



Introduction

Goals:

- Short Answer Assessment (SAA): evaluate semantic appropriateness of short (learner) answer
- Plagiarism Detection (PD):
 - extrinsic: does a document plagiarize another ?
 - intrinsic: stylistic changes inside one document ?
- Authorship Attribution (AA): detect author characteristics

Methods

- SAA: form-independent semantic classification
- PD: stylistic (and semantic) classification

⇒ Can knowledge from either domain improve the task in the other domain ?

Data

- CREG (Corpus of Reading Comprehension Exercises in German) (Ott et al., 2012)
- WRC (Wikipedia Reuse Corpus) (Clough and Stevenson, 2011)
- both “short answer” corpora contain texts, questions, target and student answers

Frequency	CREG-1032-KU	CREG-1032-OSU	WRC
student answers	610	422	95
target answers	136	87	5
questions	117	60	5

Table 1: Data distribution in CREG-1032 and Wikipedia Reuse Corpus.

System

3-step system CoMiC (Meurers et al., 2011) for both tasks:

1. **Annotation:** add linguistic markup to input
2. **Alignment:** align tokens, chunks, and dependency triples (TMA)
3. **Diagnosis:** extract (alignment) features for machine learning

- Givenness filter: material not in question only aligned
- SAA (CoMiC) + AA features (Stamatatos (2009))

Experimental Testing

- H_1 : SAA improves with features from AA
- H_1 : AA (here PD) improves with features from SAA
- WEKA ibk k=5-NN 10-fold CV with different feature sets and both corpora from different domains

Features	Data		
	KU	OSU	WRC
Baselines			
CoMiC	84.5	87.1	92.6*
all auth.att. features	86.9*	86.0	84.2
Baselines + new features			
CoMiC + all auth.att. features	85.6	87.7	90.5*
all features interpolated	78.0	87.9	62.1
CoMiC + all features interpolated	84.3	87.2	95.8*
lexical features	84.5	86.3	90.5*
CoMiC + lexical features	84.4	88.2	88.4
character features	83.3	86.3	82.2
CoMiC + character features	85.7	87.7	83.2
syntactic features	67.4	69.0	80.0
CoMiC + syntactic features	84.3	85.1	87.4
semantic features	82.1	85.0	90.6*
CoMiC + semantic features	83.8	87.0	91.6*

Table 2: Binary classification results (* for $p < \alpha = 0.1$).

Rank	CREG-1032	WRC
1	TargetSpellCorr	StudentSpellCorr
2	CharBigramFreq	Token Match
3	CharTrigramFreq	CharTrigramFreq
4	CharFourgramFreq	CharFourgramFreq
5	WordUniFreq	CharFivegramFreq
6	all features interpolated	TargetSpellCorr
7	Synonym Match	CharBigramFreq
8	CharFivegramFreq	StudentSynonym
9	StudentSpellCorr	WordUniFreq
10	TargetSynonym	TargetSynonym

Table 3: Ten most informative features (information gain) for the CREG-1032 and the WRC data set.

Discussion and Conclusion

- statistically significant improvements with domain transfer, especially PD benefits from SAA
- similar features most effective across domains
- exploration of more features (e.g. LCS, stopword patterns) needed

Feature	Description
Short Answer Assessment Features	
CoMiC baseline	
Keyword Overlap	pct keywords aligned
Target Token Overlap	pct aligned target tokens
Learner Token Overlap	pct aligned student tokens
Target Chunk Overlap	pct aligned target chunks
Learner Chunk Overlap	pct aligned student chunks
Target Triple Overlap	pct aligned target dependency triples
Learner Triple Overlap	pct aligned student dependency triples
Token Match	pct token-identical token alignments
Similarity Match	pct similarity-resolved token alignments
Semtype Match	pct type-resolved token alignments
Lemma Match	pct lemma-resolved token alignments
Synonym Match	pct synonym-resolved token alignments
Variety of Match	sum of features 8-12
Target Answer ID	target answer id
Student Answer ID	student answer id
Authorship Attribution Features	
lexical	
AvgWordLength	Average word length
TTR	Type-Token Ratio
WordNgramFreq	Word N-gram frequency similarity ($0 < N < 4$)
SpellCorr	Spell Corrected Unigram Matches
character	
CharFreq	Character frequency similarity
UpperCharFreq	Uppercase character frequency similarity
LowerCharFreq	Lowercase character frequency similarity
DigitCharFreq	Digit character frequency similarity
LetterProportion	Proportion of letters (A-Za-z) in answer
UpperProportion	Proportion of uppercase letters in answer
LowerProportion	Proportion of lowercase letters in answer
CharBigramFreq	Character bigram frequency similarity
CharTrigramFreq	Character trigram frequency similarity
CharFourgramFreq	Character fourgram frequency similarity
CharFivegramFreq	Character fivegram frequency similarity
syntactic	
Chunk	Chunk tag frequency similarity
NPChunk	Noun phrase chunk frequency similarity
PosNgram	POS tag N-gram frequency similarity ($0 < N < 6$)
semantic	
Synonym	Proportion of synonym-overlapping tokens
DepTriple	Proportion of dependency triple overlaps
combination	
all features interpolated	all features combined

Table 4: SAA and AA Features from different domains used in the experiments based on Stamatatos (2009).