

Meta Analysis within Author Verification

Benno Stein Nedim Lipka Sven Meyer zu Eissen

webis.de

Bauhaus-Universität Weimar

- Outline**
- Intrinsic Plagiarism Analysis and Authorship Verification
 - Post-Processing with Unmasking

Intrinsic Analysis and Authorship Verification

Intrinsic Analysis and Authorship Verification

Problem Setting

How to find a plagiarized section / foreign authorship without a reference corpus?

OnWeb-based Plagiarism Analysis

Alexander Kluge, Dennis Bräuer, Christoph Lorenz, Sven Mayer, or Essen
Alexander Kluge (kluge@uni-wuerzburg.de)
Dennis Bräuer (braeuer@med.uni-wuerzburg.de)
Christoph Lorenz (lorenz@uni-wuerzburg.de)
Sven Mayer (mayer@med.uni-wuerzburg.de)
Bauhaus University Witten
Faculty of Media
Media Systems
D-92521 Witten, Germany

Abstract: The paper in hand presents a Web-based application for the analysis of text documents with respect to plagiarism. Unlike most reporting experiences with classical algorithms, a new method for plagiarism analysis is introduced. Since well-known algorithms for plagiarism detection require the generation of a candidate document collection against which a suspicious document can be compared, they are unsuitable to spot potentially copied passages using only the input document. This kind of program remains independent of what paragraphs are copied from sources that are not available electronically. Our method is able to detect a change in writing style, and consequently to identify suspicious passages within a single document, apart from contributing to solve the outlined problem, the presented method can also be used to focus a search for potentially copied documents.

Key words: plagiarism analysis, style analysis, focused search, clustering, Kullback-Leibler divergence

- 1. Introduction
- 2. Plagiarism detection in the case of available data, information, literature, or writing
- 3. Other data without proper acknowledgment of the original source [1]. Recently, the growing amount of digitally available documents contributes to this possibility.
- 4. Newly find and (partially) copy text documents given a specific topic. According to [2].
- 5. In Cases in plagiarism check on 10,000 documents, about 50% of the authors selected a passage from a foreign document [3].

Plagiarism happens in several forms. The most distinctions between the following textual relationships between documents: identical copy, edited copy, reorganized document, reworded document, content-independent document, documents that include portions of other documents. However, unauthorized (partial) translations and documents that copy the structure of other documents can also be seen as plagiarized. Figure 1 depicts a taxonomy of plagiarism forms. Orthogonal to plagiarism forms are the underlying media: plagiarism may happen in articles, books or computer programs.

suspicious document

Our Web-based plagiarism analysis application takes a suspicious document from an open access domain as input. Consequently, an unprocessed, domain-independent keyword extraction algorithm that takes a single document as input would be convenient. The change independence being stable. Measure and the input propose such a method is to use a χ^2 -analysis of term co-occurrences identified in the input document.

2.2 Query Generation: Focused Search

When keywords are extracted from the suspicious document, we employ a heuristic query generation procedure, which has first presented in [1]. The K1 denotes the set of keywords that have been derived from a suspicious document. By adding synonyms, coordinate terms, and domain-specific related terms, the set K1 is extended towards a set K2 (without ground truth) by applying statistical knowledge about significant left and right neighbors and as absolute co-occurring words, yielding the set K3 [1-3]. Thus, a set of keywords is generated (and passed to search engines).

This selection step is controlled by quantitative feedback. Depending on the number of found documents using the generated keywords, a heuristic feedback is generated. Note that such a control can be realized by a heuristic, using the set K3, which considers word group sizes and word proximity classes. The result of this step is a candidate document collection $\{C_1, \dots, C_n\}$.

