

Causal Analytics for Process Scenario Mining

Anton Kolonin

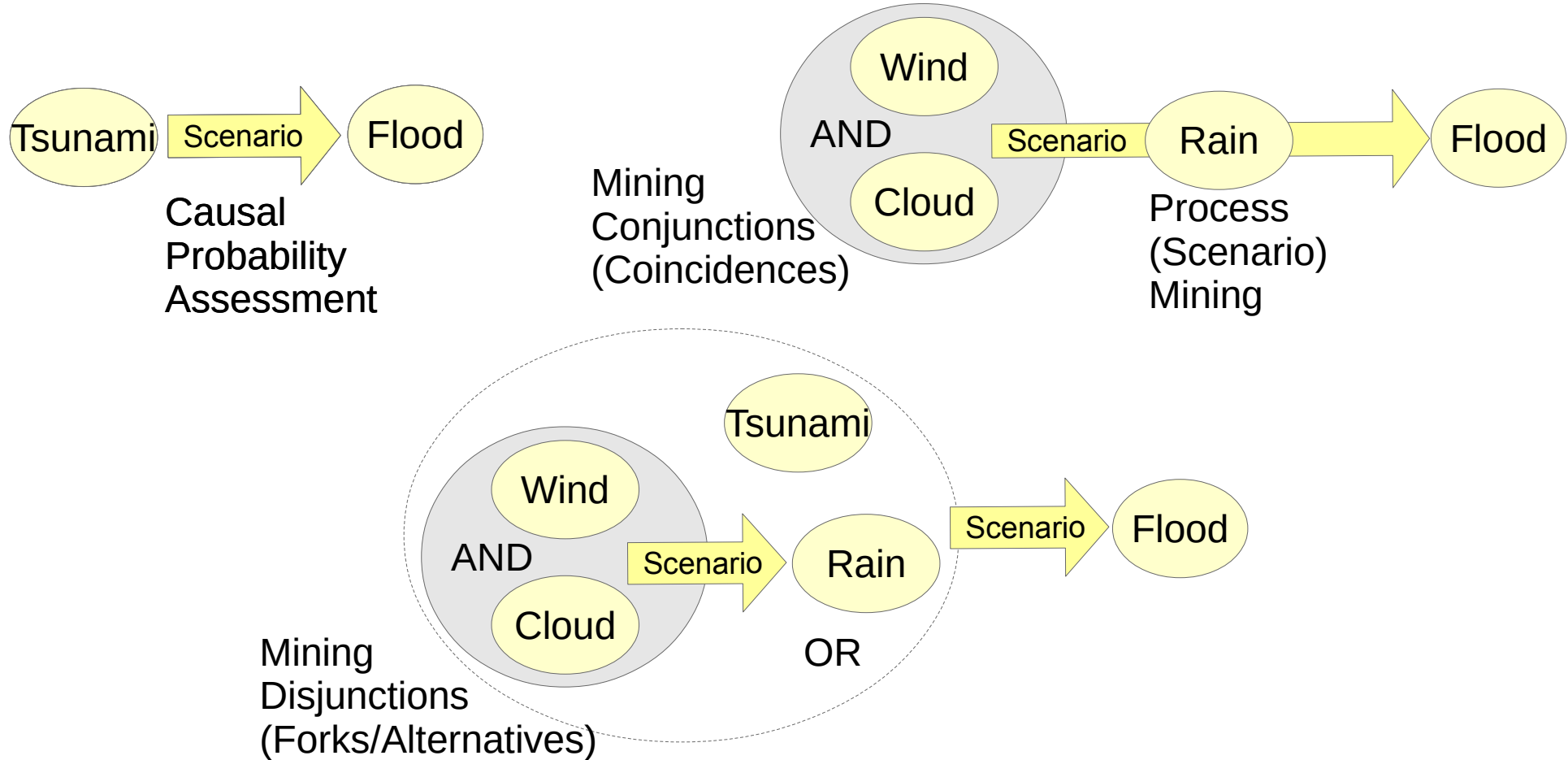
akolonin@aigents.com

Facebook: [akolonin](#)

