

Discourse-Sentiment Alignment Tool (DSAT)

Patrick Huber
huberpat@cs.ubc.ca
University of British Columbia
BC V6T 1Z4, Vancouver, BC, Canada

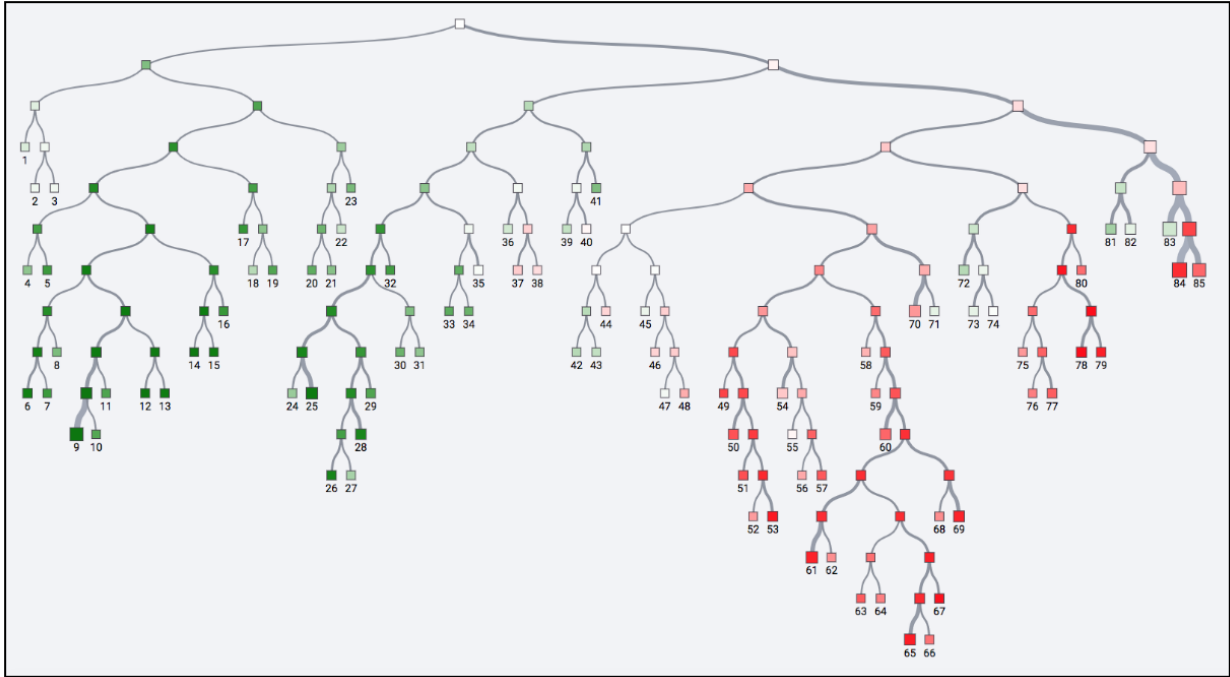


Figure 1: Discourse - Sentiment Alignment Tree produced by our system

ABSTRACT

With the rise of deep learning models and the resulting lack of intuitive explanations for the inner workings of many machine learning systems, a better understanding of those systems on a fine grained-level (single input/output combinations) is necessary to be able to trust the outcomes of the system as well as to gain a better understanding of strengths and weaknesses of certain models. This is especially true for the area of Natural Language Processing (NLP) and the important sub-area of discourse parsing, as the tree structured output is difficult to analyse without deriving visualizations.

With this intuition, we propose a novel information visualization system, designed to allow researchers and practitioners to explore automatically generated discourse data without no human intervention. Our novel system thereby supports the data exploration task on complete, binary discourse trees using multiple linked views of the data (hierarchical and sequential) and a restricted navigation idiom to support the task at hand. We show a set of use-cases in which our model can expose model drawbacks and generally verify the quality of the data.

