The Role of Semantic and Discourse Information in Learning the Structure of Surgical Procedures

Ramon Maldonado, Travis Goodwin and Sanda M. Harabagiu
Human Language Technology Research Institute
University of Texas at Dallas
Dallas, Texas 75080
Email: {ramon,travis,sanda}@hlt.utdallas.edu

Michael A. Skinner
UT Southwestern Medical Center
Childrens Medical Center of Dallas
Dallas, Texas 75235
Email: michael.skinner@childrens.com

Abstract—Electronic Operative Notes are generated after surgical procedures for documentation and billing. These operative notes, like many other Electronic Medical Records (EMRs) have the potential of an important secondary use: they can enable surgical clinical research aimed to improving evidence-based medical practice. Recognizing surgical techniques by capturing the structure of a surgical procedure requires the semantic processing and discourse understanding of operative notes. Identifying only predicates pertaining to surgical actions do not explain the various possible surgical scripts. Similarly, recognizing all actions and observations pertaining to a surgical step cannot be performed without taking into account discourse structure. In this paper we show how combining both forms of clinical language processing leads to learning the structure of surgical procedures. Experimental results on two large sets of operative notes show promising results.

I. INTRODUCTION

Operative Notes are generated after surgical procedures for documentation and billing. These operative notes are typically dictated and placed in the Electronic Medical Record (EMR). Like many other EMRs these documents exhibit the potential of an important secondary use: they can enable surgical clinical research aimed at improving evidence-based medical practice. In particular, if we had the tools and techniques to automatically process such surgical operative notes, we could extract evidence about the techniques used in each step of the procedure, the observations made during the operations, and perhaps most importantly, the management solutions devised by the surgeon in the face of unexpected or unusual situations. Unlike many other forms of EMRs, e.g. discharge summaries, few research projects have focused on the medical language processing required for capturing the information conveyed by surgeons when they author operative notes.

The vast majority of medical language processing systems (e.g. cTAKES [1], MedLEE [2], MetaMap [3]) identify in clinical texts the concepts encoded in the Unified Medical Language System (UMLS), a very large semantic network of biomedical concepts developed by the National Library of Medicine. However, a study by [4] has shown that only 11.5% of the verbal predicates describing surgical actions and observations that were mined from a large corpus of 362,310 operation narratives could be mapped to any UMLS concept. When the actions were described by nominal predicates, they could be mapped in only 58.8% of the cases in UMLS concepts. Moreover, the study of [4] showed that the verbs and nominals identified in a large corpus of operative notes they have analyzed were encoded also in several other lexical or lexico-semantic resources: in SPECIALIST, the lexicon used by UMLS, verbs from the operative notes were identified in 89.9% of the cases, whereas the nominal in 100%, and when WORDNET [5] was also considered, the lexicon for verbs was covered in 93.8% of the cases. But sadly, lexica do not provide any semantic information, thus predicates corresponding to surgical actions or observations cannot be linked to their arguments. Hence, understanding surgical techniques through the sequence of actions and observations therein is not possible without being able to reliably extract the predicates that map to the same actions and observations across multiple operative notes that describe the same type of surgical procedure. To address this problem, semantic frames need to be automatically identified in operative notes. Typically, semantic frames are identified in texts through the process of “shallow” semantic parsing, by recognizing predicates and their arguments as a classification problem, trained on annotations produced by expert linguists on vast collections of texts. These annotations are not only expensive, but they must obey semantic definitions of the frames, which state the meaning of the predicate and the roles of their arguments. Existing annotations are provided by the PROPANK [6], NOMANK [7] and FRAMENET [8] projects, giving rise to semantic parsers that identify PROPANK or NOMANK-defined predicate-argument structures or FRAMENET-defined frames. Because the analysis reported in [4] showed that only 64.2% of the verb predicates and 36.1% of the nominal predicates from operative notes are encoded in FRAMENET, it is clear that FRAMENET-defined semantic frames are not ideal for processing operative notes.

The full semantic specification of a semantic frame is provided by the definitions of the arguments. To define the arguments of the semantic frames illustrated in Figure 1, we have considered the definition of arguments provided in PROPANK and NOMANK, whenever possible, and extended them with definitions available from FRAMENET. In addition, we have considered a new type of arguments, namely ARGm-adj and ARGm-adv, corresponding to adjectival or adverbial modifiers of the predicates, to indicate the MANNER of the predicates. For example, the “prep” and “drape” predicates shared an argument ARGm-adv to indicate the manner in which the surgical actions were performed, namely “sterilely”. In PROPANK, the verbal predicate “prep” has three defined arguments: ARG0 to indicate the preparer, the ARG1, to indicate who is prepped, and ARG2 to indicate what is it prepped...
Under general anesthesia, the patient was steriley prepped and draped.

A 1-centimeter vertical incision was made through the skin and fascia of the umbilicus with a 15 blade.

