Where have I Heard this Story Before?: Identifying Narrative Similarity in Movie Remakes

Snigdha Chaturvedi
University of California
Santa Cruz
snigdha@ucsc.edu

Shashank Srivastava
Carnegie Mellon University
Pittsburgh
ssrivastava@cmu.edu

Dan Roth
University of Pennsylvania
Philadelphia
danroth@seas.upenn.edu

Abstract

People can identify correspondences between narratives in everyday life. For example, an analogy with the *Cinderella story* may be made in describing the unexpected success of an underdog in seemingly different stories. We present a new task and dataset for story understanding: identifying instances of similar narratives from a collection of narrative texts. We present an initial approach for this problem, which finds correspondences between narratives in terms of plot events, and resemblances between characters and their social relationships. Our approach yields an 8% absolute improvement in performance over an information-retrieval baseline on a novel dataset of plot summaries of 577 movie remakes from Wikipedia.

1 Introduction

The ability to automatically understand narratives has been a long-standing goal of AI. Humans routinely invoke narratives to share information, learn normative behavior, and to make sense of the world (Gottschall, 2012; Miller and Mitchell, 1983). They accept narratives that adhere to familiarity and personal experiences, and reinterpret those that appear unfamiliar (Herman, 2003). In this work, we present a new task for narrative understanding: identifying instances of similar narratives. The ability to recognize similar narratives can be valuable for tasks such as QA and information retrieval, and furnish tools towards analyzing collections of real or fictional narratives. For example, given a news story, a digital archives analyst might identify similar stories from the past.

A major bottleneck in computationally exploring narrative similarity is the limited availability of annotated data for analyses and evaluation. A contribution of this work is a dataset of plot summaries of movies, which include movie pairs that have been identified as remakes (see Sec. 3). Our

| A Dutch couple, Rex and Saskia, are on holiday in France. Saskia enters a petrol station to buy drinks and does not return. Three years after Saskia's disappearance, Rex is still searching for her. Rex's new girlfriend, Lieneke, reluctantly helps him search for Saskia. Raymond confronts Rex and admits the kidnapping; he says he will reveal what happened to her if Rex comes with him. Raymond takes Rex to the rest area. He pours Rex a cup of drugged coffee, and tells him the only way to learn what happened to Saskia is to experience it himself. Rex drinks the coffee and awakens buried in a box underground … | Jeff Harriman goes on vacation with his girlfriend Diane, who vanishes without a trace at a gas station. Three years later, Jeff is still obsessed with finding out what happened. One day, Barney Cousins arrives at Jeff's door and admits that he was responsible for her disappearance. Cousins promises to show Jeff what happened to Diane, but only if he agrees to go through exactly the same thing. Jeff is taken to the gas station and is told that if he drinks a cup of coffee, he will discover Diane's fate. He does, and wakes up to find he has been buried alive. Jeff's new girlfriend, Rita, traces him and his abductor, and discovers what has happened… |

Figure 1: Example (condensed) movie summaries with similar narratives: *Spoorloos (1988)* (left) and *The Vanishing (1993)* (right). Note that (1) plot similarity and (2) characters and their relationships are significant elements in determining this similarity.

working hypothesis is that re-tellings of similar stories would retain prominent elements in terms of narrative theme, even while they look superficially different. Figure 1 shows an example of two such movie summaries, condensed here for brevity. Our approach for identifying similar narratives infers alignments between pairs of narratives using a *story-kernel* that takes into account two kinds of likenesses: (1) plot similarity (2) correspondences between characters in the narratives (based on attributes such as name, gender, prominence in the narrative, and social relationships with other characters).[1] While our data and problem formulation do not accommodate all aspects of narrative similarity, and our approach is relatively simple (for example, it doesn't model

---

[1]Our approach is unsupervised, but we use a development set to tune parameters

temporal or sentiment trajectories), we believe they capture many substantial aspects of the phenomenon and serve as a useful starting point for research into the problem.

