

構造的整合性に基づく最小ベイズリスクテーブル生成

吉田 大城¹ 出口 祥之² 坂井 優介¹ 上垣外 英剛¹ 渡辺 太郎¹

¹ 奈良先端科学技術大学院大学 ² NTT 株式会社

yoshida.daiki.ye6@naist.ac.jp hiroyuki.deguchi@ntt.com

{sakai.yusuke.sr9, kamigaito.h, taro}@is.naist.jp

概要

Text-to-Table は、テキストからテーブル形式の構造化データを生成するタスクである。大規模言語モデルの導入により生成の網羅性や柔軟性は飛躍的に向上したが、依然として不要な情報の過剰生成や誤った数値を生成するなど、出力品質が安定しないという課題が残っている。本研究では、構造的整合性を保つための効用関数を用いた最小ベイズリスク復号に基づくテーブル生成法 TableMBR を提案する。Rotowire データセットを用いた実験の結果、TableMBR は既存手法を上回る性能を示し、F1 スコアで最大 12% の改善を達成した。

1 はじめに

Text-to-Table は自然言語で書かれた入力テキストから、構造化されたテーブルを生成するタスクである。大規模言語モデル (large language model; LLM) の登場により、行・列ヘッダといったテーブルの骨格であるスキーマの生成を含めた、柔軟なテーブル生成が可能になりつつある [1, 2, 3]。一方、正しい行・列ヘッダを必要な数だけ生成し、それらに対して適切なセル値を生成するといった、構造的整合性の担保に課題がある。Map&Make [3] は、テーブル生成をスキーマの生成とセル値の生成に分け、スキーマを確定してから各セル値を生成することで、既存手法より優れた精度を達成した。しかし、Map&Make にはテーブル全体の構造的整合性を明示的に保つ機構がないため、余分な列の追加や数値の誤りが含まれるなど、未だ出力品質が安定しない。

機械翻訳などのテキスト生成タスクにおいては、期待効用を最大化することで誤りの少ない高品質なテキストを生成する最小ベイズリスク (minimum Bayes risk; MBR) 復号が提案されている [4, 5]。MBR 復号は、複数の出力仮説を生成し、自動評価指標などを効用関数として各仮説の期待効用を推定し、こ

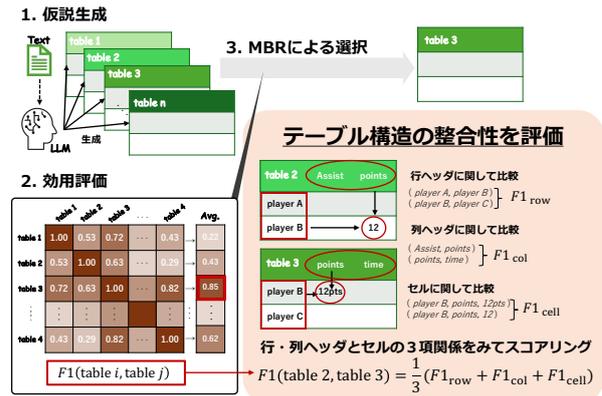


図 1: 提案法 TableMBR によるテーブル生成

れを最大化する仮説を出力する。

本研究では、テーブル生成における構造的整合性を改善するため、構造的整合性を効用とする MBR 復号に基づくテーブル生成法 (TableMBR) を提案する。図 1 に提案法 TableMBR の概要図を示す。提案法では、はじめに、Map&Make を用いて複数の仮説テーブルを生成する。続いて、正解とみなす擬似的な参照テーブルを (擬似参照) をサンプリング生成し、F1 スコアに基づく構造的効用を評価することで、各仮説の期待効用を推定する。最後に、期待効用を最大化する仮説テーブルを最終出力として選択する。提案法における効用関数は、行・列ヘッダの一致度を表す F1 スコアと、行・列ヘッダに対するセルの一致度を表す F1 スコアを合算して用い、これにより仮説と擬似参照との間のテーブル全体の構造的整合性を表現する。また、テーブル要素の一致度を表す F1 スコアの計算において、表記ゆれや曖昧性に対応するため、完全一致だけでなくテキスト間類似度に基づく柔らかい一致度も用いられるよう、類似度のアライメントに基づいて計算する。

