Appendix for Enhancing Content Selection and Planning for Table-to-Text Generation with Data Understanding and Verification

A Implementation Details

A.1 Data Statistics

We conduct experiments on ROTOWIRE (Wiseman et al., 2017) and MLB (Puduppully et al., 2019b) datasets. The statistics are listed in Table 1. For ROTOWIRE, we use the official dataset with official splits. For MLB, as the contents are not released, we retrieve the dataset via official scripts [1]. Since the script misses some important types of records that appears in reference (extracted by IE model), we add "team runs" and "bf", "out", "bs" for pitchers into the input. Corresponding statistics is in Table 1. For preprocessing on those two datasets, we follow the one used by Puduppully et al. (2019a) and Puduppully et al. (2019b)'s released code for the corresponding dataset.

A.2 Hyper-Parameters

We follow the training configurations of Puduppully et al. (2019a) and Puduppully et al. (2019b) in the base model for ROTOWIRE and MLB respectively (Table 2).

As for pre-trained task for contextual numerical value representations, we set transformer layers as 2 from $\{1, 2, 3\}$, fead-forward hidden size as 1024 from $\{512, 1024\}$, numbers of heads as 3 from $\{2, 3, 4\}$, and the ranking margin as 0.3 from $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. The pre-trained task is optimized with Adam with warm-up steps 1000 and learning rate schedule as the one in Vaswani et al. (2017). We also explore fixing or finetuning parameters from the pre-trained task jointly with parameters of the table-to-text generation model. Results show that treating pre-trained model fixed yields better performance and predicting which numerical value's contextual representation has higher value with 99.99% accuracy on held-out data.

[1] https://github.com/ratishsp/mlb-data-scripts

For content planning verification module, we discuss weights for λ_1 - λ_5 from $\{$(0.2, 0.2, 0.2, 0.2, 0.2), (0.15, 0.15, 0.15, 0.15, 0.3), (0.15, 0.25, 0.1, 0.2, 0.3)$\}$. The first gives equal importance for all, the second gives equal importance for precision, recall and content ordering and the last is based on model's performance from following aspects: Entity Importance (EI), Record Importance (RI), Entity Recall (ER), Record Recall (RR) and Record Ordering (RO). The poorer the performance, the higher the weight. We find that the last setting of λ_1 - λ_5 performs the best. We also choose λ_6 as 0.3 from $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ and $beta$ as 0.2 from $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. Hyper-parameters of DUV, discussed in this section, are chosen based on performance on development set.

	ROTOWIRE	MLB
Vocab Size	11.3K	42.8K
# Tokens	1.5M	15.7M
# Instances	4.9K	26.3K
Avg Length	337	595
# Record Types	39	108
Avg Records	628	591

Table 1: Statistics of ROTOWIRE and MLB. MLB is re-collected via official script because the contents are not released.

A.3 Training Statistics

We train our models on a NVIDIA GeForce RTX 2080 Ti with 11GB memory for ROTOWIRE. As for MLB, we train the models on a NVIDIA Tesla V100-SXM3-32GB. Please note that models for MLB can still fit into the GPU used for ROTOWIRE and the difference is merely due to the availability of GPUs at that moment.

On ROTOWIRE, DUV costs 1 minutes to train with MLE and 4 minutes for finetuing with pol-

	ROTOWIRE	MLB
Word Embeddings	600	300
Hidden state hize	600	600
LSTM Layers	2	1
Input Feeding	Yes	Yes
Dropout	0.3	0.3
Optimizer	Adagrad	Adagrad
Initial learning rate	0.15	0.15
Learning rate decay	0.97	0.97
Epochs	25	25
BPTT size	100	100
Batch size	5	12
Inference beam size	5	5

Table 2: Hyper-parameters of base model (Module 2).

icy gradient a epoch for Stage 1 (12.42M parameters). For Stage 2 (32.35M parameters) it costs 20 minutes a epoch (these two stages can be run in parallel). NCP (33.94M parameters) takes 21 minutes for a epoch (including Stage 1 and 2). On MLB, DUV costs 5 minutes to train with MLE and 14 minutes for finetuing with policy gradient for a epoch in Stage 1 (11.23M parameters). For Stage 2 (35.24M parameters), it costs 51 minutes for a epoch. NCP (42.57M parameters) takes 65 minutes a epoch, while ENT (34.27M parameters) takes 176 minutes.

B Validation Performance

We presents the comparing methods' (Sec.4.2) performance on development set in Table 3. They shows the same pattern as results on test set in the paper.

C Qualitative Example

Due to page limit, we present the qualitative example on ROTOWIRE in the paper, and the example on MLB here. Compared to base model NCP, DUV generate a more concise and meaningful text report describing both important MLB statistical and event data. Compared to ENT, DUV can include more important and correct statistical data with less redundant ones, while perform less satisfying with repect to event data.

References

Ratish Puduppully, Li Dong, and Mirella Lapata. 2019a. Data-to-text generation with content selection and planning. *Proceedings of AAAI Conference on Artificial Intelligence*.