Our contributions are:

1. We introduce the problem of characterizing narrative similarity in movie remakes, and formulate this as a ranking task.

2. We create a dataset of 577 narratives for this task, mined from plot summaries of movie remakes from Wikipedia.

3. We present a *story-kernel* that quantifies narrative similarity by considering correspondences between narratives using a character-centric approach. We empirically evaluate the *story-kernel* and its various components, and demonstrate its utility.

2   Related Work

The field of computational narratology has focused on algorithmic understanding and generation of narratives (Mani, 2012; Richards et al., 2009). Much previous work has attempted to understand narratives either from the perspective of their (i) *sequences of events* (Schank and Abelson, 1975; Chambers and Jurafsky, 2009) or plot units (McIntyre and Lapata, 2010; Goyal et al., 2010; Finlayson, 2012), or from the perspective of (ii) *characters* (Wilensky, 1978) or personas in a narrative (Propp, 1968; Bamman et al., 2013, 2014; Valls-Vargas et al., 2014). Elsner (2012) explore the plot structure of novels to distinguish original texts from novels from synthetically altered versions of the same. Some recent approaches have also focused on modeling relationships between literary characters (Chaturvedi, 2016; Iyyer et al., 2016; Chaturvedi et al., 2016), and their social networks (Elson et al., 2010; Agarwal et al., 2013; Krishnan and Eisenstein, 2015; Srivastava et al., 2016).

Other research has focused on characterizing narratives in terms of their structure. In particular, seminal formalisms such as plot units (Lehnert, 1981) and Story Grammars (Rumelhart, 1980) have been used to analyze story plots. A significant issue with almost all such frameworks is that they are either largely conceptual, or depend on careful manual annotations of features about narrative plot elements (Elsner, 2012; Elson, 2012; Finlayson and Henry Winston, 2006), which makes them unamenable to comprehensive

empirical analysis. While some of the above approaches explore prototypical patterns that characterize narratives (Nguyen et al., 2013) and narrative similarity (Fisseni and Löwe, 2012), they do not address the issue of automatically comparing narratives. Also noteworthy in this context is the Aarne-Thompson classification system (Aarne and Thompson, 1961), which has been extensively used in the analysis of folk-tales to organize types of stories, based on an index of motifs. Our work is most closely related to that of Nguyen et al. (2014) who attempt to understand the various dimensions that experts and non-experts consider while judging narrative similarity.

3   Movie Remakes Dataset

We present a dataset for evaluating narrative level similarity of texts. Our assumption is that movie remakes are re-tellings of the same story and retain prominent narrative elements. Hence, a good measure of narrative similarity should evaluate summaries of movie remakes as being similar to each other. To this end, we present a dataset of movie plots extracted from Wikipedia.

In particular, we scraped lists of movies from the '*Lists of film remakes*' page on Wikipedia, which consist of entries of movies considered remakes of previous movies. Since some movies have multiple remakes, we obtain clusters of movie plots, each of which share the same narrative theme. For each movie, we extract the text of its corresponding Wikipedia plot summary from the CMU Movie summary dataset (Bamman et al., 2013). In some cases, the remakes are close to the originals at a surface level, whereas in other cases, they diverge at a surface level, and may also significantly differ in the narrative. These clusters were then manually pruned to remove errors, and the statistics of the curated dataset are shown in Table 1. In particular, we observe that the average summary is quite long (564 words), which would make human annotations of similarity for such narratives difficult.

NLP pre-processing: We processed texts of movie summaries using the BookNLP pipeline (Bamman et al., 2014) to get dependency parses, and identify major characters. We also assigned a gender to each character which corresponded to the gender that is most frequently assigned to that character's mentions across the story using the Stanford Core NLP

Number of movies	577
Number of clusters	266
Max number of movies in a cluster	7
Avg number of words in a summary	564
Max number of words in a summary	2778
Min number of words in a summary	26

Table 1: Statistics for Movie Remakes dataset

system (Manning et al., 2014).