Rotowire [6] を用いて Gemini-2.5 Flash [7] によるテーブル生成品質を評価した結果、提案手法の TableMBR はベースラインである Map&Make よりも F1 スコアを最大 12% 改善することを確認した。

2 背景および関連研究

2.1 Text-to-Table タスクにおける生成法

Wu ら [6] は、自然言語テキストからテーブルを生成する Text-to-Table タスクを提案し、seq2seq モデル [8] を用いて複雑な 2 次元構造を表現できることを示した。近年、LLM の推論能力を活用し、段階的に生成する手法が提案されている [1, 2, 3]。中でも、Map&Make [3] は入力テキストからの事実抽出、スキーマ生成、セル値の更新を段階的に行うことで、構造的整合性や情報集約の精度を向上させている。しかし、これらの手法にはテーブル全体の構造的整合性を明示的に担保する機構が存在しない。加えて、テーブル生成に不要な情報を誤って抽出した場合、セル値の生成もその誤った構造に縛られてしまう。そのため、LLM を用いてもなお、余分な行・列の過剰生成やセル値の欠落といった誤りが残る。

2.2 最小ベイズリスク復号

MBR 復号 [4, 5] は、出力仮説の期待効用を最大化することにより、高品質なテキストを生成する手法である。入力空間を \mathcal{X} 、出力空間を \mathcal{Y} とすると、従来の最大事後確率 (maximum a posteriori; MAP) 復号は、入力テキスト $x \in \mathcal{X}$ が与えられたとき、出力確率を最大化する仮説 $y_{\text{MAP}} \in \mathcal{Y}$ を出力する。ただし、出力空間 \mathcal{Y} 全体を探索することは実行不能なため、ビーム探索やサンプリングによって得られた有限の出力仮説集合 $\mathcal{H} \subset \mathcal{Y}$ から出力を選択する。

$$y_{\text{MAP}} = \operatorname{argmax}_{y \in \mathcal{H}} p(y|x; \theta). \quad (1)$$

なお、 θ はテキスト生成モデルの訓練済みパラメータを表す。一方、MBR 復号は、出力確率ではなく期待効用を最大化する仮説を選択する。

$$y_{\text{MBR}} = \operatorname{argmax}_{h \in \mathcal{H}} \mathbb{E}_{y \sim p(\cdot|x; \theta)} [u(h, y)]. \quad (2)$$

ここで、 $u: \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ は、参照出力 y が与えられたときの仮説 h の効用を評価する効用関数を表し、 $h \succeq_y h' \iff u(h, y) \geq u(h', y)$ を満たす。ただし、 \succeq_y は参照 y の下での選好関係を表す。

出力空間全体 \mathcal{Y} における期待値計算は困難であるため、通常、MBR 復号では $p(\cdot|x; \theta)$ からサンプリングされた有限個の擬似参照の集合 $\mathcal{R} = \{r_1, \dots, r_{|\mathcal{R}|}\}$ を用い、モンテカルロ近似によ

り、期待効用を推定する。

$$y_{\text{MBR}} \approx \operatorname{argmax}_{h \in \mathcal{H}} \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} u(h, r). \quad (3)$$

近年、MBR 復号は、コード生成 [9] や Text-to-SQL [10] といった構造化データの生成タスクにも応用されている。高品質な構造化データを生成するためには、実行可能性や論理的整合性といったドメインの特性を組み込んだ効用関数の設計が鍵となる。

3 提案手法：TableMBR

テーブル生成の品質改善に向け、構造的整合性を保つ機構を導入した生成法 TableMBR を提案する。TableMBR は、図 1 に示すように、複数の仮説テーブルを生成し、構造的整合性を評価する効用関数を用いた MBR 復号により、出力テーブルを選択する。提案手法は、1) 仮説生成、2) 効用評価、3) 仮説選択の手順を通して生成する。

1) 仮説生成 まず、Map&Make [3] を用いて、 $n \in \mathbb{N}$ 個の出力仮説の集合 $\mathcal{H} = \{h_1, \dots, h_N\}$ および擬似参照の集合 $\mathcal{R} = \{r_1, \dots, r_{|\mathcal{R}|}\}$ を生成する。仮説と擬似参照に多様性を持たせるため、サンプリングを用いて生成する。