3 Plagiarism Analysis

An individual article, a document may be plagiarized in different forms. Consequently, several indicators exist to inspect a document of plagiarism. An amount of indicators that are given in [4] are:

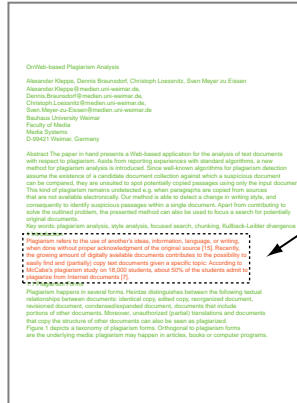
- (1) Copied text: If a passage is taken from a source that is known and it is not cited, probably then this is an obvious case of plagiarism.
- (2) Bibliography: The references in documents overlap significantly, the bibliographies and other references may be copied. A changing citing style may be a sign for plagiarism.
- (3) Change in writing style: A suspect change in the author's style may appear paragraph- or section-wise, e.g. between objective and subjective style, normal and verbiage, brilliant and baffling passages.
- (4) Paragraph formatting: In copy-and-paste plagiarism cases the formatting of the original document is inherited to pasted paragraphs, especially when content is copied from lines to text processing programs.
- (5) Textual pathway: If the line of argumentation throughout a document is consequently unchanged from the document may be a "thread plagiarist", i.e. a compilation of different sources.

corpus documents

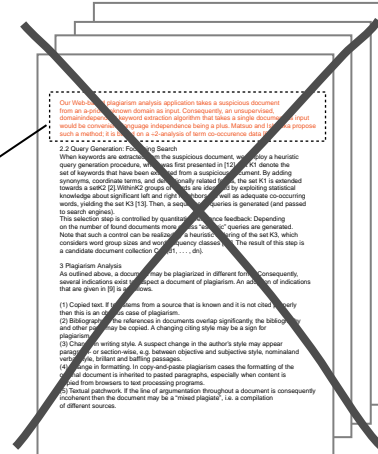
Intrinsic Analysis and Authorship Verification

Problem Setting

How to find a plagiarized section / foreign authorship without a reference corpus?



suspicious document



corpus documents

Formulated as decision problem:

Problem. AV_{FIND}

Given. A text d , allegedly written by author A .

Question. Does d contain sections written by an author B , $B \neq A$?

Intrinsic plagiarism analysis and authorship verification (AV) are two sides of the same coin.

Intrinsic Analysis and Authorship Verification

Building Blocks for Authorship Verification

Pre-analysis			Classification	Post-processing	
Impurity assessment	Decomposition strategy	Style model construction	Style outlier identification	Improvement at section level	Improvement at document level
Document length analysis	Uniform length	Formatting	Two-class discriminant analysis	Citation analysis	Confidence-based majority decision
Genre Analysis	Structural boundaries	Surface analysis			Unmasking
Analysis of issuing institution	Text element boundaries	Structure analysis	One-class classifier: density estimation		Batch means
	Topical boundaries	Complexity measures	One-class classifier: boundary estimation		Human inspection
		<i>n</i> -gram analysis	One-class classifier: reconstruction		
		Language modeling			
		Dialectic analysis			

Intrinsic Analysis and Authorship Verification

Style Model Construction: Starting Points

Selected quantifiable feature classes (from easy to difficult):