4 Identifying Narrative Similarity

Our approach's core consists of a *story-kernel*, $\mathbb{S}(s_i, s_j)$ that characterizes the similarity between two narratives, s_i and s_j. The story-kernel has the following two components: (i) Plot Kernel, which incorporates surface similarity between plots of the two stories (in terms of the principal events and entities), and (ii) Character Alignment Kernel, which considers correspondences in terms of character attributes and relationships.

Plot Kernel: A simple measure for narrative similarity can incorporate lexical similarities between textual descriptions of two narratives. However, our goal is to identify narratives that have similar plot structure, rather than incidental surface-level matches in their summaries. Therefore, we focus only on events, and entities and their properties. We model events mentioned in a story by identifying all verbs occurring in the text of the narrative. We capture entities and their properties by identifying nouns and the adjectives that modify them. As mentioned earlier, our approach specifically models characters as a separate component in the story-kernel. Hence, at this stage, we only consider text entities that do not represent a character mention. We represent the plot of a narrative using a bag-of-word representation of its events and entities (and their characteristics) as described above. We then define $\mathbb{S}_{plot}(s_i, s_j)$ as the cosine similarity between these representations for narratives s_i and s_j.

Character Alignment Kernel: This component compares two narratives by *aligning* characters of one with similar characters in the other. Specifically, we align each character, c_i, of a story, s_i, to a character, c_j, of the other story, s_j. This alignment is based on a similarity score, $\mathcal{S}(c_i, c_j)$, between the two characters (defined later). The goal of this joint alignment is to maximize the average

alignment score of characters in the narrative pair:

$$\mathbb{S}_{char}(s_i, s_j) = \max_{x_{c_i c_j}; c_i \in s_i, c_j \in s_j} \frac{\sum x_{c_i c_j} \mathcal{S}(c_i, c_j)}{N}$$

subject to alignment constraints $\sum_{c_i} x_{c_i c_j} = 1, \forall c_j$ and $\sum_{c_j} x_{c_i c_j} = 1, \forall c_i$. Here, x is a binary matrix indicating character alignments, and N is the total number of aligned characters from the two narratives. These constraints ensure that each character is aligned to one, and only one, character from the other story. This combinatorial optimization can be solved in polynomial time by modifying the Hungarian assignment algorithm (Kuhn, 1955). When two stories have different number of characters, the extra unaligned characters are aligned to a special *null* character from the other story.

In the above description, the similarity between two (non-null) characters, $\mathcal{S}(c_i, c_j) \in [0, 1]$, is defined as a convex combination of their similarities along (i) name, (ii) gender, (iii) prominence in the story, and (iv) attributes and social relationships with other characters.

$$\begin{aligned} \mathcal{S}(c_i, c_j) = \quad & \lambda_1 \cdot \mathcal{S}_{name}(c_i, c_j) + \lambda_2 \cdot \mathcal{S}_{gender}(c_i, c_j) \\ + \quad & \lambda_3 \cdot \mathcal{S}_{prom}(c_i, c_j) \\ + \quad & (1 - \lambda_1 - \lambda_2 - \lambda_3) \cdot \mathcal{S}_{reln}(c_i, c_j) \end{aligned}$$

Here, (1) $\mathcal{S}_{name}(c_i, c_j)$ is an indicator function that identifies if two character names are matching strings. It prefers aligning characters with same names, and can be a strong but shallow signal.
(2) $\mathcal{S}_{gender}(c_i, c_j)$ prefers alignments of characters with the same gender; i.e. $\mathcal{S}_{gender}(c_i, c_j) = 1$ if gender of c_i is the same as of c_j, and 0 otherwise.
(3) $\mathcal{S}_{prom}(c_i, c_j)$ aligns characters with similar prominence. E.g., it avoids matching a protagonist with a side-character. We compute the *prominence* of a character, $prom(c)$, as simply the fraction of mentions that refer to this character. We then define $\mathcal{S}_{prom}(c_i, c_j) = 1 - |prom(c_i) - prom(c_j)|$
(4) $\mathcal{S}_{reln}(c_i, c_j)$ considers how similar the two characters are in terms of attributes, and their relationship to other characters. For example, characters described with positive traits, or friends of the protagonist in one story are likely to be better matched to similar characters in other narratives.

