



Constraint-driven Evaluation in UIMA Ruta

Estimating the quality of arbitrary models on unseen documents.

Andreas Wittek, Martin Toepfer, Georg Fette, Peter Klügl, Frank Puppe

UIMA@GSCL2013









Good reasons for Constrained-driven Evaluation (CDE)

Choose one of several 3rd-party components, Debug rule-based systems, Decision support for make-or-buy



Reason 1: Development Support

• typical scenario in NLP applications:



- Constrained-driven evaluation (CDE) provides:
 - Ranking of unlabeled documents according to the expected quality of a component
 - F1 score prediction on unlabeled data sets

UNI WÜ

Reason 2: Component Selection

- typical scenario in NLP applications:
 - several 3rd-party components
 (e.g., CRF & SVM & Rule-based NER model trained on CoNLL)
 - Unlabeled data-set
 (e.g., a custom newspaper corpus)
 - Which component is the best (relatively)?



- Constrained-driven evaluation (CDE):
 - Ranking of components according to performance estimation of arbitrary models using constraints





Reason 3: make-or-buy

- typical scenario in NLP applications:
 - 3rd-party component
 (e.g., CRF NER model trained on CoNLL)
 - Unlabeled data-set
 (e.g., a custom newspaper corpus)
 - Are the results good enough for my application? Or do I have to develop my own system? (absolute quality estimation)

F1?

Dataset (unlabeled)

- Constrained-driven evaluation (CDE):
 - Automated performance estimation of arbitrary models with formalized expectations





- **F1 Prediction** task: Estimate F1-score of arbitrary models on unlabeled data
 - Reason 2 (component selection): relative performance
 - Reason 3 (make-or-buy)
 absolute performance
 - [Reason 1 (development support)]
- Ranking task: Rank documents wrt. F1
 - Reason 1 (development support)





- *Basic* operations for rapid component development:
 - 1. Gather constraints
 - **Rule Constraints**: formalize background-knowledge of domain experts
 - Annotation distribution (AD) Constraints: Collect statistics (using large data-bases)
 - 2. Constraint-driven Evaluation
 - Apply model on unlabeled data-set
 - Apply constraints on unlabeled data-set
 - Compute aggregate **CDE score** for each document



CDE Score



reflects how well the output of the component complies with the constraints:

$$\text{CDE} = \frac{1}{\alpha} \sum_{i=1}^{n} w_i C_i$$

where w_i is a weight, α is a normalizing constant, and $C_i \in [0,1]$ is either

- a rule constraint score:
 #rule matched / #rule has tried to match
- or an annotation distribution (AD) constraint score: cosine similarity between expected and observed frequency





- Formalize expectations of domain experts as UIMA Ruta rules
- Example:

Arthur M. Keller and Julie Basu. A predicate-based caching scheme for client-server database architectures. In Proceedings of PDIS-94, 1994.

Buntine, W. (1994). Operations for learning with graphical models. Journal of Artificial Intelligence Research, 2, 159-225.

Author (Title | Year); Author {-CONTAINS(NUM)}; Title {CONTAINS(W,2,200)};



Rule Constraints

Title {CONTAINS(W,2,200)};



Constraint satisfied?