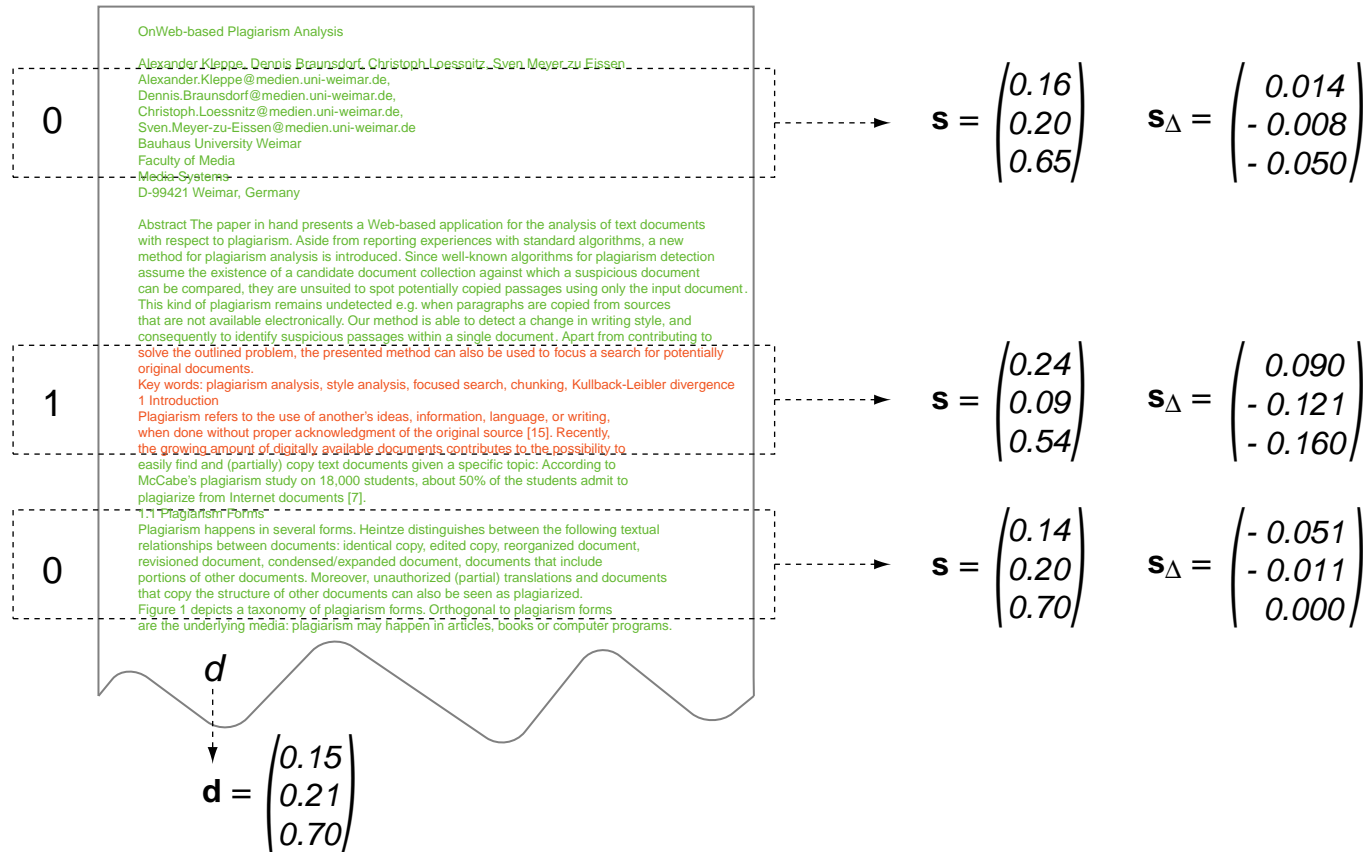
- surface features
- structure and organization
- complexity measures
 - readability
 - writing complexity
 - vocabulary richness, diction
- dialectic power
 - argumentation consistency
 - argumentation strategy

For a machine-based identification, features have to be developed and operationalized within a style model \mathcal{R} .

Intrinsic Analysis and Authorship Verification

Style Model Construction: Language Modeling

Style Outlier Identification



Supervised learning situation: given are sections s_i from both the target class (author A), where $c(s) = 0$, and the outlier class (other authors), where $c(s) = 1$.

Intrinsic Analysis and Authorship Verification

Style Outlier Identification

Compute for each section the relative differences between section-specific style feature values and document-specific style feature values.