1 INTRODUCTION

Many data driven machine learning models, especially deep learning approaches, are treated as black boxes. While researchers and practitioners know the inputs and outputs, they do know very little about the internal workings, nevertheless, deep learning methods are deployed in more and more real life applications, making important decisions in our daily lives based on those black boxes. Due to this

mismatch between the understanding of the system and the importance they play in our everyday life, approaches to understand how certain inputs effect the outputs of the system becomes a more and more important question. Some famous deep learning libraries, such as tensorflow [1], try to tackle this problem by integrating information visualization (InfoVis) systems directly into their framework, which are applicable for many standard tasks [13], but do not allow domain specific adoptions and cannot be tailored to individual use-cases beyond simple, flat data structures.

One large area of applied machine learning, where diverse deep learning approaches are effectively and efficiently implemented, is Natural Language Processing (NLP). The versatile representations of textual data (which can be sequential, hierarchical, graph-based, ...) make it notoriously hard to find a one-visualization-fits-all approach within the area of NLP compared to the data in many other domains, such as image processing, where visualizations are often straight-forward to generate. As a result, visualizations within the area of NLP often need to be custom-tailored to the task at hand. Due to this additional workload to generate meaningful information visualizations, they are currently often not considered within NLP systems. Nevertheless, there is a clear need in the area, which becomes even more significant when using completely automated pipelines, often only using weak or distant supervision, where potential biases and misalignments can be easily introduced without knowledge of the machine learning expert.

With this strong motivation for using InfoVis solutions within the area of NLP, we propose a novel approach on how to use InfoVis

to enhance the understanding of discourse-related parse trees. Discourse parsing itself is a crucial upstream task within the area of NLP, which has shown to enhance many downstream applications, like text classification [8], summarization [4] and sentiment prediction [2, 6, 11]. We believe that the use of a custom-tailored InfoVis system for discourse parsing can greatly improve the understanding of strengths and weaknesses of existing and new machine learning approaches.

The need for an effective InfoVis system has been further elevated recently, as [7] have shown that the discourse tree-structure, annotated with further features, can be retrieved using distant supervision from an auxiliary task, completely automating the complete process. While this removes the tedious and expensive manual annotation by human linguistic experts and allows the generation of large scale datasets, using this new methodology to create discourse structures from sentiment data, the human component is completely taken out of the loop.

This calls for an implementation of an exploratory information visualization system, which allows users to explore the automatically generated discourse structure dataset, to be able to draw conclusions regarding the quality and alignment of the discourse structures with the given gold-standard sentiment. We therefore propose a novel InfoVis system, which is custom-tailored to this novel and important use case, allowing the user to obtain information about the data far beyond what is possible with one-solution-fits-all approaches (e.g., [13]). The three objectives of the system we propose are thereby to analyse the automatically generated discourse trees regarding:

1. A valid tree structure, in line with the general discourse content, as defined by RST [10]
2. A well-aligned sentiment assignment for EDUs and resulting sub-trees through their aggregation
3. A reasonable relative importance score propagated along the tree branches

As the task is framed as an exploratory analysis of the data, the focus of the project lies in inspecting individual trees, rather than comparing multiple, competing discourse trees.

2 RELATED WORK

One popular information visualization system in the area of discourse parsing has been proposed by [14], where the authors present a useful InfoVis system to explore and compare multiple, parallel discourse trees regarding their differences and against the common gold-standard. The main focus of the work is to compare trees generated from competitive machine learning models with the gold standard, human-annotated discourse structure, to be able to compare the advantages and drawbacks of different discourse parsers. This work differs from the work proposed here, as it does not consider the task of exploring imperfect, automatically generated discourse trees from auxiliary tasks, making the solution proposed not applicable in our use-case scenario.

Another related work focusing on the combination of sentiment analysis and effective information visualizations for such task is conducted in [5], where the authors evaluate multiple design idioms to visualize sentiment. While one part of our task at hand is also considering the visualization of sentiment, The methodologies shown in [5] differ in a way that the sentiment visualization is the main goal, but when applied to our task, for example the proposed rose plots are not suitable to be combined with the hierarchical structure of the dataset.

