TBI-Doc: Generating Patient & Clinician Reports from Brain Imaging Data

Pamela Jordan[1], **Nancy L. Green**[2], **Christopher Thomas**[3], **Susan Holm**[1]
Learning Research & Development Center, University of Pittsburgh, Pittsburgh, PA[1]
Department of Computer Science, University of North Carolina Greensboro, Greensboro, NC[2]
Department of Computer Science, University of Pittsburgh, Pittsburgh, PA[3]
pjordan@pitt.edu, nlgreen@uncg.edu, clt29@pitt.edu,
susan.holm@gmail.com

Abstract

The TBI-Doc prototype demonstrates the feasibility of automatically producing draft case reports for a new brain imaging technology, High Definition Fiber Tracking (HDFT). Here we describe the ontology for the HDFT domain, the system architecture and our goals for future research and development.

1 Introduction

The goal of TBI-Doc is to automatically produce a draft of a traumatic brain injury (TBI) case report similar to existing expert-authored reports that interpret the results of a High Definition Fiber tracking (HDFT) procedure (Shin et al., 2012). HDFT is a new, revolutionary technology for rendering detailed images of the brain and is expected to have significant implications for TBI patients' prognosis and treatment. The typical patient for whom HDFT is indicated has suffered multiple impacts to the head over an extended period of time. Although these patients suffer significant symptoms, in most cases current imaging tools (e.g. MRI and CT) are unable to pinpoint the locations of the injuries, much less any evidence of TBI. Fortunately, HDFT is providing a wealth of details for the patient and clinician about the TBI. Unfortunately, the 25 page, expert-generated report takes up to 10 hours of effort to produce once the HDFT procedure is completed: part of the time is analysis and part is report writing.

Accordingly, TBI-Doc's success will be measured in terms of reducing the amount of human time involved in creating the final report presented to the patient and clinicians. In this paper we describe the TBI-Doc prototype which demonstrates the feasibility of the system. The main contribution at this stage of development and the focus of this paper is the ontology necessary for generating the reports and the system architecture. We conclude with our goals for future research and development.

2 The HDFT Results and Expert-Authored Case Reports

Currently HDFT produces data on 13 brain tracts. One such tract, which we focused on for the TBI-Doc prototype, is the **superior longitudinal fasciculus** (SLF) which connects regions of the frontal lobe with the parietal and temporal lobes (Fernandez-Miranda et al., 2012). The brain regions that a tract connects and the areas of the tract that appear abnormal suggest the brain functions that may be impacted (Shin et al., 2012).

To identify abnormalities, the imaging process mathematically compares the volume of the patient's right and left hemisphere for a particular tract (Shin et al., 2012) and looks for unexpected asymmetries.[1] The volume is also compared against the HDFT data of a population of individuals who have not suffered any TBI. Finally, the analyst also uses his/her knowledge of the anatomy of a healthy brain to identify abnormalities and can further characterize the density, distribution and connectivity of the fibers of the tract by visually examining a representation of it, as shown within Figure 1 for the fronto-occiptal fasciculus tract. As part of the reporting, the analyst describes the above comparisons, marks up the visual representations of the tract that illustrate his/her observations and includes graphs that represent the volume comparisons.

Because HDFT is new, currently, there are relatively few *ideal* expert-generated case reports upon which to model TBI-Doc's reports. When we began the development of TBI-Doc, an analyst had written 29 case reports but only 2-3 of these reports were considered model final reports.

[1]Some tracts normally are expected to have some asymmetries between hemispheres.

Proceedings of the 8th International Natural Language Generation Conference, pages 143–146,
Philadelphia, Pennsylvania, 19-21 June 2014. ©2014 Association for Computational Linguistics