2) 効用評価 次に、各仮説 $h \in \mathcal{H}$ について、各擬似参照 $r \in \mathcal{R}$ に対する F1 スコアに基づく効用を評価する。効用関数は、テーブル中の 2 つの要素間の類似度を測る類似度関数 sim を用い、 $u(h, r) = \text{F1}^{\text{sim}}(h, r)$ と定式化する。ヘッダによって定義される構造と内容との間の整合性を保つため、 F1^{sim} はテーブルの各構成要素である行ヘッダ row 、列ヘッダ col 、セル値 cell のそれぞれの F1 スコアの平均とする。

$$\text{F1}^{\text{sim}}(h, r) = \frac{1}{3} \left(\text{F1}_{\text{row}}^{\text{sim}}(h, r) + \text{F1}_{\text{col}}^{\text{sim}}(h, r) + \text{F1}_{\text{cell}}^{\text{sim}}(h, r) \right) \quad (4)$$

各構成要素 $\ell \in \{\text{row}, \text{col}, \text{cell}\}$ において、F1 スコア $\text{F1}_{\ell}^{\text{sim}}(h, r)$ は以下の式 (5) により算出される。

$$\text{F1}_{\ell}^{\text{sim}}(h, r) = \frac{2 \text{Prec}_{\ell}^{\text{sim}}(h, r) \text{Recall}_{\ell}^{\text{sim}}(h, r)}{\text{Prec}_{\ell}^{\text{sim}}(h, r) + \text{Recall}_{\ell}^{\text{sim}}(h, r)}. \quad (5)$$

本手法では、各構成要素を要素値の集合として扱い、要素単位で $\text{Prec}_{\ell}^{\text{sim}}$ および $\text{Recall}_{\ell}^{\text{sim}}$ を評価する。 \mathcal{E} を要素空間、 $\mathcal{E}_{\ell} \subseteq \mathcal{E}$ を構成要素 ℓ に関する要素部分空間、 $\phi_{\ell}: \mathcal{Y} \rightarrow 2^{\mathcal{E}_{\ell}}$ を与えられたテーブルから要素を抽出する関数とする。例えば、 $\phi_{\text{row}}(y) = \{\text{吉田}, \text{田中}, \dots\}$ 、 $\phi_{\text{col}}(y) = \{\text{名前}, \text{年齢}, \dots\}$ のようになる。セル (cell) については、セル値の

表 1: 提案手法 TableMBR における仮説数 $N = 32$ でのテーブル構成要素ごとの F1 スコア. 括弧内の数値 (.) は Map&Make からの改善幅を示す. 太字は, 指標ごとの最良の値を示す.

手法	効用	行ヘッダ			列ヘッダ			セル		
		$F1_{row}^{EM}$	$F1_{row}^{ChrF}$	$F1_{row}^{BERT}$	$F1_{col}^{EM}$	$F1_{col}^{ChrF}$	$F1_{col}^{BERT}$	$F1_{cell}^{EM}$	$F1_{cell}^{ChrF}$	$F1_{cell}^{BERT}$
Player テーブル										
Map&Make	–	85.27	89.44	89.38	24.35	50.46	66.56	28.31	46.28	57.56
TableMBR	$F1^{EM}$	85.55 (+0.28)	89.58 (+0.14)	89.61 (+0.23)	27.31 (+2.96)	53.30 (+2.84)	70.51 (+3.95)	30.26 (+1.95)	48.98 (+2.70)	61.00 (+3.44)
	$F1^{ChrF}$	85.58 (+0.31)	89.60 (+0.16)	89.64 (+0.26)	24.41 (+0.06)	51.63 (+1.17)	67.13 (+0.57)	27.85 (-0.46)	47.68 (+1.40)	58.13 (+0.57)
	$F1^{BERT}$	85.56 (+0.29)	89.59 (+0.15)	89.62 (+0.24)	25.91 (+1.56)	52.34 (+1.88)	69.11 (+2.55)	29.41 (+1.10)	48.48 (+2.20)	59.97 (+2.41)
Team テーブル										
Map&Make	–	9.56	60.95	30.71	8.96	33.72	40.32	0.94	16.05	9.88
TableMBR	$F1^{EM}$	9.68 (+0.12)	62.81 (+1.86)	28.70 (-2.01)	10.56 (+1.60)	34.65 (+0.93)	42.48 (+2.16)	0.80 (-0.14)	16.95 (+0.90)	10.36 (+0.48)
	$F1^{ChrF}$	8.89 (-0.67)	62.83 (+1.88)	27.63 (-3.08)	6.97 (-1.99)	33.30 (-0.42)	37.38 (-2.94)	0.69 (-0.25)	15.78 (-0.27)	9.03 (-0.85)
	$F1^{BERT}$	9.61 (+0.05)	62.71 (+1.76)	28.60 (-2.11)	9.80 (+0.84)	34.42 (+0.70)	41.02 (+0.70)	0.75 (-0.19)	16.69 (+0.64)	10.00 (+0.12)

