Cross-Document Narrative Frame Alignment

Ben Miller, Ayush Shrestha, Jennifer Olive, and Shakthidhar Gopavaram

Georgia State University
Atlanta, GA, U.S.A
miller@gsu.edu, jolive1@gsu.edu, sgopavaram1@cs.gsu.edu, ayush.shrestha@gmail.com

Abstract

Automated cross-document comparison of narrative facilitates co-reference and event similarity identification in the retellings of stories from different perspectives. With attention to these outcomes, we introduce a method for the unsupervised generation and comparison of graph representations of narrative texts. Composed of the entity-entity relations that appear in the events of a narrative, these graphs are represented by adjacency matrices populated with text extracted using various natural language processing tools. Graph similarity analysis techniques are then used to measure the similarity of events and the similarity of character function between stories. Designed as an automated process, our first application of this method is against a test corpus of 10 variations of the Aarne-Thompson type 333 story, "Little Red Riding Hood." Preliminary experiments correctly co-referenced differently named entities from story variations and indicated the relative similarity of events in different iterations of the tale despite their order differences. Though promising, this work in progress also indicated some incorrect correlations between dissimilar entities.

1998 ACM Subject Classification I.2.7 Natural Language Processing

Keywords and phrases computational narrative, natural language processing, graph theory, text mining

Digital Object Identifier 10.4230/OASIcs.CMN.2015.124

1 Introduction

Building parse tree representations of sentence-level grammars and comparing those representations to assess grammatical similarity has been an achieved goal of natural language processing (NLP), at least in English, since the development of the Penn Treebank and the success of statistical parsers in the mid-1990s [19]. Adapting this kind of parse tree comparison approach to higher-level analyses such as cross-document comparison of narrative similarity, however, remains an open challenge. The goal of this preliminary research is to advance our prior work in narrative information extraction [22] and visualization [28] for narrative similarity assessment, event alignment, and cross-document coreference using a graph comparison approach. Our method uses matrix representations of the graphs where each node is an entity, each edge is a relation, and each matrix represents one "event" as denoted by the language processing tool EVITA [26]. For this study, an entity is either a character, a location, or an organization.

Humanities scholars focus on broad problematics such as semantics, representation, narrative: problematics that frequently bridge, fracture, and co-referentially scatter throughout documents and corpora. Discourse analysis [14] and TextTiling [13] are two methods used to circumvent sentential boundaries by segmenting documents into blocks according to inferred characteristics of speaker, function, or character frequency change boundaries. As with topic

6th Workshop on Computational Models of Narrative (CMN'15).
Editors: Mark A. Finlayson, Ben Miller, Antonio Lieto, and Remi Ronfard; pp. 124–132
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

modeling methods like latent semantic analysis [8], these blocks facilitate comparisons of macro-level structures. These segmentation methods might produce blocks roughly equivalent to scenes. However, they rely on string and semantic vectors and have no particular sensitivity to features key for the structural analysis of narrative. Our research instead expands on graph comparison methods, which can more readily be made sensitive to narratological features such as events. Comparison of narrative graphs facilitates 1) alignment of event descriptions across narratives, 2) cross-document co-reference, and 3) the testing of structuralist narratological schema. To preliminarily test one and two, we implemented a method as described below.

Structural analyses of narrative successfully identified elements significant for the composition and study of narrative. Russian formalists such as Propp [25] and later work by Genette [11], Bal [1], and others yielded many complementary top-down models for deconstructing narratives. These schema generally distinguish between fabula and discourse: events to be narrated and the nature of that narration, respectively. Discourse order is the relationship between the temporality of events and their representation as part of a narrative [11]. This structural perspective serves humanists well when analyzing single narratives or small corpora but is highly subject to interpretation, and therefore operationalizes poorly. Computational models developed from formalist approaches have been the subject of compelling experiments. Like work by Finlayson on analogical story merging [9] and Fisseni on story comparison [10], our work presents a bottom-up method reliant on top-down narratological schema. Unlike theirs, our work focuses on unsupervised cross-document comparison of events and characters.

