Detecting Interesting Event Sequences for Sports Reporting

François Lareau, Mark Dras and Robert Dale

Centre for Language Technology
Macquarie University

Goal

We are developing an NLG system for generating descriptions of Australian football games in English and Arrernte (an Australian language). One aspect of this project is **content selection**. We are concerned with **aggregative inferences** such as:

> *The Swans kicked seven goals from 16 entries inside their forward 50 to open a 30-point advantage at the final change—to that point the largest lead of the match.*

This goes beyond aggregating similar facts. It involves an inference on the scoring events table (see below) and other data to identify a strong moment of arbitrary duration that forms a unit of discourse.

Available data

		Event		Score		
Time	**Player**	**H**	**A**	**H**	**A**	**M**
1′40″	Jesse White	G		6	0	6
4′42″	Jarrad McVeigh	B		7	0	7
10′05″	Patrick Ryder		B	7	1	6

Scoring events
(H=home team, A=away team, M=margin, G=goals, B=behinds)

Player	**K**	**M**	**H**	**G**	**B**	**T**
Jude Bolton	16	3	20	0	0	12
Adam Goodes	11	5	5	2	4	1
Heath Grundy	8	2	8	0	0	1

In-game player statistics
(K=kicks, M=marks, H=handballs, G=goals, B=behinds, T=tackles)

More data is available from `www.afl.com.au`, `stats.rleague.com/afl` and Champion Data.

Available texts

We collected 39 English-language commentaries (around 500 words each) written by professional journalists. They focus on in-game events, in particular scoring events, and typically consist of a gist of the game, a chronological presentation of the highlights, comments on key players and injuries, and a teaser about upcoming matches.

Text content analysis

More than half of in-game information conveyed in texts required reasoning over the data. We identified three main types of such propositions:

Raw data: data readily available from the database, e.g., the margin in *The Swans led by 33 points at the final break.*

Homogeneous aggregative inferences: reasoning over one type of data, e.g., *the Tigers kicked eight of the last 10 goals* or *the result was never in doubt*.

Heterogeneous aggregative inferences: inferences on data of different types, e.g., *Melbourne physically dominated the Swans*.

We manually annotated ten texts using this typology, further dividing the homogeneous aggregative inferences into score-based ones vs. others.

Type	**#**	**%**
Raw data	120	38.8
Homogeneous aggregative infer.		
Score-based	68	21.7
Other	13	4.2
Heterogeneous aggregative infer.	112	35.8
Total	313	100.0

Types of information conveyed in annotated texts

Score-based aggregative inferences are calculated from a sequence of goals and behinds, and are not detectable by existing approaches to aggregation such as [1, 3]. Most refer to intervals between events, which are of four types:

Type	**#**	**%**
A team is on a roll	22	55.0
There is a tight struggle	7	17.5
There are many lead changes	5	12.5
Other	6	15.0
Total	40	100.0

Types of intervals referred to in texts

Hence, detecting when a team is on a roll is an important task for this genre. It requires a curve segmentation algorithm that can handle variable time granularity. We propose a baseline technique for this task, and also try to tackle the detection of tight struggles, since it is a similar problem.

