

Benchmark: Extracting Table Information from Scientific Documents

Valda Seminar

Marijan Soric



May 16, 2025

Outline

Introduction

Context

Goal

Table extraction methods

Evaluation

Results and analysis

Conclusion

Project Context

- Inria Valda** (Inria Paris, DI ENS, CNRS) Topics: management of complex data, data generated by human activity
- Inria Cedar** (Inria Saclay, LIX, CNRS) Topics: Cloud-scale analysis of rich data
- BRGM** (*Bureau de Recherches Géologiques et Minières*) French National Geological Survey: Earth science applications for managing soil and subsoil resources and risks

Working environment

- GéolAug project collaboration Inria & BRGM
- Helping geologists prepare their missions, facilitating access to knowledge
- Thesis: “Exploitation and Structuring of Heterogeneous Geological Data and Knowledge”
- Heterogeneous data:
 - Geological maps, diagrams
 - Databases
 - Text
 - Tables

Outline

Introduction

Context

Goal

Table extraction methods

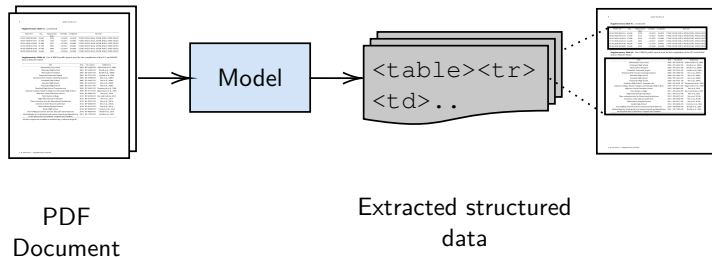
Evaluation

Results and analysis

Conclusion

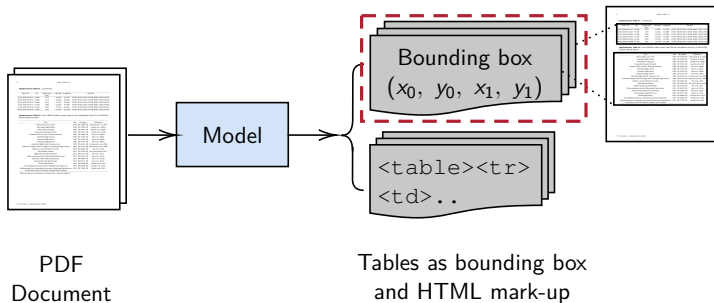
Definitions

Automatic structured table extraction from PDF documents



Task definition (1/2)

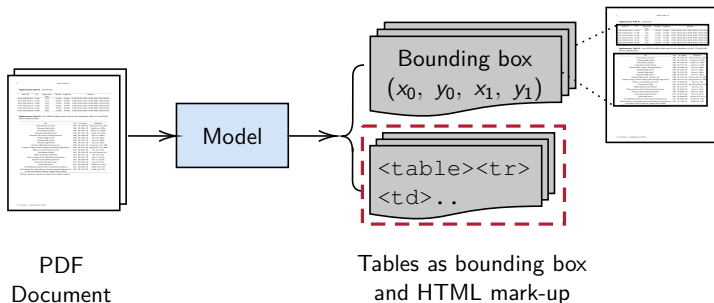
Table Detection: Find all tables within a document



Hard: various table styles (with or without borders)

Task definition (2/2)

Table Structure Recognition: Extract content from tables while keeping their structures



Hard: various cell styles (empty, alignment. . .)

Table Extraction: Detection + Structure

Outline

Introduction

Table extraction methods

Baseline

Object detection

Evaluation

Results and analysis

Conclusion

Baseline

PdfPlumber Python library, PDF parser, rule-based heuristics

Camelot Python library, rule-based heuristics

Baseline

PdfPlumber Python library, PDF parser, rule-based heuristics

Camelot Python library, rule-based heuristics

GROBID PDF parser, used in HAL <https://hal.science>

Baseline

PdfPlumber Python library, PDF parser, rule-based heuristics

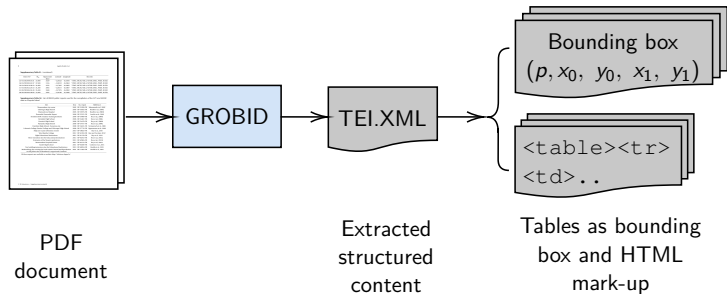
Camelot Python library, rule-based heuristics

GROBID PDF parser, used in HAL <https://hal.science>

LLM-Vision GPT-4o-mini with OpenAI API

GROBID

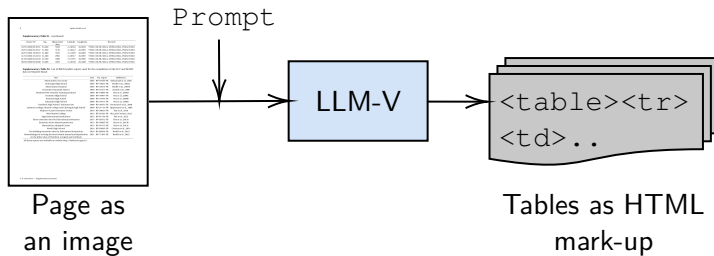
GROBID (Lopez, 2008) PDF parser, used in HAL¹.