This method facilitates cross-document narrative analysis by indicating the similarity of a character's relationships across different tellings of a particular story and by allowing for the comparison of event language. Although much work remains and the anaphora resolution task was manually verified, this method would work with larger corpora as extraction, lookup, and comparison operate in an unsupervised manner.

2 Method

Comparison of events across documents relies on the production of structured representations of events. In the case of this study, that structure is a matrix of entity-entity relations for each event. Generalizing the specific language of a story is necessary as abstracted language facilitates comparison. This study used event hypernym sequences to generalize from the specific language of a given event. After identifying language features that are indicative of events, identifying the entities present in that event, and finding the hypernym of the lexical feature identified as the verb or state of the event, matrices were produced. Some language features indicative of events include finite clauses, event-referring nouns, and nominalized noun phrases [26]. Comparison via a neighborhood similarity function provided our primary comparison method to highlight event and character similarities.

2.1 Extraction

Events were automatically marked in the narratives using the Events in Text Analyzer (EVITA). EVITA uses statistical and linguistic approaches to identify and classify the language denoting orderable dynamic and stative situations [18]. EVITA's overall accuracy in event recognition was found by [18] to be $80.12\% F\_\{\beta\} = 1$ over TimeBank, with 74.03% precision and 87.31% recall. [18] summarizes evaluations of related work in automatic event detection, including TimeML [5], STEP [3], and event recognition using a multiclass classifier [20]. Their summary findings showed that EVITA either outperformed or was competitive

Table 1 Adjacency matrix created from one version of "Little Red Riding Hood". An edge (in the graph) or 1 (in the adjacency matrix) between two entities signify that these entities interacted within the given set of events.

	lrrh	wolf	grandmother	woodcutters	forest	gm_house
lrrh	1	0	0	1	1	0
wolf	0	1	0	0	1	0
grandmother	0	0	1	0	0	0
woodcutter	1	0	0	0	1	0
forest	1	1	0	1	0	0
gm_house	0	0	0	0	0	1

with other automated solutions. A more robust theoretical model for what constitutes an event is being developed for implementation by the NewsReader project in [31].

EVITA sequentially numbers events. That sequence must stand in for discourse order because fiction frequently lacks the dates and timestamps necessary to identify story order. They features are also necessary for discrete temporal language taggers like SUTime [7] and GUTime [32]. Entity extraction and anaphora resolution was accomplished using the Stanford Named Entity Recognizer (NER) followed by manual verification; entity classification was not relevant for this method as all three types of NE were identically represented in the matrices.

2.2 Graph Creation

Given an extracted set of events from a document, E_1 to E_n, we first divide them into k subsets ordered according to the story time. Event subsets can be defined in various ways: by manual adudication according to various criteria, or automatically by document section, by prevalent entities, by location shifts, or by prevalent event types. For this experiment, we ran the process two with manually defined event subsets based on location shifts, and with no event subsets.The number of events is determined by the event analyzer. The number of subsets is variable but currently must match from story to story. All entities (characters and locations) associated with the events are listed on a per-event basis. Each version of the story included a subset of some version of Little Red Riding Hood, mother, home, wolf, grandmother, woodcutters, forest, and grandmother's house as key entities.

Following this process, we create a graph with these entities for every event subset. We begin by treating each entity as a vertex and adding an edge between verticies if both are present in the same event within an event subset. An adjacency matrix representation of a subset is shown in Table 1. In this subset of events, Little Red Riding Hood and the woodcutters are present in the forest in a particular event (the value is 1). In the same subset, the wolf is also in the forest. However, the wolf does not meet Little Red Riding Hood in any of the events in this subset, thereby resulting in no edge between them (the value is 0).

