAAI*, pp. 3450–3456, 2017.

[4] Tiancheng Zhao, Ran Zhao, and Maxine Eskenazi. Learning Discourse-level Diversity for Neural Dialog Models using Conditional Variational Autoencoders. In *ACL (1)*, pp. 654–664, 2017.

[5] Biao Zhang, Deyi Xiong, jinsong su, Hong Duan, and Min Zhang. Variational Neural Machine Translation. In *EMNLP*, pp. 521–530, 2016.

[6] Zhaopeng Tu, Yang Liu, Lifeng Shang, Xiaohua Liu, and Hang Li. Neural Machine Translation with Reconstruction. In *AAAI*, pp. 3097–3103, 2017.

[7] Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew Dai, Rafal Jozefowicz, and Samy Bengio. Generating Sentences from a Continuous Space. In *CoNLL*, pp. 10–21, 2016.

[8] 齋藤純, 坂口智洋, 柴田知秀, 河原大輔, 黒橋禎夫. 述語項構造に基づく言語情報の基本単位のデザインと可視化. 言語処理学会, pp. 93–96, 2018.

[9] Zhongyang Li, Xiao Ding, and Ting Liu. Generating Reasonable and Diversified Story Ending Using Sequence to Sequence Model with Adversarial Training. In *COLING*, pp. 1033–1043, 2018.