VeriCliG: Extraction and Verification of Clinical Guidelines via Syntactic and Semantic Annotation

Camilo Thorne[1], Marco Montali[1], Diego Calvanese[1], Claudio Eccher[2] and Elena Cardillo[2]

[1] KRDB Research Centre for Knowledge and Data
39100, Bolzano, Italy
{cthorne,mmontali,calvanese}@inf.unibz.it

[2] Fondazione Bruno Kessler
Trento, Italy
{eccher,cardillo}@fbk.eu

http://www.inf.unibz.it/~cathorne/vericlig

ClinText 2013, Potsdam, 19/3/2013

Outline

Problem - CIG Extraction

- We want to extract process representations from clinical guidelines

  1. basic tool in hospitals and clinics [Got12]
  2. describe state-of-the art therapies
  3. implemented in workflow and decision support systems
  4. processed manually
  5. require loads of costly expert knowledge

Problem - CIG Extraction

- We want to extract process representations from clinical guidelines

  1. basic tool in hospitals and clinics [Got12]
  2. describe state-of-the art therapies
  3. implemented in workflow and decision support systems
  4. processed manually
  5. require loads of costly expert knowledge

- Quality assurance can be enhanced via verification [BCMT12]

Problem - CIG Extraction

- We want to extract process representations from clinical guidelines
  1. basic tool in hospitals and clinics [Got12]
  2. describe state-of-the art therapies
  3. implemented in workflow and decision support systems
  4. processed manually
  5. require loads of costly expert knowledge
- Quality assurance can be enhanced via verification [BCMT12]
- We need computer interpretable guidelines (CIGs) representing therapies as
  1. activities (e.g., surgery)
  2. resources (e.g., a drug)
  3. actors (e.g., patients, doctors, nurses)
  4. control flows (e.g., conditional, sequential, parallel)
  5. several formalisms: BPMN, Asbru, Glare, Proforma (PROTOCURE EU Project)

Problem - CIG Extraction

- We want to extract process representations from clinical guidelines

  1. basic tool in hospitals and clinics [Got12]
  2. describe state-of-the art therapies
  3. implemented in workflow and decision support systems
  4. processed manually
  5. require loads of costly expert knowledge

- Quality assurance can be enhanced via verification [BCMT12]

- We need computer interpretable guidelines (CIGs) representing therapies as

  1. activities (e.g., surgery)
  2. resources (e.g., a drug)
  3. actors (e.g., patients, doctors, nurses)
  4. control flows (e.g., conditional, sequential, parallel)
  5. several formalisms: BPMN, Asbru, Glare, Proforma (PROTOCURE EU Project)

QUE: Use (clinical) NLP annotation to extract CIGs?

Process Extraction - Diabetes 2 NCCN

> 1.5.1.2 Emphasise advice on healthy balanced eating that is applicable to the general population when providing advice to people with type 2 diabetes.
>
> 1.5.1.3 Continue with metformin if blood glucose control remains inadequate and another oral glucose-lowering medication is added.

$\Rightarrow$ Clinical guidelines describe therapies or treatment processes

Process Extraction - Diabetes 2 NCCN

1.5.1.2 Emphasise advice on healthy balanced eating that is applicable to the general population when providing advice to people with type 2 diabetes.

1.5.1.3 Continue with metformin if blood glucose control remains inadequate and another oral glucose-lowering medication is added.

⇒ Clinical guidelines describe therapies or treatment processes

▷ Data scarse to train state-of-the-art supervised annotators !

Process Extraction - Diabetes 2 NCCN

1.5.1.2 Emphasise advice on healthy balanced eating that is applicable to the general population when providing advice to people with type 2 diabetes.

1.5.1.3 Continue with metformin if blood glucose control remains inadequate and another oral glucose-lowering medication is added.

⇒ Clinical guidelines describe therapies or treatment processes

▷ Data scarse to train state-of-the-art supervised annotators !