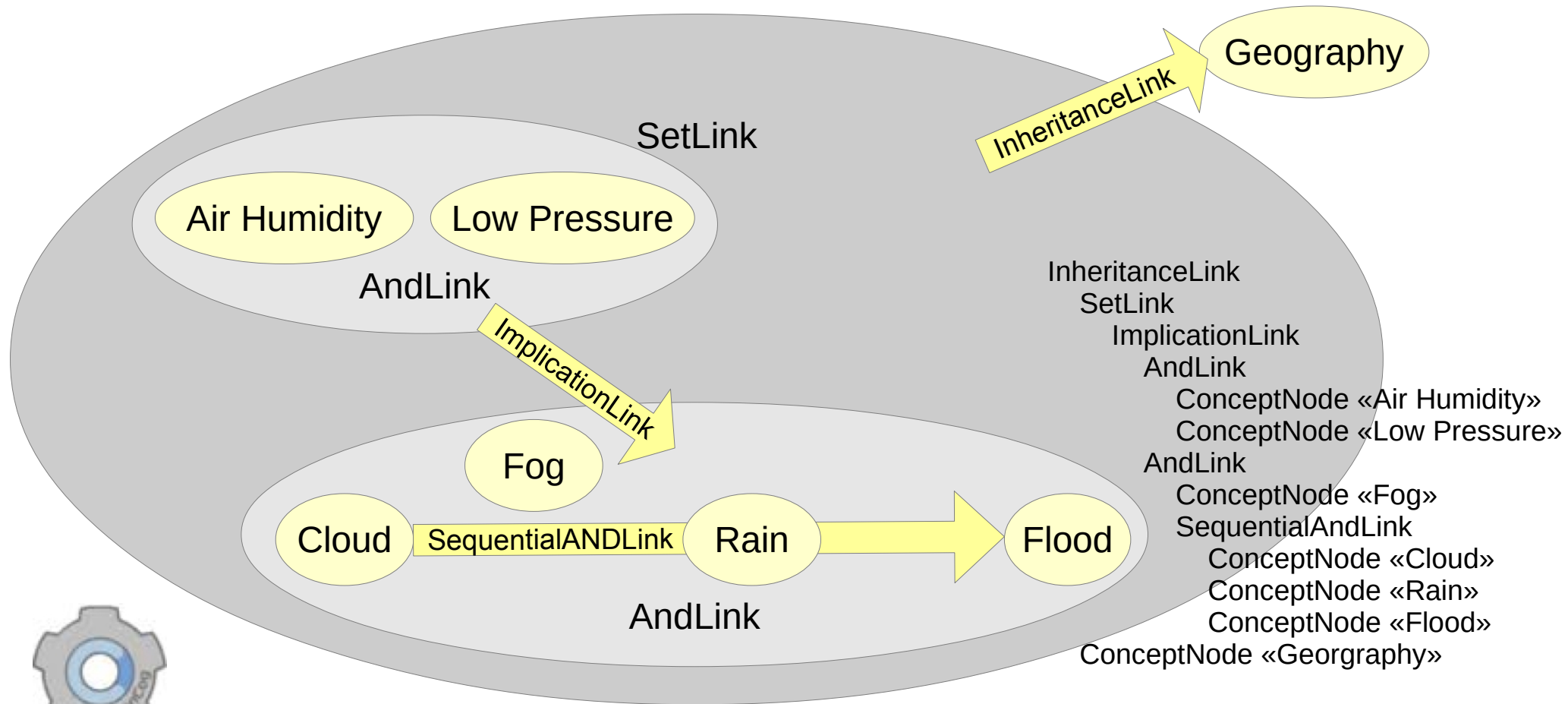
Telegram: [akolonin](#)