Blunt dissection was used to enter the abdomen through this incision.

A 10-millimeter trocar was placed through the incision into the abdomen and connected to CO2 gas to create a pneumoperitoneum.

Fig. 1: Semantic Frames and Discourse in Operative Notes.

for (e.g. if in the first sentence, the surgeon would have written “prepped for appendectomy”, “appendectomy” would have been the ARG2 for the predicate “prep”, as well as a nominal predicate). Therefore, the semantic frame of the predicate “prep” also comprises the argument “the patient”, having the meaning of the person being prepped. The predicate “drape” for the first sentence is not encoded in PROPANK, but it is encoded in WORDNET, where it has three semantic senses as a verb and another three as a noun. The third semantic sense of the noun “drape” is glossed as “a sterile covering arranged over a patient’s body during a medical examination or during surgery in order to reduce the possibility of contamination”, indicating the appropriate sense for surgery. The genus of the gloss “covering” is a nominalization of the verb “cover”, which is encoded in PROPANK, having three arguments, Arg1 being defined as the thing covered, which corresponds to the semantic role of “the patient” in the first sentence. In Figure 1, relations between the three predicate-argument structures of the first sentence are also defined: the “anesthesia” is a condition that enables the “prepping” and “draping” of “the patient”.

Unlike the first sentence, in which three predicates are discerned, the second sentence corresponds to a single predicate, namely “make incision”, a predicate which is not encoded in PROPANK. This predicate requires definitions for its multiple arguments, including the direction of the incision (e.g. “vertical”), the path of the incision through the various organs and tissues, the instruments used as well as the size of the incision. At the discourse level, the predicate “make incision” from sentence 2 is coherently connected to the nominal predicate “dissection” from the sentence 3. This predicate, performed in a manner indicated by its argument “blunt”, was motivated by the goal of entering the abdomen. In PROPANK, the verb “enter” is encoded with two arguments, corresponding to the thing entering and the place it enters for Arg1-LOC, the semantic role of “the abdomen” in sentence 3. However, in the same sentence, the predicate “enter” is also connected to “this incision”, and thus it needs a definition for the semantic role of “this incision”. Because the verb “enter” is a lexical unit of the FRAMENET frame Path-Shape, which has a frame Element Means, the definition of the argument for “this incision” is provided by the definition of Means, incorporated from FRAMENET. Moreover, the expression “this incision” is an event coreference to the entire surgical action represented by the predicate-argument structure of the predicate “make incision”, and interestingly enough, it does not represent an identity-coreference, but rather a resultative-coreference. Moreover, the argument “this incision” is also involved in an entity-coreference with the argument “the incision” of the predicate “place” from the sentence 4 illustrated in Figure 1. The predicate “place” has two arguments in the last sentence illustrated in Figure 3: “trocar”, which as a surgical instrument can also be semantically argumented by a modifier indicating its size “10-millimeter”, and “the incision”. While the role of “trocar” is defined in PROPANK as having the semantics of the thing being placed, there is no definition for the role of “the incision”, requiring the use of the definition of Path from the FRAMENET frame of Placing, a frame having the verb place as a lexical unit.

As shown in Figure 1 in the fourth sentence, the predicates
“place” and “connect” share the argument “trocar”, a form of intra-sentential discourse coherence to indicate a condition relation. The semantic role of “trocar” for the predicate “connect” is defined as being the first thing connected, while the “CO2 gas” instillation tubing is the second thing connected. The final predicate, “create pneumoperitoneum” is the goal of the previous two surgical actions. This illustration of the analysis of surgical actions derived from operative notes indicates that (1) the semantic specifications are not readily available in a single semantic resource, such as PROPBank; (2) relations between semantic frames also require some form of discourse processing, and (3) surgical techniques vary across narratives from operative notes. Thus we claim that in order to analyze surgical procedures we need to consider both semantic and discourse information to (a) first discover which actions and observations correspond to each surgical step; and (b) semantically align the predicate argument structures to be able to cover all their arguments and to capture both their semantic variability and discern their paraphrases.

When the discourse of the operative notes was previously considered, [9] showed that the structure of surgical procedures discovered through active learning performed on operative notes was revealed by a simple categorization of the surgical actions, without needing to rely on the full semantic specification of the surgical actions. However, the technique reported in [9] requires new annotations for each type of surgical procedure, and re-training though active learning. We are more interested in developing a framework of learning the semantic and discourse structure without any new annotations, enables us to discover the structures of the two types of surgical procedures we have analyzed. To our knowledge, this is the first attempt to generate a semantic and discourse representation for expressing surgical actions and observations.

The remainder of this paper is organized as follows. In section II we describe the two surgical operations we have analyzed and the corpus of operative notes on which we conducted our experiments. Section III details the semantic processing of the narratives from the operative notes, while Section IV describes the discourse processing we produced. Section V presents the multiple sequence alignments of the Pred2Vec semantic representations and the surgical structures they enabled, while Section VI discusses the evaluations of our experiments. Section VII summarizes the conclusions.