¹<https://hal.science>

LLM-Vision

LLM-Vision GPT-4o-mini with OpenAI API



Note: LLM-Vision does not output coordinates

Outline

Introduction

Table extraction methods

Baseline

Object detection

Evaluation

Results and analysis

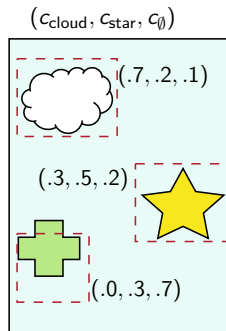
Conclusion

Object detection

Object detection: detect the instances in an image

Instance: Object to find (“table”, “column”, “row”, “cell”).

Detection: Locations (bounding boxes)
+ probability distribution on labels
(confidence score).



Object detection: instances (cloud, star, no object)

Baseline methods do not output confidence scores.

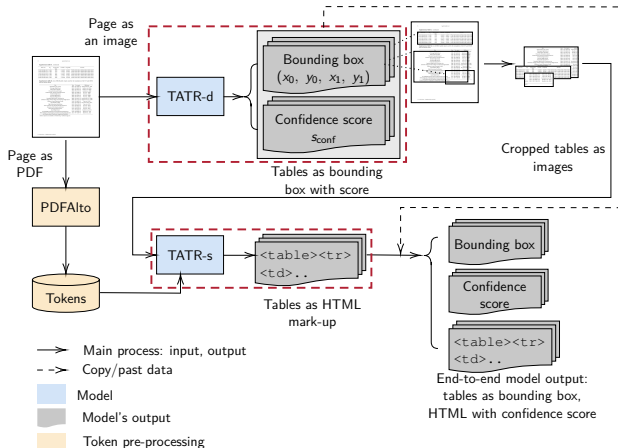
Two-step extraction method

Specialized models are assembled for each task.

Table Extraction Method	Table Detection	Table Structure Recognition
TATR-extract (Smock et al., 2022)	TATR-detect	TATR-structure
VGT+TATR-structure	VGT (Da et al., 2023)	TATR-structure
XY+TATR-extract	XY+TATR-detect	TATR-structure

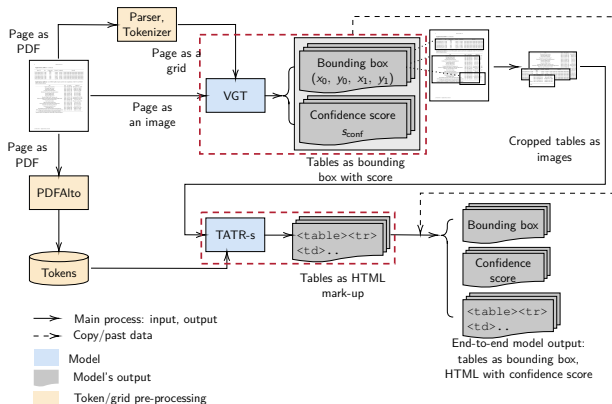
Extraction method (1/3)

TATR-extract is composed of two models: TATR-detect and TATR-structure, using DETR (Carion et al., 2020) (transformer encoder-decoder) architecture.



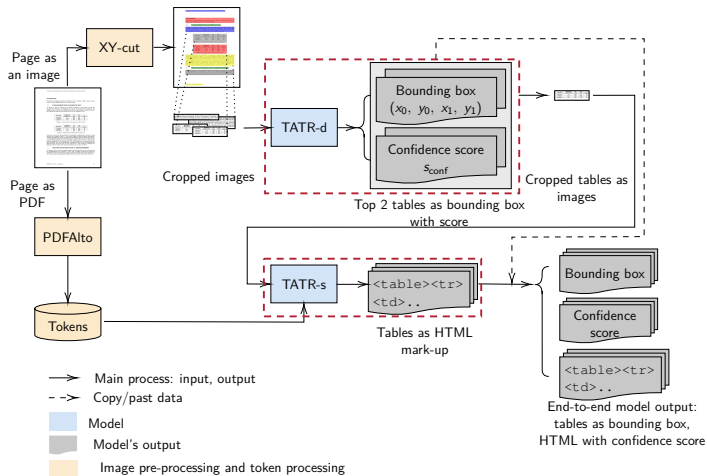
Extraction method (2/3)

VGT+TATR-structure uses VGT for table detection: specialized in document layout detection (including TD). VGT is **multimodal**: it operates on visual and textual content.



Extraction method (3/3)

XY+TATR-extract adds pre-processing with X-Y cut algorithm (Ha et al., 1995).



Outline

Introduction

Table extraction methods

Evaluation

- Table Detection Metrics

- Table Structure Recognition Metrics

- Table Extraction Metrics

- Datasets

Results and analysis

Conclusion

Table Detection

Usual metrics: Precision, Recall, based on *positive* predictions

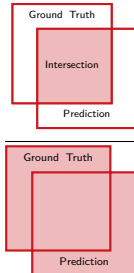
Positive	“There is a table”
True Positive (TP)	Table correctly identified
False Positive (FP)	Detected table is not real
False Negative (FN)	True table not detected

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

TP and FP (with bounding boxes)

We decide if a positive (prediction) is a TP or a FP using Intersection-over-Union (IoU) with a threshold θ_J .

$$\text{IoU} = \frac{\text{area of overlap}}{\text{area of union}} = \frac{\text{Intersection}}{\text{Union}}$$


If $\text{IoU} > \theta_J$ then the positive is a TP, otherwise a FP.

