Automated Activity Recognition in Clinical Documents

C. Thorne, M. Montali, D. Calvanese
Free University of Bozen-Bolzano, Bolzano, Italy
{thorne.montali,calvanese}@inf.unibz.it

E. Cardillo, C. Eccher
Fondazione Bruno Kessler, Povo, Italy
{cleccher,cardillo}@fbk.edu

A. Problem

Clinical guidelines are documents describing the state-of-the-art on clinical therapies [3]; building a careflow from a clinical guideline is time consuming and error prone.

Question (1): Can NLP be used to automatically extract careflow fragments?

Question (2): Should the techniques leverage on guideline syntax or semantics?

B. Activity Recognition

(1) Let $\vec{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$ be n input content words or entities of a sentence.

(2) Let $\vec{c} = (c_1, \ldots, c_n)^T$ denote n entity types drawn from the set:

$$\{\text{activity}, \text{resource}, \text{actor}, \text{other}\}.$$

(3) In the clinical entity recognition task [1] we want to find the entities s.t.

$$\vec{c}^* = \arg\max_{\vec{c}} \mu(\rho(\vec{\alpha}, \vec{c}))$$

- $\mu(\cdot)$ denotes a classifier;
- $\rho(\cdot, \cdot)$ is a feature extraction function, that maps $\vec{c}$ and $\vec{\alpha}$ into a high-dimensional space of features.

D. Features & Entities

(1) Types harvested from entities by mapping MetaMap and UMLS [2, 7] concepts to entity types:

activity	actor	resource	other
laboratory procedure	professional society	manufactured object	qualitative concept

(2) Semantic features for each entity extracted also with MetaMap and the UMLS Metathesaurus:

- compute the raw frequency *freq* of entity type c;
- compute the (repeated) entity types *labs* of the entity's noun phrase (**NP**);
- compute the rel. frequency *lf* of entity type c:

$$lf = \frac{||labs \Cap \{c\}||}{||labs||};$$

- compute the overlap *hd* of *labs* and the types *labsh* of its **NP**'s head noun, and the overlap *ls* of *labs* and entity subtypes *sub(c)* (in the UMLS taxonomy):

$$hd = \frac{||labs \Cap labsh||}{||labs|| + ||labsh||}, \quad ls = \frac{||labs \Cap sub(c)||}{||labs|| + ||sub(c)||}$$

($||.||$ and $\Cap$: bag cardinality and intersection, resp.).

(3) Syntactic features for each entity extracted with the Stanford parser [6]:

- compute position *pos* in sentence, subordination *sub* and nesting level *nest*.

feature F	description	value f
nest	nesting level in tree	$n \in \mathbb{N}$
pos	position w.r.t. verb	subject, object
sub	occurs in clause?	yes, no
freq	freq. of label in corpus	$n \in \mathbb{N}$
lf	rel. freq. of type	$r \in [0, 1]$
hd	head/entity overlap	$r \in [0, 1]$
ls	type/entity overlap	$r \in [0, 1]$
type	entity type	activity, actor, resource, other

(7 predictors, and 1 predicted feature: type)

6th Int. Joint Conference on Natural Language Processing Nagoya, Japan, October 14-18, 2013