Covering many different NLP tasks and their dominant idioms, the work described in [9] reinforces the important role that discourse parsing plays as a sub-area of NLP.

Another closely related paper [3] describes multiple models for syntax parsing. While this task is still different from discourse

parsing, which focuses primarily on semantic connections between text spans rather than syntactic relationships between words, some of the general intuition shown in this paper, such as the representation of the tree in a vertical way will be used in our system.

Another line of related work is located in the area of discourse parsing. As the upstream NLP task is our target domain for the visualization, following the Rhetorical Structure Theory (RST) [10], the data used within this work, generated by Huber and Carenini (2019), contains complete, projective, binary trees representing the textual discourse of a complete document. We will extend this work by showing opportunities to further evaluate their generated data beyond the dataset-level measures previously used, by allowing for a more fine-grained evaluation of the generated data.

The remainder of this work will be structured as follows: We will start with the description of our novel visualization approach for discourse-sentiment alignment, describing the available data as well as the task itself. Furthermore, we will outline the used design goals and technique on a high level, before reporting the visual and interactive idioms used in this work. The section on the visualization approach will finish on the implementation description and the outline of major usage scenarios of the system. Before discussing this work and pointing towards future directions, we will have a conclusion summarizing the contributions.

3 THE VISUALIZATION APPROACH

Our novel visualization approach to allow for the evaluation of discourse-sentiment alignment is described in this section, starting with the structure of the underlying data and the task we want to solve with this visualization approach. Subsequently we are describing the high-level design goals and techniques before showing the detailed usage of visual and interaction idioms, the implementation based on the d3 framework and finally the usage-scenarios we envision with the system.

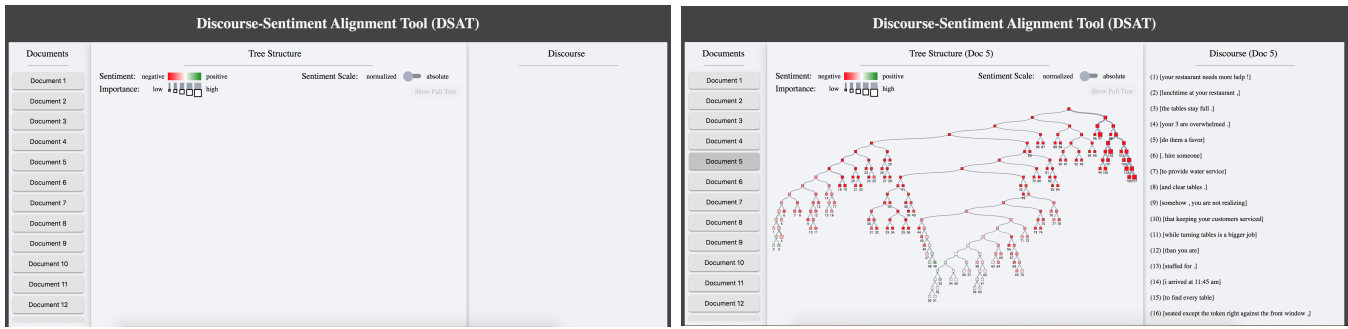
3.1 Data Structures and Task

The proposed project is within the domain of Natural Language Processing (NLP), concerned with discourse parsing, a fundamental task within the area, which is working with one out of two discourse theories, RST [10] or PDTB [12], containing complete or shallow discourse trees respectively. For this work, we will follow the previous approach by [7] and consider only RST-style discourse trees, while in general, the methodology shown is discourse-theory agnostic and can be adopted with small changes in the framework.