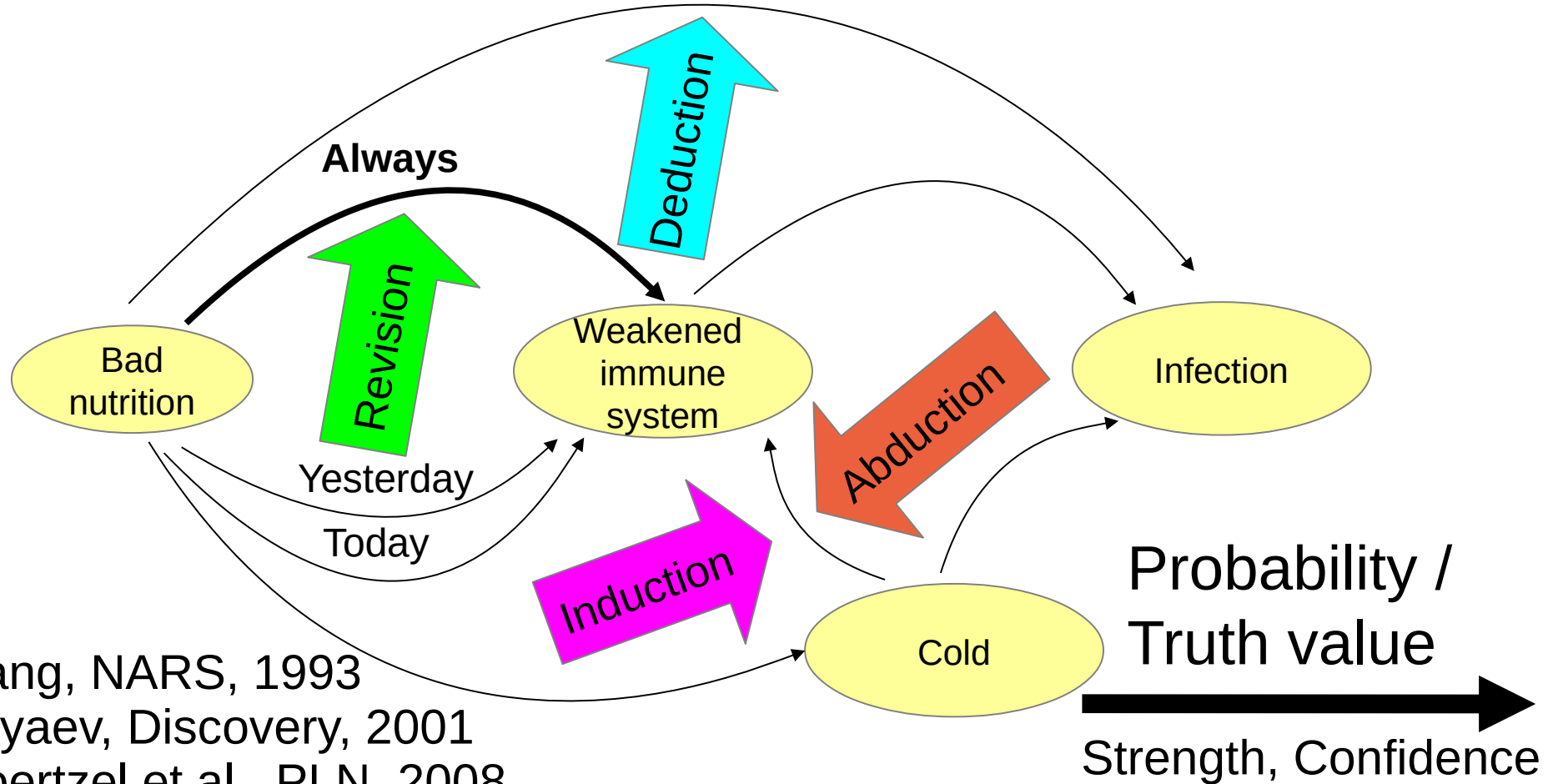
Causal Analytics Ingredients



OpenCog Hyper/Meta-Graphs

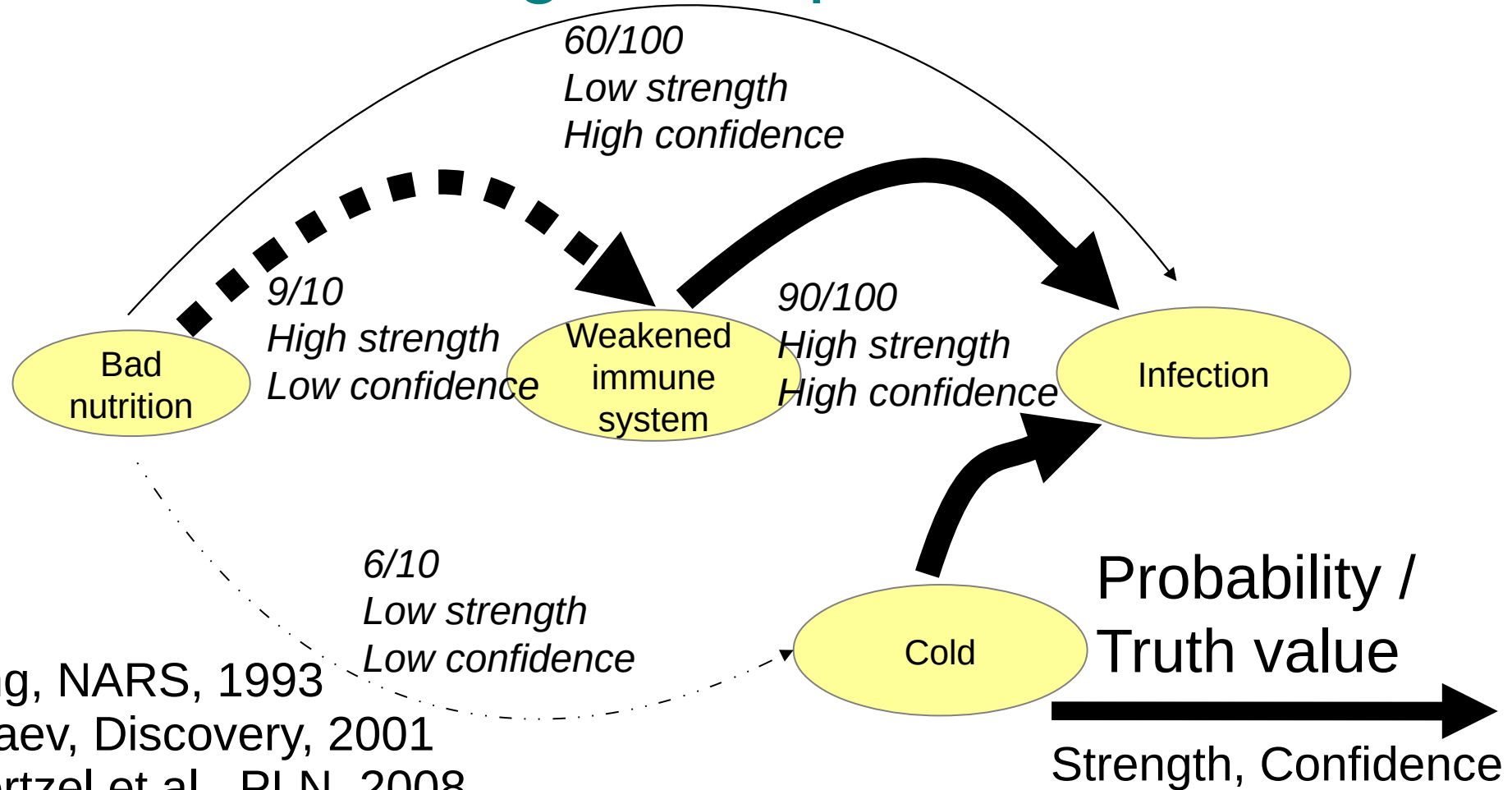


Probabilistic/Non-Axiomatic Inference



P. Wang, NARS, 1993
E. Vityaev, Discovery, 2001
B. Goertzel et al., PLN, 2008

Non-Axiomatic Logic Compound Truth Values

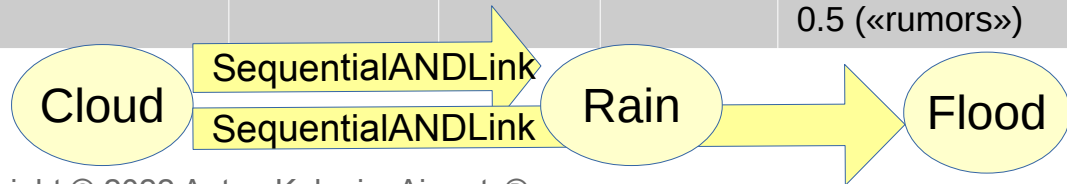


P. Wang, NARS, 1993
E. Vityaev, Discovery, 2001
B. Goertzel et al., PLN, 2008

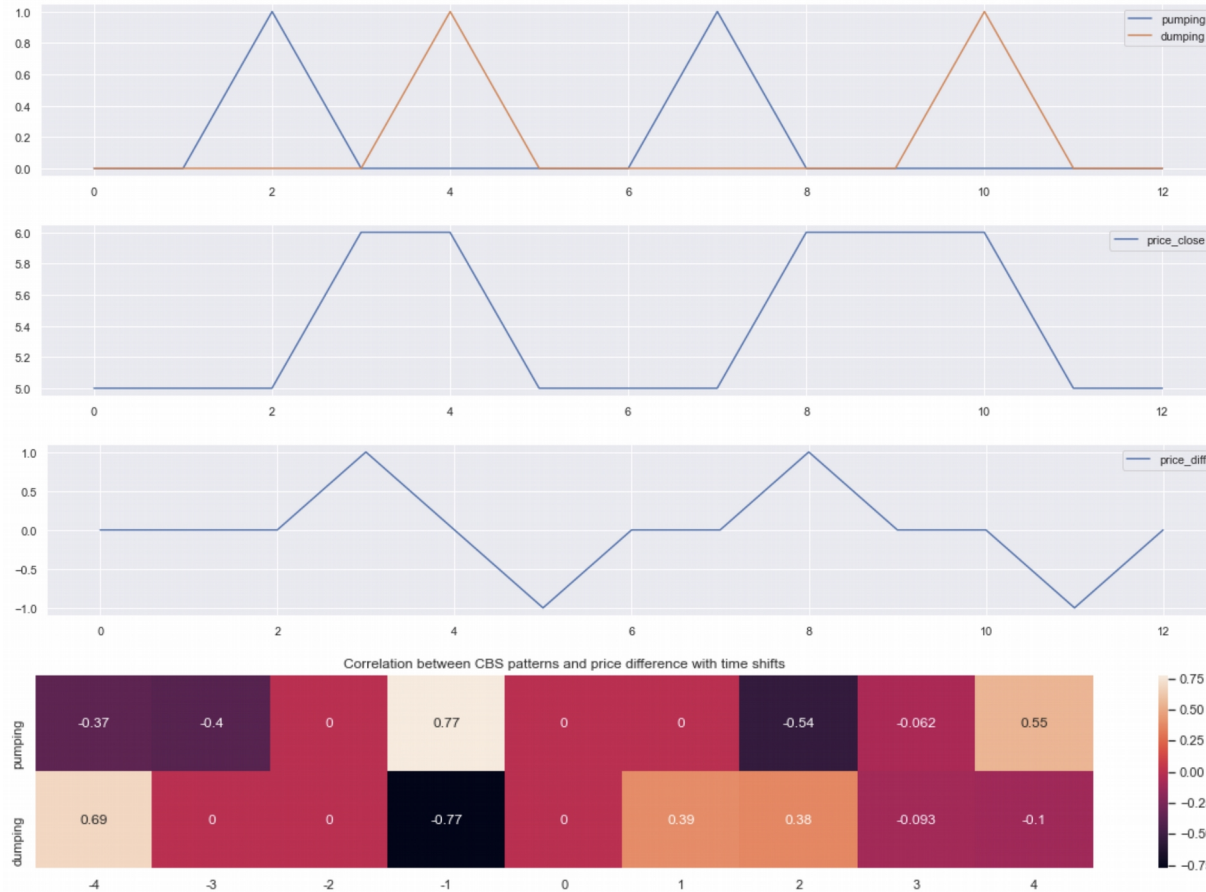
OpenCog PLN for Temporal Reasoning

Id	Type	Space	Name (Label)	Level in Meta-Graph	Arity	Arguments	Truth Value: Count (Evidence), Strength, Confidence
11	SequentialANDLink	5	-	2	2	13, 14	22 0.5 (22/44=50%) 1.0 («surely»)
12	SequentialANDLink	5	-	2	3	13, 14, 15	11 0.25 (11/44=25%) 0.5 («probably»)
13	ConceptNode	5	«Cloud»	0	0	-	44 1.0 1.0 («fact»)
14	ConceptNode	5	«Rain»	0	0	-	22 1.0 1.0 («fact»)
15	ConceptNode	5	«Flood»	0	0	-	11 1.0 0.5 («rumors»)

<https://www.youtube.com/watch?v=yky-9rZVZEQ>
AGI-2021: Uncertain Temporal Reasoning