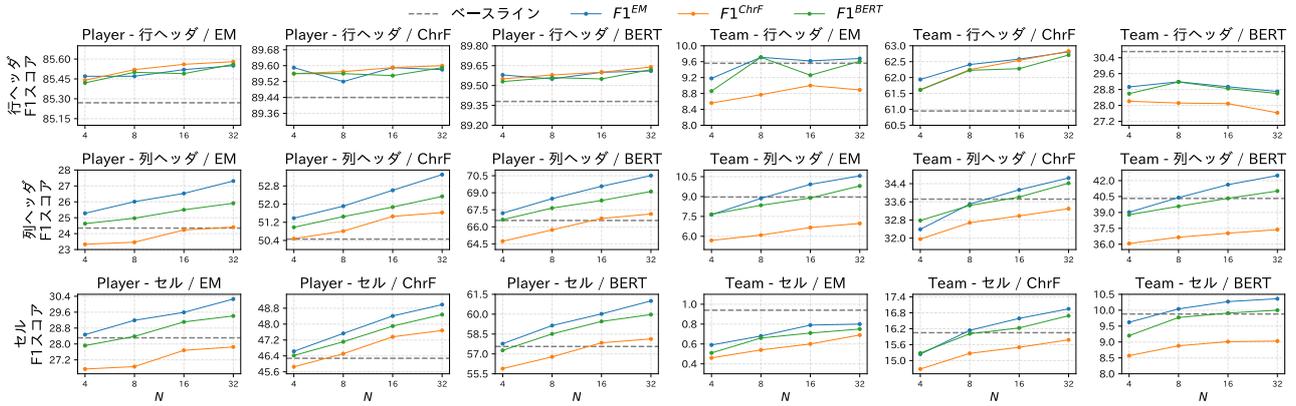


図 2: 仮説数 N を変化させたときの F1 スコア

みの集合では「どの行・列のセルであるか」という位置情報を特定できないため, 対応する行ヘッダと列ヘッダを用いた位置情報をセル要素の表現に含める. 例えば, $\phi_{cell}(y) = \{(\text{吉田}, \text{年齢}, 25), \dots \}$ となる. 適合率 $\text{Prec}_{\ell}^{\text{sim}}: \mathcal{Y} \times \mathcal{Y} \rightarrow [0, 1]$ および再現率 $\text{Recall}_{\ell}^{\text{sim}}: \mathcal{Y} \times \mathcal{Y} \rightarrow [0, 1]$ は, 類似度関数 sim を用い, 要素集合内から類似度を最大化する要素のライメントをとって計算する.

$$\text{Prec}_{\ell}^{\text{sim}}(h, r) = \frac{1}{|\phi_{\ell}(h)|} \sum_{e_h \in \phi_{\ell}(h)} \max_{e_r \in \phi_{\ell}(r)} \text{sim}(e_h, e_r) \quad (6)$$

$$\text{Recall}_{\ell}^{\text{sim}}(h, r) = \frac{1}{|\phi_{\ell}(r)|} \sum_{e_r \in \phi_{\ell}(r)} \max_{e_h \in \phi_{\ell}(h)} \text{sim}(e_h, e_r) \quad (7)$$

ここで, $\text{sim}: \mathcal{E} \times \mathcal{E} \rightarrow [0, 1]$ は, 要素間の一致度を表す類似度関数である. なお, セル (cell) については, 各要素がセル値だけでなくセルの位置情報まで含むため, 2つのセル間の類似度は, 要素を構成する位置情報とセル値それぞれの類似度の積として算出する.