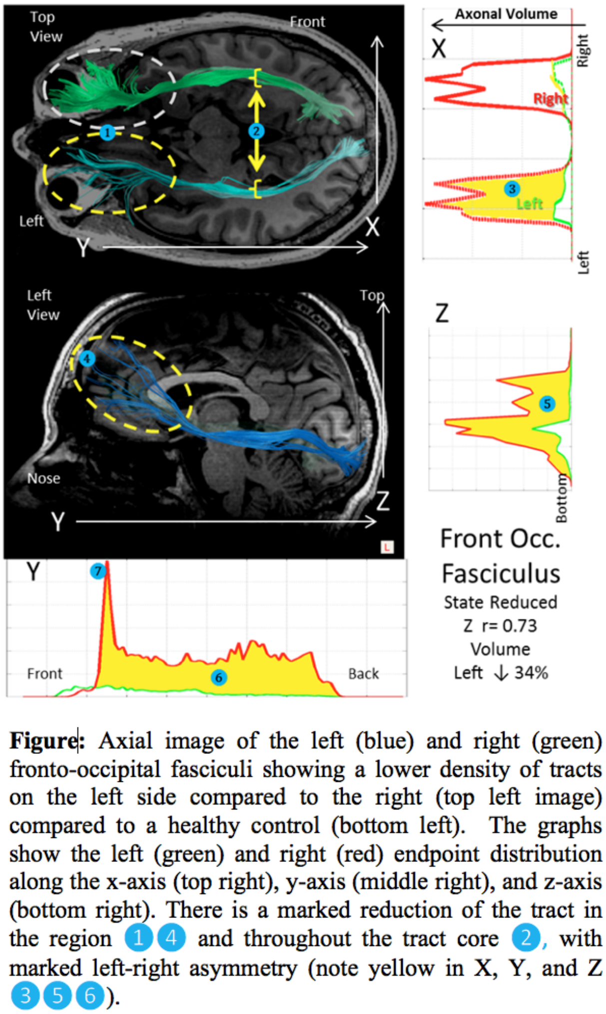


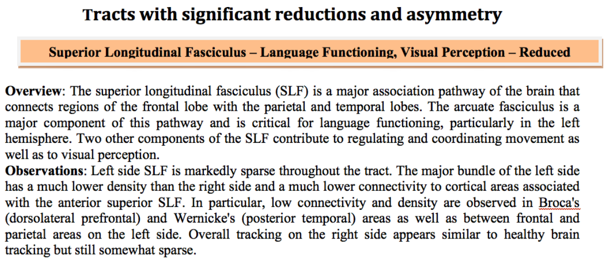
Figure 1: Tract Report Excerpt: shows HDFT images of fibers representing the FOF Tract

The content and format of the final case report is continuing to evolve as the primary physician and nurse on the HDFT team provide feedback on the type of report they believe will be most beneficial. As of yet, there has been no feedback from patients or other types of treating clinicians (e.g. speech therapists, physical therapists, etc.) on the content and format that they find most helpful.

For the prototype we focused on modeling the tract section of one of the case reports but used the remaining reports for determining the ontology. An excerpt of the SLF tract section of that model case report is shown in Figure 2.

3 The TBI-Doc System Design

Given the existing analyst workflow, we designed the TBI-Doc system process (see Figure 3) as follows; after interpreting and manually annotating HDFT images of tracts, and creating and annotating data graphics that show quantitative HDFT results, the analyst uses TBI-Doc's graphical user interface (GUI) to provide his qualitative evaluation of the HDFT results and preferences for tailoring the report. Using the analyst's specifications provided through the TBI-Doc GUI and the annotated tract images and data graphics, TBI-Doc automatically produces a first draft of the case report. The draft is then manually reviewed and edited by the analyst before delivery. The case report is delivered to the clients as a file that can be printed on paper and viewed on a tablet.



Tracts with significant reductions and asymmetry

Superior Longitudinal Fasciculus – Language Functioning, Visual Perception – Reduced

Overview: The superior longitudinal fasciculus (SLF) is a major association pathway of the brain that connects regions of the frontal lobe with the parietal and temporal lobes. The arcuate fasciculus is a major component of this pathway and is critical for language functioning, particularly in the left hemisphere. Two other components of the SLF contribute to regulating and coordinating movement as well as to visual perception.

Observations: Left side SLF is markedly sparse throughout the tract. The major bundle of the left side has a much lower density than the right side and a much lower connectivity to cortical areas associated with the anterior superior SLF. In particular, low connectivity and density are observed in Broca's (dorsolateral prefrontal) and Wernicke's (posterior temporal) areas as well as between frontal and parietal areas on the left side. Overall tracking on the right side appears similar to healthy brain tracking but still somewhat sparse.