2.3 Similarity Analysis

Many domain-specific algorithms to compute similarity have been developed. Most are based on neighborhood analysis. Considering the problem of narrative frame alignment in this context treats a narrative as a directed graph; each event leads to the next and each set of events constitutes a group or neighborhood. That perspective allows for event or story analogy to be considered using the more robust methods applied to network similarity problems. In this paper, we propose our own similarity analysis method inspired by the work of Blondel et al [4].

Given a document, A, let p be the total number of entities in the document. If the set of events in this document are divided into k parts, we can represent the events in the document as a $3D$ matrix: $A_{p,p,k}$. The number of parts is some number equal to or less than the total number of event segments. Let $B_{q,q,r}$ be another document with q entities and r parts. Likewise, the number of parts is some number equal to or less than the number of events in that story. We compare each adjacency matrix in A with the corresponding adjacency matrix in B. In cases where $k \neq r$, we reduce to zero and pad the smaller matrix to the bigger size. For each adjacency matrix, as in the hyperlink-induced topic seach (HITS) inspired algorithm [15] proposed by [16], we compute

$$X \leftarrow BXA^T + B^T XA \tag{1}$$

and normalize X after each iteration. HITS was developed to facilitate search on the web by assessing the authority and role of nodes in large graphs; [16] extended that algorithm to the problem of identifying topological similarities in large, sparse, isomorphic graphs. That structure corresponds to the graphs that result from our event and entity extraction processes. The even iterations converge to a final similarity matrix. To simplify and speed up this process, we use the Kronecker product and the vec(.) operator. This process results in

$$x \leftarrow (A \bigotimes B + A^T \bigotimes B^T)x \tag{2}$$

where $x = vec(X)$. This set of equations give a similarity score frame per scene (part), which is then aggregated to produce a final similarity score between the stories.

3 Preliminary Experiment

For the purposes of testing our methodology, we selected 10 of the 58 known iterations [29] of the Aarne-Thompson type 333 story (ATU333), "Little Red Riding Hood." Those 10 iterations are from [12, 33, 27, 21, 24, 2, 30, 6]. This corpus of 10 was compiled and selected to represent the canonical versions of the ATU333 story and significant variations from that story (e.g., where the wolf was the hero). The purpose of compiling and using this corpus was to begin our testing with a story featuring a high degree of narrative overlap. That overlap let us test the method on fine-grain distinctions between re-tellings. While our method benefits from such homogeneous narrative content, we believe that analyses of other narrative corpora with overlapping sets of events would be equally viable because of the highly granular event segmentation, the hypernym language abstraction procedure, and the binning of entity classifications into a single entity category.

$1,384$ events were extracted via this method across 10 story versions. Numbering 8,450 tokens, including titles and authorship information, the overall density of extracted events to tokens is high. Contrasted to event detection methods reliant on temporal expressions such as SUTime, which only identified two events in the corpus, this density of event detection provides a good basis on which to compare narrative structure. Generalizing event keywords from specific tokens to hypernyms of those tokens (e.g., event 41 from [6]: "armed" lemmatized to "arm" of which the hypernym found via WordNet [23] is "supply") preserves the function of each event within the story but allows for storytelling variation. The current method for finding the hypernym looks for agreement across all results returned by WordNet. In the case of disagreement, the hypernym most frequently returned is selected; in the case of a tie, the first hypernym is used. The automatically produced matrices for this work are exemplified by Table 2. The stack corresponds to the "Oh, grandmother, what big ears you have!" to "[a]nd with that he jumped out of bed, jumped on top of poor Little Red Cap, and ate her up" sequence from [17].

Table 2 shows six layers from the $3D$ event matrix stack. The current language processing pipeline finds the events hypernyms but does not use them to assess narrative similarity. Results of functions (1) and (2) on the adjacency matrices are exemplified below in Table 3. Column headings correspond to entities from [12] for event 3, and row headers correspond to entities from [17] for event 4.