3) 仮説選択 最後に, 擬似参照集合を用いて期待効用を推定し, これを最大化する仮説を選択する.

$$\text{argmax}_{h \in \mathcal{H}} \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} u(h, r). \quad (8)$$

4 実験

TableMBR の有効性を, 構造的整合性の観点から評価する. 特に入力テキストにテーブル生成に無関係な情報が含まれる場合に, 生成の過不足がどの程度抑制されるかを評価する.

実験設定 Rotowire [11] を用いて, テーブル生成実験を行う. Rotowire は, 入力テキストとなる NBA の試合要約から, プレイヤーとチームの情報に関する Player テーブルおよび Team テーブルをそれぞれ予測するタスクである¹⁾. 仮説生成には, Gemini-2.5 Flash [7] を用い, 各入力ごとに $N \in \{1, 4, 8, 16, 32\}$ 個の仮説を生成したときの生成品質をそれぞれ評価する. なお, $N = 1$ は Map&Make [3] に相当する. 多様な仮説を生成するためサンプリングを用い, サンプリングのパラメータは, 温度は 1.0 に, top- k の k

1) 詳細は付録 A を参照

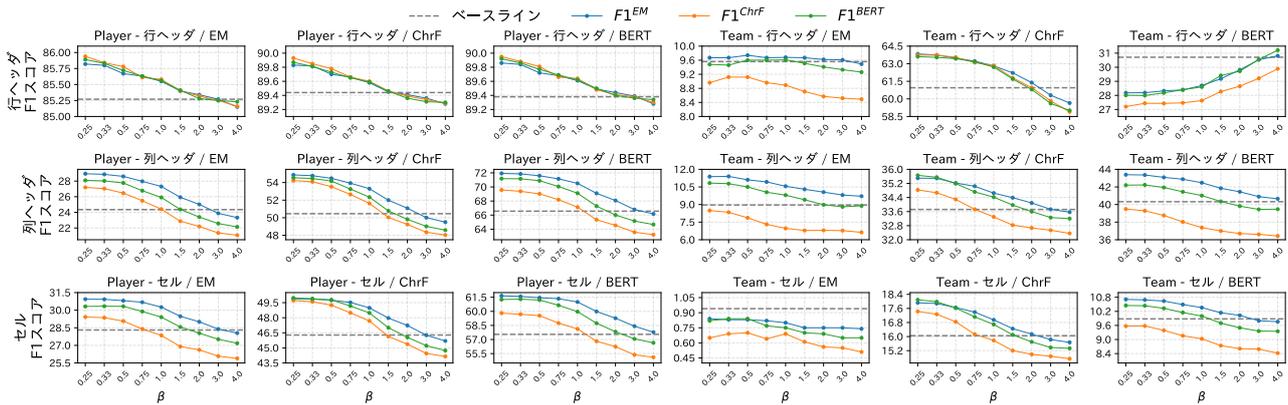


図 3: 効用関数に $F\beta$ を用いたときの β の影響

は 40 に, $\text{top-}p$ の p は 0.95 にそれぞれ設定する. 仮説集合と擬似参照集合には, 同一の集合を用いる. 構造的整合性の効用関数および評価指標には, 完全一致 (EM), ChrF [12], および BERTScore²⁾ (BERT) [14] に基づく F1 スコアを使用する.

実験結果 表 1 に仮説数 $N = 32$ における各テーブル構成要素の F1 スコアを示す. 表からわかるように, TableMBR は Player テーブルにて大幅な F1 スコアの向上が確認できた. 具体的には, 効用関数として $F1^{EM}$ を用いた場合, $F1^{BERT}_{col}$ が 2.96% 改善した. また, 図 2 に示すように, 仮説数 N を増やすにつれ TableMBR の F1 スコアが改善することを確認した. 一方, ベースラインのスコアが極めて低かった Team テーブルでは, TableMBR の効果は一貫しなかった. 一部の設定ではわずかな改善が見られたものの, ベースラインの性能を下回るものもあり, 仮説集合の品質が極めて低い場合には TableMBR の利点が限定的であることが示唆された.