n —

n

n

V

n

V

•

A. Said a	Id W.A. Pearlman, "A new, fast, and efficient image codec based on set
partitioni	ig in hierarchical trees," IEEE Transactions on Ciruits and Systems for
Video lechi	10logy, vol. 6, no. 3, pp. 243249, June 1996.
M. Yang a	Id L. M. NI, "Incremental design of scalable interconnection networks
using basi	building blocks," Symposium of Parallel and Distributed Processing,
pp. 252	259, Oct. 1995.
L. Hernqu	ist, "Vectorization of Tree Traversals," Journal of Computational Physics,
vol. 87,	1990.
J. Hipp,	I. Lober, "Plastering: All-Hexahedral Mesh Generation Through Connectivity
Resolution	" Proc. 3 rd Int. Meshing Roundtable, 1994.
P. Sarkar	, "An Iterative Algorithm for Optimal Style-Conscious Field Classification,"
Proc. 16 ⁻	ch Int'l Conf. Pattern Recognition, pp. 243-246, Aug. 2002.
C. Rosene	, "Incremental dependence analysis," Ph.D. thesis, Rice University,
March 19	20.
H. J. Lan	au and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis,
and uncert	ainty-III: The dimension of the space of essentially time- and band-limited
signals,"	Bell Syst. Tech. J., vol. 41, no. 4, pp. 1295–1336, Jul. 1962.
7 6 1	
J. Carter	and M. Wegman, "Universal classes of hash functions," J. of Computer
and System	Sciences, vol. 18, no. 2, pp. 143-154, 1979.
D. E. Whi	They, "Force feedback control of manipulator fine motions," ASME J.
Dyn. Syst.	Meas. Contr., Vol. 99, no. 2, pp. 91-97, 1977.
C. Norris	and L. Pollock, A scheduler-sensitive global register allocator,
In Proc. (of Supercomputing, pages 804-813, 1993.
m. J. Hari	'Old and M. L. Sotta, An incremental approach to unit testing during
maintenance	e, Proceedings of the international conference on Software Maintenance,
pp. 362-	10/, UCTODER 1988.
J. H. M.	J. Dally and K. Kelser, Detecting obstacles in range imagery, Proc.
OT TARPAT	Image Understanding workshop, pp. 87-97, 1987.
T Vlan	white 0 E Densky "Dendie Courses of Dend Turnels at 000 and 4000 MU-

Constrained-driven

Evaluation

13:

 \rightarrow score: 4/10 (4 out of 10 titles have \geq 2 words)

Some errors are detected correctly by this Rule Constraint; Some errors are not detected

AD constraints may help in such cases (see next slide)

title	





- Statistical assumptions gathered from databases, e.g., Bibsonomy dump
- Example: in DB:
 "VHDL": Author 0.001, Title 0.4, ...
 Word Label Freq.
- in Document: "VHDL"/Author: 1/1 $C_i = \cos\left(\binom{1.0}{10}, \binom{0.001}{10}\right) = 0.51$

R. Lipsett, C. Schaefer, C. Ussery, VHDL: Hardware Descri	ion and Design, Kluwer
Academic Publishers, 1989.	
T. Imielinski, S. Viswanathan, Adaptive Wire less Informa	on Systems, Proc.
of SIGDBS Conf., Tokyo, October, 1994.	
K. Chandy and J. Misra, Parallel Program Design: A Founda	on, Addison-Wesley,
1988	
1900.	itle 📃, 🛛 author 📃
LUNA @CCCL 2012, Constrained driver Evolution	

UN WÜ

Constrained-driven Evaluation



• Extended workflow:

Gather constraints

• ...

- Constraint Development
- Create a small dev.-set
- Test constraints; specify weights
- Constraint-driven Evaluation

• ...





- measures to compare CDE score against ground-truth F1-score/ranking on a set of labeled documents:
 - Spearman (ranking)
 - Pearson (linear dependency)





Experimental Study

Segmentation of References of Scientific Papers





- D_{ruta} : used to <u>develop rules for reference segmentation</u> (219 references in 8 docs)
- *D*_{dev} : labeled by the constructed rules; used for <u>developing</u> <u>the constraints</u> (192 references in 8 docs)
- D_{test} : labeled by the constructed rules; used to <u>evaluate the</u> <u>constraints</u> (155 references in 7 documents)
- D_{crf} : labeled by training and applying <u>CRFs</u> (5-fold-cross) (ruta+dev+test: 566 references)
- *D*_{gen} : <u>different source</u>; unknown style guides; labeled by the constructed rules (452 references in 28 documents)





- Three constraint sets:
 - C_{ruta}: 15 <u>Rule constraints (Author, Title, Date); weight of each</u> constraint is 1
 - C_{ruta+bib}:
 C_{ruta} + AD constraints (entity distribution of words extracted from Bibsonomy); weight of each constraint is set to 1
 - 3. $C_{ruta+5*bib}$: weight of each <u>AD</u> constraint set to <u>5</u>





Table 3. Spearman's ρ and Pearson's r given for the predicted CDE score (for each document) compared to the actual F_1 score.