We model the relationship between two characters (from the same narrative) by extracting features describing actions in which they participate and adjectives describing character attributes (Chaturvedi et al., 2017). E.g. we identify the actions using verbs that have the two character
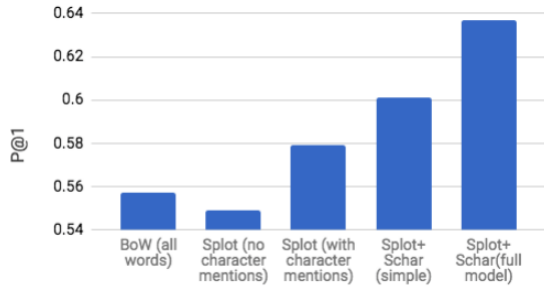
Figure 2: Performance of various approaches on Narrative Similarity task

Component	Weights
$\mathbb{S}_{plot}$	0.7
$\mathbb{S}_{character}$	0.3
$\mathbb{S}_{character} - \text{name}$	0.4
$\mathbb{S}_{character} - \text{gender}$	0.1
$\mathbb{S}_{character} - \text{prominence}$	0.1
$\mathbb{S}_{character} - \text{relationship}$	0.4

Table 2: Parameter values for various components of the story-kernel (tuned on development set)

mentions as their agents (identified using 'nsubj' and 'agent' dependency relations), and patients (using 'dobj' and 'nsubjpass' relations). We then represent a character's relationship with all other characters in the narrative using these features. Finally, we compute the relationship-based similarity, $\mathcal{S}_{reln}(c_i, c_j)$ between two characters, c_i and c_j as their cosine similarity in this feature space.

The **story kernel** is then defined as:
$$\mathbb{S}(s_i, s_j) = \alpha\, \mathbb{S}_{plot}(s_i, s_j) + (1 - \alpha)\, \mathbb{S}_{char}(s_i, s_j).$$

5 Evaluation

For our experiment, we tune parameters on 20% of the data and use the remaining data (466 movies) for the test set. In order to keep the test set completely distinct from the types of stories used for parameter tuning, this split was performed at the cluster level. Given a test story, we output the most similar story from the dataset. For evaluation, we compute P@1 (precision at 1) as follows: The output is deemed correct only if the predicted movie belongs to the same remake cluster, and incorrect otherwise.

Figure 2 shows the performance of our approach in identifying narrative similarity. Here, BoW refers to a baseline approach that uses all words in the movie summary (after stopword removal and lemmatization) as feature representation, and uses cosine similarity for retrieval. This approach achieves a P@1 performance of 0.558. The second column corresponds to using $\mathbb{S}_{plot}$ alone (does not include character mentions), and performs slightly worse than BoW, suggesting that character names do indicate remakes in our data. Adding character names in the plot kernel (third column) improves the performance significantly above BoW, indicating substantial value in focusing on narrative elements such as events and entities, rather than the entire text. The fourth column shows an ablated variant of the approach that separately adds a character kernel score to $\mathbb{S}_{plot}$ based on character names alone, but does not incorporate correspondences based on gender, prominence and character relationships. We combine this alternative character-kernel with the plot-based kernel in the same manner (using a mixing parameter α tuned on the development set). This indicates that it helps to have a separate component dedicated to characters while solving this task. The final column shows the full model, which leads to a significant further improvement in performance to 0.637, reflecting an 8% absolute improvement over the baseline model. This indicates significant value in modeling multiple facets of character attributes and relationships. We observed similar trends on the development set.