5 考察：適合率と再現率のバランス

効用関数に F1 ではなく $F\beta$ を用いて β を変化させたときの適合率と再現率のバランスが生成品質に与える影響を分析する. 具体的には, $\beta \in \{0.25, 0.33, 0.5, 0.75, 1.0, 1.5, 2.0, 3.0, 4.0\}$ と変化させる. なお, $\beta < 1$ は適合率重視の設定, $\beta > 1$ は再現率重視の設定に対応する. 図 3 に示すように, 両テーブルにおいて, β が小さいほど, すなわち適合率重視の設定ほど, 高い F1 スコアが得られることを確認した. これは自動要約のような再現率が重視されるテキスト生成タスクとは異なり, テーブル生成においては, 誤った行・列の混入は全体の整合性を損なうため, 過剰生成を抑制する適合率重視の

設定が重要であることを示唆する.

また, この結果の傾向を, Rotowire データセットの特徴と Map&Make の生成プロセスの関係からも分析する. Map&Make を用いる利点は, 定義済みのテーブル構造を必要とせず, 入力テキストから自動的にスキーマを抽出できる点にある. しかし, この高い汎用性の副作用として, テキスト内のさまざまな情報を不要な列として拾い上げてしまい, 過剰生成を招くことがある. 実際に, ベースラインの Map&Make を用いたテーブル生成の結果を確認したところ, 不要な列が生成されていた. これに対し, TableMBR において適合率重視の設定 ($\beta = 0.25$) では, こうした無関係な列の生成を抑制できていた³⁾. したがって, 適合率重視の TableMBR は Map&Make と比べ, 過剰生成を招くノイズの影響を軽減することに成功しているといえる.

6 おわりに

本稿では, 構造的整合性を明示的に保つ Text-to-Table 生成手法である TableMBR を提案した. 提案手法の TableMBR は, 複数の仮説間におけるテーブルの行, 列, およびセルの合致に基づき, 期待効用を最大化するテーブルを選択する. Rotowire を用いたテーブル生成実験の結果, 提案法の TableMBR が, 従来手法より最大 12% F1 スコアを改善した. さらに, 効用関数の設計において, 適合率を重視する設計が, 構造化データ生成において有効であることを明らかにした.

今後は, 高い信頼性が求められる医療・法律ドメインなど, 他のドメインにおける TableMBR の有効性を調査し, テーブル生成における致命的な誤りの削減を目指す.

2) google-bert/bert-base-uncased [13]