The available data for the project is generated in a previous research project published in 2019 [7], containing over 100,000 distinct discourse trees, each representing a document ranging from 2 to 150 atomic, clause-like textual elements, so called elementary discourse units (EDUs). While there has been previous work [14] to compare different discourse trees, the combination of discourse structure trees and sentiment (but no available gold-standard data) is novel and has therefore not been explored before. The specific structure of the data used for this project is hierarchical and contains 4 information-units per node in the tree:

1. A sentiment value within the range of $(-1, \dots, 1)$, which is probabilistically determined and encodes the range between very negative sentiment (-1) and positive sentiment (1)
2. An non-normalized attention value, encoding the local importance of a sub-tree within the hierarchical representation, ranging from $(0, \dots, 1)$
3. Two pointers to the two children nodes in the binary tree

Every internal tree node (not being a leaf of the tree) is assigned an aggregated value of the two children sub-trees. With this aggregation, the sentiment and attention value are propagated through the tree



(a) Initial starting state of the interface

(b) Interface after document has been selected

Figure 2: The interface before and after a document has been selected for further analysis

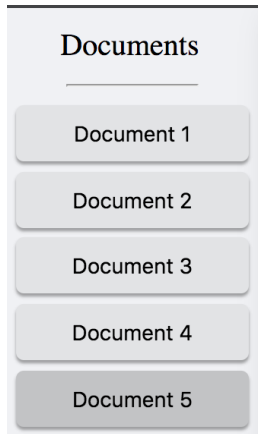


Figure 3: Selection Panel to browse through the dataset

up to the root node. This way, the information visualization system can be used to not only compare the sentiment and tree structure dependent on the leaves, but the internal nodes “summarize” the local sub-tree giving additional information on the alignment.

Based on the available data and the underlying data-structure, the discourse tree-structure itself obviously entails a central meaning for the task of discourse-sentiment alignment. As such, the tree structure does not only define the aggregation of internal nodes, but should further also reflect the different parts of the document, covering textual information of different sentiment in separate sub-trees. The task we want to solve with this system is to allow the user to find misalignments between the discourse structure, the sentiment and the textual representation effectively and efficiently. The main components required for this are:

- An easy comparison of discourse sub-trees and sentiment scores
- A clear connection between textual clauses (EDUs) and their representation in the tree
- A method to allow evaluations of special sub-trees of interest within the overall (possible up to 150 nodes) tree

To allow the user to fulfill those tasks, we describe the overall design goals and techniques in the next section and focus on the idioms after.

3.2 Design Goals and Techniques

In this section, we want to give a high-level overview of the interface and the relationship between high-level components.

The interface consists of three major components, one selection panel and two detail panels, horizontally aligned in a column design (see Figures 2a and 2b). Initially on startup of the system, only the selection panel contains information on all the documents in the corpus, while the details views are kept empty per default (Figure 2a). Once the user selects one of the discourse documents on the left, the detail-views are propagated with the hierarchical discourse tree visualization (center panel) as well as the textual representation of the document on the right (Figure 2b). With this alignment of the components, our goal is to put a clear preference on the tree structure, as this is the main contribution of this work, allowing more than just a textual comparison, which is often difficult, especially for long documents. However, to not only be able to check the alignment of discourse-trees and sentiment, we further augment the interface with the textual representation on the right, giving the full information stored in the data and allowing the user to further also investigate on the alignment of the textual representation with the sentiment scores.

3.3 Visual and Interactive Idioms

Our solution idiom will be evaluated in this section. We go panel to panel and explain our decisions on the visual encoding as well as the interaction idioms.

Selection Panel The selection panel on the left, shown in Figure 3, helps the user to explore the available documents in the corpus and select a single document to be shown in detail in the other two panels. On clicking the document in the list of available documents, a linked action is triggered, retrieving the detailed information of the document and showing it in the detail views (see Figures 2a and 2b).

Discourse Tree Panel The center panel of the interface is the discourse tree representation is initially propagated when a document in the selection panel is selected. Once displayed, the originally textual data is represented in a vertical node-link diagram (Figure 4). Having a vertical node-link diagram is thereby motivated by previous work [3] mentioned in section 2. We further align with this intuition, as the natural way to read text in Western countries is from left to right, which is imitated with this vertical tree layout. The node-link idiom itself is well suited for this task, as the spacial information is crucial, which would be not sufficiently encoded in a matrix view. Furthermore, the maximum size of 150 EDUs strictly limits the amount of data visualized, which avoids common pitfalls for node-link diagrams, such as the hairball effect. Our visualization uses the spacial dimension to encode the relationship between nodes and links. To additionally show the sentiment dimension as well as the attention (encoding importance), we use a diverging, sequential color scheme for the sentiment between red and green, with neutral sentiment represented as white nodes (see Figure 4). While the assignment of sentiment to nodes is done with a probabilistic model,