### II. Data

We performed semantic and discourse analysis on a corpus of 3,546 operative notes, provided by a pediatric surgeon from The Childrens Hospital Medical Center in Dallas, TX. The notes document two common types of pediatric operations, namely appendectomy and humerus repair. An appendectomy is the surgical removal of the vermiform appendix. This procedure is normally performed as an emergency procedure, when the patient is suffering from acute appendicitis. The humerus repair surgery is performed to repair the fracture of the humerus. The humerus is the only bone in the upper arm, running from shoulder to elbow. It is commonly fractured, often by sports injuries, accidents or falls. In our data we had 2,816 appendectomy notes and 730 humerus repair notes. The appendectomy notes were written by 12 different surgeons and the humerus repair notes were generated by 14 different
An infraumbilical incision was made, and the peritoneal cavity was entered.

Fig. 3: (a) sentence and its dependency parse; (b) the semantic parse of the same sentence, with predicates highlighted and argument roles specified; (c) lexical parsing and normalization; (d) semantic parsing customized for the surgical domain.

Surgeons. All notes were de-identified to protect the privacy of the patients. The study was performed under an IRB exemption granted by the Institutional Review Board of the University of Texas Southwestern Medical Center.

When performing a surgical procedure, a sequence of steps are typically followed. Figure 2 illustrated the steps of the two procedures documented in our corpus, described by a pediatric surgeon with 30+ years of experience. As evidenced by the three sentences illustrated in Figure 1, not all operative notes describe all the steps of the surgical procedure. For example, the first step of the appendectomy documented in the operative note that starts with the three sentences illustrated in Figure 1 describe the actions performed in the second and third steps of the surgical procedure, while the first step is not documented. Moreover, some of the typical actions of step 2 of appendectomy, such as performing a time-out, administering pre-operative antibiotics or inserting a catheter are not mentioned either, although being typical to this step of an appendectomy. In the step 3 of the note illustrated in Figure 1, the typical action of insufflation is described by its resultant, namely “creating pneumoperitoneum” which allows the surgeon to visualize the surgical area after the camera is introduced. The enablement produced by the resultant was expressed as “A 5-millimeter camera was placed into the trocar and the contents of the abdomen were visualized” in the operative note partially illustrated in Figure 1. Therefore, the surgical action “insert camera” was expressed by its paraphrase “place camera into the trocar”. As in Figure 1, discourse coreference is used, as the surgical instrument “trocar” refers to the “10-millimeter” trocar that was placed in the previous action. The coherence of the discourse is captured by the fact that the trocar is the means for inserting the camera in the patients abdomen, allowing the surgeon to visualize it and make observations. This logical entailment is made possible by access to medical knowledge, indicating the usage of the trocar as a surgical instrument. Therefore, not only the semantic processing of surgical actions is complicated by the semantic definitions of the arguments of each predicate (as we have seen in Figure 1), but also discourse plays an important role in capturing the knowledge about the surgeons actions, and discourse coherence is informed by sequences of predications from the same surgical step.

In the corpus of operative notes that we have processed, we noticed that some verbs were predominantly used to describe an action. Much fewer nominal predications were observed. In our corpus, we have discovered a total of 114,422 verbal predicates and 22,458 nominals. Verbal predicates were identified after producing a syntactic parse of the corpus (the parser is described in Section III) and verbs that were heads of verbal phrases were recognized. Nominals were identified by (1) requiring the noun to be deverbal (using information from WORDNET, which lists derivational information, e.g. connecting “incision” to “making an incision”). Surgical actions were always described by some verbal or nominal predications.

In addition to surgical actions, operative notes document observations. For example, in the appendectomy notes, the state of the appendix is nearly always noted by an observation, e.g. “The appendix was inflamed and nonperforated.” Similarly, in the humerus fracture notes, surgeons make note of the state of the blood flow in the arm, providing observations of the form “There was an intact radial artery pulse”. Unlike surgical actions, the observations are expressed by the attributes of an anatomical location or bodily function which conform to the expected state or indicate a finding that is atypical, usually requiring additional surgical actions. Hence, in processing operative notes, we are also interested in capturing the observations that lead to a surgical maneuver involving actions which are not listed in Figure 2; such findings may suggest an unexpected phenomenon or complication. To generate automatic processing techniques that capture the semantics and discourse of the operative notes we constrained the predicate-arguments structures to involve only information about the patient, anatomical locations, surgical instruments and surgical supplies, including the operative table and the operative room. A predication was included in our analysis if it was coordinated syntactically with another predication which was considered of interest based on the constraints presented above. This assumption allowed us to produce semantic information of 114,422 verbs and 22,458 nominals, out of which 108,332 and 14,925 were included due to syntactic co-ordination.
III. SEMANTIC PROCESSING