TP and FP (without bounding boxes)

We decide if a positive (prediction) is a TP or a FP using Jaccard-index according with a threshold θ_J .

$$\text{Jaccard}(S_P, S_{GT}) = \frac{|S_P \cap_{\text{multi}} S_{GT}|}{|S_P \cup_{\text{multi}} S_{GT}|}$$

Where S are multisets of 2-grams, where tokens are 2-characters (non-empty) strings from table content (HTML tags not included).

If $J > \theta_J$ then the positive is a TP, otherwise a FP.

Metrics

Precision	P_{θ_J} measures how precise the model is in its predictions
Recall	R_{θ_J} measures how much the model misses real tables

But these metrics are sensitive to the choice of threshold θ_J , that is why we use metrics aggregating $\mathbb{E}_{\theta_J}[X_{\theta_J}]$ w.r.t $\theta_J \sim f$.

Average Precision	Area under the Precision–Recall curve for models with confidence scores.
Model miscalibration	Makes sure that confidence scores are probabilities.

Expected metrics

Before: binary value

$$TP^i = \mathbb{1}_{[IoU_i > \theta_J]}$$

After: score

$$\widetilde{TP}^i = \mathbb{E}_{\theta_J}[\mathbb{1}_{[IoU_i > \theta_J]}]$$

Expected metrics

Before: binary value

$$TP^i = \mathbb{1}_{[\text{IoU}_i > \theta_J]}$$

After: score

$$\widetilde{TP}^i = \mathbb{E}_{\theta_J}[\mathbb{1}_{[\text{IoU}_i > \theta_J]}]$$

Example:

$$\mathbb{E}_{\theta_J}[P_{\theta_J}] = \frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} \mathbb{E}_{\theta_J}[\mathbb{1}_{[\text{IoU}_i > \theta_J]}]$$

With $f(\theta_J) \propto \theta_J$ and $f(\theta_J) \propto \theta_J \mathbb{1}_{[0.5, 1]}$

Outline

Introduction

Table extraction methods

Evaluation

Table Detection Metrics

Table Structure Recognition Metrics

Table Extraction Metrics

Datasets

Results and analysis

Conclusion

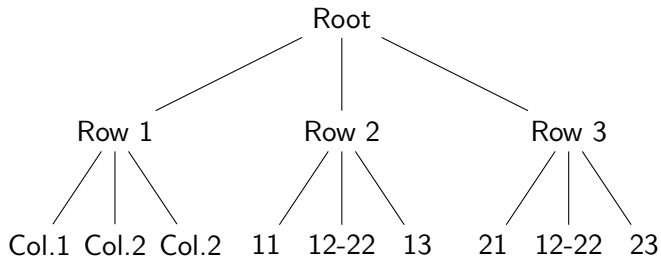
Table Structure Recognition

- Available metrics:
 - Structure **absolute coordinates** (rows, columns, cells) as we did for TD.
 - Cells **relative positions** and global structure, like **TEDS** (Li et al., 2020) and **GriTS** (Smock et al., 2023).
- Evaluation of extraction methods as a whole: the TSR part depends on the TD part.

Metrics (1/2)

TEDS measures the similarity of tables viewed as trees

Col.1	Col.2	Col.3
11	12-22	13
21		23



$$\text{TEDS}(T_P, T_{GT}) = 1 - \frac{\text{EditDist}(T_P, T_{GT})}{\max(|T_P|, |T_{GT}|)}$$

Metrics (2/2)

GriTS represents tables as matrices and computes different similarity types

GriTS Content

$$\begin{pmatrix} \text{Col.1} & \text{Col.2} & \text{Col.3} \\ 11 & 12 - 22 & 13 \\ 21 & 12 - 22 & 23 \end{pmatrix}$$

GriTS Topology

$$\begin{pmatrix} (0, 0, 1, 1) & (0, 0, 1, 1) & (0, 0, 1, 1) \\ (0, 0, 1, 1) & (0, 0, 1, 2) & (0, 0, 1, 1) \\ (0, 0, 1, 1) & (0, -1, 1, 1) & (0, 0, 1, 1) \end{pmatrix}$$

$$\text{GriTS}_f(T_P, T_{GT}) = 2 \frac{\sum_{i,j} f(\tilde{T}_{P,i,j}, \tilde{T}_{GT,j})}{|T_P| + |T_{GT}|}$$

With 2D most similar substructures $(\tilde{T}_P, \tilde{T}_{GT}) = \text{2D-MSS}_f(T_P, T_{GT})$

Outline

Introduction

Table extraction methods

Evaluation

Table Detection Metrics

Table Structure Recognition Metrics

Table Extraction Metrics

Datasets

Results and analysis

Conclusion

Table Extraction Metrics

Need to evaluate end-to-end methods on TE (not TD and TSR independently)

Before: binary value

$$TP^i = \mathbb{1}_{[IoU_i > \theta_J]}$$

After: score

$$\widetilde{TP}^i = s_i^{\text{TSR}} \mathbb{1}_{[IoU_i > \theta_J]}$$

Table Extraction Metrics

Need to evaluate end-to-end methods on TE (not TD and TSR independently)