1. Let $\sigma_1, \dots, \sigma_m$ denote style feature functions.

2. For each section $s \subseteq d$:

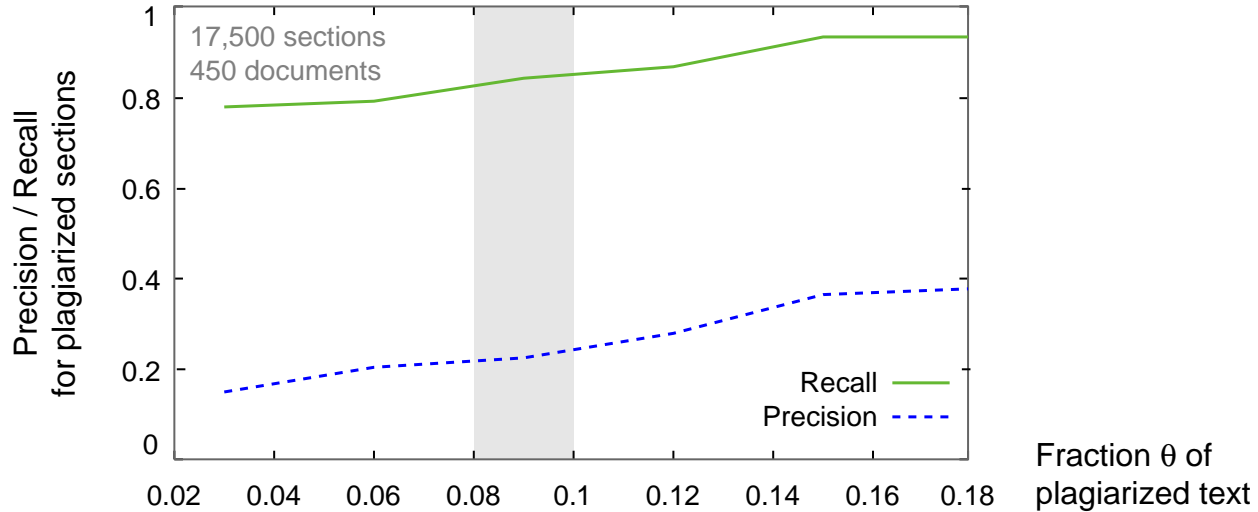
□ compute style model $\mathbf{s} = \begin{pmatrix} \sigma_1(s) \\ \vdots \\ \sigma_m(s) \end{pmatrix} \in \mathbf{R}^m$

□ compute relative deviations $\mathbf{s}_\Delta = \begin{pmatrix} \frac{\sigma_1(s) - \sigma_1(d)}{\sigma_1(d)} \\ \vdots \\ \frac{\sigma_m(s) - \sigma_m(d)}{\sigma_m(d)} \end{pmatrix} \in \mathbf{R}^m$

3. Learn an outlier hypothesis h from a sample $\{(\mathbf{s}_\Delta, c(s))\}$, $c(s) \in \{0, 1\}$.

Intrinsic Analysis and Authorship Verification

Evaluation: Style Model Performance

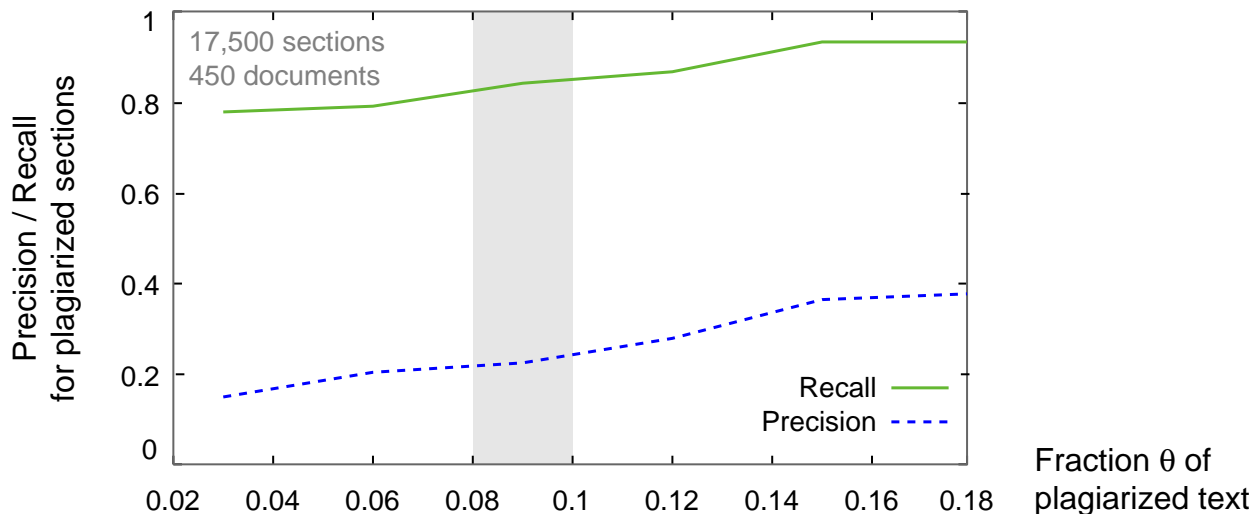


The unsatisfying precision is rooted in the class imbalance.