Figure 2: Expert Observation being Modeled

The architecture of TBI-Doc (shown by the remainder of Figure 3) follows the standard NLG pipeline (Reiter et al., 2000) and is similar to the architecture of the healthcare-related systems described in (Green et al., 2011; Hunter et al., 2012; Scott et al., 2013). The TBI-Doc GUI represents the TBI-Doc ontology and its columns (an excerpt is shown in Table 1) cue the analyst to enter his/her qualitative judgments about the data for a tract at the region level, which is the lowest judgment level as it describes the endpoints of subsections of a tract, bundles between regions, hemisphere level and overall. The ontology was derived by analyzing existing reports to understand what is being described across all of the reports and by interviewing HDFT team members. The ontology identifies states (e.g. measures), relations (e.g. similar

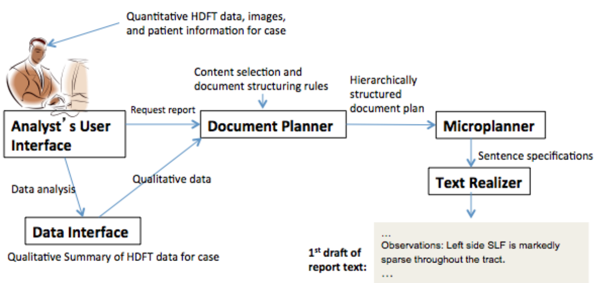


Figure 3: The TBI-Doc Process

Table 1: Example Input for SLF Tract Assessment

Tract Assessment	SPARSE						
Tract	hemi-sphere	Area Type	Area Name	Second End Point	Measure	Evaluation	Kind of Comparison
SLF	left	OverallTract	tract		density	very sparse	healthy
SLF	left	BundleBtwnRegions	DLPFC	pTemporal	density	sparse	right
SLF	left	BundleBtwnRegions	DLPFC	pTemporal	connectivity	reduced	right
SLF	left	SpecificRegion	DLPFC		connectivity	little	right
SLF	left	SpecificRegion	pTemporal		connectivity	little	right
SLF	left	SpecificRegion	DLPFC		density	sparse	right
SLF	left	SpecificRegion	pTemporal		density	sparse	right
SLF	left	BundleBtwnRegions	DLPFC	pParietal	connectivity	little	right
SLF	left	BundleBtwnRegions	DLPFC	pParietal	density	sparse	right
SLF	right	OverallTract	tract		density	normal	healthy
SLF	right	OverallTract	tract		density	some sparse	healthy

to), and entities (e.g. tract(s), regions(s), connection(s), measurements such as density) relevant to the HDFT reporting domain.

When the qualitative judgment entries or updates are complete, the analyst requests that a report be generated. The TBI-Doc Document Planner (logic implemented in Java) selects appropriate content from the database using the TBI-Doc Data Interface and adds messages constructed from that content as leaves of the Document Plan. For the parts of the report that are not dependent on values in the database, the Document Planner also adds English (canned) text as leaves of the Document Plan. The Document Planner is a set of rules for what content to select and how to order that content. For example, an abbreviated excerpt of the content selection rules follow.

```
GENERATETRACTSECTION(tract, patientId)
    GETTRACTSTATUS(tract,patientId)
    If status=reduced then
        GETTRACTFUNCTION(tract)
        GENTRACTOBSERVATION(tract,patientId)

GENTRACTOBSERVATION(tract, patientId)
  If Evaluation=reduced & TractOverall then
      GENTRACTSUMMARYSENTENCE
  ElsIf hemi not both & AreaType=OverallTract then
      GENHEMISPHERESUMMARYSENTENCE
  regions=
   GATHERREGIONS(SpecificRegion,BundleBtwnRegions)
  orderedRegions=ORDERREGIONS(regions)
  For region in orderedRegions do
     GENSENTENCE(region)
```

While there is often just a single sentence for a tract or hemisphere summary, region descriptions are generally multi-sentential. Currently the orderRegions function is designed as a default set of guidelines for ordering the region descriptions. The output of the Document Plan is then a series of predicates that represent the content to be realized. Some content, such as getTractFunction, is static and does not pass through the pipeline to the Microplanner.

The TBI-Doc Microplanner transforms the predicates output from the Document Plan into SimpleNLG sentence specifications (in Java) via a set of mapping rules. The Microplanner selects mapping rules based on the predicates to be realized and any context variables that are available. The mapping rules indicate what syntactic structures to create for a predicate and where to attach them in the sentence being built. Currently, for this demonstration prototype we have not yet addressed lexical realization and sentence aggregation. In the final step of the pipeline, SimpleNLG (Gatt and Reiter, 2009) renders the sentence specifications as English sentences. Once the pipeline is complete, the TBI-Doc Formatter combines all the sentences from SimpleNLG and the canned text into an HTML document which can then be displayed by a browser and edited via an XML editor.