A. Identifying Predicate Argument Structures

Shallow semantic processing of narratives aims to identify predicates and their arguments in texts. To be able to recognize predicates and their arguments, semantic parsers rely on (a) semantic definitions of predicate-argument structures and (b) annotations that are provided to exemplify the definitions. The PropBank project [10] had the goal of documenting the syntactic realization of verbal predicates and their arguments by annotating a newswire corpus (mainly from the Wall Street Journal) with semantic roles. In parallel, the NomBank Project [7] performed the same annotations on the same newswire corpus, targeting the nominal predicates. These annotations enable the design of several semantic parsers, capable of detecting verbal and nominal predicates and their arguments in any new texts. From the initial efforts of producing automatic shallow semantics, the role of the syntactic parse results in detecting the predicate-argument structure was evident [11]. Moreover, joint learning of the syntactic and semantic parsers was shown to be optimal [12]. State-of-the-art parsers that discover predicate-argument structures follow the joint learning framework for identifying both the syntactic dependency parse and the semantic role labeling of the predicates from texts. In our experiments, we have used and enhanced the parser described in [13], which has the pipeline architecture illustrated in Figure 4.

The operative notes were tokenized using the GENIA Tagger for biomedical text [14], lemmatized using the shortest edit script and part-of-speech tagging was produced using the Margin Infused Relaxed Algorithm (MIRA) [15], after which a dependency parse is produced as a tree consisting of all the words (and punctuations signs) of a sentence and all syntactic dependencies between them as edges. The dependency parse is generated by learning to extract features and produce the dependencies in parallel, using a hash kernel. To produce the semantic parse that identified both predicates and their arguments, four different classifiers are used. First a binary classifier decides which words express a predicate, while a second classifier selects the semantic sense of predicates that have multiple senses in PropBank or NOMBank. After the correct sense of the predicate is known, the words that belong to arguments of each predicate are identified by a binary classifier and the final classification decides the type of arguments for each predicate. All classifiers are using the L2-regularized linear logistic regression from the LIBLINEAR package [16]. In Figure 3(a) and (b) we illustrated the dependency and semantic parses obtained by this procedure. Figure 3(c) illustrates the lexical parse and normalization which allowed us to produce predicate argument structures similar to those illustrated in Figure 1.

B. Learning Embeddings of Predicate Argument Structures

Predicates describing surgical actions and observations recorded by surgeons do not occur in isolation in the operative notes. As illustrated in Figure 2, surgical actions and observations are generally related to a particular step of the surgical procedure. To capture the semantic context of the predications characterizing a surgical step, we have designed a vector representation of the semantic context of each predicate by making use of the architectures for efficient learning in neural language processing. Inspired by the work of [17], we used the Skip-gram model to find representations of predicate-argument structures identified automatically through the method detailed in Section III-B such that we can predict the surrounding semantic context in an operative note. For each predicate and its arguments identified by semantic parsing, we considered a predicate argument structure \( \text{PAS} = \{\text{predicate}, \text{argument1}, \text{argument2}, \ldots\} \). The semantic parser provides not only information about the semantic roles of the arguments, but because it jointly learns the syntactic dependency parse, it also provides the order of the arguments in the sentence where the predicate is identified. Thus in \( \text{PAS}_i \), argument1 is the first one encountered in the sentence, while its semantic role is identified by the argument classification detailed in Figure 4. More importantly, any \( \text{PAS} \) does not appear in isolation, it has its own semantic context which is represented by a “window of \( \text{PAS}_s \)” of size \( C \) centered on \( \text{PAS}_i \), representing the \( C \) \( \text{PAS} \)s identified in an operative note before \( \text{PAS}_i \) as well as the \( C \) \( \text{PAS} \)s identified after \( \text{PAS}_i \). For example, if the \( \text{PAS}_i = \{\text{enter}, “\text{the abdomen}=\text{Arg1-Loc,” “this incision”}\} = \text{Arg-Means(Path-Shape)} \), one of the \( \text{PAS} \)s illustrated in Figure 1, and \( C=2 \), then, based on the semantics of the example illustrated in Figure 1, the semantic context consists of 2 \( \text{PAS} \)s identified prior to \( \text{PAS}_i \) as well as 2 \( \text{PAS} \)s identified after \( \text{PAS}_i \), namely Semantic-Context(\( \text{PAS}_i \)) = \{\( \text{PAS}_{i-2}, \text{PAS}_{i-1}, \text{PAS}_{i+1}, \text{PAS}_{i+2} \)\}, with:

- \( \text{PAS}_{i-2} = \{\text{make incision}, “1-centimeter” = \text{ARG}-\text{size}, “vertical” = \text{Arg-Direction,” “skin” = \text{Arg-Path,” fascia of the umbilicus” = \text{Arg-Path,” “15-in blade” = \text{Arg-Instrument}}\}
- \( \text{PAS}_{i-1} = \{\text{dissection,” blunt” = \text{Argm-adv}}\}