The Gretchenfrage: Are parts of d plagiarized, if we find an outlier?

Intrinsic Analysis and Authorship Verification

Evaluation: Style Model Performance



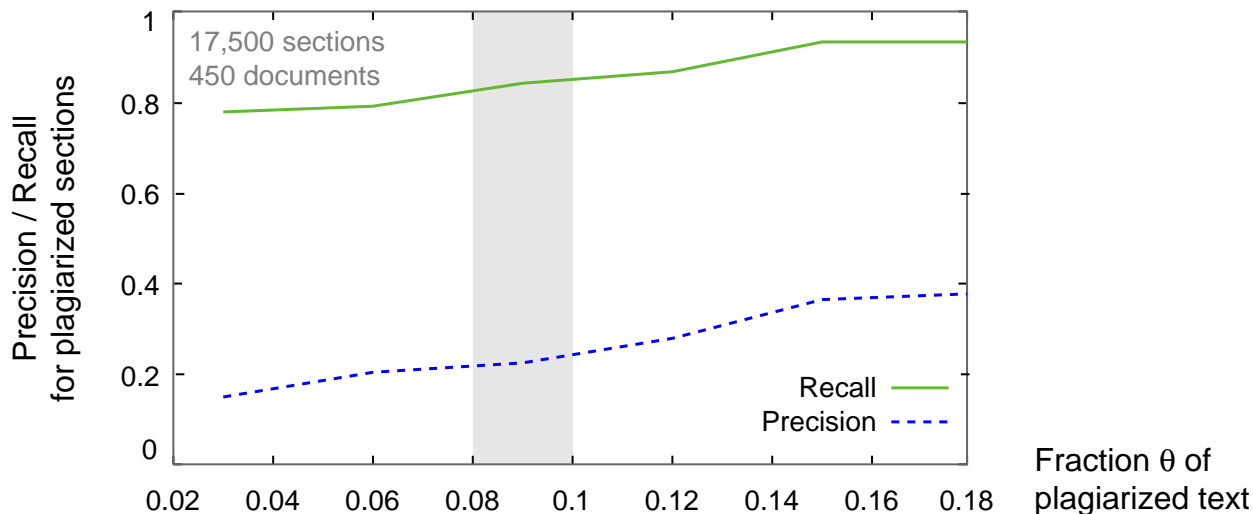
The unsatisfying precision is rooted in the class imbalance.

The Gretchenfrage: Are parts of d plagiarized, if we find an outlier?

# Outliers	Strategy	→	Hypothesis
0	minimum risk	→	not plagiarized
1	minimum risk	→	plagiarized
2	minimum risk	→	plagiarized
3	minimum risk	→	plagiarized

Intrinsic Analysis and Authorship Verification [Building Blocks]

Evaluation: Style Model Performance



The unsatisfying precision is rooted in the class imbalance.

The Gretchenfrage: Are parts of d plagiarized, if we find an outlier?

# Outliers	Strategy	→	Hypothesis
0	minimum risk	→	not plagiarized
1	minimum risk	→	plagiarized
2	minimum risk	→	plagiarized
3	minimum risk	→	plagiarized

Strategy	→	Hypothesis
post-processing	→	not plagiarized
post-processing	→	not plagiarized
post-processing	→	not plagiarized
post-processing	→	plagiarized

Post-Processing with Unmasking [Building Blocks]

Post-Processing with Unmasking

Reliable Interpretation of Outliers

Problem. AVOUTLIER (an easier variant of AVFIND)

Given. A set of texts $D = \{d_1, \dots, d_n\}$, allegedly written by author A .

Question. Does D contain texts written by an author B , $B \neq A$?

Post-Processing with Unmasking

Reliable Interpretation of Outliers

Problem. AVOUTLIER (an easier variant of AVFIND)

Given. A set of texts $D = \{d_1, \dots, d_n\}$, allegedly written by author A .

Question. Does D contain texts written by an author B , $B \neq A$?