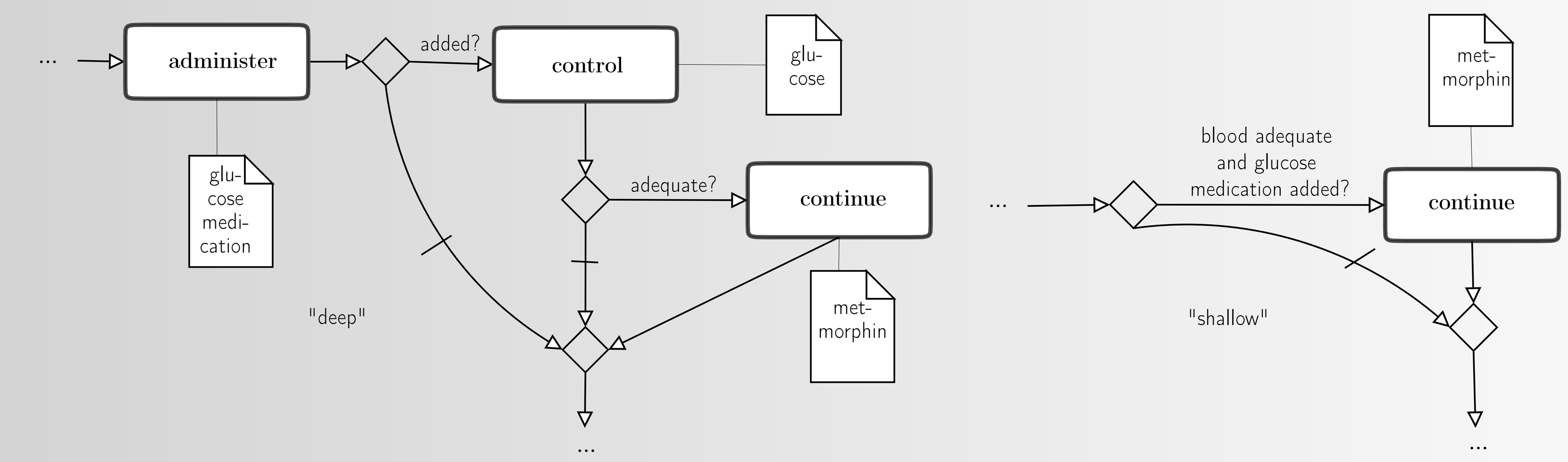
C. Biomedical Thesauri & Careflow Fragments

(1) MetaMap UMLS (automated) annotations of a type 2 diabetes guideline recommendation [4]; boxes surround entities, annotations are MetaMap's:

`Continue` (⇓ regulatory activity) `with` `metformin` (⇓ pharmacological substance) `if` `blood glucose control` (⇑ clinical attribute; ⇓ laboratory procedure) `remains`

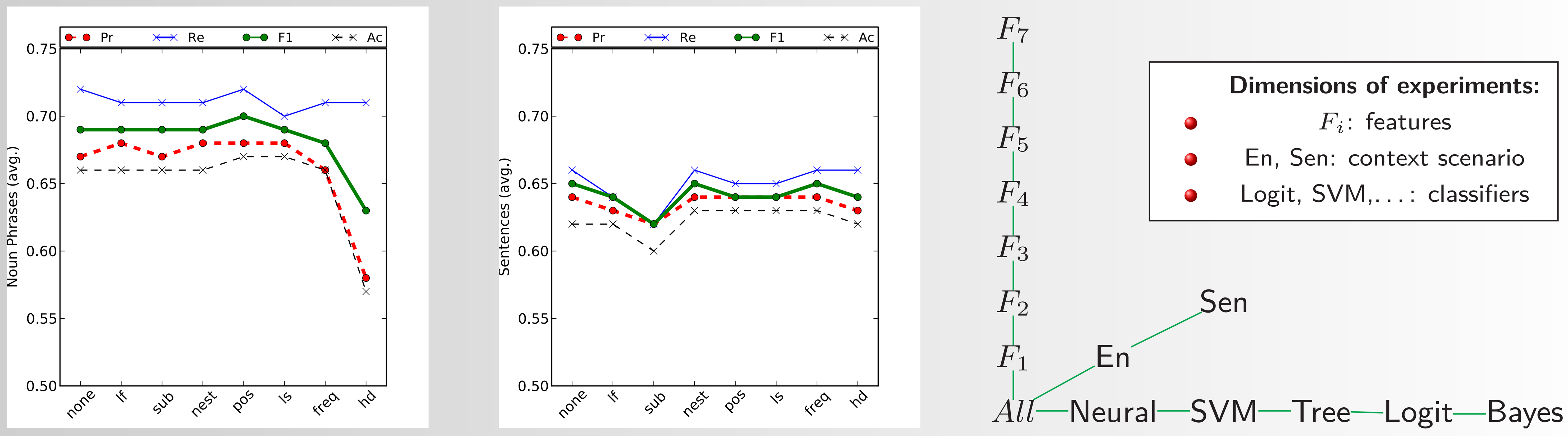
`inadequate` (⇓ qualitative concept) `and another` `oral glucose-lowering medication` (⇑ clinical attribute; ⇓ therapeutic procedure) `is` `added` (⇓ functional concept) .