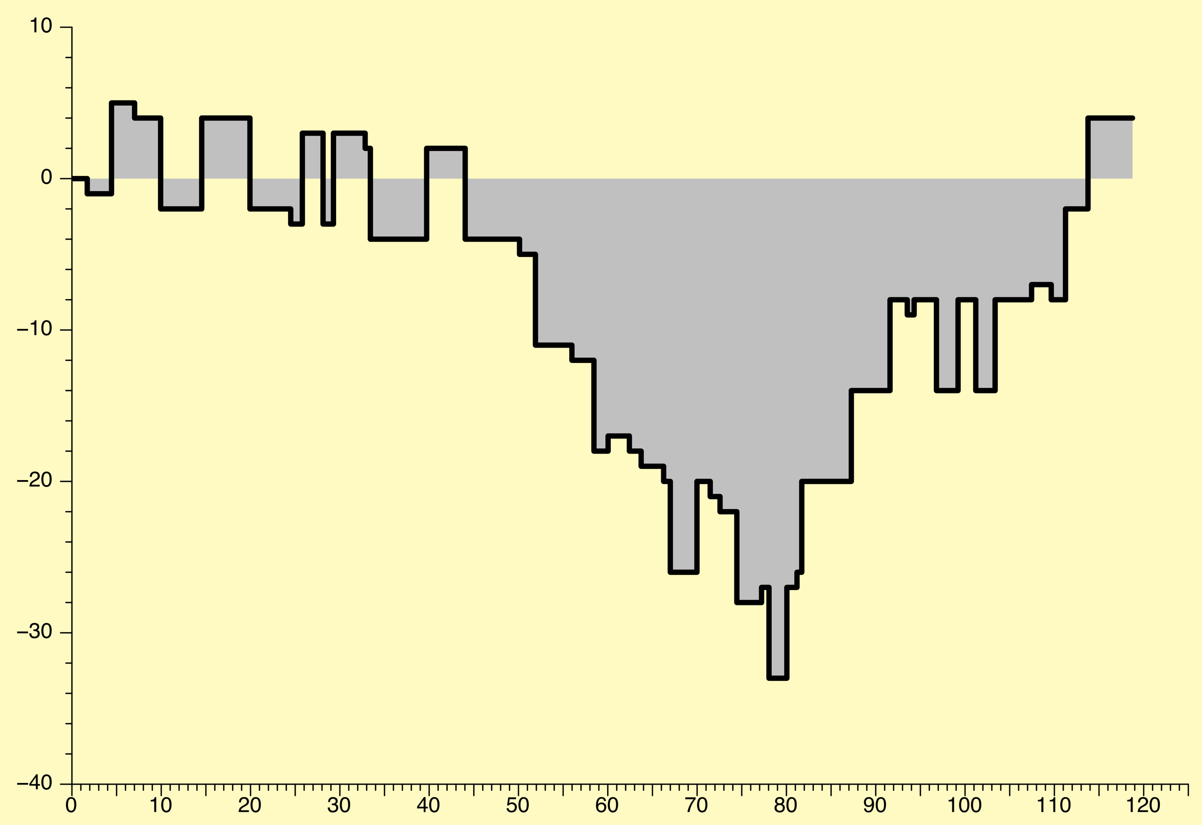
A curve segmentation technique

We look for intervals in the score margin curve where the slope is either steep or rather flat, but we do not know how steep or flat is significant, how long the interval should be, and where it should start.

There are natural time anchors for intervals (the boundaries of the game or quarters, and peaks in the curve), but human reporters also select intervals that are not bound to these anchors.

Unlike, for instance, [4], we want to do without set thresholds, and given the small number of data points to consider, a technique such as [2] seems an overkill. We propose the following method:

1. For each game, we trace the score margin curve. Every change in value corresponds to a scoring event (six points for a goal, one for a behind).



Sample score margin curve

2. We calculate the absolute value of the slope between any two scoring events (the direction of the slope is irrelevant). This yields a matrix of $n \times \frac{n-1}{2}$ slopes for a game with n scoring events (a typical match has about 50 goals and behinds).

3. The slopes are normalised relative to all other slopes that span the same number of events in the same game.

4. The normalised values are ranked in comparison with the other values for the game, and the ranks are expressed as percentiles.

Evaluation

We applied this technique to the data that corresponds to the texts we had annotated, and checked how many of the rolls and struggles mentioned in the texts received a rank that made sense (high ranks for rolls, low ones for struggles).

Rolls			Struggles		
Rank	**#**	**%**	**Rank**	**#**	**%**
≥ 0.9	15	68.2	≤ 0.1	3	42.9
≥ 0.8	17	77.3	≤ 0.2	3	42.9
≥ 0.7	20	90.9	≤ 0.3	4	57.1
Total	22	100.0	Total	7	100.0
Median: 0.956			Median: 0.204		

Percentile ranks for normalised interval slopes

Conclusion

This technique works for rolls, and could be used as a baseline and a starting point for a stochastic reranking approach, taking, say, the top 30%, and reranking them based on additional local score context. However, the results are not as promising for struggles; this is probably because struggles tend to be in games with a generally flat curve. One possible alternative is to use a different score-related measure, e.g., a matrix of lead changes per time period. A second is to compare intervals with other intervals of the same duration in all games, rather than in the same game.

Our method is only viable for curves with a limited number of data points, since it must consider all possible sub-segments of the curve.

References (see the paper for more references)

[1] R. Barzilay and M. Lapata. Collective content selection for concept-to-text generation. In *Proceedings of HLT/EMNLP'05*, pp. 331–338, 2005.

[2] S. Boyd. TREND: a system for generating intelligent descriptions of time series data. In *Proceedings of ICIPS'98*, 1998.

[3] C. Kelly, A. Copestake and N. Karamanis. Investigating content selection for language generation using machine learning. In *Proceedings of ENLG'09*, pp. 130–137, 2009.

[4] J. Yu, E. Reiter, J. Hunter and C. Mellish. Choosing the content of textual summaries of large time-series data sets. *Natural Language Engineering*, 13(1):25–49, 2006.