Ratish Puduppully, Li Dong, and Mirella Lapata. 2019b. Data-to-text generation with entity modeling. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008.

Sam Wiseman, Stuart M. Shieber, and Alexander M. Rush. 2017. Challenges in data-to-document generation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

ROTOWIRE (RW)	**RG**		**CS**			**CO**	**BLEU**
	P%	**#**	**P%**	**R%**	**F1%**	**DLD%**	
TEMP	**99.92**	**54.23**	26.60	**59.13**	36.69	14.39	8.62
ED+CC	75.10	23.95	28.11	35.86	31.52	15.33	14.57
NCP+CC (NCP)	87.51	33.88	33.52	51.21	40.52	18.57	16.19
ENT	91.97	31.84	36.65	48.18	41.63	19.68	15.97
HETD	91.84	32.11	35.39	48.98	41.09	20.70	16.24
NCP(R)	85.60	26.60	36.29	44.68	40.05	19.68	15.16
S-NCP	84.07	26.78	35.38	45.16	39.68	19.36	15.32
S-NCP+V (S-N+V)	84.70	25.40	37.17	44.68	40.58	19.47	14.60
DU	87.23	28.81	39.03	51.64	44.46	22.97	**16.64**
DUV	87.35	26.11	**42.00**	50.63	**45.91**	**24.86**	16.29

MLB	**RG**		**CS**			**CO**	**BLEU**
	P%	**#**	**P%**	**R%**	**F1%**	**DLD%**	
TEMP	**98.02**	**57.46**	23.15	**66.49**	34.34	10.54	2.77
ED+CC	91.67	17.35	63.08	48.50	54.84	25.97	9.64
NCP+CC (NCP)	88.38	15.99	**63.40**	52.12	57.21	27.51	8.24
ENT	85.41	22.12	55.18	61.53	58.18	24.07	**12.84**
S-NCP	88.08	16.74	62.22	54.43	58.07	27.75	9.47
S-NCP+V (S-N+V)	88.32	16.72	62.57	54.70	58.37	28.21	9.53
DU	88.41	16.88	62.41	54.91	**58.42**	27.92	9.58
DUV	88.72	16.63	62.95	54.50	**58.42**	**28.34**	9.45

Table 3: Automatic evaluation results on development set. We use the same Information Extraction (IE) models as described in the automatic evaluation results on test set in the paper.

KANSAS CITY , Mo . - - Adam Dunn and the Chicago White Sox are the only team in the majors that has n't been in first place in the AL Central . Adam Dunn and Adam Dunn *homered* and *doubled* in the go - ahead run in the 12th inning to lift the White Sox to a **9** - **8** victory over the Kansas City Royals on Friday night . The White Sox have won five of six and are a major league - best 17 - 5 since the All-Star break . The White Sox have lost six of seven . They have won seven of eight . The White Sox have lost six of seven . They have won seven of eight . The White Sox have lost six of seven . They have won seven of eight . The White Sox have lost five of six . They have won seven of eight . The White Sox have lost five of six . They have won seven of eight . The White Sox have lost five of six . They have won seven of eight . The White Sox have lost five of six . They have won seven of eight . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost six of seven . They have lost five of six . The White Sox have lost five of six . The White Sox have lost five of six . The White Sox have lost five of six . The White Sox have lost five of six ....

NCP+CC (NCP)

KANSAS CITY , Mo . - - There was nobody White Sox manager Robin Ventura would rather have had at the plate with runners on the corners in the 14th inning than his gritty new third baseman . Kevin Youkilis hit a **sacrifice fly** on the 14th pitch of his marathon at - bat , scoring Gordon Beckham and giving Chicago a **9** - **8** victory over the Kansas City Royals in a game that stretched into the wee hours of Saturday morning . Youkilis earned the praise , but it was Beckham who got the winning rally going . He drew a one - out walk off Everett Teaford (**1** - **3**) , and Alejandro De Aza sent a **single** down the third - base line to put runners on the corners for Youkilis , who muscled a fly ball deep to center to drive in the eventual winning run . Dylan Axelrod (**1** - **2**) worked around trouble in the 13th for Chicago , and then retired the side in the 14th to end a game that lasted 5 hours , 23 minutes . Both teams had chances to end it much earlier . Royals closer Jonathan Broxton blew a **save** in the ninth inning and White Sox closer Addison Reed did the same in the 12th . Both teams burned through nine pitchers , setting a franchise record for the Royals . Teaford was scheduled to start Tuesday night before being pressed into relief duty . The game also came close to ending in the 13th inning , when Chicago 's Dayan Viciedo **earned a walk** from Tim Collins , and Beckham **doubled** to score pinch - runner Jordan Danks from second base . Jeff Francoeur answered with a **double** of his own leading off the bottom of the inning , and after Brayan Pena **grounded out** , Lorenzo Cain **singled** to put runners on the corners . Alex Gordon sent a chopper up the middle that scored Francoeur and kept the game alive . Of course , that was a microcosm of the entire back - and - forth affair . Adam Dunn and Alex Rios **homered** to give Chicago a **3** - **0** lead in the first inning , Mike Moustakas **homered** to help the Royals score three times in the second , and Kansas City added **two** more runs in the fourth - - one coming on a **homer** by Francoeur down the left - field line . The White Sox charged back in the fifth . Dunn and A.J . Pierzynski **singled** off Royals starter Bruce Chen , and Viciedo connected for a three - run **homer** and a **6** - **5** lead . The Royals were down to their final out in the eighth inning when Cain **doubled** and Gordon **walked** . Alcides Escobar came through with a clutch two - run **triple** . Broxton , who seems to make every outing an adventure , gave up a **single** to Youkilis to start the ninth . Dunn and Paul Konerko **walked** to load the bases with no outs , and after Rios **flied out** to center , Pierzynski sent **an RBI single** to right field . Francoeur came up throwing as pinch - runner Orlando Hudson tried to score . The throw was well up the first - base line , but it arrived in plenty of time for catcher Salvador Perez to lunge back across home and make the tag - - even though replays showed that Hudson may have been safe . Ventura argued to no avail with plate umpire Chris Guccione . The Royals put a runner aboard with one out in the 10th and could n't get him home , and loaded the bases with one out in the 11th before Yuniesky Betancourt popped out and Moustakas **grounded out** to shortstop to keep the game going well into the night .