	C_{rut}	a	C_{ruta}	+bib	$C_{ruta+5xbib}$		
Dataset	ho	r	ho	r	ho	r	
D_{dev}	0.8708	0.9306	0.9271	0.9405	0.8051	0.6646	
D_{test}	0.9615	0.9478	0.9266	0.8754	0.8154	0.6758	
D_{crf}	0.6793	0.7881	0.7429	0.8011	0.7117	0.7617	
D_{gen}	0.7089	0.8002	0.7724	0.8811	0.8150	0.9504	

- Strong correlation ($\rho \ge 0.6, r \ge 0.6$ for all D_i, C_j):
 - constraint rules really estimate the performance of the models on the new unlabeled data
- Even with different data (new styles):
 - use AD constraints in this case





UIMA Ruta CDE Plugin

Constrained-driven Evaluation in the UIMA Ruta Workbench

UNI UIMA Ruta CDE Plug-in WÜ - Constraint Specification -

🕒 UIM	A Ruta CDE - CDE2/data/processed/fold3/1471-2105-12	-43.pdfbox.txt.xmi (CDE2/descriptor/typ	es/Bibtex	.xml) - Eclipse Platfo	orm		l	- 0 X
<u>F</u> ile <u>E</u>	dit <u>N</u> avigate Se <u>a</u> rch <u>P</u> roject <u>R</u> un <u>W</u> indow <u>H</u> el	p						
1	• 🖬 🖻 💁 • 😥 🛷 • 📎 🖢 •	월 ▼ 🌾 🔶 ▼ 🔶 ▼				🖹 🎲 UIMA Ruta C.	🥜 UIMA R	luta ि Resou »
·····	J05-3001.txt.xmi 🗍 A97-1010.txt.xmi 🗍 147	1-2105-12-43.pdfb 🛛 🔭	- 8	品 CDE Documen	ts 🖾 🔡 (Outline 📎 Annotatio	n Browser View	
l	evolutionary history of the first	Documents: work\workspace-ruta\CDE2\data\processed\fold3						
	biosynthesis. Bioessays 1993, 15:	157-164.		Test Data: D:\work\workspace-ruta\CDE2\data\com			a∖complete	
	Andrade , (Ed.): Bioinformatics a	nd Genomes: Current Per	spec	Type System:	ı\CDE2\descr	riptor\uima\ruta\exam	ple\IETypeSyst	em.xml 🛅
	Heidelberg, Germany: Horizon Scie	entific Press; 2003.		mse=0.0 spearn	mans=0.9615 p	pearsons=0.9478 cosin	e=1.0	
	Pruitt KD, Tatusova T, Maglott DR	NCBI Reference Sequen	ce r	Document		CDE	F1	
1-				kdml12.pdft	oox.txt.xmi	0.9846	0.9807	
	CDE Constraint 🕴 🔨 Selection View 🥜 TextRule	er 🦻 Annotation Tes 🎲 Ruta Query		 J05-3001.txt. 	xmi	0.9867	0.9857	
				A97-1010.txt	.xmi	0.9911	0.995	
				• 1471-2105-1	2-43.pdfbox.t	0.9914	0.9871	
	Constraint	Weight		J05-1003.txt.	xmi	1.0	1.0	
	Reference{OR(CONTAINS(Author), CONTAINS(Editor	1		C04-1024.txt		1.0	1.0	
	Reference{CONTAINS(Title)};	1		A88-1009.txt	təxmi	1.0	1.0	
	Reference{CONTAINS(Date)};	1						
	Author{STARTSWITH(Reference)};	1	60	CDE Result 🔀				
	Author (Date Title);	1						
	Title{-CONTAINS(PublisherInd));	1		Constraint		Result		^
	Title ANY{-PARTOF(EditionMarker)};	1		Title ANY{-PAR	TOF(Editio	1.0		
	Date{-PARTOF(NumInW)};	1		Date{-PARTOF(I	NumInW)};	1.0		
	Date ANY{OR(-IS(SPECIAL), IS(Quote))};	1		Date ANY{OR(-I	S(SPECIAL)	1.0		
	Title{OR(STARTSWITH(W), STARTSWITH(Quote))};	1		Title{OR(START	SWITH(W),	0.9565217391304348		
	Author{-CONTAINS(NUM)};	1		Author{-CONTA	AINS(NUM)};	1.0		
	Author{CONTAINS(CW,1,100)};	1		Author{CONTA	INS(CW,1,1	1.0		E
	Author{CONTAINS(W,2,100)};	1		Author{CONTA	INS(W,2,10	0.9787234042553191		
	Title{CONTAINS(W,2, 200)};	1		Title{CONTAINS	S(W,2, 200)};	0.9782608695652174		
	Date{CONTAINS(NUM)};	1		Date{CONTAINS	5(NUM)};	1.0		-