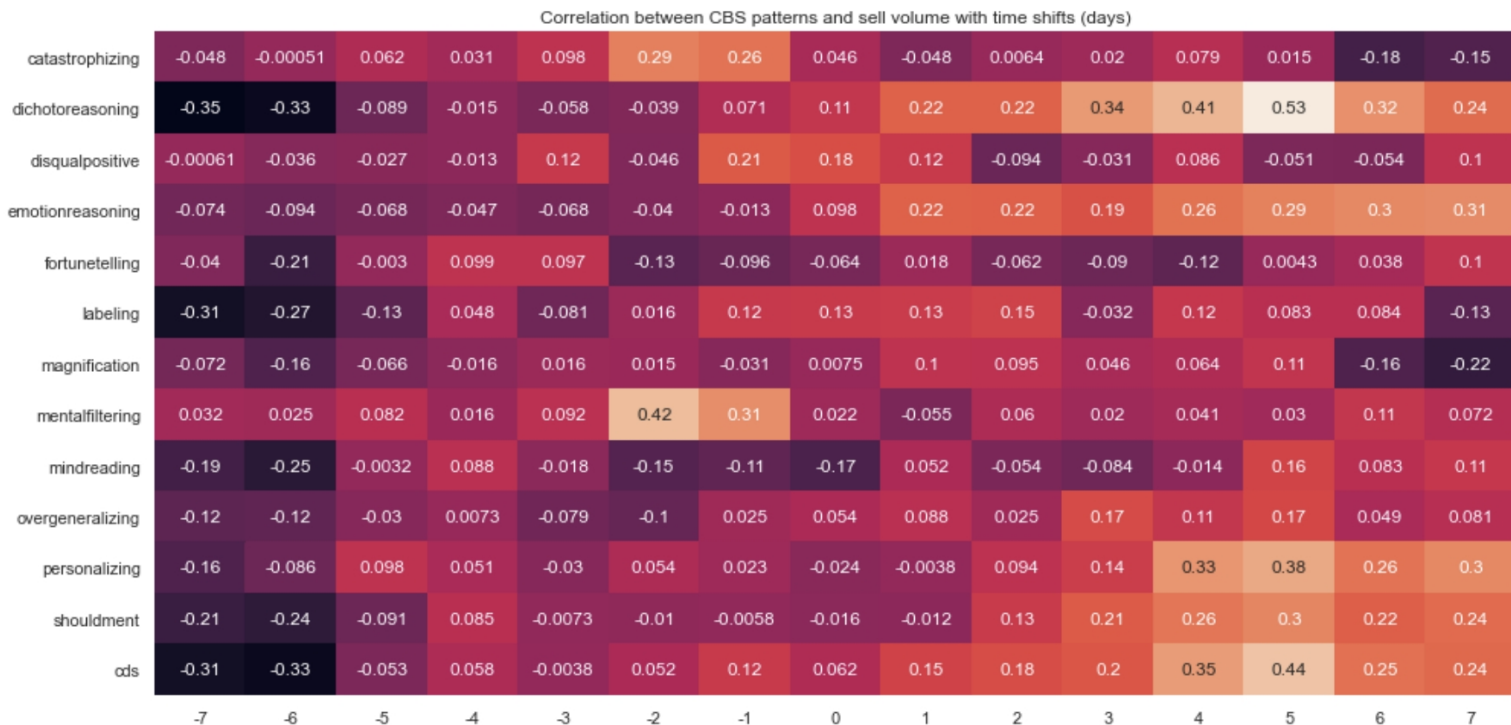


Causal Indicator Mining for Price Moves



Reverse temporal correlation as causal connection between sentiment patterns and price change

CBS patterns vs. Market Dynamics (Volume)



Tracking overall Twitter and Reddit cognitive distortions as a predictor for BTC/USD trade volume

