Narrative similarity detection using character roles

Text2Story@ECIR'26, 29 March 2026
Alexandra Barancová & Giovanni Sileno
Informatics Institute, University of Amsterdam, The Netherlands

Background

Story comparison is a longstanding problem: what makes stories similar?

Background

Story comparison is a longstanding problem: what makes stories similar?

Recent work has framed narrative similarity detection as an **information retrieval** task, identifying proxies that can be used as ground truth labels.

Background

Story comparison is a longstanding problem: what makes stories similar?

Recent work has framed narrative similarity detection as an **information retrieval** task, identifying proxies that can be used as ground truth labels.

Along one dimension, narrative research examines the **roles of characters**; their functions, archetypes, tropes, etc.

Aim

(1) We examine how structured narrative elements, extracted with LLMs, can complement document-level embeddings for narrative similarity detection.

Aim

(1) We examine how structured narrative elements, extracted with LLMs, can complement document-level embeddings for narrative similarity detection.
(2) We focus on character role information to assess whether it improves performance on similar story retrieval for medium- and long-form narratives.

Aim

(1) We examine how structured narrative elements, extracted with LLMs, can complement document-level embeddings for narrative similarity detection.
(2) We focus on character role information to assess whether it improves performance on similar story retrieval for medium- and long-form narratives.

Extendible proof of concept.

Data: Tell me again!

A dataset of over 96k film summaries
from Wikipedia across 29k stories.

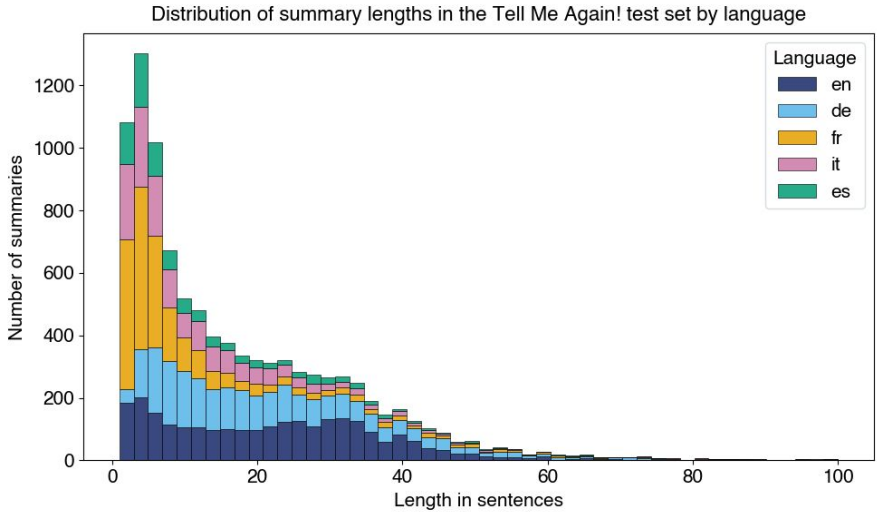
H. O. Hatzel, C. Biemann, Tell Me Again! a Large-Scale Dataset of Multiple Summaries for the Same Story,
in: Proceedings of LREC-COLING 2024, ELRA and ICCL, Torino, Italia, 2024, pp.15732–15741.

Data: Tell me again!

A dataset of over 96k film summaries from Wikipedia across 29k stories.

Test set:

- 9,778 summaries across 2,951 unique stories
- mean=17.4, median=12 sentences



H. O. Hatzel, C. Biemann, Tell Me Again! a Large-Scale Dataset of Multiple Summaries for the Same Story, in: Proceedings of LREC-COLING 2024, ELRA and ICCL, Torino, Italia, 2024, pp.15732–15741.

Experimental setup

Using film summaries, original & anonymised.