Table 2 shows the weights for individual components of our kernel (tuned on the development set). These results validate our assumption that both plot and character similarity are distinct and important facets in evaluating narrative similarity.

Qualitative Results and Error Analysis: Figure 3 shows an illustrative example of character alignment using our *story-kernel* for the movie-summaries shown in Figure 1. Note that the stories do not share any character names. Our approach aligns the protagonists of the two narratives, Rex and Jeff. It also aligns their respective kidnapped girlfriends, Saskia and Diane, and their new girlfriends, Lieneke, and Rita. However, it aligns Saskia's kidnapper, Raymond, with a *null* character, even though the movie's summary mentions Diane's kidnapper, Barney Cousins. In this case, the NLP pipeline does not identify Barney Cousins as an animate character, possibly due to his unusual name. As a result, the method received as input a summary in which only three characters were identified for the story on the right. Nevertheless, it correctly identifies the story on the right
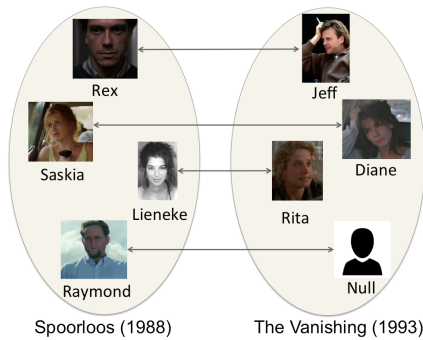
Figure 3: Example of aligned characters from the two movies in Figure 1

as most similar to the story on the left. An error analysis reveals that apart from missed character-identification, other NLP pipeline errors such as missed coreference, are major sources of errors.

6 Conclusion

We introduce an objective task, dataset and approach for quantitative evaluation of narrative similarity. Our approach, which compares narratives based on plot and character correspondences, takes a step towards addressing this problem. However, the general problem of narrative similarity can have further complexities. For example, narrative similarity can be abstract and rely on deeper reasoning (e.g., the subliminal resistance for temptation of power in 'The Lord of The Rings'). Such aspects are beyond the scope of current NLP tools, but may guide future explorations. Future work can also explore other domains (e.g., newswire and literary fiction) and evaluate character and event alignments between narratives based on established ground truths.

Acknowledgments

References

A. Aarne and S. Thompson. 1961. *The types of the folktale: a classification and bibliography*. FF communications. Academia Scientarum Fennica.

Apoorv Agarwal, Anup Kotalwar, and Owen Rambow. 2013. Automatic extraction of social networks from literary text: A case study on alice in wonderland. In *Proceedings of IJCNLP 2013, Nagoya, Japan, October 14-18, 2013*. pages 1202–1208.

David Bamman, Brendan O'Connor, and Noah A. Smith. 2013. Learning latent personas of film characters. In *Proceedings of ACL (Volume 1: Long Papers)*. Association for Computational Linguistics, Sofia, Bulgaria, pages 352–361.

David Bamman, Ted Underwood, and Noah A. Smith. 2014. A bayesian mixed effects model of literary character. In *Proceedings of ACL (Volume 1: Long Papers)*. Association for Computational Linguistics, Baltimore, Maryland, pages 370–379.

Nathanael Chambers and Dan Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In *Proceedings of the Joint Conference of ACL and AFNLP: Volume 2-Volume 2*. Association for Computational Linguistics, pages 602–610.

Snigdha Chaturvedi. 2016. *Structured Approaches for Exploring Interpersonal Relationships in Natural Language Text*. Ph.D. thesis, University of Maryland, College Park.

Snigdha Chaturvedi, Mohit Iyyer, and Hal Daumé III. 2017. Unsupervised Learning of Evolving Relationships Between Literary Characters. In *Proceedings of the Thirty First AAAI Conference on Artificial Intelligence*. AAAI'17.

Snigdha Chaturvedi, Shashank Srivastava, Hal Daumé III, and Chris Dyer. 2016. Modeling evolving relationships between characters in literary novels. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA..* AAAI'16, pages 2704–2710.