Before: binary value

$$TP^i = \mathbb{1}_{[IoU_i > \theta_J]}$$

After: score

$$\widetilde{TP}^i = s_i^{\text{TSR}} \mathbb{1}_{[IoU_i > \theta_J]}$$

Finally,

$$P_{\theta_J}^{\text{TSR}} = \frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} s_i^{\text{TSR}} \mathbb{1}_{[IoU_i > \theta_J]}$$

$$R_{\theta_J}^{\text{TSR}} = \frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{P}} s_i^{\text{TSR}} \mathbb{1}_{[IoU_i > \theta_J]}$$

Outline

Introduction

Table extraction methods

Evaluation

Table Detection Metrics

Table Structure Recognition Metrics

Table Extraction Metrics

Datasets

Results and analysis

Conclusion

Datasets

Table-BRGM manually annotated, PDF from geological reports from BRGM²

PubTables scientific articles from PubMed Central Open Access³

Table-arXiv synthetically generated from arXiv⁴ paper source code. Use anchor and LaTeXML⁵

Dataset	# Pages	# Tables
Table-BRGM	499	124
PubTables	46 942	55 990
Table-arXiv	36 869	6 308

²<https://infoterre.brgm.fr/rechercher/>

³<https://pmc.ncbi.nlm.nih.gov/tools/openftlist/>

⁴<https://arxiv.org>

⁵<http://dlmf.nist.gov/LaTeXML/>

Outline

Introduction

Table extraction methods

Evaluation

Results and analysis

Table Detection

Table Structure Recognition

Table Extraction

Conclusion

Confidence scores

Baseline No confidence scores. $\mathcal{P}^+ = \mathcal{P}$. We can compute $P, R \dots$

Object detection With confidence scores. We have to define a set of positive predictions $\mathcal{P}_{\theta_c}^+$ from \mathcal{P} . We can then compute $P, R \dots$

We use a threshold θ_c to define *positive* predictions from models with confidence scores.

$$\mathcal{P}_{\theta_c}^+ := \{\hat{y} \mid (\hat{y}, c) \in \mathcal{P}, c_{\text{table}} > \theta_c\}$$

Then we obtain tuples $(P^{\theta_c}, R^{\theta_c})_{\theta_c}$

Precision–Recall curves with bounding boxes

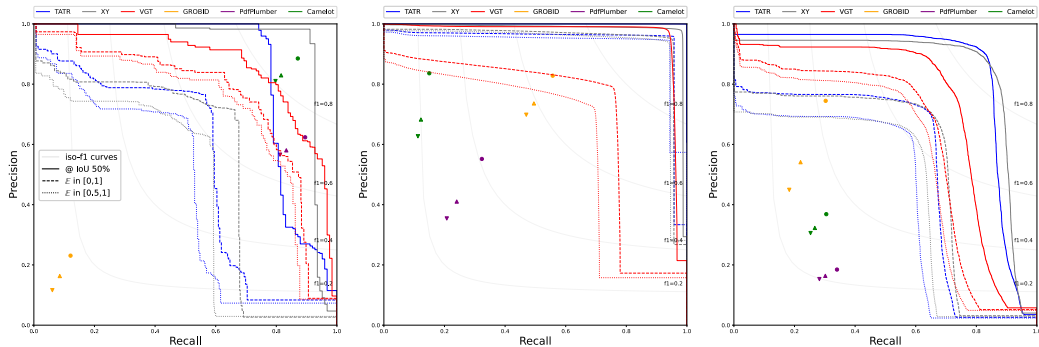


Figure: Table-BRGM, PubTables, Table-arXiv

Precision-Recall curves without bounding boxes

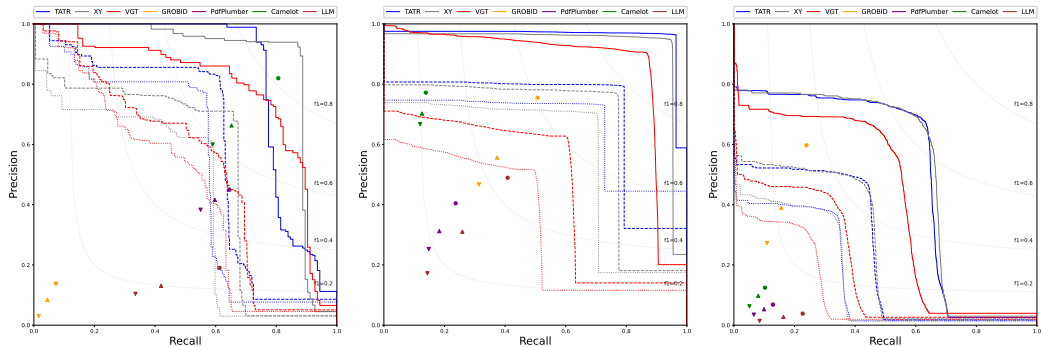
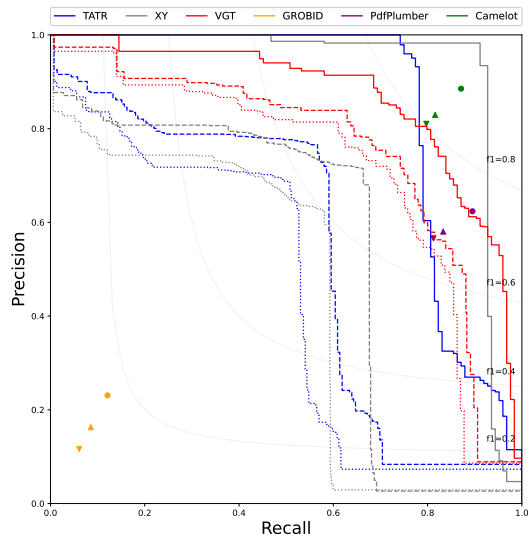
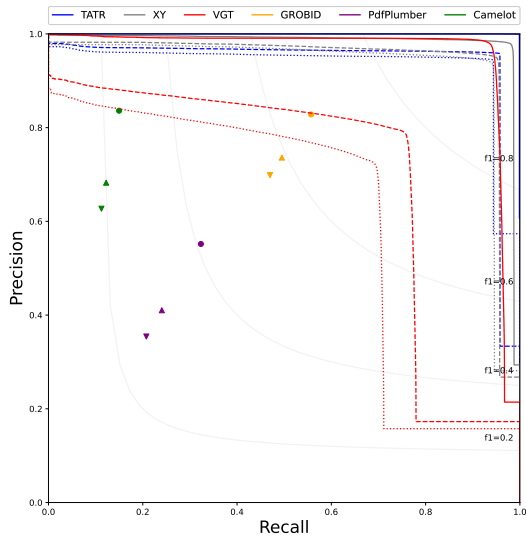