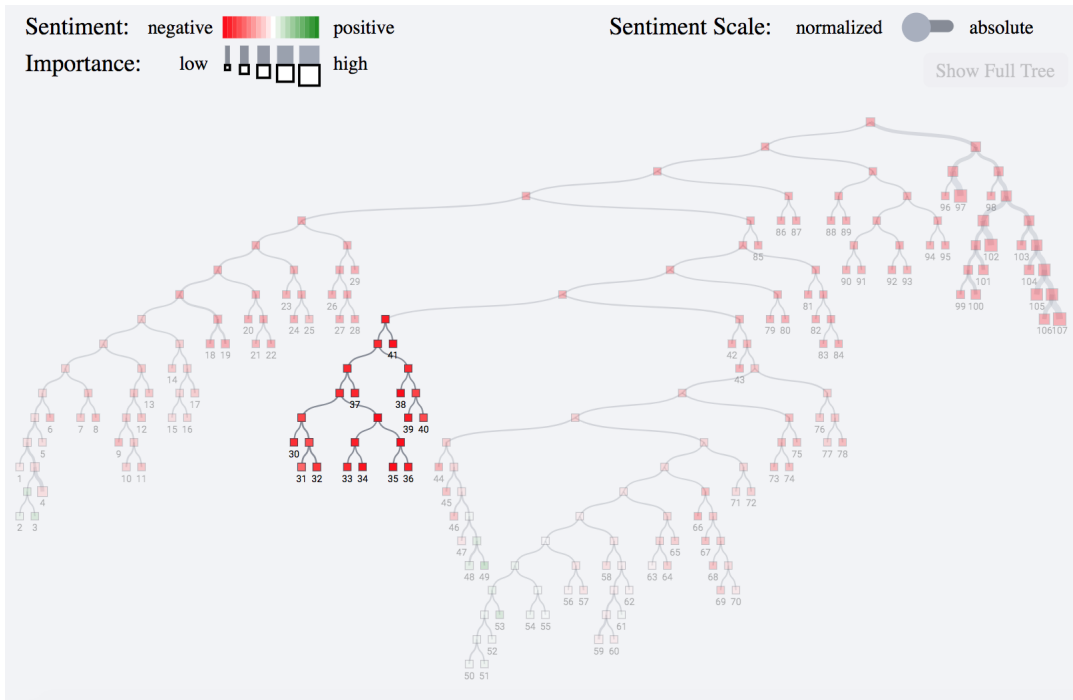


Figure 4: Center panel showing a vertical node-link diagram of the discourse tree with an interactively selected sub-tree highlighted

theoretically assigning every node a value between $(-1, \dots, 1)$, the nature of the system generates sentiment scores closer to 0, rather than strong positive and negative sentiment (due to the probabilistic properties). To alleviate the problem of the nodes to be very faint in color (see Figure 5a compared to Figure 5b), we additionally offer a normalized sentiment scale, where the strongest positive and strongest negative sentiment are used to define the coloring of the nodes. To ensure no sentiment to be flipped, we do not allow both bounds to be above or below 0. To also be able to show the propagation of the attention property through the tree, we encode this feature as the size of the nodes and the links, with a larger node/link representing a more important sub-tree than a small node with a small connection. We decided to redundantly represent the attention feature with the size of the node and the link, as the node size is easy to compare between neighbours, while the link width nicely shows the attention value being propagated through the tree and giving a good intuition where the actual importance for a sub-tree originates.