Fig. 4: Joint Learning for Syntactic and Semantic Parsing.
Given a predicate-argument structure \( (PAS) \) of a surgical action or surgical observation, we learned a high-dimensional vector representation of the \( PAS \) that can be used to predict its semantic context. Learning such representations is important because they enable us to identify the same surgical action or observation that is reported in different operative notes, even when using different words or expressions. We hypothesize that the same predications used to document the same type of surgical procedure have very similar semantic contexts. Thus, knowing the context of a predication, we could predict the most likely surgical action or observation which can be performed by maximizing the average log probability of the \( PAS \)s from the semantic context:

\[
\frac{1}{S} \sum_{s=1}^{S} \sum_{C \leq j \leq C, j \neq 0} \log p(PAS_{s+j}|PAS_j)
\]

where \( C \) defines the size of the window of the semantic context and \( S \) represents the total number of \( PAS \)s identified in the corpus of operative notes used for training the learning system. In the basic Skip-gram formulation reported in [18], the conditional probabilities \( p(PAS_{s+j}|PAS_j) \) are computed using the softmax function as defined by:

\[
p(PAS_O|PAS_I) = \frac{\exp (v_{PAS-O}^T v_{PAS-I})}{\sum_{p=1}^{P} \exp (v_{PAS-O}^T v_{PAS-I})}
\]

where \( P \) represents the number of \( PAS \)s identified in the entire corpus and \( v_{PAS-I} \) or \( v_{PAS-O} \) are the “input” and “output” vector representations of any \( PAS \), while \( PAS_I \) and \( PAS_O \) represent the “input” and “output” vector representations of a \( PAS \). More specifically, as illustrated in Figure 5, to learn the vector representations of the \( PAS \)s, or the predicate structure “embeddings”, we use a neural network model whose underlying principle is based on the assumption that similar predicate argument structures should have similar semantic contexts. In the Skip-gram model, as illustrated in Figure 5, a sliding window is used on the sequence of \( PAS \)s identified in an operative note to generate the training samples. In each sliding window, the model tries to use the central \( PAS \) to predict its semantic context (i.e. the surrounding \( PAS \)s). Specifically, as illustrated in Figure 5, the \( PAS \) is represented in the 1-of-S format (with \( S \) the total number of \( PAS \)s observed in the training corpus) and each \( PAS \) is represented by a long vector with only one non-zero element. Learning of the \( PAS \) embeddings in the neural network architecture represented in Figure 5 takes place in two phases: the feed-forward process and the back-propagation process. In the feed-forward process, the input \( PAS \) is first mapped into its embedding vector by the weight matrix \( M \). After that, the embedding vector is mapped back into the 1-of-S space by another weight matrix \( M' \) and the resulting vector is used to predict the surrounding \( PAS \)s using the softmax function. As training examples are used, the errors from the prediction to the training labels are computed and the prediction errors are propagated back to update the neural network in the back-propagation process. This leads to updates to the \( M \) and \( M' \) matrices. When the training process converges, the weight matrix \( M \) is regarded as the learned \( PAS \) representations in a multi-dimensional space.

Computing the prediction errors for back-propagation entails computing the derivative of \( p(PAS_{s+j}|PAS_j) \) whose computational cost is proportional to the size of the vocabulary of \( PAS \)s. As this is impractical, as the vocabulary is quite large, we have used Huffman codes to encode the vocabulary and thus used the hierarchical softmax solution reported in [18] on a context size \( C=5 \). The embeddings of the \( PAS \)s discovered automatically in the operative notes enabled us to discover which \( PAS \)s were most similar, based on cosine distance between their embeddings.

IV. DISCOURSE PROCESSING

Operative notes describe the procedure by using a discourse which documents the steps of the surgical procedure. However, not all operative notes are created in the same way. Some of them contain more elaborations then others, and some do not discuss all steps of the procedure. Moreover, each procedure exhibits a particular course, thus the actions and observations recorded may be quite unique. To automatically capture the steps of the operation which are described in the corpus we are studying, we have considered two methods. The first one is based on the assumption that predicate argument structure embeddings model a semantic context, which can be seen as a portion of the description of a surgical step. When similar embeddings are clustered, they provide a semantic representation of the surgical step, as it emerges...
From all the operative notes. The second method considers that each operative note provides a separate discourse that can be automatically segmented to account for the description of each surgical step. Finally, we combined the strengths of each method using the expectation-maximization algorithm to (a) change the clusters of embeddings based on evidence of the discourse segments, and (b) correct the discourse segments based on the content of the updated clusters of predicate embeddings.