Table 3 shows that the measure of similarity between Little Red Riding Hood ("lrrh") and Little Red Cap ("lrc") is 0.32. Although low, that score was calculated only based on entity-entity connections and the sequence of those connections. When examined on the basis of an individual event, of which [17] contains 122, the correlations are unremarkable. Effectively, the wolf could be seen as similar to Rotkäppchen as to the woods. It is only when aggregates of events are compared that the method begins to correctly indicate entity similarities across documents.

Table 4 shows the potential for this method to align characters from different versions based upon their position within the story. It presents the similarity comparison for all events across two iterations of the story, summing all event matrices for two variations. Version 1 occupies the columns (Little Red Riding Hood, Wolf, Grandmother, Woodcutters, Home, Forest, and Old Woman's House) and version 2 the rows (Little Red Cap, Wolf, Grandmother, Huntsman, Home, Woods, Grandmother's House). Name independent character similarity is demonstrated by the 0.94 correspondence between the two wolves.

The event matrix suggests that certain characters function dissimilarly between variations: most notably, Grandmother. The corresponding value between the Grandmother characters is only 0.31, suggesting that they share some event associations but not as many as are held by other cross-document pairings. That assessment is accurate as, in version 1, the story concludes upon the wolf's consumption of both Little Red Riding Hood and Grandmother. In version 2, both survive to boil a second hungry wolf. Table 5 compares version 2 and version 6, a more modern iteration, showing promising albeit imperfect results.

In Table 5, we see the method correctly correlate two principal characters in the story, a process we refer to as alignment. It also suggests strong correlations between each of those two characters and their respective wolves. However, for many of the other principal characters, it is not the highest similarity score that suggests correct character alignment, but rather the second highest similarity. The wolf in version 6 is seen as 0.86 similar to Rotkäppchen but only 0.62 similar to the wolf from version 2. Other less well-documented characters simply do not seem to show up frequently enough to be susceptible to alignment. One takeaway from this preliminary work is that it may only be a viable method for characters that frequently appear in stories. Another compelling way to read this table, however, is to compare the similarity of two characters from two different works against each other. For example, version 6's Little Golden Hat is seen as more similar to both the wolf and the woods than her counterpart, Rotkäppchen. That way of reading the results of our method suggests that we can both identify which characters are most similar between two versions of a story and compare the varying similarity of a character between versions of a story.

4 Conclusion and further work

This preliminary work resulted in a viable method for narrative alignment and for the cross-document coreference of characters bearing different names but similar story functions.

Table 2 Six matrix layers from 3*d* stack of event matrices.

Event		LRRH	Grandmother	Wolf
106 – undergo	Bed	1	1	1
107 – perceive	Bed	1	1	1
108 – undergo	Bed	1	1	1
109 – seize	Bed	1	1	1
110 – undergo	Bed	1	1	1
111 – consume	Bed	1	1	1

Table 3 Character similarity across “Little Red Riding Hood” and “Rotkäppchen”.

	LRRH	Wolf	Grandmother	Woodcutters	Home	Woods	OWH
LRC	.32	.25	0	.25	0	.32	0
Wolf	.32	.25	0	.25	0	.32	0
Grandmother	0	0	0	0	0	0	0
Huntsman	0	0	0	0	0	0	0
Home	0	0	0	0	0	0	0
Forest	.32	.25	0	.25	0	.32	0
Grandmother’s	0	0	0	0	0	0	0

Table 4 Character similarity across all events for “Little Red Riding Hood” and “Rotkäppchen”.

	LRRH	Wolf	Grandmother	Woodcutters	Home	Forest	OWH
LRC	**.67**	.76	.31	.14	.14	.48	.37
Wolf	.79	**.94**	.42	.14	.14	.56	.5
Grandmother	.35	.47	**.31**	0	0	.16	.37
Huntsman	.23	.28	.18	**0**	0	0	.26
Home	0	0	0	0	**0**	0	0
Woods	.48	.53	.16	.14	.14	**.48**	.16
Grandmother’s	.39	.52	.34	0	0	.16	**.42**

Table 5 Character similarity across all events for “Little Golden Hat" and “Rotkäppchen".