Regarding the interaction idioms in this panel, we propose a special form of restricted navigation/zooming on clicking a node. When comparing to the standard pan&zoom idiom, where the user can zoom into any arbitrary part of the interface, we only allow the user to zoom and navigate along complete sub-trees, as a node does not have any meaning without its complete sub-tree, which it is derived from. With this approach on restricting user actions, we make sure that the area shown to the user always makes sense to evaluate regarding our objectives. To be able to focus attention of a user to certain sub-trees without zooming into that region by clicking the nodes, we also allow the user to simply hover over any node in the tree and will highlight the complete sub-tree beneath the node. This interaction is thereby not restricted to only the central discourse tree panel, but also has an impact on the second detail view on the right.

Textual Discourse Representation The third panel in our new information visualization system shows the discourse itself as a vertical list of text clauses (EDUs). To be able to manually refer to EDUs in the discourse tree representation in the center

panel, we augment every unit in the discourse with an identifier and consistently match them with the nodes in the discourse tree (see Figure 6).

To allow users to easily navigate between the textual representations and the discourse tree representation, we:

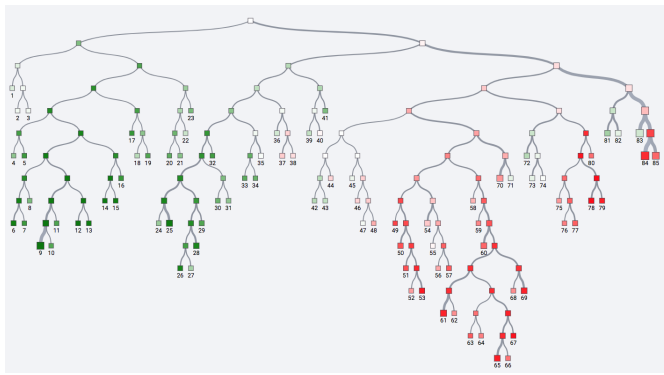
- (1) Dynamically scroll to the selected EDU in the textual representation, once the user hovers over a node in the tree which is not in the field of view already
- (2) Link the two panels with bidirectional linking between discourse units in either views. When the user hovers over one of the discourse units in the textual representation, the connected node in the tree representation is highlighted. The same happens if the user hovers over a node in the discourse tree, highlighting the EDU in the textual panel.
- (3) To be consistent with the restricted zooming approach in the tree layout, textual units which are not zoomed into are automatically disabled and lose their hovering property until they are back in focus (Figure 8)

3.4 Implementation

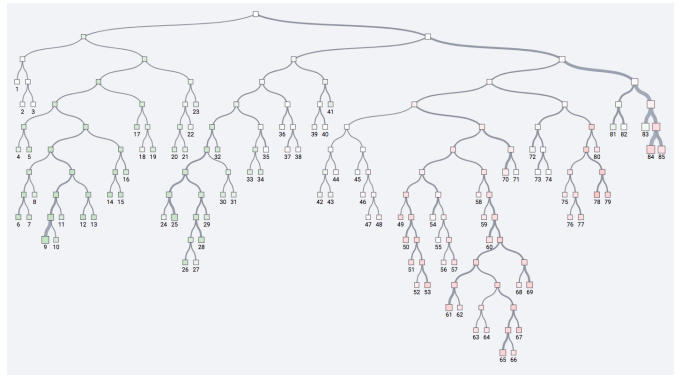
The information visualization system is completely written in native d3, with no other frameworks used. Within the d3 package, the following pre-defined functionalities are used:

```
d3.tree()
d3.hierarchy()
```

where the `d3.hierarchy()` method generates a hierarchical d3 object from a nested array, while the `d3.tree()` functionality allows the automatic assignment of tree nodes in a defined space, spreading them out as far as possible. Besides those helper functions, the complete system has been implemented from scratch. We want to especially mention the restricted navigation and zooming, which required a significant amount of time and effort to implement correctly as well as the hierarchical highlighting of the tree, requiring recursive calls on the tree to find complete sub-trees in the data. Another difficult task was the correct size adaptations regarding different discourse tree



(a) Normalized sentiment defined by the min and max sentiment polarity



(b) Interface after document has been selected

Figure 5: Non-normalized sentiment within the range $(-1, \dots, 1)$

Discourse (Doc 5)