Experimental setup

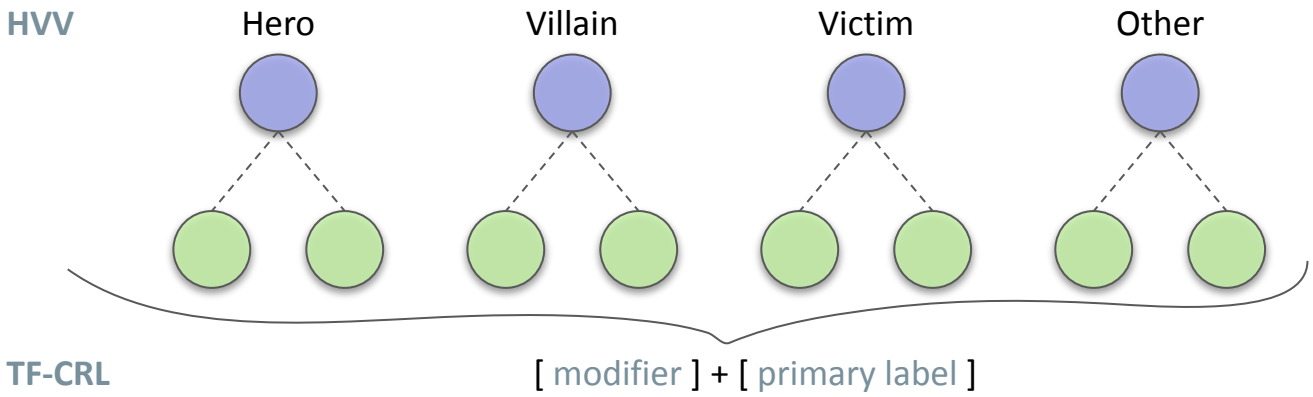
Using film summaries, original & anonymised.

(1) Entity detection & coreference resolution → character list per summary

Experimental setup

Using film summaries, original & anonymised.

(1) Entity detection & coreference resolution → character list per summary
(2) Character role extraction → HVV and taxonomy-free label per character



D. G. Hobson, D. Ruths, A. Piper, Evaluating Taxonomy Free Character Role Labeling (TF-CRL) in News Stories using Large Language Models, in: Proceedings of EMNLP 2025, ACL, Suzhou, China, 2025, pp.14828–14850.

Experimental setup

Using film summaries, original & anonymised.

(1) Entity detection & coreference resolution → character list per summary
(2) Character role extraction → HVV and taxonomy-free label per character
(3) Encoding → document & character role embedding per summary

Psychologist Kris Kelvin arrives after an interstellar journey to the Earth space station suspended a few kilometers above the ocean of Solaris, on board of which there are three scientists. As soon as he arrives, Kelvin makes contact with his colleague Snaut, who behaves illogically, as if he has something to hide. From him Kelvin learns that his colleague Gibarian, whose pupil and collaborator he was, has just committed suicide. Kelvin occupies the deceased's apartment and notices with disbelief that inside the space station there is an alien presence: a tall black woman, dressed very casually, who crosses in a corridor and then finds her in the refrigerated cell next to Gibarian's corpse.

It soon becomes apparent that Snaut and his third partner, Sartorius, are also hiding someone in their apartments but in the morning when Kelvin wakes up he finds in the room his young wife, Harey, who took her own life a few years earlier. Kelvin is shocked but realizes that the other colleagues are also hiding an anthropomorphic materialization extracted from their subconscious. According to Snaut, the "presences" appeared after Sartorius subjected the sea surface to a massive dose of X-rays for experimental reasons. Getting rid of the "guests" is impossible: someone out of desperation and horror has tried to eliminate them physically, but the next day they return, with no memory of what happened. Even Kelvin, devastated by the horror, gets rid of Harey by sending her into orbit with a missile to find her in his room when he wakes up the next day.

The scientist spirit takes over. Kelvin discovers that Harey's body is not composed of normal matter but of neutrinos, and that it is able to regenerate damaged tissue within minutes. After the initial rejection, Kelvin finds himself emotionally involved in the unreal

Psychologist Kris Kelvin arrives after an interstellar journey to the Earth space station suspended a few kilometers above the ocean of Solaris, on board of which there are thr
makes cont
illogical
Kelvin l
and coll
Kelvin c
disbelie
presence
crosses
refriger
It soon
partner,
apartmer
finds ir
life a f
that the
anthropo
subconsc
appeared
massive
rid of t
desperat
physical
of what
gets rid
missile to
day.
The scientist spirit takes over. Kelvin discovers that Harey's body is not composed of normal matter but of neutrinos, and that it is able to regenerate damaged tissue within minutes. After the initial rejection, Kelvin finds himself emotionally involved in the unreal