	LGH	Mother	Grandmother	Wolf	Wood	Grandmother’s	Woodcutters
LRC	**1.00**	0.06	0.45	0.86	0.06	0.24	0.10
Mother	0.04	**0.01**	0.07	0.03	0.00	0.03	0.00
Grandmother	0.61	0.09	**0.32**	0.55	0.07	0.12	0.01
Wolf	0.79	0.05	0.21	**0.62**	0.05	0.23	0.01
Woods	0.21	0.03	0.06	0.13	**0.04**	0.05	0.01
Grandmother’s	0.05	0.00	0.12	0.04	0.01	**0.04**	0.00
Huntsman	0.10	0.00	0.00	0.09	0.00	0.00	**0.00**

Story function is being used here principally to describe the social function of a character or location relative to other characters and locations. It was determined by segmenting the story into a series of events, then identifying character-character and character-location relations and the order of those relations. The event segmentation, relation extraction, and matrix comparison methods are implemented and tested. The hypernym extension of our method will divide the event hypernyms into overlapping three-window sequences of two-to-four terms each corresponding to past, present, and future states. Those sequences will be used as weighting functions on the Kronecker product for the cross-document comparison of narrative frame similarity. For example, the entity relationships in the matrix representing a sequence of three events in document A and the entity relationships in the matrix representing a sequence of three events in document B will be factored against each other with the relative similarity multiplied by the similarity score of the hypernym sequence. Three identical terms in each window frame of past, present, and future will score as a 1. No common hypernyms across that frame would score a 0. Our current method describes narrative similarity as a proxy for character relation similarity; this extension will enrich that description. Next stages for this research include refining the comparison algorithm, applying it to a corpus of dissimilar narratives, implementing the role of the hypernym in event comparisons, and assessing the method's ability to cluster stories by narrative similarity.

Acknowledgements. This work is supported in part by NSF award 1209172.

References

1 Mieke Bal. *Narratology: Introduction to the theory of narrative.* University of Toronto Press, 1997.

2 Clara Doty Bates. *Little Red Riding-Hood.* D. Lothrop and Co., 1883.

3 Steven Bethard and James H Martin. Learning semantic links from a corpus of parallel temporal and causal relations. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, pages 177–180. Association for Computational Linguistics, 2008.

4 Vincent D. Blondel, Anahí Gajardo, Maureen Heymans, Pierre Senellart, and Paul Van Dooren. A measure of similarity between graph vertices: Applications to synonym extraction and web searching. *SIAM review*, 46(4):647–666, 2004.

5 Branimir Boguraev and Rie Kubota Ando. Timeml-compliant text analysis for temporal reasoning. In *IJCAI*, volume 5, pages 997–1003, 2005.

6 Angela Carter. The werewolf. In *The Bloody Chamber*, pages 108–110. Penguin Books, 1993.

7 Angel X Chang and Christopher D Manning. Sutime: A library for recognizing and normalizing time expressions. In *LREC*, pages 3735–3740, 2012.

8 Susan T Dumais. Latent semantic analysis. *Annual review of information science and technology*, 38(1):188–230, 2004.

9 Mark Alan Finlayson. *Learning narrative structure from annotated folktales.* PhD thesis, Massachusetts Institute of Technology, 2012.