⇒ Apply business process extraction techniques [FMP11]

Process Extraction - Diabetes 2 NCCN

1.5.1.2 Emphasise advice on healthy balanced eating that is applicable to the general population when providing advice to people with type 2 diabetes.

1.5.1.3 Continue with metformin if blood glucose control remains inadequate and another oral glucose-lowering medication is added.

⇒ Clinical guidelines describe therapies or treatment processes

▷ Data scarse to train state-of-the-art supervised annotators !

⇒ Apply business process extraction techniques [FMP11]

⇒ Distinguish

1. control-flow words (= discourse relations)
2. activity, resources and actor words (= content words)

NLP Annotation - Combining Resources

$\boxed{\textit{Continue}}$ $\boxed{\textit{with}\ \texttt{metformin}}$ *if* $\boxed{\texttt{blood glucose control}}$ $\boxed{\textit{remains}}$

Continue ⇓ reg. activity — with metformin ⇓ pharm. substance — blood glucose control ⇓ laboratory procedure — remains ⇓ ql. concept

$\boxed{\texttt{inadequate}}$ *and* $\boxed{\textit{another}\ \texttt{oral glucose-lowering medication}}$ *is* $\boxed{\textit{added}}$.

inadequate ⇓ ql. concept — another oral glucose-lowering medication ⇓ therapeutic procedure — added ⇓ fc. concept

```
(S (VP (VB Continue) (PP (IN with) (NP (NN metformin)))
     (SBAR (IN if)
        (S (S (NP (NN blood) (NN glucose) (NN control))
              (VP (VBZ remains)(ADJP (JJ inadequate))))
           (CC and)
           (S (NP (DT another) (JJ oral) (JJ glucose-lowering)
              (NN medication))
              (VP (VBZ is) (VP (VBN added))))))))
```

Legend
activity evoking
resource evoking
control evoking

NLP Annotation - Combining Resources

$\boxed{\textit{Continue}}$ $\boxed{\underline{\textit{with}}\ \texttt{metformin}}$ $\underline{\textit{if}}$ $\boxed{\texttt{blood glucose control}}$ $\boxed{\textit{remains}}$

Continue ⇓ reg. activity; with metformin ⇓ pharm. substance; blood glucose control ⇓ laboratory procedure; remains ⇓ ql. concept

$\boxed{\texttt{inadequate}}$ $\underline{\textit{and}}$ $\boxed{\underline{\textit{another}}\ \texttt{oral glucose-lowering medication}}$ $\underline{\textit{is}}$ $\boxed{\textit{added}}$.

inadequate ⇓ ql. concept; another oral glucose-lowering medication ⇓ therapeutic procedure; added ⇓ fc. concept

```
(S (VP (VB Continue) (PP (IN with) (NP (NN metformin)))
     (SBAR (IN if)
       (S (S (NP (NN blood) (NN glucose) (NN control))
             (VP (VBZ remains)(ADJP (JJ inadequate))))
          (CC and)
          (S (NP (DT another) (JJ oral) (JJ glucose-lowering)
             (NN medication))
             (VP (VBZ is) (VP (VBN added))))))))
```

Legend
activity evoking
resource evoking
control evoking

⇒ MetaMap & UMLS [AL10] finds activities, resources and actors

NLP Annotation - Combining Resources

Continue	*with* metformin	*if*	blood glucose control	*remains*
⇓	⇓		⇓	⇓
reg. activity	pharm. substance		laboratory procedure	ql. concept

inadequate	*and*	*another* oral glucose-lowering medication	*is*	*added* .
⇓		⇓		⇓
ql. concept		therapeutic procedure		fc. concept