Figure: Table-BRGM, PubTables, Table-arXiv

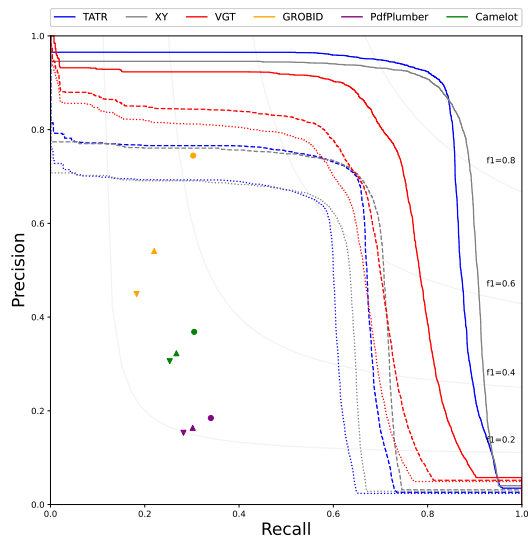
Precision–Recall curves with bboxes (Table-BRGM)



Precision–Recall curves with bboxes (PubTables)



Precision–Recall curves with bboxes (Table-arXiv)



Example: comparison TATR-detect / VGT

Evaluation de l'axe statique pour 8 sites de futurs établissements scolaires à Mayotte

Echantillon		Critères d'exclusion	
SP4 - 2 à 3 m		$D_{50} < 14 \mu\text{m}$ et $U > 10$	non
		$D_{50} < 14 \mu\text{m}$ et $U > 10$	non
		Critères de suspension	
		Si $U < 10\%$	oui
		$C_u = \frac{D_{10}}{D_{60}} < 15$	non
		$U < 0,075 < 1,5 \text{ mm}$	non
		$\sigma' < 300 \text{ kPa}$	oui
		$D_{50} > 5 \mu\text{m}$	non
		$w < 35\%$	non
		$w > 5,0\%$	non
		Soit au-dessus de la droite A du diagramme de plasticité	non

Tableau 33 - Critères de susceptibilité à la liquéfaction de l'échantillon n° 5 issu de SP43

Echantillon		Critères d'exclusion	
SP4 - 2 à 3 m		$D_{50} < 14 \mu\text{m}$ et $U > 10$	non
		Critères de suspension	
		Si $U < 10\%$	oui
		$C_u = \frac{D_{10}}{D_{60}} < 15$	non
		$U < 0,075 < 1,5 \text{ mm}$	non
		$\sigma' < 300 \text{ kPa}$	oui
		$D_{50} > 5 \mu\text{m}$	non
		$w < 35\%$	non
		$w > 5,0\%$	non
		Soit au-dessus de la droite A du diagramme de plasticité	non

Tableau 34 - Critères de susceptibilité à la liquéfaction de l'échantillon n° 7 issu de SP44

Echantillon		Critères d'exclusion	
SP4 - 3 à 6 m		$D_{50} < 14 \mu\text{m}$ et $U > 10$	oui
		Critères de suspension	
		Si $U < 10\%$	non
		$C_u = \frac{D_{10}}{D_{60}} < 15$	non
		$U < 0,075 < 1,5 \text{ mm}$	non
		$\sigma' < 300 \text{ kPa}$	oui
		$D_{50} > 5 \mu\text{m}$	non
		$w < 35\%$	non
		$w > 5,0\%$	non
		Soit au-dessus de la droite A du diagramme de plasticité	non

Tableau 35 - Critères de susceptibilité à la liquéfaction de l'échantillon n° 2 issu de SP44

Echantillon		Critères d'exclusion	
SP4 - 3 à 6 m		$D_{50} < 14 \mu\text{m}$ et $U > 10$	non
		Critères de suspension	
		Si $U < 10\%$	oui
		$C_u = \frac{D_{10}}{D_{60}} < 15$	non
		$U < 0,075 < 1,5 \text{ mm}$	non
		$\sigma' < 300 \text{ kPa}$	oui
		$D_{50} > 5 \mu\text{m}$	non
		$w < 35\%$	non
		$w > 5,0\%$	non
		Soit au-dessus de la droite A du diagramme de plasticité	non