The belief into an answer depends on the number of found outliers:

# Outliers	Strategy	→	Hypothesis
0	minimum risk, post-processing	→	not plagiarized
2	minimum risk	→	plagiarized
2	post-processing	→	not plagiarized
4	minimum risk, post-processing	→	plagiarized

Post-Processing with Unmasking

Reliable Interpretation of Outliers

Problem. AVOUTLIER (an easier variant of AVFIND)

Given. A set of texts $D = \{d_1, \dots, d_n\}$, allegedly written by author A .

Question. Does D contain texts written by an author B , $B \neq A$?

The belief into an answer depends on the number of found outliers:

# Outliers	Strategy	→	Hypothesis
0	minimum risk, post-processing	→	not plagiarized
2	minimum risk	→	plagiarized
2	post-processing	→	not plagiarized
4	minimum risk, post-processing	→	plagiarized

Post-process **borderline situations** to gain further evidence for accepting or rejecting a hypothesis.

Idea: Interpret AVOUTLIER results under the Unmasking framework.

Post-Processing with Unmasking

Unmasking for Authorship Verification [Koppel/Schler 2004]

Problem. AV

Given. Two documents d_1, d_2 .

Question. Are d_1 and d_2 written by the same author?

Procedure Unmasking:

1. *Chunking.*
2. *Model Fitting.*
3. *Impairing.*
4. Goto Step 2 until the feature space is sufficiently reduced.

Post-Processing with Unmasking

Unmasking for Authorship Verification [Koppel/Schler 2004]

Problem. AV

Given. Two documents d_1, d_2 .

Question. Are d_1 and d_2 written by the same author?

Procedure Unmasking:

1. *Chunking.* Decompose d_1, d_2 into two sets of sections, D_1, D_2 .
2. *Model Fitting.* With the 250 most frequent words in d_1, d_2 build a VSM for each s in D_1, D_2 . Learn a classifier that discriminates between D_1, D_2 .
3. *Impairing.* Drop the 3 most discriminating features from the VSMs.
4. Goto Step 2 until the feature space is sufficiently reduced.

Post-Processing with Unmasking

Unmasking for Authorship Verification [Koppel/Schler 2004]

Problem. AV

Given. Two documents d_1, d_2 .

Question. Are d_1 and d_2 written by the same author?

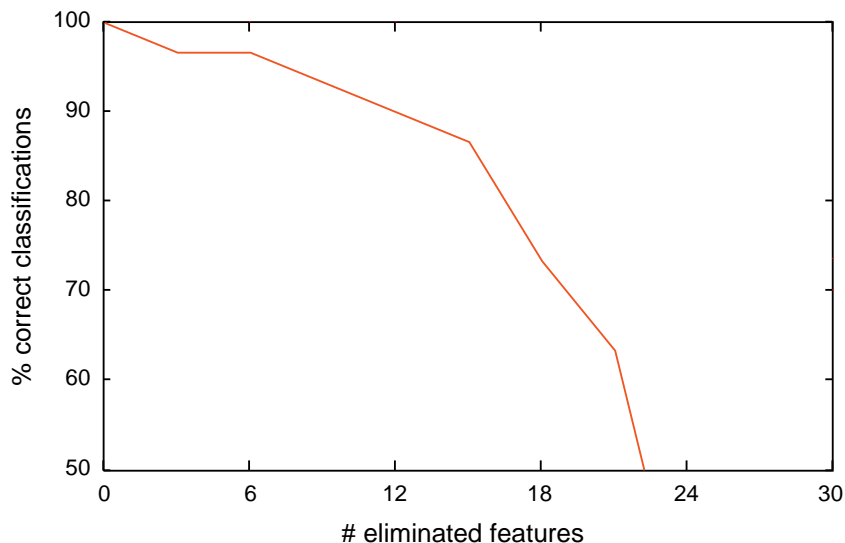
Procedure Unmasking:

1. *Chunking.* Decompose d_1, d_2 into two sets of sections, D_1, D_2 .
2. *Model Fitting.* With the 250 most frequent words in d_1, d_2 build a VSM for each s in D_1, D_2 . Learn a classifier that discriminates between D_1, D_2 .
3. *Impairing.* Drop the 3 most discriminating features from the VSMs.
4. Goto Step 2 until the feature space is sufficiently reduced.
5. *Meta Learning.* Analyze the degradation in the quality of the model fitting.