#Catastrophizing: Exaggerating the importance of negative events

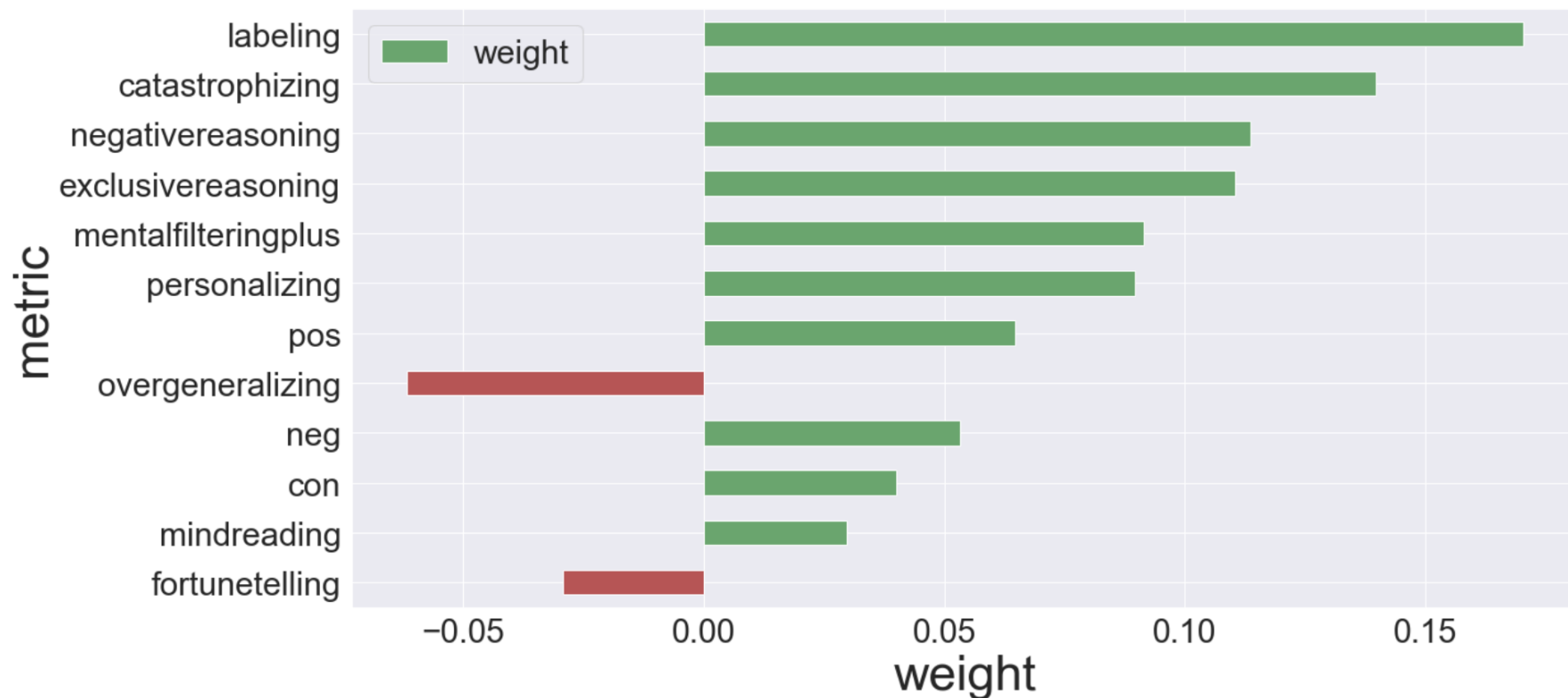
distortions['catastrophizing'] = "will fail, will go wrong, will end, will be impossible, will not happen..."

#Mental Filtering: Paying too much attention to negative details instead of the whole picture

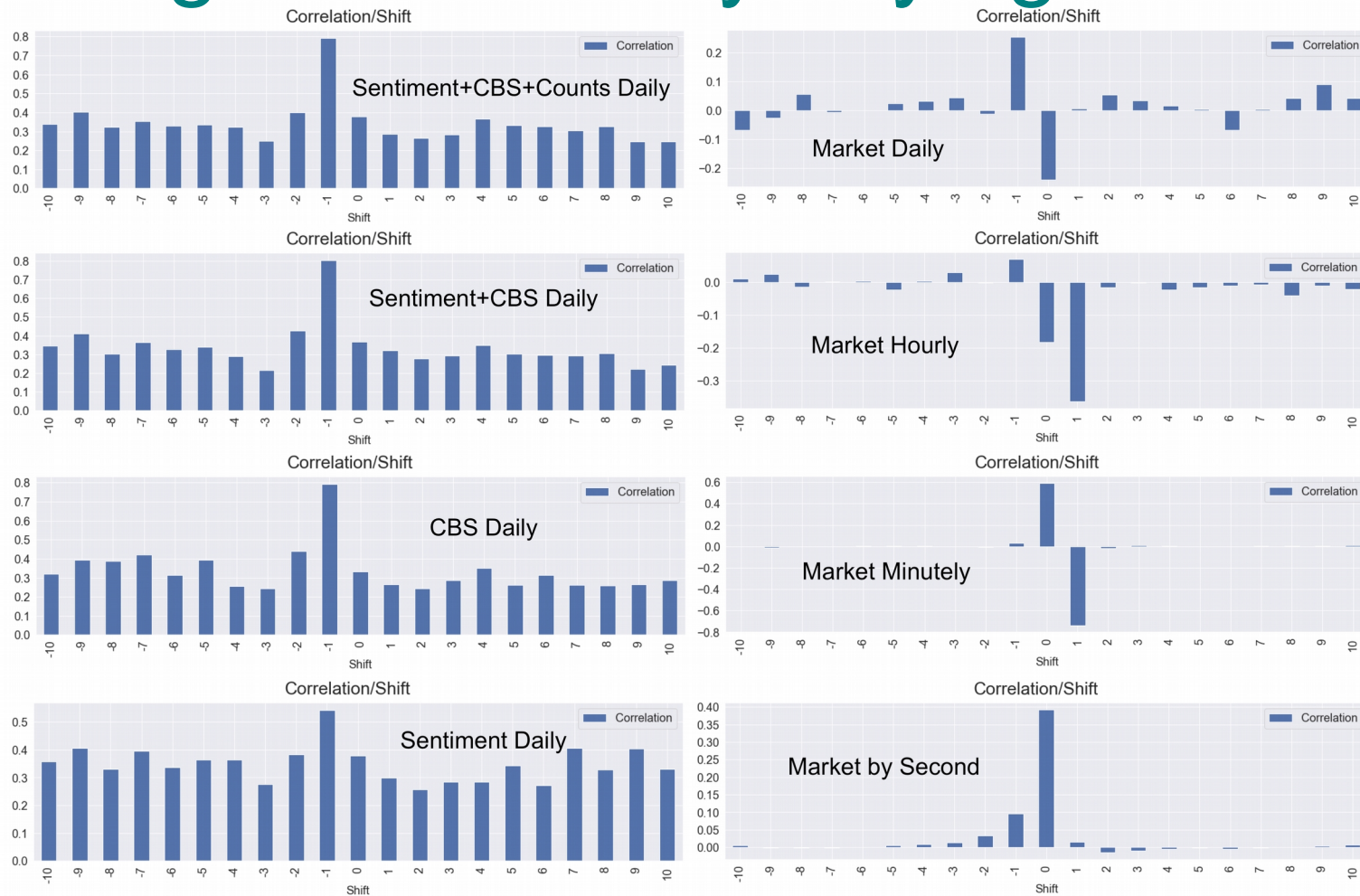
distortions['mentalfiltering'] = "I see only, all I see, all I can see, can only think, nothing good..."