Tableau 36 - Critères de susceptibilité à la liquéfaction de l'échantillon n° 3 issu de SP44

BPMMP-61175-FR - Rapport final

71

Evaluation de l'axe statique pour 8 sites de futurs établissements scolaires à Mayotte

Echantillon		Critères d'exclusion	
table 83%		$D_{50} < 14 \mu\text{m}$ et $U > 10$	non
		Critères de suspension	
		Si $U < 10\%$	oui
		$C_u = \frac{D_{10}}{D_{60}} < 15$	non
		$U < 0,075 < 1,5 \text{ mm}$	non
		$\sigma' < 300 \text{ kPa}$	oui
		$D_{50} > 5 \mu\text{m}$	non
		$w < 35\%$	non
		$w > 5,0\%$	non
		Soit au-dessus de la droite A du diagramme de plasticité	non

Tableau 33 - Critères de susceptibilité à la liquéfaction de l'échantillon n° 5 issu de SP43

Echantillon		Critères d'exclusion	
table 83%		$D_{50} < 14 \mu\text{m}$ et $U > 10$	non
		Critères de suspension	
		Si $U < 10\%$	oui
		$C_u = \frac{D_{10}}{D_{60}} < 15$	non
		$U < 0,075 < 1,5 \text{ mm}$	non
		$\sigma' < 300 \text{ kPa}$	oui
		$D_{50} > 5 \mu\text{m}$	non
		$w < 35\%$	non
		$w > 5,0\%$	non
		Soit au-dessus de la droite A du diagramme de plasticité	non

Tableau 34 - Critères de susceptibilité à la liquéfaction de l'échantillon n° 7 issu de SP44

Echantillon		Critères d'exclusion	
table 83%		$D_{50} < 14 \mu\text{m}$ et $U > 10$	non
		Critères de suspension	
		Si $U < 10\%$	oui
		$C_u = \frac{D_{10}}{D_{60}} < 15$	non
		$U < 0,075 < 1,5 \text{ mm}$	non
		$\sigma' < 300 \text{ kPa}$	oui
		$D_{50} > 5 \mu\text{m}$	non
		$w < 35\%$	non
		$w > 5,0\%$	non
		Soit au-dessus de la droite A du diagramme de plasticité	non

Tableau 35 - Critères de susceptibilité à la liquéfaction de l'échantillon n° 2 issu de SP44

Echantillon		Critères d'exclusion	
table 87%		$D_{50} < 14 \mu\text{m}$ et $U > 10$	non
		Critères de suspension	
		Si $U < 10\%$	oui
		$C_u = \frac{D_{10}}{D_{60}} < 15$	non
		$U < 0,075 < 1,5 \text{ mm}$	non
		$\sigma' < 300 \text{ kPa}$	oui
		$D_{50} > 5 \mu\text{m}$	non
		$w < 35\%$	non
		$w > 5,0\%$	non
		Soit au-dessus de la droite A du diagramme de plasticité	non

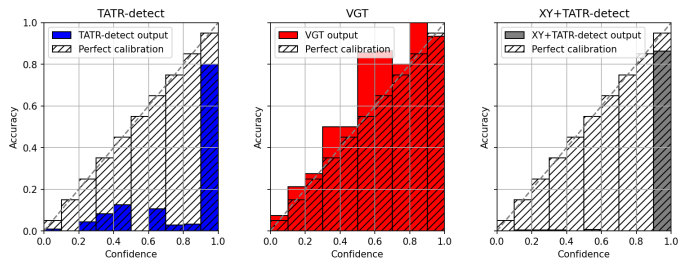
Tableau 36 - Critères de susceptibilité à la liquéfaction de l'échantillon n° 3 issu de SP44

BPMMP-61175-FR - Rapport final

71

Model calibration

Should we trust confidence scores from models?



Reliability diagrams (Niculescu-Mizil & Caruana, 2005)

Outline

Introduction

Table extraction methods

Evaluation

Results and analysis

Table Detection

Table Structure Recognition

Table Extraction

Conclusion

On which dataset should we evaluate TSR?

- The TSR part **depends** on the TD for evaluation.
- We decided to compute average TSR score on the **set of *True Positive***: tuples (predicted table, ground truth table), setting with IoU @ 50%.

TSR histograms scores

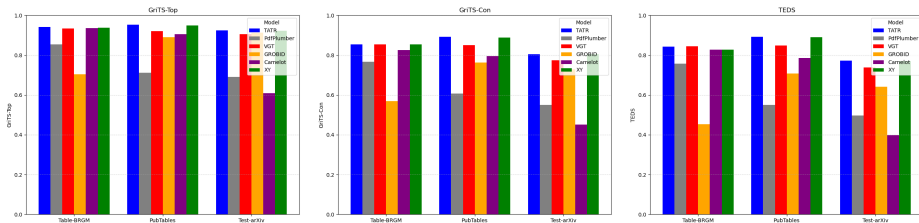


Figure: GrITS Topology, GrITS Content and TEDS.

Example: End-to-end extraction with TATR-extract

Suzanne Beech et al.