Post-Processing with Unmasking

Unmasking for Authorship Verification

Characteristic of a typical outcome:



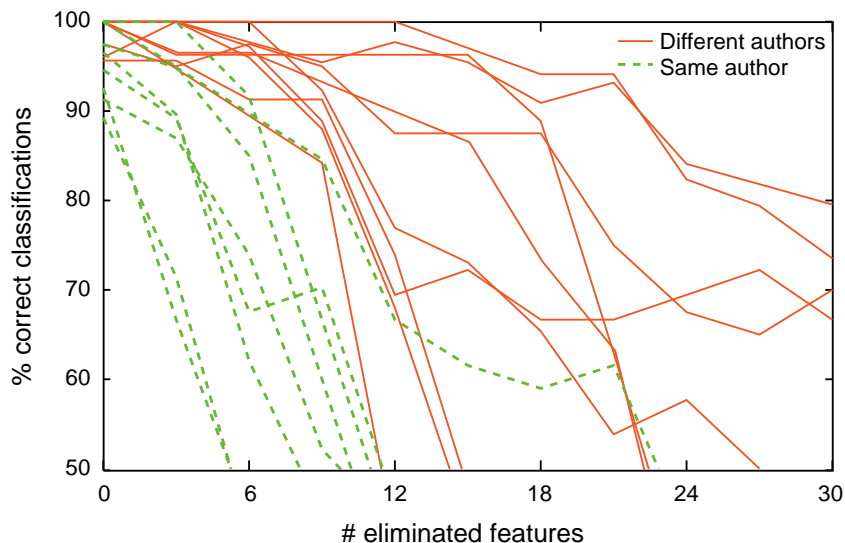
Rationale:

- ❑ A large fraction of the 250 words are function words and stop words.
- ❑ Only few of the words are related to topic.
- ❑ Only few words do the discrimination job—the topic words for a large part.
- ❑ Different authors can be distinguished by their use of function words.

Post-Processing with Unmasking

Unmasking for Authorship Verification

Characteristic of a typical outcome:



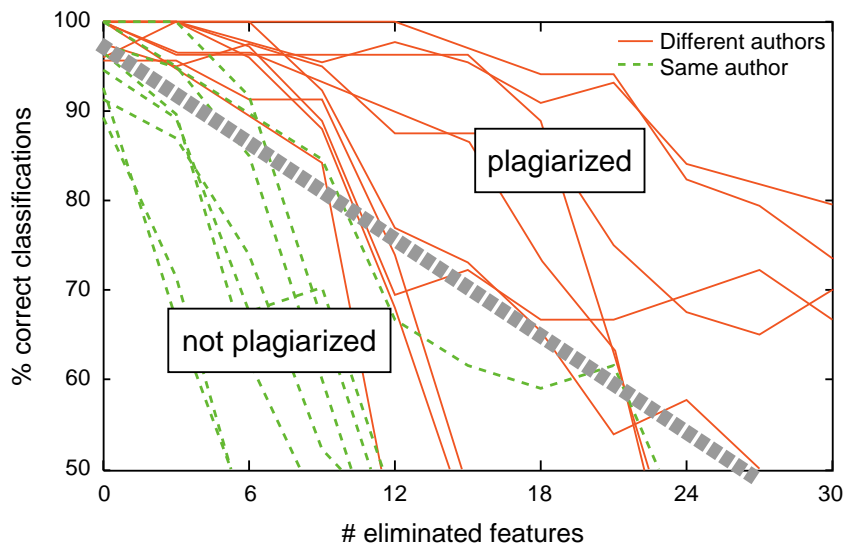
Rationale:

- ❑ A large fraction of the 250 words are function words and stop words.
- ❑ Only few of the words are related to topic.
- ❑ Only few words do the discrimination job—the topic words for a large part.
- ❑ Different authors can be distinguished by their use of function words.