Causal Weighting on Feature Basis

Weight by metric: twitter/BTCTN (1.0)



Assessing Predictability relying on Causality



Everything is a Process of a Scenario!

	Spoken language recognition	Written language recognition	Identification of patterns in the text	Causal analysis
Actor	Coefficient on spectrogram for particular frequency	Property of specific stroke: On the top, at the bottom, long, short, skewed, etc.	Object property value: Name "John"	Specific actor: John Doe
Event	Combination of coefficients on spectrogram	Period or stroke composing the letter: .	Object class instance: Name "John", surname "Doe"	Specific event: John Doe cleaning the window on the second floor.
Coincidence	Specific sound	Coincidence: i	Co-occurrence of object of class person properties: "John Doe"	Specific coincidence: Window on the second floor is dirty, John Doe is cleaning it.
Process	Specific spoken word	Specific written word: ping	Specific phrase: "John Doe cleans the window"	Specific process: Window on the second floor was dirty, John Doe has cleaned it and now it is clean.
Role	Pitch frequency	Property of symbol: orientation, extent, symmetry, etc.	Domain of the object class property: person's surname	Typical role: Cleaner
Appearance	Spectral cluster on the spectrogram	Element of symbol: .	Class of the variable object: person	Typical appearance: Someone cleaning something
Situation	Sound of speech	Symbol: i	Pattern variable: \$subject	Typical situation: Someone is cleaning something which is dirty.
Scenario	Spoken word accordingly to the language model	Written word: ping	Phrase pattern: "\$subject cleans \$object"	Typical scenario: Something was dirty, someone has cleaned it, it is clean now.