Solaris summary 1 (it)

Hero(es)
- Obsessive Investigator (Kris Kelvin)
- Compassionate Psychologist (Kelvin)

Villain(s)
- Fanatical Experimenter (Sartorius)

Victim(s)
- Despondent Suicide (Gibarian)
- Ephemeral Reconstruction (Harey)

Other
- Manipulative Deceiver (Snaut)
- Alien Presence (Solaris)
- Absent Witness (Berton)

Solaris summary 2 (es)

Hero(es)
- Determined Investigator (Kris Kelvin)

Villain(s)
- Obsessive Guardian (Snaut)

Victim(s)
- Lost Echo (Harey)
- Suicidal Lost Soul (Gibarian)

Other
- Obsessive Researcher (Sartorius)

Experimental setup

Using film summaries, original & anonymised.

(1) Entity detection & coreference resolution → character list per summary
(2) Character role extraction → HVV and taxonomy-free label per character
(3) Encoding → document & character role embedding per summary
(4) Combining → combined embedding per summary

Evaluation

Narrative similarity assessed using cosine similarity of embeddings.

Evaluation

Narrative similarity assessed using cosine similarity of embeddings.

Metrics:

- Hits at K (H@1, H@5, H@10)
- Mean reciprocal rank (MRR)

Similarity detection is trickier when summaries are anonymised

model	H@1 original	H@1 anonymised
bge-base-en-v1.5	0.876	0.382
nomic-embed-text-v1.5	**0.946**	0.448
sentence-t5-large	0.909	**0.555**
snowflake-arctic-embed-m	0.892	0.447
entity BoW	0.855	-

Similarity detection is trickier when summaries are anonymised

model	H@1 original	H@1 anonymised
bge-base-en-v1.5	0.876	0.382
nomic-embed-text-v1.5	**0.946**	0.448
sentence-t5-large	0.909	**0.555**
snowflake-arctic-embed-m	0.892	0.447
entity BoW	0.855	-

Performance improves if texts are grouped by length

		H@1 original				H@1 anonymised			
length	items	bge	nomic	sentence	snowflake	bge	nomic	sentence	snowflake
1≤n<5	1345	0.868	**0.897**	0.833	0.781	0.474	0.494	**0.543**	0.371
5≤n<10	970	0.949	**0.977**	0.960	0.951	0.533	0.577	**0.669**	0.537
10≤n<20	1039	0.950	**0.993**	0.971	0.988	0.501	0.614	**0.724**	0.643
20≤n<40	1600	0.948	**0.995**	0.977	0.994	0.534	0.642	**0.760**	0.740
40≤n<60	190	0.953	**0.995**	0.989	0.989	0.726	0.842	**0.884**	0.868
60≤n	53	0.943	0.868	0.962	**1.000**	0.698	0.566	**0.792**	**0.792**

Structured character role labels *only* perform okay-ish

	summaries				character roles				combined			
model	H@1	H@5	H@10	MRR	H@1	H@5	H@10	MRR	H@1	H@5	H@10	MRR
bge	0.403	0.567	0.627	0.481	0.211	0.324	0.378	0.269	0.485	0.629	0.683	0.554
nomic	0.495	0.669	0.732	0.575	**0.237**	**0.367**	**0.424**	**0.301**	0.604	0.758	0.809	0.675
sentence	**0.610**	**0.765**	**0.817**	**0.682**	0.213	0.348	0.412	0.281	**0.634**	**0.782**	**0.831**	**0.704**
snowflake	0.538	0.688	0.740	0.609	0.227	0.353	0.413	0.291	0.603	0.744	0.793	0.669