UNI UIMA Ruta CDE Plug-in WÜ - Constraint Development & Results -



Edit Navigate Search Project Run Window He		on, cypes, bibiex.	anny compact lat				
• 🗟 🙆 • 🗀 🖉 • 😒 ½ •	यी ▼ №		-		E 🕼 UIMA Ruta C		ita 🕒 Kes
J05-3001.txt.xmi A97-1010.txt.xmi	1-2105-12-43.pdfb 🛛 🔭		CDE Documer	nts 🛿 📴	Outline S Annotation	Browser View	
Daviuson JN, Chen KC, Jamison KS,	Plusinanno LA, Kern C	D. me					
evolutionary history of the first	three enzymes in py	/rimidir	Documents:	work\works	pace-ruta\CDE2\data\pr	ocessed/fold3	
blosynthesis. Bloessays 1993, 15:	157-164.		Test Data: D:\work\workspace-ruta\CDE2\data\			\complete	
Andrade (Ed.): Bioinformatics a	und Genomes: Current	Perspec	Type System:	\CDE2\des	rintor\uima\ruta\exam	le\IFTvneSvste	m xml
Heidelberg, Germany: Horizon Scie	entific Press: 2003.	r cr spec	Type System.	I/CDL2/UES	chptor (unna (ruta (examp	ne (ie i ypesyste	
incruciper g, der manyr nor izon sere			mse=0.0 spear	mans=0.9615	pearsons=0.9478 cosine	=1.0	
Pruitt KD, Tatusova T, Maglott DF	: NCBI Reference Sec	uence r	Document		CDE	F1	
			kdml12.pdf	box.txt.xmi	0.9846	0.9807	
CDE Constraint XX Selection View // TextRul	er 🦻 Annotation Tes 🐲 Ruta 🤇	luery	J05-3001.txt	t.xmi	0.9867	0.9857	
			A97-1010.tx	t.xmi	0.9911	0.995	
[• 1471-2105-1	12-43.pdfbox.t	0.9914	0.9871	
Constraint	Weight		J05-1003.txt	txmi	1.0	1.0	
Reference{OR(CONTAINS(Author), CONTAINS(Editor	1	(E)	C04-1024.tx	t.xmi	1.0	1.0	
Reference{CONTAINS(Title)};	1		A88-1009.tx	t.xmi	1.0	1.0	
Reference{CONTAINS(Date)};	1						
Author{STARTSWITH(Reference)};	1		🗅 CDE Result 🔀				
Author (Date Title);	1						
Title{-CONTAINS(PublisherInd)};	1		Constraint		Result		
Title ANY{-PARTOF(EditionMarker)};	1		Title ANY{-PAR	TOF(Editio	1.0		
Date{-PARTOF(NumInW)};	1		Date{-PARTOF((NumInW)};	1.0		
Date ANY{OR(-IS(SPECIAL), IS(Quote))};	1		Date ANY{OR(-	IS(SPECIAL)	1.0		
Title{OR(STARTSWITH(W), STARTSWITH(Quote))};	1		Title{OR(START	SWITH(W),	0.9565217391304348		
Author{-CONTAINS(NUM)};	1		Author{-CONT	AINS(NUM)};	1.0		
· · · ·	1		Author{CONTA	INS(CW,1,1	1.0		
Author{CONTAINS(CW,1,100)};			Author{CONTA	INS(W,2,10	0.9787234042553191		
Author{CONTAINS(CW,1,100)}; Author{CONTAINS(W,2,100)};	1				0.0700000000000000000000000000000000000		
Author{CONTAINS(CW,1,100)}; Author{CONTAINS(W,2,100)}; Title{CONTAINS(W,2, 200)};	1		Title{CONTAIN	S(W,2, 200)};	0.9782608695652174		