10 Bernhard Fisseni and Benedikt Löwe. What Makes Stories Similar? Report on a Research Project, 2011-2014 (Invited Report). In Mark A. Finlayson, Jan Christoph Meister, and Emile G. Bruneau, editors, *2014 Workshop on Computational Models of Narrative*, volume 41 of *OpenAccess Series in Informatics (OASIcs)*, pages 9–12, Dagstuhl, Germany, 2014. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

11 Gérard Genette. *Narrative discourse: An essay in method.* Cornell University Press, 1983.

12 Jacob Grimm and Wilhelm Grimm. Rotkäppchen. In *Kinder- und Hausmärchen*, pages 113–118. Realschulbuchhandlung, v. 1, no. 26 edition, 1812. translated by D. L. Ashliman.
13 Marti A. Hearst. Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational linguistics*, 23(1):33–64, 1997.
14 Marianne W. Jørgensen and Louise J. Phillips. *Discourse analysis as theory and method.* Sage, 2002.
15 Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.
16 Giorgos Kollias, Shahin Mohammadi, and Ananth Grama. Network similarity decomposition (nsd): A fast and scalable approach to network alignment. *Knowledge and Data Engineering, IEEE Transactions on*, 24(12):2232–2243, 2012.
17 Andrew Lang. Little red riding hood. In Andrew Lang, editor, *The Blue Fairy Book*, pages 51–53. Longmans, Green, and Company, 5th edition, 1891.
18 Hector Llorens, Estela Saquete, and Borja Navarro-Colorado. Timeml events recognition and classification: learning crf models with semantic roles. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 725–733. Association for Computational Linguistics, 2010.
19 David M. Magerman. Statistical decision-tree models for parsing. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, pages 276–283. Association for Computational Linguistics, 1995.
20 Olivia March and Timothy Baldwin. Automatic event reference identification. In *Australasian Language Technology Association Workshop 2008*, volume 6, pages 79–87, 2008.
21 Charles Marelles. The true history of little golden-hood. In *The Red Fairy Book*, pages 215–219. Longmans, Green, and Company, 1895.
22 Ben Miller, Ayush Shrestha, Jason Derby, Jennifer Olive, Karthikeyan Umapathy, Fuxin Li, and Yanjun Zhao. Digging into human rights violations: Data modelling and collective memory. In *Big Data, 2013 IEEE International Conference on*, pages 37–45. IEEE, 2013.
23 George A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
24 Beatrix Potter. *The Tale of Jemima Puddle-Duck.* Frederick Warne and Company, 1908.
25 Vladimir Iakovlevich Propp. *Morphology of the Folktale*, volume 9 of *Publications of the American Folklore Society. Bibliographical and special series.* University of Texas Press, 1968.
26 Roser Saurí, Robert Knippen, Marc Verhagen, and James Pustejovsky. Evita: a robust event recognizer for qa systems. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 700–707. Association for Computational Linguistics, 2005.
27 Christian Schneller. *Märchen und Sagen aus Wälschtirol: Ein Beitrag zur deutschen Sagenkunde*, chapter Das Rothhütchen, pages 9–10. Verlag der Wagner'schen Universitäts-Buchhandlung, 1867. Translated by D. L. Ashliman.
28 Ayush Shrestha, Ying Zhu, Ben Miller, and Yi Zhao. Storygraph: Telling stories from spatio-temporal data. In *Advances in Visual Computing*, pages 693–702. Springer, 2013.
29 Jamshid J. Tehrani. The phylogeny of little red riding hood. *PloS one*, 8(11):e78871, 2013.
30 James Thurber. The little girl and the wolf. In *Fables for Our Time and Famous Poems Illustrated*, page 3. Harper Collins, 1983.
31 Sara Tonelli, Rachele Sprugnoli, and Manuela Speranza. Newsreader guidelines for annotation at document level. Technical report, Technical Report NWR-2014-2, Fondazione Bruno Kessler, 2014.
32 Marc Verhagen, Inderjeet Mani, Roser Sauri, Robert Knippen, Seok Bae Jang, Jessica Littman, Anna Rumshisky, John Phillips, and James Pustejovsky. Automating temporal

annotation with tarsqi. In *Proceedings of the ACL 2005 on Interactive poster and demonstration sessions*, pages 81–84. Association for Computational Linguistics, 2005.

33 A. H. Wratislaw. Little red hood. In *Sixty Folk-Tales from Exclusively Slavonic Sources*, pages 97–100. Elliot Stock, no. 15 edition, 1889.