3) 詳細は付録 B を参照

参考文献

- [1] Zheyang Deng, Chunkit Chan, Weiqi Wang, Yuxi Sun, Wei Fan, Tianshi Zheng, Yauwai Yim, and Yangqiu Song. Text-tuple-table: Towards information integration in text-to-table generation via global tuple extraction. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, **Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing**, pp. 9300–9322, Miami, Florida, USA, November 2024. Association for Computational Linguistics.
- [2] Anirudh Sundar, Christopher Richardson, and Larry Heck. gtbls: Generating tables from text by conditional question answering, 2024.
- [3] Naman Ahuja, Fenil Bardoliya, Chitta Baral, and Vivek Gupta. Map&make: Schema guided text to table generation. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, **Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 30249–30262, Vienna, Austria, July 2025. Association for Computational Linguistics.
- [4] Shankar Kumar and William Byrne. Minimum Bayes-risk decoding for statistical machine translation. In **Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004**, pp. 169–176, Boston, Massachusetts, USA, May 2 - May 7 2004. Association for Computational Linguistics.
- [5] Bryan Eikema and Wilker Aziz. Is MAP decoding all you need? the inadequacy of the mode in neural machine translation. In Donia Scott, Nuria Bel, and Chengqing Zong, editors, **Proceedings of the 28th International Conference on Computational Linguistics**, pp. 4506–4520, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics.
- [6] Xueqing Wu, Jiacheng Zhang, and Hang Li. Text-to-table: A new way of information extraction. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 2518–2533, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [7] Gheorghe Comanici, Eric Bieber, Mike Schaeckermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, Luke Marris, Sam Petulla, Colin Gaffney, Asaf Aharoni, Nathan Lintz, Tiago Cardal Pais, Henrik Jacobsson, Idan Szpektor, Nan-Jiang Jiang, Krishna Haridasan, Ahmed Omran, Nikunj Saunshi, Dara Bahri, Gaurav Mishra, Eric Chu, Toby Boyd, Brad Hekman, Aaron Parisi, Chaoyi Zhang, Kornraphop Kawintiranon, Tania Bedrax-Weiss, Oliver Wang, Ya Xu, Ollie Purkiss, Uri Mendlovic, Ilaï Deutel, Nam Nguyen, Adam Langley, Flip Korn, Lucia Rossazza, Alexandre Ramé, Sagar Waghmare, Helen Miller, Nathan Byrd, Ashrith Sheshan, Raia Hadsell, Sangnie Bhardwaj, Pawel Janus, Tero Rissa, Dan Horgan, Alvin Abdagic, Lior Belenki, James Allingham, Anima Singh, Theo Guidroz, Srivatsan Srinivasan, Herman Schmit, Kristen Chiafullo, Andre Elisseeff, Nilpa Jha, Prateek Kolhar, Leonard Berrada, Frank Ding, Xiance Si, Shrestha Basu Mallick, Franz Och, Sofia Erell, Eric Ni, Tejasi Latkar, Sherry Yang, Petar Sirkovic, Ziqiang Feng, Robert Leland, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities, 2025.
- [8] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 7871–7880, Online, July 2020. Association for Computational Linguistics.
- [9] Freda Shi, Daniel Fried, Marjan Ghazvininejad, Luke Zettlemoyer, and Sida I. Wang. Natural language to code translation with execution. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 3533–3546, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- [10] Anonymous. Exploring minimum bayes risk decoding for text-to-SQL ensemble. In **Submitted to The Fourteenth International Conference on Learning Representations**, 2025. under review.
- [11] Sam Wiseman, Stuart Shieber, and Alexander Rush. Challenges in data-to-document generation. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing**, pp. 2253–2263, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.
- [12] Maja Popović. chrF: character n-gram F-score for automatic MT evaluation. In Ondřej Bojar, Rajan Chatterjee, Christian Federmann, Barry Haddow, Chris Hokamp, Matthias Huck, Varvara Logacheva, and Pavel Pecina, editors, **Proceedings of the Tenth Workshop on Statistical Machine Translation**, pp. 392–395, Lisbon, Portugal, September 2015. Association for Computational Linguistics.
- [13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [14] Tianyi Zhang*, Varsha Kishore*, Felix Wu*, Kilian Q. Weinberger, and Yoav Artzi. BERTscore: Evaluating text generation with bert. In **International Conference on Learning Representations**, 2020.

正解データ

Player Name	Assists	3-pointers attempted	3-pointers made	Field goals attempted	Field goals made	Points	Total rebounds
John Wall	10	None	None	None	None	30	None
Bradley Beal	None	None	None	None	None	23	None
Gary Harris	None	9	6	15	10	26	None
Wilson Chandler	None	None	None	None	None	21	None
Mason Plumlee	None	None	None	None	None	19	10

$\beta = 0.25$

Player Name	Assists	Three Pointers Attempted	Three Pointers Made	Field Goals Attempted	Field Goals Made	Points	Rebounds	Team
John Wall	10	None	None	36	21	30	None	Washington Wizards
Bradley Beal	None	None	None	None	None	23	None	Washington Wizards
Gary Harris	None	9	6	15	10	26	None	Denver Nuggets
Wilson Chandler	None	None	None	None	None	21	None	Denver Nuggets
Mason Plumlee	None	None	None	None	None	19	10	Denver Nuggets

$\beta = 1.0$

Player Name	Assists	Three Pointers Attempted	Three Pointers Made	Field Goals Attempted	Field Goals Made	Points	Rebounds	Team	Field Goal Percentage	Double Double
John Wall	10	None	None	36	21	30	None	Washington Wizards	58%	No
Bradley Beal	None	None	None	None	None	23	None	Washington Wizards	None	No
Gary Harris	None	9	6	15	10	26	None	Denver Nuggets	None	No
Wilson Chandler	None	None	None	None	None	21	None	Denver Nuggets	None	No
Mason Plumlee	None	None	None	None	None	19	10	Denver Nuggets	None	Yes