- (1) [your restaurant needs more help !]
- (2) [lunchtime at your restaurant .]
- (3) [the tables stay full .]**
- (4) [your 3 are overwhelmed .]
- (5) [do them a favor]
- (6) [. hire someone]
- (7) [to provide water service]
- (8) [and clear tables .]
- (9) [somehow , you are not realizing]

Figure 6: Discourse representation of the document, linked to the discourse tree by indexes as well as bidirectional linking

sizes, as the node-, link- and text-size is dynamic depending on the number of nodes in the tree.

3.5 Usage Scenarios

To show the capabilities of the above proposed and described system, we give a user scenario, in which the information visualization systems significantly helps to identify some issues with the underlying machine learning model, which generates the data itself. To start the use-case, we assume that the user pressed on one of the documents in the left panel to select the element from the dataset and show the detailed information in the center and right panel. The user will thereby find a tree like the one shown in Figure 1. Initially the user is able to see a good separation between positive and negative sentiment in the tree, where the leaf nodes 1 – 41 are in the strictly positive part of the tree, while the second part of the document, from EDU 42 – 85 contains mixed sentiment, with a tendency to negative. When looking at the textual discourse clauses, it can be seen that there is also a shift in topic between the EDUs 1 – 41, mostly referring to the service received in the restaurant compared to EDUs 42 – 85 complaining about the mediocre food. While the first part of the tree seems well aligned, the mixed sentiment in the second part is worth taking a closer look into. To do so, the user will zoom into the subtree containing EDUs 42 – 85, to get a narrowed down view of the document (see Figure 7). In this close-up view, it can be seen that for some of the EDUs, which should have neutral sentiment, such as EDU 75 “I can only say”, the predicted sentiment

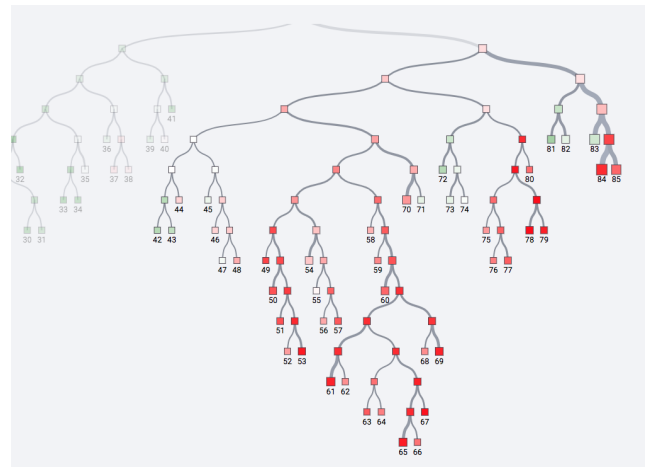


Figure 7: Selection of sub-tree to investigate further on the mixed sentiment assignment

Sentiment: ■ negative ■ positive

Importance: low ■ ■ ■ high

Sentiment Scale: ■ normalized ■ absolute

Show Full Tree

(26) [in advance]

(27) [I was .]

(28) [the asked]

(29) [what I wanted .]

(30) [my order was for som tom .]

(31) [green pepaya salad made this hot .]

(32) [and leave the green beans off , along with a bowl of steamed rice .]

(33) [the waiter did not know the Thai name for green pepaya salad .]

(34) [so he was going to open the menu]

(35) [when I told him]

(36) [what som tom is .]

(37) [he repeated back my order , with no Thai but no steamed rice .]

(38) [I repeated the order .]

Figure 8: Bidirectional linking between the detail views with additional disabling of out-of-focus textual units

is negative. Similar examples can be found at multiple places in the sub-tree. With this finding, a potential cause can be inferred to be the underlying machine learning model, which is contextual and might therefore be influenced by the neighbouring sentiment, not accurately predicting the neutral polarity. With this new insight, the machine learning approach can be enhanced and another round of information visualization can show the effects of the changes made.