A. Clustering Predicate Embeddings

The predicate-argument structures (PAS) embeddings are vectors of real numbers that can be clustered. A variety of clustering methods can be used, but the most appealing one is the K-Means clustering method [19]. Given that for each surgery type we know the numbers of surgical steps (as illustrated in Figure 2, we have 8 steps for appendectomies and 7 steps for humerus fracture repairs), a flat clustering method such as K-Means can be used, as the number of resulting clusters (K = number of surgical steps) is known. The vector representation of the embeddings enables the computation of the distance between embeddings by using the cosine similarity metric, introduced in Information Retrieval vector models. The K-Means algorithm produced clusters of embeddings corresponding to the PAS listed in Table 2. One problem posed by this representation of the surgical steps stems from the inability to link back any PAS to the operative note where it was identified. We have resolved this problem with the learning framework described in Section IV-C.

B. Discourse Segmentation

In addition to the semantic context of predicate argument structures (PASs), we took into account the observation that the discourse from each operative note has its own structure. Lexical cohesion was considered as a strong indicator of the discourse structure in [20], informing the TextTiling algorithm [21], which automatically segments any discourse. Using the identification of multiple simultaneously occurring “themes”, the algorithm discovers the structure of an operative note by dividing it into sentences and computing the word overlap between those sentences. The central idea is to consider the structure of an operative note as a function of the connectivity patterns of the clinical terms that comprise it, a viewpoint also advocated by [22]. The TextTiling algorithm consists of three steps performed after sentence boundaries are identified: (1) term tokenization; (2) lexical score determination and (3) segmentation boundary identification. The automatic identification of the PASs in the operative notes has already determined the sentence boundaries and performed tokenization. To compute lexical scores, the notes are first divided into blocks which consist of several token sequences. Each token sequence has a length of 20, while the blocks consist of 6 such sequences. Each token of a block receives a weight $w_{t,b}$ computed as the frequency of the token in the block. These weights enable the computation of the similarity between blocks of the operative note:

$$sim(b_1, b_2) = \frac{\sum_t w_{t,b_1} \times w_{t,b_2}}{\sqrt{\sum_t w_{t,b_1}^2 \times \sum_t w_{t,b_2}^2}}$$

(3)

The similarity score between blocks informs the identification of segments in each of the operative notes. A segment boundary is identified when the gap in similarity between two blocks exceeds the difference between the average gap and the standard deviation. Clearly, one of the problems of this text segmentation method is that it produced a number of segments that is different than the number of surgical steps.

C. Learning to Identify Surgery Steps in Each Operative Note

Ideally, we would like that each vector representation of a predicate-argument structure that was assigned to a cluster $C_l$ by the K-Means algorithm would represent one of the surgical actions or observations performed during step $i$ of the operation. Taking into account the fact that the steps of a surgical procedure are ordered sequentially, and so are the discourse segments of each operative note, we designed a simple and efficient framework for learning the PASs corresponding to each surgical step, which also enable us to re-assign the discourse segments of each operative note. The by-product of this learning framework is that we generate improved semantic representations of the surgical steps and improved segmentations of the operative notes. The latter allow us to align all operative notes with the methodology detailed in Section V-A and to infer the structure of the operations, with methods detailed in Section V-B. The semantic representations of the surgical steps consist of clusters of predicate argument structures (PASs) pertaining to the same surgical step. We denote as $CL_i$ the set of clusters, where $CL = \{C_{l1}, C_{l2}, \ldots, C_{lk}\}$, with $k$ = the number of steps of a surgical procedure. The PASs were identified automatically from the corpus of operative notes (or reports), which we denote as $R$. If the cardinality of $R$ is $N$, from all the $N$ operative notes in the corpus we have identified a vocabulary $V$ of PASs, such that each distinct PAS has one entry in the vocabulary $V$. To be able to learn the optimal assignment of a PAS from $V$ into one of the clusters from $CL$, we define a likelihood function $L$ that uses as arguments $R$, the corpus, which is observable, as well as a mapping function $Z$ which assigns each PAS to a certain cluster:

$$L(Z, R) = \prod_{r_i \in R} \prod_{PAS_j \in r_i} p(Z_{i,j})$$

(4)

where $Z_{i,j}$ is the result of the assignment of $PAS_j$ from report $r_i$ to one of the clusters from $CL$, which we wanted to learn. The log of the likelihood (log likelihood) is therefore:

$$\log L(Z, R) = \sum_{r_i \in R} \sum_{PAS_j \in r_i} \log p(Z_{i,j})$$

(5)

We used the EM algorithm [23] iteratively to maximize the expected log likelihood of the joint assignment $Z$, given $R$. The E-Step of EM finds the expected value of the log likelihood:

$$E\left[\log L(Z^{(t)}, R)\right] = \sum_{r_i \in R} \sum_{PAS_j \in r_i} \log p(PAS_j | Z_{i,j}^{(t-1)}) + \log p(Z_{i,j})$$

(6)

To compute the expected value of (6), we estimated $p(PAS_j | Z_{i,j})$ as the ratio of how often $PAS_j$ was assigned
and in each note PAS that each cluster contains a representation of multiple PAS order between clusters, we took into account the observation of which step of the operation they model, we first need to clusters produced by both the structure of the operative notes provided by discourse initialized the cluster assignments $Z$ by the clusters.