Micha Elsner. 2012. Character-based kernels for novelistic plot structure. In *Proceedings of EACL*. Association for Computational Linguistics, pages 634–644.

David K Elson. 2012. Detecting story analogies from annotations of time, action and agency. In *Proceedings of the LREC 2012 Workshop on Computational Models of Narrative*.

David K. Elson, Nicholas Dames, and Kathleen McKeown. 2010. Extracting Social Networks from Literary Fiction. In *Proceedings of ACL 2010*. pages 138–147.

Mark Finlayson and Patrick Henry Winston. 2006. Analogical retrieval via intermediate features: The goldilocks hypothesis. Technical report, MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA.

Mark Alan Finlayson. 2012. *Learning Narrative Structure from Annotated Folktales*. Ph.D. thesis, Massachusetts Institute of Technology.

Bernhard Fisseni and Benedikt Löwe. 2012. Which dimensions of narratives are relevant for human judgments of story equivalence. In *The Third Workshop on Computational Models of Narrative*. pages 114–118.

Jonathan Gottschall. 2012. *The storytelling animal: How stories make us human.* Houghton Mifflin Harcourt.

Amit Goyal, Ellen Riloff, and Hal Daumé III. 2010. Automatically Producing Plot Unit Representations for Narrative Text. In *Proceedings of EMNLP 2010*. pages 77–86.

David Herman. 2003. *Narrative Theory and the Cognitive Sciences.* Stanford Univ Center for the Study.

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan L. Boyd-Graber, and Hal Daumé III. 2016. Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. In *Proceedings of NAACL HLT 2016, San Diego California, USA, June 12-17, 2016*. pages 1534–1544.

Vinodh Krishnan and Jacob Eisenstein. 2015. "You're Mr. Lebowski, I'm the Dude": Inducing Address Term Formality in Signed Social Networks. In *Proceedings of NAACL HLT 2015*. pages 1616–1626.

Harold W. Kuhn. 1955. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly* 2:83–97.

Wendy G. Lehnert. 1981. Plot Units and Narrative Summarization. *Cognitive Science* 5(4):293–331.

Inderjeet Mani. 2012. *Computational Modeling of Narrative.* Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, System Demonstrations*. pages 55–60.

Neil McIntyre and Mirella Lapata. 2010. Plot Induction and Evolutionary Search for Story Generation. In *Proceedings of ACL 2010*. pages 1562–1572.

Owen Miller and WJT Mitchell. 1983. On narrative.

Dong Nguyen, Dolf Trieschnigg, and Mariët Theune. 2013. Folktale classification using learning to rank. In *ECIR*. Springer, pages 195–206.

Dong Nguyen, Dolf Trieschnigg, and Mariët Theune. 2014. Using crowdsourcing to investigate perception of narrative similarity. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014, Shanghai, China, November 3-7, 2014*. pages 321–330.

Vladimir Iakovlevich Propp. 1968. *Morphology of the Folktale.* University of Texas.

Whitman Richards, Mark Alan Finlayson, and Patrick Henry Winston. 2009. Advancing computational models of narrative. *MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, Tech. Rep* 63:2009.

David E Rumelhart. 1980. On evaluating story grammars. *Cognitive Science* 4(3):313–316.

Roger C Schank and Robert P Abelson. 1975. Scripts, plans, and knowledge. In *IJCAI*. pages 151–157.

Shashank Srivastava, Snigdha Chaturvedi, and Tom M. Mitchell. 2016. Inferring interpersonal relations in narrative summaries. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA.*. pages 2807–2813.

Josep Valls-Vargas, Jichen Zhu, and Santiago Ontanón. 2014. Toward automatic role identification in unannotated folk tales. In *Proceedings of the Tenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*. AAAI Press, pages 188–194.

Robert Wilensky. 1978. *Understanding Goal-Based Stories.* Outstanding Dissertations in the Computer Sciences. Garland Publishing, New York.