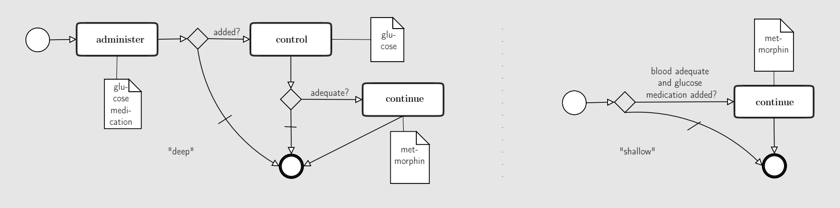
```
(S (VP (VB Continue) (PP (IN with) (NP (NN metformin)))
     (SBAR (IN if)
       (S (S (NP (NN blood) (NN glucose) (NN control))
            (VP (VBZ remains)(ADJP (JJ inadequate))))
          (CC and)
          (S (NP (DT another) (JJ oral) (JJ glucose-lowering)
            (NN medication))
            (VP (VBZ is) (VP (VBN added))))))))
```

Legend
activity evoking
resource evoking
control evoking

⇒ MetaMap & UMLS [AL10] finds activities, resources and actors

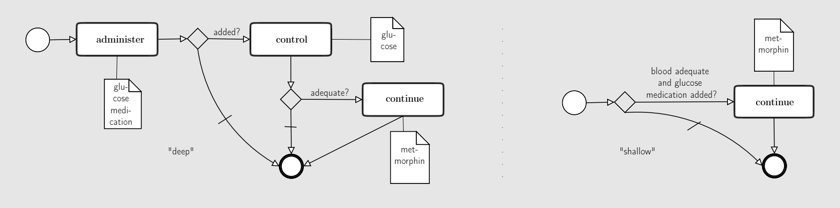
⇒ Stanford parser [dMMM06] finds discourse relations

CIG & Process Granularity



⇒ Two possible CIGs in BPMN notation due to English ambiguity

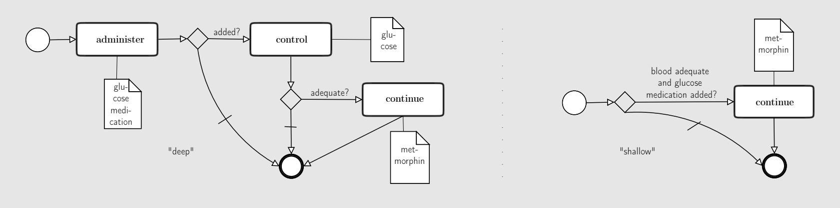
CIG & Process Granularity



⇒ Two possible CIGs in BPMN notation due to English ambiguity

▷ Limitations of rule-based CIG extraction

CIG & Process Granularity

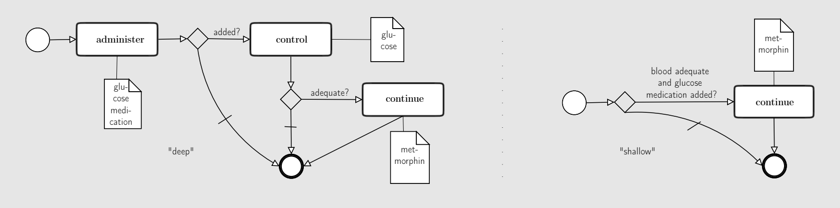


⇒ Two possible CIGs in BPMN notation due to English ambiguity

▷ Limitations of rule-based CIG extraction

1. no clear clinical meaning for control-flow evoking words !
2. no clear evaluation methodology !

CIG & Process Granularity



⇒ Two possible CIGs in BPMN notation due to English ambiguity

▷ Limitations of rule-based CIG extraction

1. no clear clinical meaning for control-flow evoking words !
2. no clear evaluation methodology !