UNI UIMA Ruta CDE Plug-in WÜ - Document Details -

UIMA Ruta CDE - CDE2/data/processed/fold3/1471-2105-12	-43.pdfbox.txt.xmi (CDE2/descriptor/types/B	libtex.)	(ml) - Eclipse Platfo	orm					
<u>File E</u> dit <u>N</u> avigate Se <u>a</u> rch <u>P</u> roject <u>R</u> un <u>W</u> indow <u>H</u> e	lp								
📬 🕶 🕼 🖕 🖉 🗸 🛸 📎 🖢 🗸	월 ▼ № 🔶 ▼ 🔿 ▼				🖹 🎲 UIMA Ruta C	🥜 UIMA R	uta ि Resou »		
J05-3001.txt.xmi 🗍 A97-1010.txt.xmi 🗍 147	1-2105-12-43.pdfb 🛛 🔭 1 📑		品 CDE Documen	ts 🕱 🔚 🕻	Dutline 📎 Annotatior	Browser View			
evolutionary history of the first	Daviuson JN, chen KC, Jamison KS, Musmanno LA, Kern CD. me					Documents: work\workspace-ruta\CDE2\data\processed\fold3			
biosynthesis. Bioessays 1993, 15:	157-164.								
			Test Data:	D:\work\wor	kspace-ruta\CDE2\data	\complete			
Andrade , (Ed.): Bioinformatics a	and Genomes: Current Perspe	ec	Type System:	I\CDE2\descr	iptor\uima\ruta\examp	ole\IETypeSyst	em.xml 🛅		
Heidelberg, Germany: Horizon Scie	entific Press; 2003.								
			mse=0.0 spearn	nans=0.9615 p	earsons=0.9478 cosine	=1.0			
Pruitt KD, Tatusova T, Maglott DR	: NCBI Reference Sequence	r -	Document		CDE	F1			
			kdml12.pdft	nox.txt.xmi	0.9846	0.9807			
CDE Constraint X Selection View // TextRul	er 🦻 Annotation Tes 👘 Ruta Query 🦳		J05-3001.txt.	xmi	0.9867	0.9857			
			A97-1010.txt	xmi	0.9911	0.995			
	r		I471-2105-1	2-43.pdfbox.t	. 0.9914	0.9871			
Constraint	Weight	•	 J05-1003.txt. 	xmi	1.0	1.0			
Reference{OR(CONTAINS(Author), CONTAINS(Editor	1 [C04-1024.txt	txmi	1.0	1.0			
Reference{CONTAINS(Title)};	1	6	A88-1009.txt	t.xmi	1.0	1.0			
Reference{CONTAINS(Date)};	1 (A							
Author{STARTSWITH(Reference)};	1		🛛 CDE Result 🕺						
Author (Date Title);	1				D				
Title{-CONTAINS(PublisherInd)};	1		Constraint		Kesult				
Title ANY{-PARTOF(EditionMarker)};	1	Ð	Title ANY{-PAR	TOF(Editio	1.0				
Date{-PARTOF(NumInW)};	1		Date{-PARTOF(I	NumInW)};	1.0				
Date ANY{OR(-IS(SPECIAL), IS(Quote))};	1		Date ANY{OR(-I	S(SPECIAL)	1.0				
Title{OR(STARTSWITH(W), STARTSWITH(Quote))};	1		Title{OR(START	SWITH(W),	0.9565217391304348				
Author{-CONTAINS(NUM)};	1		Author{-CONTA	AINS(NUM)};	1.0				
Author{CONTAINS(CW,1,100)};	1		Author{CONTA	INS(CW,1,1	1.0		E		
Author{CONTAINS(W,2,100)};	1		Author{CONTA	INS(W,2,10	0.9787234042553191				
Title{CONTAINS(W,2, 200)};	1		Title{CONTAINS	S(W,2, 200)};	0.9782608695652174				
Date{CONTAINS(NUM)};	1		Date{CONTAINS	S(NUM)};	1.0		-		
		1							