The combined representation improves performance

	summaries				character roles				combined			
model	H@1	H@5	H@10	MRR	H@1	H@5	H@10	MRR	H@1	H@5	H@10	MRR
bge	0.403	0.567	0.627	0.481	0.211	0.324	0.378	0.269	0.485	0.629	0.683	0.554
nomic	0.495	0.669	0.732	0.575	**0.237**	**0.367**	**0.424**	**0.301**	0.604	0.758	0.809	0.675
sentence	**0.610**	**0.765**	**0.817**	**0.682**	0.213	0.348	0.412	0.281	**0.634**	**0.782**	**0.831**	**0.704**
snowflake	0.538	0.688	0.740	0.609	0.227	0.353	0.413	0.291	0.603	0.744	0.793	0.669

A “flat” representation works better if using roles only

	summaries				character roles				combined			
model	H@1	H@5	H@10	MRR	H@1	H@5	H@10	MRR	H@1	H@5	H@10	MRR
bge	0.403	0.567	0.627	0.481	0.211	0.324	0.378	0.269	0.485	0.629	0.683	0.554
nomic	0.495	0.669	0.732	0.575	**0.237**	**0.367**	**0.424**	**0.301**	0.604	0.758	0.809	0.675
sentence	**0.610**	**0.765**	**0.817**	**0.682**	0.213	0.348	0.412	0.281	**0.634**	**0.782**	**0.831**	**0.704**
snowflake	0.538	0.688	0.740	0.609	0.227	0.353	0.413	0.291	0.603	0.744	0.793	0.669

	taxonomy-free character roles				combined			
model	H@1	H@5	H@10	MRR	H@1	H@5	H@10	MRR
bge	0.192	0.325	0.389	0.259	0.485	0.651	0.710	0.563
nomic	**0.270**	**0.416**	**0.483**	**0.343**	0.610	0.767	0.817	0.682
sentence	0.244	0.383	0.449	0.315	**0.626**	**0.772**	**0.819**	**0.694**
snowflake	0.242	0.374	0.434	0.308	0.590	0.732	0.778	0.657

…but in the combined setup HVV data remains useful

	summaries				character roles				combined			
model	H@1	H@5	H@10	MRR	H@1	H@5	H@10	MRR	H@1	H@5	H@10	MRR
bge	0.403	0.567	0.627	0.481	0.211	0.324	0.378	0.269	0.485	0.629	0.683	0.554
nomic	0.495	0.669	0.732	0.575	**0.237**	**0.367**	**0.424**	**0.301**	0.604	0.758	0.809	0.675
sentence	**0.610**	**0.765**	**0.817**	**0.682**	0.213	0.348	0.412	0.281	**0.634**	**0.782**	**0.831**	**0.704**
snowflake	0.538	0.688	0.740	0.609	0.227	0.353	0.413	0.291	0.603	0.744	0.793	0.669

	taxonomy-free character roles				combined			
model	H@1	H@5	H@10	MRR	H@1	H@5	H@10	MRR
bge	0.192	0.325	0.389	0.259	0.485	0.651	0.710	0.563
nomic	**0.270**	**0.416**	**0.483**	**0.343**	0.610	0.767	0.817	0.682
sentence	0.244	0.383	0.449	0.315	**0.626**	**0.772**	**0.819**	**0.694**
snowflake	0.242	0.374	0.434	0.308	0.590	0.732	0.778	0.657

Conclusion

Combined story representation offers a promising improvement that could be extended to other narrative features.

Conclusion

Combined story representation offers a promising improvement that could be extended to other narrative features.

- Differences in performance based on anonymisation and length
  → overcoming differences in granularity

Conclusion

Combined story representation offers a promising improvement that could be extended to other narrative features.

- Differences in performance based on anonymisation and length
  → overcoming differences in granularity
- Differences between character role models
  → investigating which (parts of) schemas work

Discussion

More/other narrative schemas

- Fine-grained closed taxonomies
- Clustering open taxonomies
- Temporal dimension

Discussion

More/other narrative schemas

- Fine-grained closed taxonomies
- Clustering open taxonomies
- Temporal dimension

Fusion strategies

- Weighted combination, learned fusion

Discussion

More/other narrative schemas

- Fine-grained closed taxonomies
- Clustering open taxonomies
- Temporal dimension

Fusion strategies

- Weighted combination, learned fusion

Human perception

- For validation, evaluation, and interpretability

thank you!

a.barancova@uva.nl