Rather than implementing each of the above steps one-by-one to cover all possible cases, each step was implemented to focus on replicating the observation section of one case report. This allowed us to perform an end-to-end demonstration of the feasibility of this design. Thus many of the rules described above are incomplete for alternative pathways. TBI-Doc can currently generate from input data an observation section such as the one shown below. The judgments entered on behalf of the analyst for this demonstration are shown in Table 1 and represent what was expressed in the expert-written observations section in Figure 2:

Observations Left SLF is particularly sparse throughout the tract. The left tract from the DLPFC to the pTemporal region when compared to the right has a sparse density and a reduced connectivity. In particular little connectivity and sparse density are observed in the DLPFC and pTemporal regions as well as between the DLPFC and pParietal regions on the left. Overall the right tract appears similar to a healthy tract but still appears somewhat sparse.

4 Future Work

The current TBI-Doc is a demonstration of the feasibility of generating case reports and the main contribution of the work thus far has been to define an ontology for the HDFT domain. However, because HDFT is a new technology that is continuing to be improved rapidly and the reporting goals are still evolving, the ontology is not yet complete. Because the ontology drives the rest of the system, it follows that the rest of the system components still need more development.

For the demonstration we focused on reporting on one of the 13 existing types of brain tracts. While we anticipate that the ontology will generalize well to the other tract types, each tract type may introduce some extensions to the ontology. In addition the HDFT developers anticipate providing data on additional tract types over time.

Since knowledge acquisition is still ongoing, the Document Planner logic is still very shallow. As a result, the demonstration version of TBI-Doc is currently limited to reacting to descriptor changes and does not yet alter the document structure or intelligently alter content selection. The Microplanner currently does some context checking to select the appropriate set of transformation rules to apply but this will need expansion as the Document Planner becomes more complete. More specifically, the sentence structure needs to vary depending on the choices made by the Document Planner. In addition, lexical selection in which internal abbreviations are mapped to user preferred forms needs more work (e.g. depending on user preferences, DLPFC could map to dorsolateral prefrontal cortex and pParietal to posterior Parietal).

Our longer term interest is to explore ways to appropriately adapt the reports for different clients. A patient for whom the HDFT results indicate cognitive processing issues may find a different style of report and reading level more suitable than a supporting family member or a treating clinician. Different treating clinicians may prefer reports with different content selected. For example, a speech therapist may prefer a report that focuses on the injuries that relate to a patient's speech and language goals, while a sleep specialist may prefer a different focus.

Acknowledgments

This work was supported by US Army Combat Casualty Care Research Program contracts CDMRP PT110773 (W81XH-12-2-0140) & US Army 12342013 (W81XWH-12-2-0319). We thank Walter Schneider, Kevin Jarbo, Lauren Wagener, Will Bird & the HDFT team for their expertise and for involving us in their research.

References

Juan C Fernandez-Miranda, Sudhir Pathak, Johnathan Engh, Kevin Jarbo, Timothy Verstynen, Fang-Cheng Yeh, Yibao Wang, Arlan Mintz, Fernando Boada, Walter Schneider, et al. 2012. High-definition fiber tractography of the human brain: neuroanatomical validation and neurosurgical applications. *Neurosurgery*, 71(2):430–453.

Albert Gatt and Ehud Reiter. 2009. Simplenlg: A realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 90–93. Association for Computational Linguistics.

Nancy Green, Rachael Dwight, Kanyama Navoraphan, and Brian Stadler. 2011. Natural language generation of biomedical argumentation for lay audiences. *Argument & Computation*, 2(1):23–50.

James Hunter, Yvonne Freer, Albert Gatt, Ehud Reiter, Somayajulu Sripada, and Cindy Sykes. 2012. Automatic generation of natural language nursing shift summaries in neonatal intensive care: Bt-nurse. *Artificial intelligence in medicine*, 56(3):157–172.

Ehud Reiter, Robert Dale, and Zhiwei Feng. 2000. *Building natural language generation systems*, volume 33. MIT Press.

Donia Scott, Catalina Hallett, and Rachel Fettiplace. 2013. Data-to-text summarisation of patient records: Using computer-generated summaries to access patient histories. *Patient education and counseling*, 92(2):153–159.

Samuel S Shin, Timothy Verstynen, Sudhir Pathak, Kevin Jarbo, Allison J Hricik, Megan Maserati, Sue R Beers, Ava M Puccio, Fernando E Boada, David O Okonkwo, and Walter Schneider. 2012. High-definition fiber tracking for assessment of neurological deficit in a case of traumatic brain injury: finding, visualizing, and interpreting small sites of damage: Case report. *Journal of neurosurgery*, 116(5):1062–1069.