Studying the Impact of Filling Information Gaps on the Output Quality of Neural Data-to-Text

Craig Thomson, Zhijie Zhao, Somayajulu Sripada
Department of Computing Science
University of Aberdeen, UK

Example generated basketball summary

The Memphis Grizzlies (5-2) defeated the Phoenix Suns (3-2) 102-91 on Monday at the Talking Stick Resort Arena in Phoenix.

*Isaiah Thomas added 15 points, he is **averaging 19 points on the season so far**.*

The Suns' next game will be on the road against the Boston Celtics on Friday.

Not in in the data!

This text is generated by a system using the Rotowire dataset (Wiseman et al. 2017), but some facts are not available in the data.

How can we find what is missing and:

- Add it from another source.
- **Note the types of data we cannot add**.

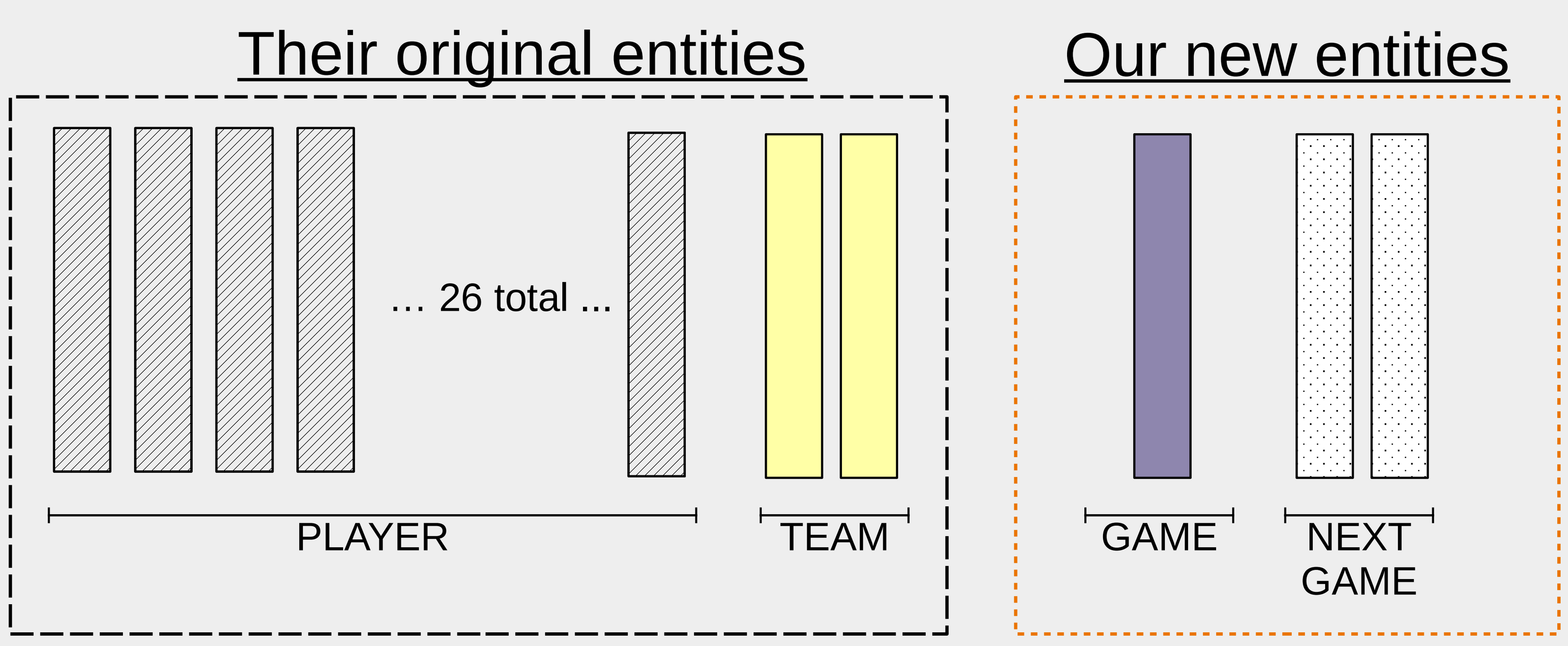
Identifying gaps using assisted corpus analysis

We performed a corpus analysis using spaCy, replacing named entities (and some domain specific syntax) in the text with type tokens before grouping sentences and ordering by the most common pattern.

"The Memphis Grizzlies (5-2) defeated the Phoenix Suns (3-2) 102-91 on Monday at the Talking Stick Resort Arena in Phoenix."
becomes
"The PROPN-ORG (X-Y) defeated the PROPN-ORG (X-Y) on NOUN-DATE at the PROPN-FAC in PROPN-GEP"

Sentences with a similar verb and/or a subset of these entities appear 100s of times in the corpus!

Hierarchical encoder - Rebuffel et al. 2020



Entities added to encoder input

Added entities from SportSett database (Thomson et al. 2020) database:

- Stadium name, capacity, city, and state.
- Day of the week.
- Team names and who the host is.
- Divisions and conferences for both teams
- Final score (GAME only).
- Stadium attendance (NEXT-GAME only).

Ablation study results

All metrics increased:

- Regardless of stopping metric
- $p < 0.005$ (except for BLEU)

Datasets:
D1 = Original entities
D2 = D1 + New entities

RG metric (Wiseman et al. 2017) adjusted to detect new entities.

Dataset	Stopping Metric	BLEU	RG	CS-PREC	CS-REC	CO
D1	BLEU	17.18 ± 0.386	0.70 ± 0.021	0.39 ± 0.015	0.38 ± 0.009	0.19 ± 0.006
D2	BLEU	17.39 ± 1.189	0.75 ± 0.034	0.43 ± 0.033	0.40 ± 0.019	0.21 ± 0.015
D1	RG	16.97 ± 0.435	0.71 ± 0.016	0.38 ± 0.015	0.38 ± 0.009	0.18 ± 0.008
D2	RG	17.00 ± 1.207	0.77 ± 0.029	0.42 ± 0.028	0.40 ± 0.013	0.21 ± 0.009
D1	CS-PREC	17.08 ± 0.358	0.71 ± 0.018	0.39 ± 0.012	0.38 ± 0.009	0.19 ± 0.007
D2	CS-PREC	17.30 ± 1.301	0.76 ± 0.026	0.44 ± 0.034	0.41 ± 0.015	0.21 ± 0.015
D1	CS-REC	17.12 ± 0.314	0.71 ± 0.017	0.39 ± 0.011	0.38 ± 0.007	0.19 ± 0.005
D2	CS-REC	17.27 ± 1.191	0.77 ± 0.026	0.43 ± 0.029	0.41 ± 0.015	0.21 ± 0.013
D1	CO	17.09 ± 0.540	0.70 ± 0.022	0.39 ± 0.012	0.38 ± 0.010	0.19 ± 0.003
D2	CO	17.34 ± 1.348	0.76 ± 0.025	0.44 ± 0.034	0.41 ± 0.014	0.22 ± 0.011
Gold	N/A	—	0.92	—	—	—

Aggregated data

Many sentences express complex aggregations of data. For example: **"He is averaging 19 points on the season so far"**.

Expecting neural models to learn database-like transformations such as GROUP, LIMIT, and MEAN (from strings) seems unreasonable. But it also is impractical to include in the input data every possible aggregate data combination.

Future work

Further ablation studies with:

- Different data combinations.
- Different systems / system components.

How can we include season/league structure, as well as play-by-play data from the SportSett database?

Can we better define types of information gap?