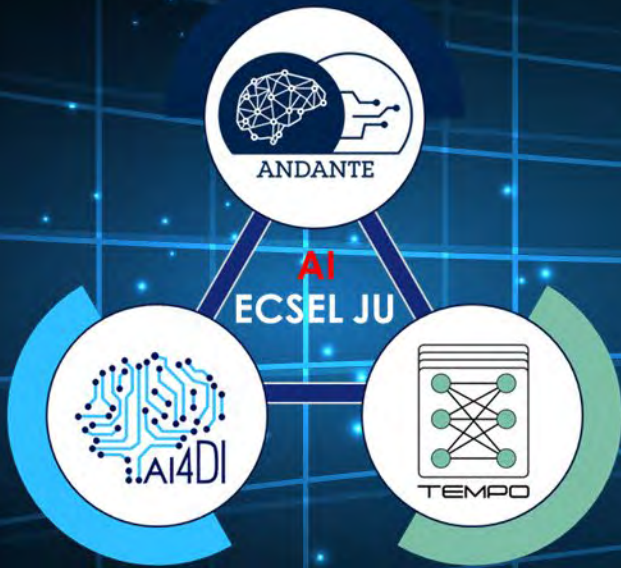
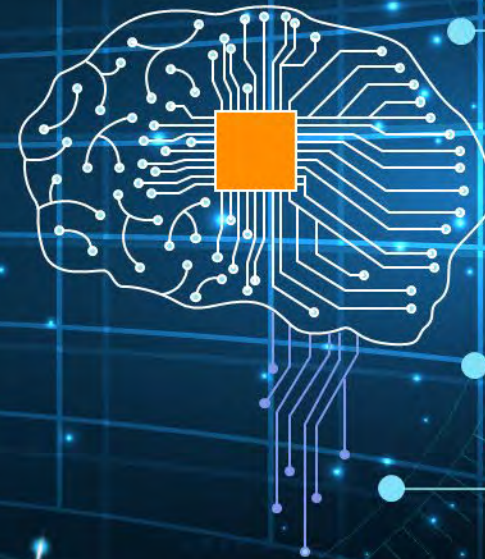


The International Workshop on Edge Artificial Intelligence for Industrial Applications (EAI4IA)

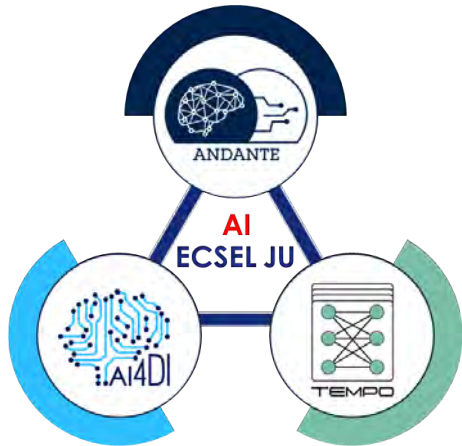


AI



Vienna, Austria
25-26 July 2022

The International Workshop on Edge Artificial Intelligence for Industrial Applications (EAI4IA)




S2ORC-SemiCause: Annotating and Analysing Causality in the Semiconductor Domain

Lan Liu, Eileen Salhofer, Anna Safont Andreu, and Roman Kern



Vienna, Austria 25-26 July 2022

Introduction: FMEA

		RPNbase:	84	If you change O and D, please make sure you have defined corresponding recommended actions for reduction of both O and D and documented it		Potential Failure Mode and Effects Analysis restricted Template version 5.0														
		FE-Process-FMEA																		
Product Name / Process Flow Name		Code / ID / Rev / Package		Requirement & Specs		FMEA Title & Version		Original FMEA Date		FMEA Revision Date		Prepared by:								
Team Members (with job titles)				Scope		Plant / Resp. Organization														
Element ID	Characteristic: (D) Customer Requirements / Function / Module	Potential Failure Mode	Potential Effect(s) of Failure Mode	Severity	Class	Potential Cause(s) / Mechanism(s) of Failure	Occurrence	Prevention Measure Design or Process for Prevention	Target Completion Date	Severity	Occurrence	Detection	RPN target	Actions Taken / Evidence of improvement / Comments (C:)		Date of Actions Taken	Severity	Occurrence	Detection	RPN
								P: D:												
								P: D:												
								P: D:												
								P: D:												
								P: D:												
								P: D:												

- What can go wrong
- What effect will it have
- ...

FMEA (Failure Mode and Effect Analysis):

- Semi-structured data for cause-effect relation
- Fundamental Tool for Quality Management.

Decreasing the Ar flow rate causes a decrease of the etch rate.

Motivation: extract cause-effect relation from free text to:

- Convert free text to structured data
- Easier access to large number of cause-effect relations
- Eventually provide input for FMEA analysis

- **Goal**: Extract cause-effect pairs from free texts

Decreasing the Ar flow rate causes a decrease of the etch rate.