Table 3. Description of the EC2 soil classes used in this paper

1	0	0	Description of soil profile	$F_{1,0}$ parameter (%)
A			Back or other rock (geological horizon), including at least 3 m of similar material at surface	>100
B			Deposits of very dense ground, or very stiff clay, at least several tens of cm in thickness, characterized by a gradual increase of mechanical properties with depth	300-800
C			Deep deposits of dense or medium dense sand, gravel or stiff clay with thickness from several tens to many hundreds of cm	100-300
D			Deposits of loose to medium cohesionless soils, or of soft to medium stiff cohesive layers, or of predominantly soft to stiff fine-grained soils with cohesionless soil	<100
E			A soil profile consisting of a surface alluvium layer with $V_{1,0}$ values of type C or D and thickness of up to several tens of cm to 20 m, overlain by a stiffer material with $V_{1,0}$ values of type B or C	>100

Table 4. Statistical values for f_0 and $V_{1,30}$ parameters distribution (data count, Q25, Q50 and Q75) and 90% soil class according to the classified medium

[illegible]

The V_{1300} parameter distribution (Figure 12) shows a quite similar trend for the four analyzed geological formations with most of the V_{1300} distribution (i.e., the interval between Q25 and Q75) ranging from 230 to 300 ms. The autochthonous altered volcanic formations (alluvites and tuffites), which form the upper weathering profile on top of the fractured lava formations (see Figure 2 for a conceptual presentation of a characteristic weathering profile in Mayotte Island), present values around 270–320 ms. Both their lithological characteristics, geometry, and V_{1300} values lead us to consider them as a C soil class. The

aliothous formations classified as slope formations (mainly colluvium) present higher $V_{30,0}$ values between 380 and 390 m/s. These formations grade the sides of the relief; their thickness is prone to increase downstream and can reach 30 m at particular areas of accumulation. They present different facies from fine colluvium to boulder colluvium with strong lateral variations of facies and thickness. At the bottom of the relief, these formations tend to rest directly on bedrock with a clear contrast between the two formations. Those formations are thus difficult to classify following EC8 criteria since they can

Table Table (rotated)

(a) TATR-detect

Table 3. Description of the EC3 soil classes used in this paper

Soil class	Description of soil profile	V ₅₀ range (m/s)
A	Rock or other rock-like geological formation, including at most 5 m of weaker material on surface	>800
B	Deposits of very dense gravel, or very stiff clay, at least several tens of m in thickness. Characterized by a gradual increase of mechanical properties with depth	360–630
C	Deep deposits of dense or medium-dense sand, gravel or stiff clay with stiffness from several tens to many hundreds of MPa	180–360
D	Deposits of loose to medium cohesionless soil (both with and without some soft cohesive layers), or of predominantly soft-to-firm cohesive soil	<180
E	A soil profile consisting of a surface alluvium layer with V ₅₀ values of type Cor D and stiffness varying between about 5 and 10 to 20 m. Underlain by stiffer material	

Table 4. Statistical values for ξ_0 and V_{∞} parameters distribution (data count, Q25, Q50 and Q75) and

■ Data cell ■ Column header cell ■ Projected row header cell

(b) TATR-structure

Soil class	Description of soil profile	V 5.30 parameter (m/h)
A	Rock or other rock-like geological formation, including at most 5 m of weaker material at surface	800
B	Deposits of very dense gravel, or very stiff clay, at least several tens of m in thickness, characterized by a gradual increase of mechanical properties with depth	260-800
C	Deep deposits of coarse- to medium-size sand, gravel or stiff clay with thickness from several tens to many hundreds of m	180-360
D	Deposits of loose- to medium- cohesionless soil (with or without some soft cohesive layers), or of predominantly soft-to-firm cohesive soil	~180
E	Soil profile consisting of a surface alluvium layer with V 5.30 values of type C or D and thicknesses varying between about 5 m to 50 m, underlain by stiffer material with V 5.30 ~ 800 m	

(c) Extracted table (HTML)

Outline

Introduction

Table extraction methods

Evaluation

Results and analysis

Table Detection

Table Structure Recognition

Table Extraction

Conclusion

TE Precision–Recall curves with bboxes (Table-BRGM)

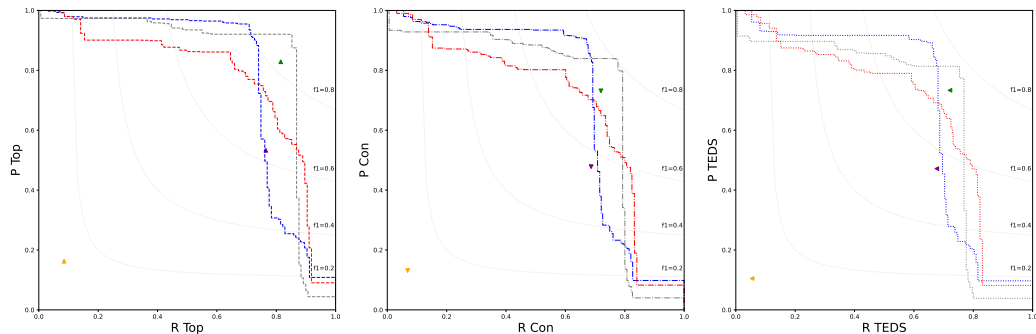


Figure: $P^{\text{TSR}} - R^{\text{TSR}}$ curves for GriTS Topology, GriTS Content and TEDS.