⇒ Surface-level clinical language peculiarities deceptive

Linguistic Patterns - PEWs and NEGs

⇒ We checked control-flow evoking words (PEWs)

1. conjunctions and prepositions **IN**s (subordinating prepositions, e.g., "if") **CC**s (coordinating conjunctions, e.g., "and", "or") **CC**s (subordinating conjunctions, e.g., "then")
2. adverbs: **RB**s (base adverbs, e.g., "after") **RBR**s (comparative adverbs, e.g., "later") **RBT**s (superlative adverbs, e.g., "latest") **RN**s (nominalized adverbs, e.g.,) **RP**s (adverbial particles, e.g.,)

⇒ We also checked negative rules (NEGs)

1. "not", of category ***** (i.e., negation); "nobody", "none" and "nothing", of category **PN**, the negative determiner "no" of category **AT**
2. negated modal verbs of category **MD*** (e.g., "cannot", "will not").

Linguistic Patterns - PEWs and NEGs

⇒ We checked control-flow evoking words (PEWs)

1. conjunctions and prepositions **IN**s (subordinating prepositions, e.g., "if") **CC**s (coordinating conjunctions, e.g., "and", "or") **CC**s (subordinating conjunctions, e.g., "then")
2. adverbs: **RB**s (base adverbs, e.g., "after") **RBR**s (comparative adverbs, e.g., "later") **RBT**s (superlative adverbs, e.g., "latest") **RN**s (nominalized adverbs, e.g.,) **RP**s (adverbial particles, e.g.,)

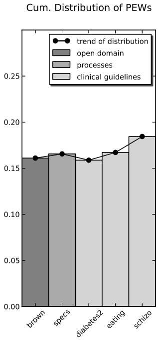
⇒ We also checked negative rules (NEGs)

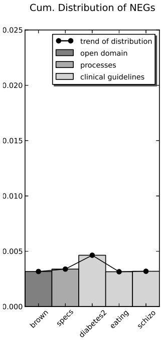
1. "not", of category * (i.e., negation); "nobody", "none" and "nothing", of category **PN**, the negative determiner "no" of category **AT**
2. negated modal verbs of category **MD*** (e.g., "cannot", "will not").

QUE: Distribution uniform across domains or correlated to clinical domain?

PEWs and NEGs - Uniform & Uncorrelated

Distrib.	χ^2-ind.	p (< 0.001 sig.)	df.	t-one way	μ_0	p (< 0.01 sig.)	df.
PEWs	9.39	0.009	2	1.02	0.20	0.36	4
NEGs	1.96	0.375	2	1.02	0.03	0.36	4





(Corpus)	(# Words)	(Domain)
Brown	1,391708	Open
Business	3,824	Business
Diabetes 2 guid.	7,109	Clinical
Eating dis. guid.	5,078	Clinical
Schizophr. guid.	5,367	Clinical

Current Work Plan

- Refine control-flow extraction methodologies
- Develop a methodology to evaluate process-mining rules vis-à-vis supervised techniques
- Use existing annotated corpora (e.g., SemRep) to understand how to disambiguate MetaMap/UMLS
  1. activity annotations
  2. resource annotations
  3. actor annotations
- Design ways to merge/integrate/link annotated datasets to experiment with supervised annotators

Thank you :-)

References

Alan R. Aronson and François-Michel Lang.
And overview of MetaMap: Historical perspective and recent advances.
J. of the American Medical Informatics Association, 17(3):229–236, 2010.

A. Bottrighi, F. Chesani, M. Montali, and P. Terenziani.
Conformance checking of executed clinical guidelines in presence of basic medical knowledge.
In *Proc. of the 2011 Business Process Management Work.*, 2012.

Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning.
Generating typed dependency parses from phrase structure parses.
In *Proc. of the 5th Int. Conf. on Language Resources and Evaluation (LREC 20006)*, 2006.

Fabian Friederich, Jan Mendling, and Frank Puhlmann.
Process model generation from natural language text.
In *Proc. of the 23rd Int. Conf. on Advanced Information Systems Engineering (CAiSE 2011)*, 2011.

Gregory Goth.
Analyzing medical data.
Comm. of the ACM, 55(6):13–15, 2012.