- **Task**: information extraction (IE)
 - Extract and classify span
 - Extract relation
- **Method**: supervised learning
 - Requires annotated data

BECauSE

We are in serious economic trouble because of inadequate regulation

BioCause

In the case of PmrB, a normal response to mild acid pH requires not only a periplasmic histidine but also several glutamic acid residues. Therefore, regulation of PmrB activity may involve protonation of one or more of these amino acids.

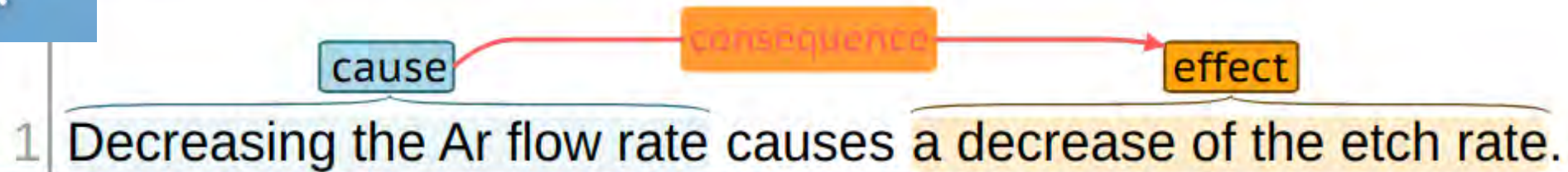
S2ORC-SemiCause

Decreasing the Ar flow rate causes a decrease of the etch rate.

- Existing annotated data in different domain:
 - News: BECauSE
 - Bio-medical: BioCause
- No existing data in semiconductor domain --> S2ORC-SemiCause

Annotation: Inter-annotator-agreement (IAA)

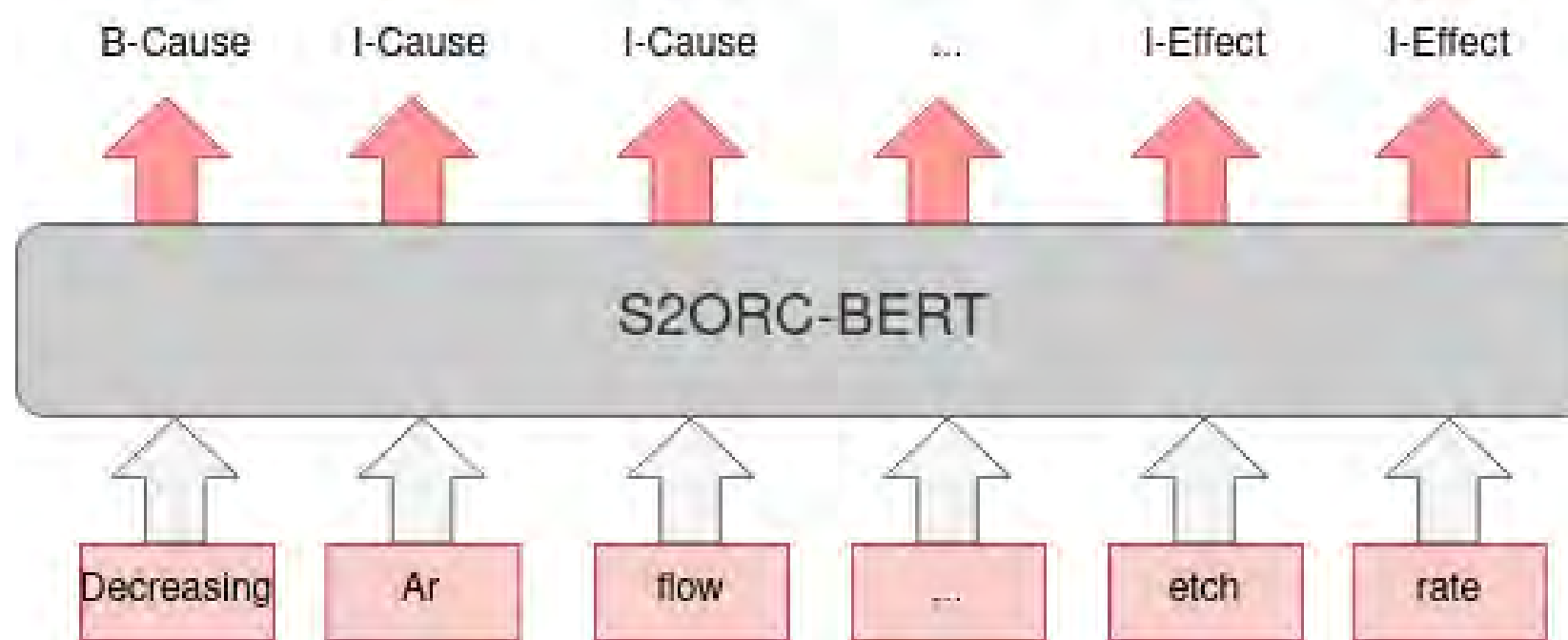
brat



	Iteration 1	Iteration 2
Relation Cohen's κ	0.65	0.80
Cause F1	0.55	0.71
Effect F1	0.60	0.81
BEcauSE span F1	0.83	Reference for machine learning performance

- Improved IAA for iter 2, due to updated guideline
- IAA as measured by span F1 comparable to BEcauSE

2 Annotators on same data; 2 iterations



Decreasing the Ar flow rate causes a decrease of the etch rate.