4 CONCLUSION

We have shown an information visualization design study of our new interface to tackle an important problem for deep learning algorithms in the domain of discourse parsing: The lack of understanding and fine-grained evaluation to enhance model performance.

Our new system is build in a three column design and especially features interaction idioms custom tailored to discourse parse trees, such as restricted navigation of full discourse sub-trees and bidirectional, hierarchical highlighting. We chose the node-link visualization idiom due to the key features of the data. In the usage scenario we show a valuable use-case for the system based on real-world data, confirming the need for a visualization system to better understand the generated data on a document-level, rather than corpus-level.

5 DISCUSSION AND FUTURE WORK

Our novel information visualization tool is able to execute most of the intended use-case scenarios, especially focusing on the alignment of discourse structure, sentiment and the textual representation. Even though one of hte goals in the creation of the system was to allow novice users to be able to use the system, the inherit difficulty to understand textual semantic structure requires some basic knowledge of discourse related tasks and therefore does not apply to total novice users, such as reote workers on Amazon Mechanical Turk (AMT), which could have reduced the cost of human evaluation through crowd-sourcing. Another weakness of the model is the selection panel on the left, which does not contain any information on the document at the moment. Some key metrics on the document itself could significantly enhance the selection speed and would additionally enable filtering, which is not possible at the moment. Some of the strengths of the system are the solid visual representation, which is adaptable along the complete range of possible inputs. However, for future use with far beyond 150 EDUs common problems like the hairball problem might occur and need to be evaluated in a separate study.

Our main directions for future work include the extension of the selection panel to contain key metrics on the document and allow filtering to find interesting or misaligned documents faster, reducing the process-time. We want to extend the bidirectional highlighting to also allow the selection of a consecutive range of textual representations and visualize the selection in the tree-view. Another large area of future work is the ability to collapse sub-trees and overlay multiple, similar documents to see general trends easier.

REFERENCES

- [1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pp. 265–283, 2016.
- [2] P. Bhatia, Y. Ji, and J. Eisenstein. Better document-level sentiment analysis from rst discourse parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2212–2218, 2015.
- [3] C. Culy, C. Dima, and E. Dima. Through the looking glass: Two approaches to visualizing linguistic syntax trees. In *2012 16th International Conference on Information Visualisation*, pp. 214–219. IEEE, 2012.
- [4] S. Gerani, Y. Mehdad, G. Carenini, R. T. Ng, and B. Nejat. Abstractive summarization of product reviews using discourse structure. In

- Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1602–1613, 2014.
- [5] M. L. Gregory, N. Chinchor, P. Whitney, R. Carter, E. Hetzler, and A. Turner. User-directed sentiment analysis: Visualizing the affective content of documents. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, pp. 23–30. Association for Computational Linguistics, 2006.
 - [6] A. Hogenboom, F. Frasinca, F. De Jong, and U. Kaymak. Using rhetorical structure in sentiment analysis. *Commun. ACM*, 58(7):69–77, 2015.
 - [7] P. Huber and G. Carenini. Predicting discourse structure using distant supervision from sentiment. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 2306–2316, 2019.
 - [8] Y. Ji and N. A. Smith. Neural discourse structure for text categorization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 996–1005, 2017.
 - [9] K. Kucher and A. Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 117–121. IEEE, 2015.
 - [10] W. C. Mann and S. A. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281, 1988.
 - [11] B. Nejat, G. Carenini, and R. Ng. Exploring joint neural model for sentence level discourse parsing and sentiment analysis. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 289–298, 2017.
 - [12] R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber. The penn discourse treebank 2.0. *LREC*, 2008.
 - [13] J. Wexler, M. Pushkarna, T. Bolukbasi, M. Wattenberg, F. Viégas, and J. Wilson. The what-if tool: Interactive probing of machine learning models. *IEEE transactions on visualization and computer graphics*, 2019.
 - [14] J. Zhao, F. Chevalier, C. Collins, and R. Balakrishnan. Facilitating discourse analysis with interactive visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2639–2648, 2012.