Summary: Constraint-driven Evaluation in UIMA Ruta



Predict performance of arbitrary models on unseen data by constraints

- Rule-constraints
- Annotation Distribution constraints
- UIMA Ruta workbench plug-in
 - Supports rapid prototyping
 - and component selection

Experimental study on reference segmentation shows usefulness





Thank you for your attention!

Try CDE in UIMA Ruta 2.1.0: <u>http://uima.apache.org/ruta.html</u>

UNI WÜ References



1. Bellare, K., Druck, G., McCallum, A.: Alternating Projections for Learning with Expectation Constraints. In: Proceedings of the Twenty-Fifth Conference on Uncertainty in AI. pp. 43-50. AUAI Press (2009)

2. Cohn, D.A., Atlas, L., Ladner, R.: Improving generalization with active learning. Machine Learning 15, 201-221 (1994)

3. Councill, I., Giles, C.L., Kan, M.Y.: ParsCit: an Open-source CRF Reference String Parsing Package. In: Proceedings of the Sixth International Language Resources and Evaluation (LREC'08). ELRA, Marrakech, Morocco (2008)

4. Culotta, A., McCallum, A.: Condence Estimation for Information Extraction. In: Proceedings of HLT-NAACL 2004: Short Papers. pp. 109-112. HLT-NAACL-Short '04, Association for Computational Linguistics, Stroudsburg, PA, USA (2004)

5. Ferrucci, D., Lally, A.: UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment. Natural Language Engineering 10(3/4), 327-348 (2004)

6. Graca, J., Ganchev, K., Taskar, B.: Expectation Maximization and Posterior Constraints. In: Platt, J., Koller, D., Singer, Y., Roweis, S. (eds.) NIPS 20, pp. 569-576. MIT Press, Cambridge, MA (2008)

7. Kluegl, P., Atzmueller, M., Puppe, F.: TextMarker: A Tool for Rule-Based Information Extraction. In: Chiarcos, C., de Castilho, R.E., Stede, M. (eds.) Proceedings of the 2nd UIMA@GSCL Workshop. pp. 233-240. Gunter Narr Verlag (2009)

8. Kluegl, P., Hotho, A., Puppe, F.: Local Adaptive Extraction of References. In: 33rd Annual German Conference on Articial Intelligence (KI 2010). Springer (2010)

9. Kluegl, P., Toepfer, M., Lemmerich, F., Hotho, A., Puppe, F.: Collective Information Extraction with Context-Specic Consistencies. In: Flach, P.A., Bie, T.D., Cristianini, N. (eds.) ECML/PKDD (1). Lecture Notes in Computer Science, vol. 7523, pp. 728-743. Springer (2012)

10. Lafferty, J., McCallum, A., Pereira, F.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. Proc. 18th International Conf. on Machine Learning pp. 282-289 (2001)

11. Mann, G.S., McCallum, A.: Generalized Expectation Criteria for Semi-Supervised Learning with Weakly Labeled Data. J. Mach. Learn. Res. 11, 955-984 (2010)

12. McCallum, A., Nigam, K.: Employing EM and Pool-Based Active Learning for Text Classication. In: Shavlik, J.W. (ed.) ICML. pp. 350-358. Morgan Kaufmann (1998)

13. Savova, G.K., Masanz, J.J., Ogren, P.V., Zheng, J., Sohn, S., Kipper-Schuler, K.C., Chute, C.G.: Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. Journal of the American Medical Informatics Association : JAMIA 17(5), 507-513 (Sep 2010)