Baseline model:

transfer learning from pretrained-S2ORC-BERT for token-classification task

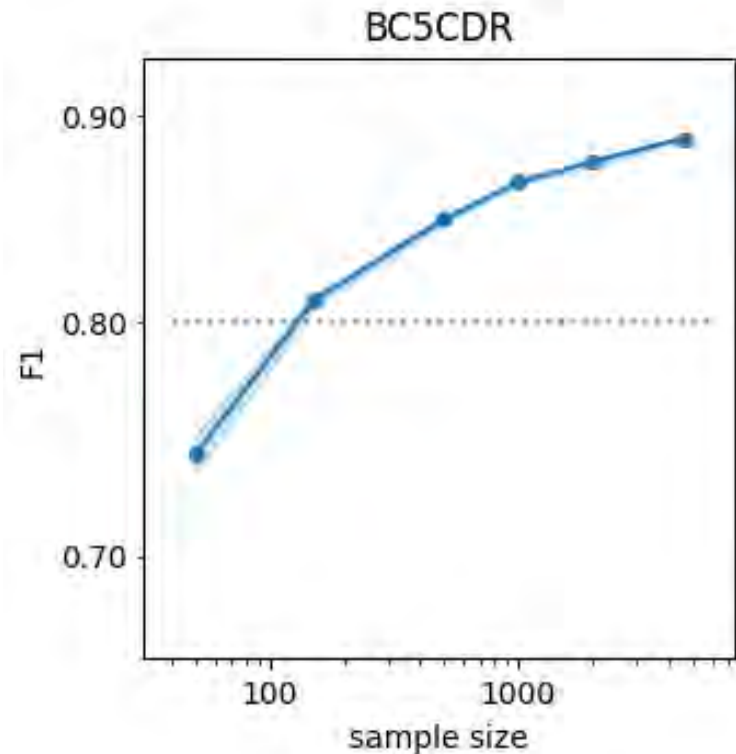
	F1	F1 partial overlap
Cause	0.48 ± 0.02	0.59 ± 0.01
Effect	0.50 ± 0.03	0.62 ± 0.02

Model

Annotation

These safe zones are provided to a model predictive controller as reference to generate feasible trajectories for a vehicle

- Baseline performance stable against random initialization
- Model performance lower than other datasets (token classification)



- Finetuning BERT typically reach ~90% F1 for token classification
 - Lower performance for smaller training data size (e.g. S2ORC-SemiCause)

	Avg span length
S2ORC-SemiCause	9.4 ± 7.2
BC5CDR	1.4 ± 1.3

- S2ORC-SemiCause dataset has much longer span length

- S2ORC-SemiCause dataset: annotated cause-effect relations for free text in semiconductor domain (scientific publications)
- Baseline model available for extracting causes and effects
- Long span length challenging for state-of-the-art token classification models

Research Areas

TRUSTWORTHY DATA-DRIVEN ARTIFICIAL INTELLIGENCE

- ML Systems for Data Science (<https://systemds.apache.org/>)
- Integrated Data Analysis Pipelines

DATA MANAGEMENT

- Cryptography (Homomorphic Encryption)
- Privacy Enhancing Technologies
- IoT Security

DATA SECURITY

- Machine Learning (Deep (Reinforcement) Learning)
- Time Series Analysis
- Natural Language Processing
- Information Retrieval

KNOWLEDGE DISCOVERY

- Recommender Systems
- (Social) Network Analysis
- Social Data Science (Fairness, Transparency)

SOCIAL COMPUTING

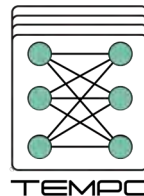
KNOWLEDGE VISUALIZATION

- Data-driven Business Models
- Process & Decision Support
- Learning Technologies

DATA-DRIVEN BUSINESS

- Visual Analytics
- Immersive Technologies
- Human-Computer Interaction

Event Organisers



The Key Digital Technologies Joint Undertaking - the Public-Private Partnership for research, development and innovation – funds projects for assuring world-class expertise in these key enabling technologies, essential for Europe's competitive leadership in the era of the digital economy. KDT JU is the successor to the ECSEL JU programme. www.kdt-ju.europa.eu

The AI4DI project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826060. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the national authorities. www.ai4di.eu

The TEMPO project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826655. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Belgium, France, Germany, The Netherlands, Switzerland. www.tempo-ecsel.eu

The ANDANTE project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876925. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Belgium, France, Germany, The Netherlands, Portugal, Spain, Switzerland. www.andante-ai.eu



Thank You

For your attention



lliu@know-center.at