After inferring the order of the clusters in $Z$, we computed an average segment order a temporal script graph to represent the structure of surgical procedures. By first computing a Multiple Sequence Alignment (MSA) of all the sequences of surgical actions or observations reported in the surgical notes, and then converting the MSA into a graph by taking into consideration also (i) the information about surgical steps recognized in the clusters learned with the EM algorithm as well as (ii) information about the segments of the surgical notes.

A. Multiple Sequence Alignment

We computed a Multiple Sequence Alignment of all the surgical notes of a specific operation to be able to capture all paraphrases of the same surgical action or observation, as expressed by different surgeons throughout the corpus. A MSA algorithm uses as input some sequences $S_1, \ldots, S_n \in \Sigma^*$ over an alphabet $\Sigma$ along with a cost function $C: \Sigma \times \Sigma \rightarrow \mathbb{R}$ for substitutions and a gap cost $C_{GAP} \in \mathbb{R}$ for insertion or deletion. The problem of MSA originated in bioinformatics, where it was used to find corresponding elements in protein sequences or DNA [25]. In bioinformatics, the elements of $\Sigma$ can be nucleotides and a sequence can be a DNA sequence. In our case, $\Sigma$ contains elements from the vocabulary $V$ of PASs identified in the corpus. Given the set of $N$ surgical notes $\mathcal{R}$, an MSA of $\mathcal{R}$ is a matrix $A$ in which the j-th column of $A$ represents the sequence $S_j$ containing PASs identified in note $r_j \in \mathcal{R}$, possibly with some gaps $G$ intersected between the PASs of $r_j$ such that each row of $A$ contains at least one non-gap. If a row of $A$, $A_i$, contains two non-gaps, we consider those PASs aligned; while aligning a non-gap with a gap is interpreted as an insertion or deletion. The cost of a MSA $A$ is provided by:

$$Cost(A) = \sum_{j \in A_r} \sum_{i=1}^{n_a} \sum_{k=1}^{n_r} \cos(E(PAS_j^i), E(PAS_k^r))$$

where $E(PAS_j^i)$ refers to the embedding vector of PAS $j^i$. Because the range of the $\cos$ similarity function for vectors is $[-1, 1]$, we chose a gap cost $C_{GAP}$ of 0. To calculate the lowest cost MSA, we used the polynomial time algorithm for pairwise alignment reported in [26] and recursively aligned these pairwise alignments, considering each alignment as a single sequence whose elements are pairs as in [27].

B. Building Surgical Script Structures

Given that we have produced several forms of semantic information, (e.g. predicate argument structures (PASs), their embeddings, and their alignments) as well as discourse information (e.g. the segments of operative notes), we can induce a temporal script graph to represent the structure of surgical procedures. Such a graph consists of nodes, representing a set of related (or paraphrased) surgical actions or observations, and edges between nodes, which represent a possible temporal evolution of the surgical procedure, as induced from the operative notes. The generation of the graph consists of three steps: Step 1: initialize the nodes; Step 2: decide which nodes should be connected; and Step 3: simplify the graph by merging similar
In Step 1, we started by considering each row of the matrix $A$ (representing the MSA of the corpus $R$), and using only those $PAS_i$ having sufficient frequency in the corpus ($> 50$). In Step 2, considering that the MSA also induces a temporal ordering, we allowed edges between the nodes $N_i$ and $N_j$ if $i < j$. In addition, we used two constraints. The first constraint is for all $PAS_i \in N_j$, there must be a $PAS_i \in N_i$ that precedes $PAS_j$ in at least one note from $R$. Because each $PAS$ was assigned to a cluster by the EM algorithm, we compute for each Node $N_i$ the average cluster number $(ACN)$ by taking into account all $PAS_i$ aligned into $N_i$. For $i < j$, to connect a node $N_i$ to $N_j$, the second constraint we imposed is $0 \leq ACN(N_j) - ACN(N_i) \leq 1$. In Step 3, we merged similar nodes by considering the centroid vector of all the embeddings of $PAS_i$ from the same node. We merged two nodes if the cosine similarity between their centroids was greater than 0.8. We considered merging nodes $N_i$ and $N_j$ only when (i) there was an edge between them in the graph; and (ii) $0 \leq ACN(N_j) - ACN(N_i) \leq 1$. Figure 6 illustrates the resulting structure of the appendectomy procedure documented in our corpus.