Gold

KANSAS CITY , Mo . - - The Chicago White Sox had a chance to take the lead in the bottom of the 12th inning . The White Sox scored three runs in the bottom of the ninth to take a two - run lead , but the White Sox rallied for an ***8*** - **8** victory over the Kansas City Royals on Friday night . The White Sox scored three times in the 12th to take a *6* - *5* lead , but the White Sox tied it in the bottom half on a two - run **double** by Alex Gordon and a two - run **triple** by Alcides Escobar . Dylan Axelrod (**1** - **2**) pitched **two** scoreless innings for the win and Jonathan Broxton gave up a two - out *RBI double* to A.J . Pierzynski in the ninth that tied it **7** - **7** . The White Sox tied it at **7** in the 12th on a two - run *homer* by Jeff Francoeur and a two - run *triple* by Alcides Escobar . The White Sox tied it at **7** in the bottom half on a *sacrifice fly* by A.J . Pierzynski and a two - run **homer** by Jeff Francoeur , who had **three** hits . Dayan Viciedo also **homered** and drove in *three* runs for the White Sox , who have won five of six . Chicago starter Bruce Chen gave up **six** runs and **nine** hits in **4 2/3** innings . He walked **three** and struck out **five** . Jose Quintana gave up **five** runs and **eight** hits in **five** innings for Kansas City . He walked **three** and struck out **four** . Alex Rios and Mike Moustakas **homered** for the Royals , who have lost six of seven . The White Sox scored three runs in the first inning off Bruce Chen and three in the first . Dunn hit a two - out , two - run **homer** in the first , and Viciedo hit a three - run shot in the fifth to give the White Sox a **6** - **5** lead . The White Sox tied it in the bottom half on Kevin Youkilis ' **sacrifice fly** and A.J . Pierzynski 's two - out **single** in the ninth .

ENT

KANSAS CITY , Mo . - - The Chicago White Sox had a chance to win it . Dayan Viciedo hit a go - ahead *double* in the 12th inning and the White Sox rallied for a **9** - **8** victory over the Kansas City Royals on Tuesday night . The White Sox have won four straight and eight of nine . The White Sox have lost four straight and eight of 10 . Everett Teaford (**1** - **3**) picked up the win with two innings of scoreless relief . Dylan Axelrod (**1** - **2**) pitched two scoreless innings for his first major league win . Mike Moustakas **homered** for the White Sox , who have lost four straight and seven of eight . The White Sox scored three runs in the second inning to take a *5* - **3** lead . Moustakas led off with a *single* , advanced to third on a wild pitch and scored on a **single** by Alcides Escobar . The White Sox answered with three runs in the bottom half . Alcides Escobar hit an RBI single and Alcides Escobar followed with an RBI single to give the White Sox a 5 - 3 lead . Alex Rios hit a solo **home run** in the second for Kansas City , but the White Sox responded with three runs in the bottom half . Moustakas led off the inning with a single , advanced to third on a wild pitch and scored on a single by Alcides Escobar . The White Sox answered with three runs in the bottom of the inning . Mike Moustakas led off with a home run to right , his third of the season , to give the White Sox a 4 - 3 lead . The White Sox tied it in the bottom of the inning on a two - out , three - run homer by Mike Moustakas .

DUV (Ours)

Figure 1: An example on MLB with the generated result NCP, gold, ENT, and our model's text. Important/unimportant records are in red/blue. Text that accurately/incorrectly reflects the statistics in table is in bold/italic.