(2) Candidate careflow fragments (represented in BPMN): to the left, the intended "deep" careflow, to the right a "shallow" one:



E. Experiments & Results

Goal: To extract the intended "deep" fragment we need to recognize, e.g., "blood glucose control" as an activity (therapeutic procedure) instead of a resource (clinical attribute), and understand if this choice depends on syntax or semantics:



corpus	size (words)	domain	rel. freq.
Brown	1,391,708	news	0.16
Friederich	3,824	processes	0.17
SemRep	13,948	clinical	0.18
diabetes 2	7,109	clinical	0.16
eating disorder	5,078	clinical	0.17
schizophrenia	5,367	clinical	0.18

- remove feature F_i from predictors $\{F_1, \ldots, F_7\}$;
- consider sentence context (Sen scenario) or not (En scenario);
- evaluate the classifiers via a 10-fold cross-validation over the Gold-standard UMLS-annotated SemRep clinical corpus [5], and measure average classifier precision (Pr), recall (Re), F1-measure and accuracy (Ac) per each (F, S) feature-scenario pair.

(1) Performance drops if semantic features (*ls*, *freq*, *hd*) are disregarded and we ignore sentence context.
(2) When we consider sentence context, syntax is more determinant (*sub*), but performance drops overall.
(3) Corpus analysis shows no significant difference in syntax between clinical and non clinical text.

Complete results: `http://www.inf.unibz.it/~cathorne/vericlig/ijcnlp2013-exp.pdf`

F. Conclusions & Further Work

(1) Conducted a preliminary experiment on automatic clinical activity recognition using MetaMap.
(2) Experimented on the SemRep gold standard UMLS-annotated corpus.
(3) Experiments suggest that the semantic environment of an entity is more useful for this task.
(4) Corpus analysis seems to confirm this observation.
(5) In the future, we plan to consider more powerful classification models for NLP.
(6) We also plan to consider larger UMLS-annotated corpora.

G. References

[1] Asma Ben Abacha and Pierre Zweigenbaum. Medical entity recognition: A comparison of semantic and statistical methods. In *Proceedings of the BioNLP 2011 Workshop*, 2011.
[2] Alan R. Aronson and François-Michel Lang. And overview of MetaMap: Historical perspective and recent advances. *Journal of the American Medical Informatics Association*, 17(3):229–236, 2010.
[3] A. Bottrighi, F. Chesani, M. Montali, and P. Terenziani. Conformance checking of executed clinical guidelines in presence of basic medical knowledge. In *Proceedings of the 2011 Business Process Management Workshop*, 2012.
[4] National Institute for Health and Clinical Excellence (UK). *Type 2 Diabetes*. 2008. Available from `http://www.nice.org.uk/nicemedia/pdf/CG87NICEGuideline.pdf`.
[5] Halil Kilicoglu, Graciela Rosenblat, Marcelo Fiszman, and Thomas C. Rindfleisch. Constructing a semantic predication gold standard from the biomedical literature. *BMC Bioinformatics*, 12(486), 2011.
[6] Dan Klein and Christopher D. Manning. Accurate unlexicalized parsing. In *Proceedings of the 41st Meeting of the Association for Computational Linguistics ACL 2003*, 2003.
[7] US National Library of Medicine (NLM - NIH). *UMLS© Reference Manual*. 2009. Available from: `http://www.ncbi.nlm.nih.gov/books/NBK9676/`.

Acknowledgments

This work was supported by the VeriCliG project, funded by a Free University of Bozen-Bolzano Foundation grant, see: `http://www.inf.unibz.it/~cathorne/vericlig`.