Post-Processing with Unmasking

Unmasking for Authorship Verification

Characteristic of a typical outcome:



Rationale:

- ❑ A large fraction of the 250 words are function words and stop words.
- ❑ Only few of the words are related to topic.
- ❑ Only few words do the discrimination job—the topic words for a large part.
- ❑ Different authors can be distinguished by their use of function words.

Post-Processing with Unmasking [Results]

Strategy Overview

1. Solve AVOUTLIER with one-class classifier. For borderline situations:
2. Construct AVBATCH from the classified target and outlier sections.
3. Apply Unmasking to solve AVBATCH.

Post-Processing with Unmasking ^[W]

Evaluation: Artificial Data

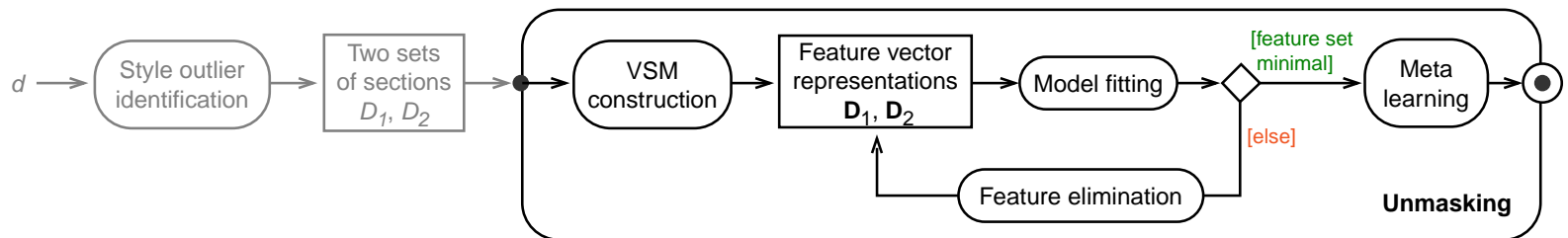
Impurity	Classification			Post-processing					
	AVOUTLIER Minimum risk			AVBATCH Majority			AVBATCH Unmasking		
	<i>prec</i>	<i>rec</i>	<i>F</i>	<i>prec</i>	<i>rec</i>	<i>F</i>	<i>prec</i>	<i>rec</i>	<i>F</i>
0.20	0.12	1.00	0.56	0.71	0.83	0.77	0.73	0.90	0.82
0.30	0.20	1.00	0.60	1.00	0.56	0.78	1.00	0.93	0.97
0.40	0.18	1.00	0.59	1.00	0.83	0.92	1.00	0.87	0.94

Post-Processing with Unmasking ^[W]

Evaluation: Artificial Data

Impurity θ	Classification			Post-processing					
	AVOUTLIER Minimum risk			AVBATCH Majority			AVBATCH Unmasking		
	<i>prec</i>	<i>rec</i>	<i>F</i>	<i>prec</i>	<i>rec</i>	<i>F</i>	<i>prec</i>	<i>rec</i>	<i>F</i>
0.20	0.12	1.00	0.56	0.71	0.83	0.77	0.73	0.90	0.82
0.30	0.20	1.00	0.60	1.00	0.56	0.78	1.00	0.93	0.97
0.40	0.18	1.00	0.59	1.00	0.83	0.92	1.00	0.87	0.94

Strategy overview:



Summary

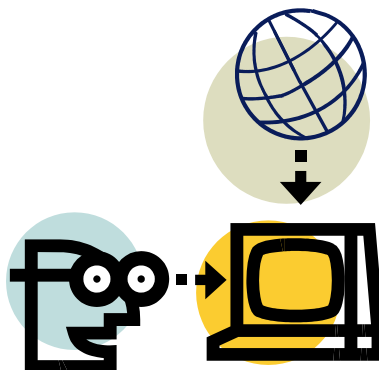
Summary

Authorship verification happens within three steps:

1. Pre-processing. Text decomposition + style model construction
2. Classification. Style outlier identification / one-class classification
3. Post-processing. Improve reliability of the classification step.

Main contribution:

A post-processing strategy for borderline situations, based on unmasking.



Thank you!