In Figure 6, the nodes automatically identified by the procedure detailed in the section are labeled according to (a) the surgical procedure step number illustrated in Figure 2(a); and (b) the group of aligned actions or observations produced by the MSA detailed in Section V-A. For example, for the third step of appendectomies, the temporal script graph has induced three nodes: 3.1, 3.2, and 3.3. Within a node, we list the $PAS$s that were aligned, which often indicate paraphrases (e.g. “identify appendix”, “visualize appendix”, “note appendix inflamed”) as well as temporal sequences (e.g. “place port”, “place trocar camera”). In general, we represent $PAS$s in the node as described in Section III-B, i.e. predicate, first-argument, second-argument, . . . Figure 6 also illustrates the edges between nodes which were induced by the procedure detailed therein.

VI. EXPERIMENTAL RESULTS & ANALYSES

We evaluated our approach for learning the structure of surgical procedures on a corpus of 3,546 operative notes provided by Childrens Medical Center Research Institute at UT Southwestern. These operative notes contain 2,816 appendectomy notes and 730 humerus repair notes authored by a total of 26 different surgeons. As described in section IV-C we discovered surgical structures using unsupervised clustering techniques. This allows us to evaluate the quality of our induced surgical structures by leveraging a number of popular techniques from the field of cluster analysis [28].

In order to evaluate the individual and combined impact of both semantic and discourse information, we considered the quality of clusters (and thus surgical structures) discovered with and without semantic information (Section III) and with and without discourse information (Section IV-C). When evaluating the role of semantic information, we contrasted the performance of our Pred2Vec approach against a baseline approach in which the vector representation of a $PAS$ is simply a linear interpolation of the individual word embedding vectors learned by Word2Vec [18]. When evaluating the role of discourse information, we contrasted the quality of clusters refined from discourse segmentation using the EM algorithm (Section IV-C) against a straightforward $k$-means implementation (Section IV-A). Thus, we measured the quality of learned clusters for four configurations: (1) using syntactic information without discourse processing, (2) using syntactic information with discourse processing, (3) using semantic information without discourse processing, and (4) using semantic information with discourse processing. Table I presents the performance of each of these four approaches according to three measures of cluster quality: (i) the Davies-Bouldin index, (ii) the Dunn Index, and (iii) the Silhouette Coefficient.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Semantic</th>
<th>Discourse</th>
<th>DBI</th>
<th>DI</th>
<th>SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>X</td>
<td>X</td>
<td>1.580</td>
<td>0.607</td>
<td>-0.640</td>
</tr>
<tr>
<td>(2)</td>
<td>X</td>
<td>✓</td>
<td>1.443</td>
<td>0.560</td>
<td>-0.358</td>
</tr>
<tr>
<td>(3)</td>
<td>✓</td>
<td>X</td>
<td>1.067</td>
<td>1.009</td>
<td>-0.736</td>
</tr>
<tr>
<td>(4)</td>
<td>✓</td>
<td>✓</td>
<td>0.573</td>
<td>1.451</td>
<td>-0.612</td>
</tr>
</tbody>
</table>

TABLE I: Quality of Clusters Learned for four configurations, where DBI refers to the Davies-Bouldin Index, DI refers to the Dunn Index, and SC refers to the Silhouette Coefficient.
most similar neighbor; (ii) the Dunn index [30], the ratio between the minimal inter-cluster distance and maximal intra-cluster distance; and (iii) the Silhouette coefficient, which contrasts the average distance between PASs assigned to the same cluster against the average distance to PASs assigned to other clusters [31]. The “best” surgical structure is that with the smallest Davies-Bouldin index, the highest Dunn index, and the highest Silhouette coefficient.

Clearly, the best performance was achieved using both semantic and discourse processing (approach (4)). Without access to discourse information, the Dunn and Davies-Bouldin indices drop by 30.5% and 46.3%, respectively. This highlights the impact of not only discourse information, but also of the power of our EM-based approach. Interestingly, using discourse information without semantic information achieves the third best Dunn and Davies-Bouldin indices, but achieves the best Silhouette coefficient (-0.358). This suggests that while Pred2Vec provides significantly improved cluster quality, further research is needed to determine the optimal number of clusters when Pred2Vec is used because Pred2Vec reduces the sparsity in the data. Unsurprisingly, without access to semantic or discourse information, approach (4) performs the worst, illustrating the importance of semantic and discourse information for discovering the structure of surgical procedures.

VII. CONCLUSION

In this paper, we presented a novel method of learning the structure of a type of surgical procedure when a corpus of operative notes is available. We have shown how this structure can be learned when (a) semantic information is available in the form of predicate argument structures that identify knowledge about surgical actions and observations; (b) neural learning of multi-dimensional embeddings of predicate argument structures (Pred2Vec) is tried; and (c) discourse information indicating the segments of an operative note is provided. Moreover, we have presented a novel way of semantically representing the surgical steps and provided a framework of learning the most likely representation of the surgical steps and their segmentation into the operative notes. The experimental evaluations on a large corpus documenting two types of surgical procedures have yielded promising results.

REFERENCES