TE Precision–Recall curves with bboxes (PubTables)

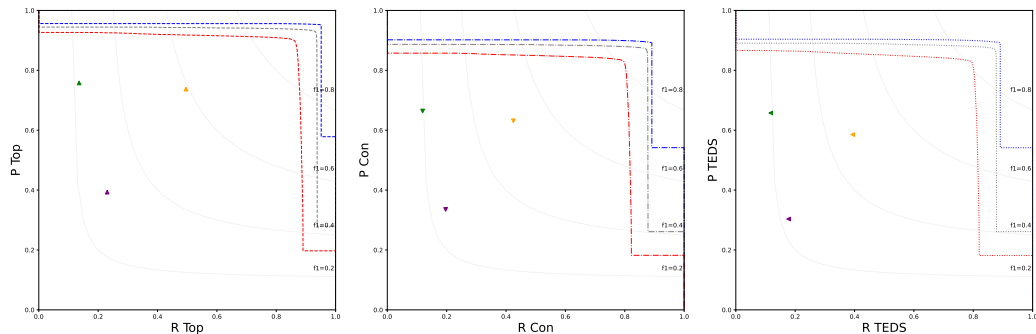


Figure: $P^{\text{TSR}} - R^{\text{TSR}}$ curves for GriTS Topology, GriTS Content and TEDS.

TE Precision–Recall curves with bboxes (Table-arXiv)

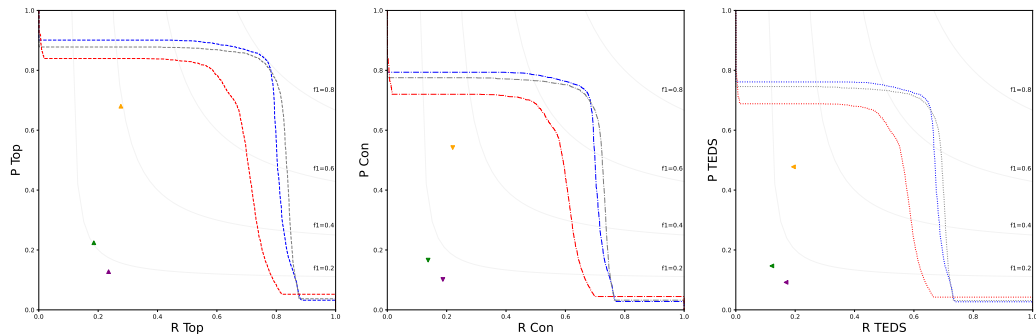


Figure: $P^{\text{TSR}} - R^{\text{TSR}}$ curves for GriTS Topology, GriTS Content and TEDS.

TE evaluation for models with scores

	Models	AP	AP ^{Top}	AP ^{Con}	AP ^{TEDS}
BRGM	TATR	0.84	0.77	0.69	0.66
	VGT	0.86	0.76	0.67	0.65
	XY	0.92	0.83	0.71	0.67
PubTab	TATR	1.00	0.91	0.80	0.80
	VGT	0.96	0.81	0.69	0.70
	XY	0.97	0.88	0.78	0.78
arXiv	TATR	0.84	0.73	0.56	0.52
	VGT	0.73	0.60	0.44	0.40
	XY	0.85	0.73	0.56	0.51

TE evaluation for models without scores

	Models	F_1	F_1^{Top}	F_1^{Con}	F_1^{TEDS}
BRGM	Camelot	0.88	0.82	0.73	0.73
	GROBID	0.16	0.11	0.09	0.07
	PdfPlumber	0.73	0.63	0.56	0.56
PubTab	Camelot	0.25	0.23	0.20	0.20
	GROBID	0.67	0.59	0.51	0.47
	PdfPlumber	0.41	0.29	0.25	0.22
arXiv	Camelot	0.33	0.20	0.15	0.13
	GROBID	0.43	0.39	0.31	0.28
	PdfPlumber	0.24	0.17	0.13	0.12

Conclusion

- Problem not solved
- We developed new framework for TE evaluation, built datasets and compared various methods
- In downstream tasks, two choices:
 - Use a threshold θ_c in order to define \mathcal{P}^+
 - Trust confidence scores




Outlook





This work:

- Semantize tables through their content, context and captions
- Perform Q/A on tables

My thesis:

- Focus on other data types
- Exploite **heterogeneous** data through **multi-modal** methods
- Locate knowledge **spatially and temporally**

-  Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers.
<https://arxiv.org/abs/2005.12872>
-  Da, C., Luo, C., Zheng, Q., & Yao, C. (2023). Vision grid transformer for document layout analysis. <https://arxiv.org/abs/2308.14978>
-  Ha, J., Haralick, R., & Phillips, I. (1995). Recursive x-y cut using bounding boxes of connected components. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 2, 952–955 vol.2.
<https://doi.org/10.1109/ICDAR.1995.602059>

-  Li, M., Cui, L., Huang, S., Wei, F., Zhou, M., & Li, Z. (2020, May). TableBank: Table benchmark for image-based table detection and recognition. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the twelfth language resources and evaluation conference* (pp. 1918–1925). European Language Resources Association.
<https://aclanthology.org/2020.lrec-1.236/>
-  Lopez, P. (2008). Grobid.
-  Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. *Proceedings of the 22nd International Conference on Machine Learning*, 625–632. <https://doi.org/10.1145/1102351.1102430>
-  Smock, B., Pesala, R., & Abraham, R. (2022). Pubtables-1m: Towards comprehensive table extraction from unstructured documents. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4634–4642.



Smock, B., Pesala, R., & Abraham, R. (2023). Grits: Grid table similarity metric for table structure recognition. <https://arxiv.org/abs/2203.12555>