$\beta = 4.0$

Player Name	Assists	Three Pointers Attempted	Three Pointers Made	Field Goals Attempted	Field Goals Made	Points	Rebounds	Team	Field Goal Percentage	Double Double	Game Date
John Wall	10	None	None	36	21	30	None	Washington Wizards	58%	No	Wednesday
Bradley Beal	None	None	None	None	None	23	None	Washington Wizards	None	No	Wednesday
Gary Harris	None	9	6	15	10	26	None	Denver Nuggets	None	No	Wednesday
Wilson Chandler	None	None	None	None	None	21	None	Denver Nuggets	None	No	Wednesday
Mason Plumlee	None	None	None	None	None	19	10	Denver Nuggets	None	Yes	Wednesday

テキスト

The Washington Wizards defeated the Denver Nuggets, 123 - 113, Wednesday at Pepsi Center. The Wizards (39 - 24), who have now won three straight road games to improve their record away from home to 13 - 15, won **Wednesday's** matchup in fairly effortless fashion, as they opened up the game early and went into halftime with a 17 - point lead. That was thanks in large part to an incredibly hot shooting performance from Washington, as they shot 57 percent from the field and 42 percent from behind the arc. The Wizards were once again led by a pair of great performances from their starting backcourt, with John Wall posting a game - high 30 points and 10 assists, while Bradley Beal scored 23 points of his own. The two shot a combined 21 - of - 36 (**58 percent**) from the field. For the Nuggets (29 - 35), **Wednesday's** defeat marked their second loss in their last three games, and it was thanks in large part to their lack of defense, as Denver was pretty solid on the offensive side of the ball. As a team, the Nuggets shot 53 percent from the field and a scorching 46 percent from behind the arc, but the 123 points they gave up ended up being too much to overcome. Gary Harris led the way for Denver, shooting 10 - of - 15 from the field and 6 - of - 9 from behind the arc en route to scoring a game - high 26 points. Wilson Chandler added 21 points of his own, while Mason Plumlee added 19 points and 10 rebounds for a **double - double**. Up next, the Wizards will head to Sacramento on Friday to take on the Kings, while the Nuggets will head home Friday and play the Celtics.

図 4: Rotowire データセットにおける生成例の比較。上段は正解データ, 下段は MBR 復号のパラメータ β を変化させた際の選択結果を示す。図に示す通り, 再現率を重視する ($\beta = 4.0$) 設定では, 入力テキストに含まれるスタッツ以外の情報 (試合日や定性評価など) を反映した不要な列が増加する傾向が見られた。対照的に, 適合率を重視した設定 ($\beta = 0.25$) では, これらの過剰な列生成が抑制され, 正解データに近いコンパクトなスキーマが選択されている。

A データセットの詳細

本研究では使用する Rotowire データセットは NBA の試合結果に関する記述と, Team および Player の 2 種類の統計テーブルで構成される。本データセットは, 以下の特性から「TableMBR」が目的とする構造的整合性の評価に極めて適している: (1) **冗長性とノイズ**: 入力テキストには試合展開や選手の感情, 他試合の成績といったテーブル生成には不要な冗長な情報が多数含まれており, 必要な事実のみを選択し誤った行・列の生成を抑制する能力 (事実的・構造的整合性) が厳密に評価される。(2) **表現の多様性とエンティティ識別**: チーム名 (Knicks) とその本拠地 (New York) のような別名や類義語が頻出するため, これらを正確に同一エンティティとして識別し, テーブルの行・列ヘッダとして一貫性を保つ能力も求められる。

B 定性評価

図 4 は Rotowire における生成例の比較である。Rotowire の入力テキストには, “Wednesday’s matchup” や “double-double” のように, テーブル化すべきでない

記述が多く含まれている。入力テキストから動的にスキーマを抽出する Map&Make はこれらの要素を等価に抽出するため, 図 4 の上段に示すように, “Game Date” や “Double Double” といった不要な列やハルシネーションによるヘッダを頻繁に生成する。これに対し, 図 4 の中段に示す $\beta = 0.25$ の適合率重視設定の結果では, 無関係な列を抑制しつつ, 適切な列のみを保持することに成功している。これは, 適合率重視の TableMBR が, 複数の仮説間で合意に至らない冗長なセルを含む仮説に対して強いペナルティを課すためである。このように, TableMBR は Map&Make の汎用性を維持しつつ, 構造的整合性の観点から過剰なノイズをフィルタリングすることに成功している。結論として, 誤情報の排除を優先する適合率重視の効用関数は, 構造化データ生成において不可欠であり, 汎用的なプロンプティング手法の実用性を高める鍵となると言える。