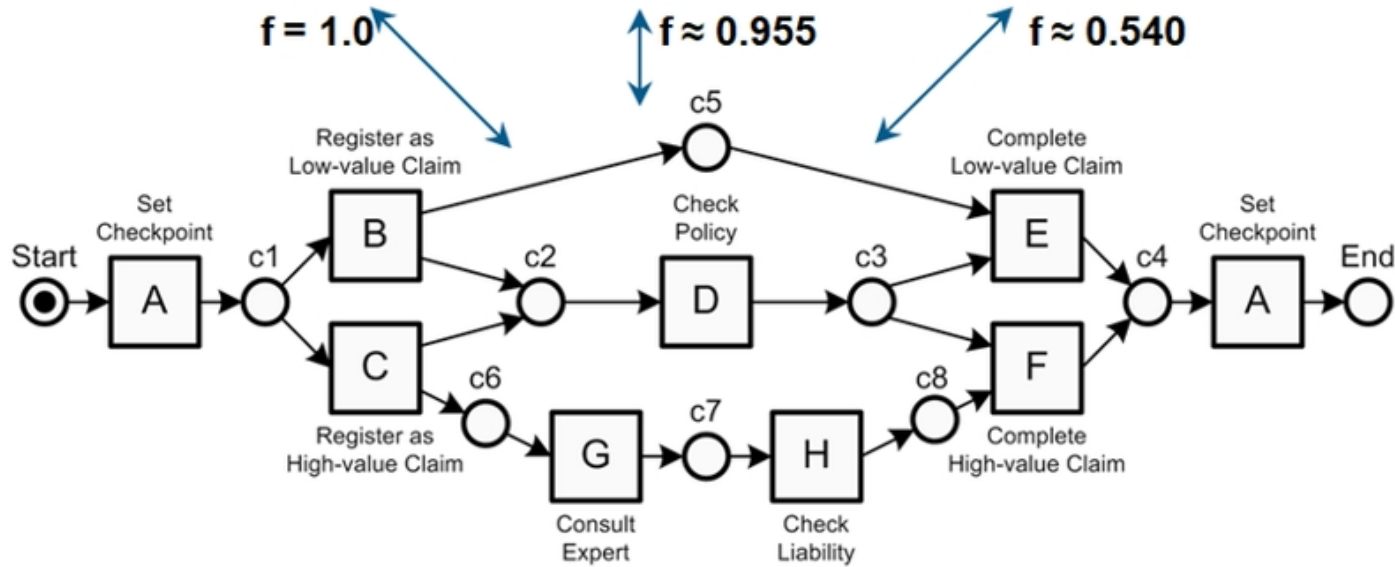
<https://arxiv.org/abs/1807.02072>

Process Mining in Business Log Data

No. of Instances	Log Traces
4070	ABDEA
245	ACDGHFA
56	ACGDHFA

No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA

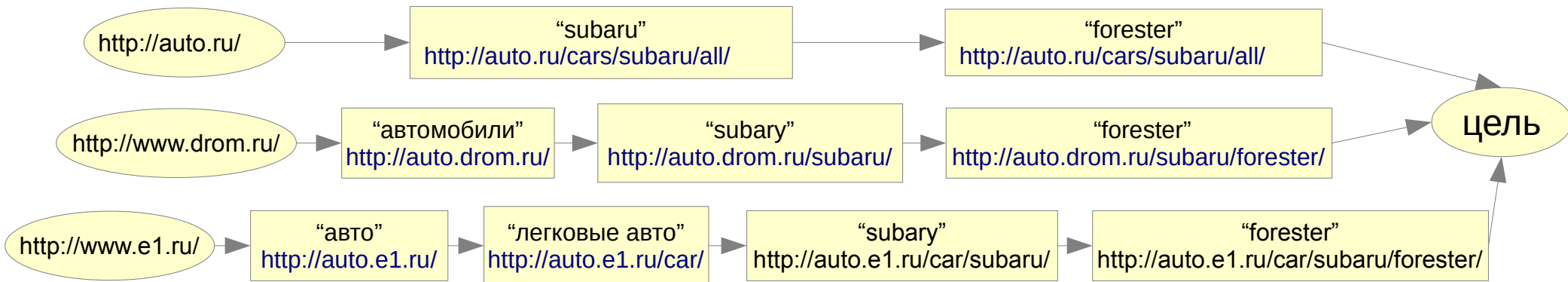
No. of Instances	Log Traces
24	BDE
7	AABHF
15	CHF
6	ADBE
1	ACBGDFAA
8	ABEDA



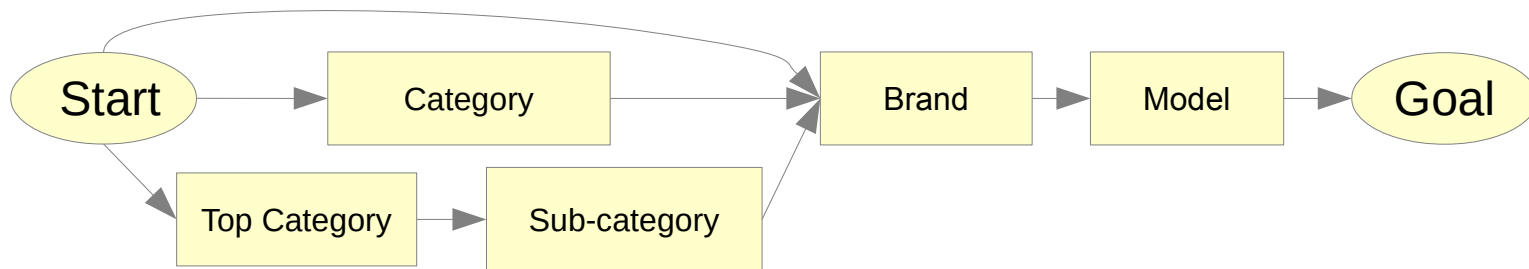
https://www.puzzledata.com/process-mining_eng/

Scenario Mining in Process Data Example

Маршруты интернет-навигации при поиске автомобилей марки “Forester”



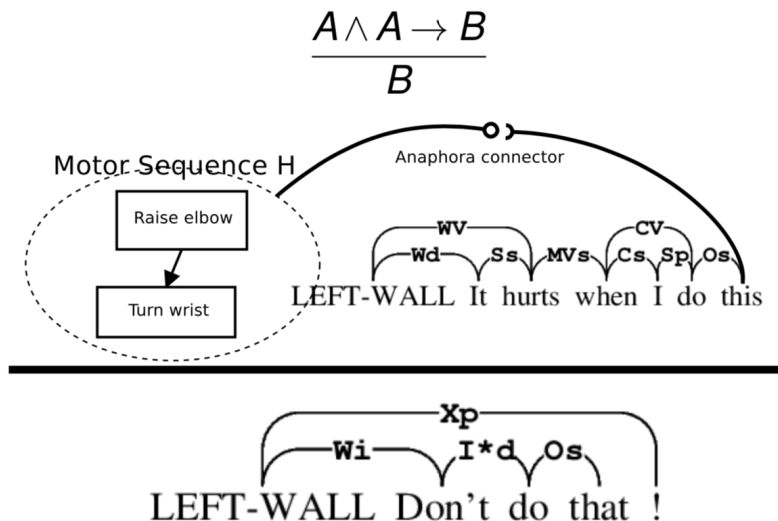
Обобщенный сценарий переходов по ссылкам при поиске товаров



Text vs. Non-Text

Common Sense Reasoning

Rules, laws, axioms of reasoning and inference can be learned.



Naively, simplistically: Learned Stimulus-Response AI (SRAI)⁹

⁹Metaphorical example: Mel'čuk's Meaning Text Theory (MTT) SemR + Lexical Functions (LF) would be better.

Applied to NLP and Language Modeling

Contrastive Evaluation: Test Specific Phenomena

To test if your LM knows something very specific, you can use contrastive examples. These are the examples where you have several versions of the same text which differ only in the aspect you care about: one correct and at least one incorrect. A model has to assign higher scores (probabilities) to the correct version.

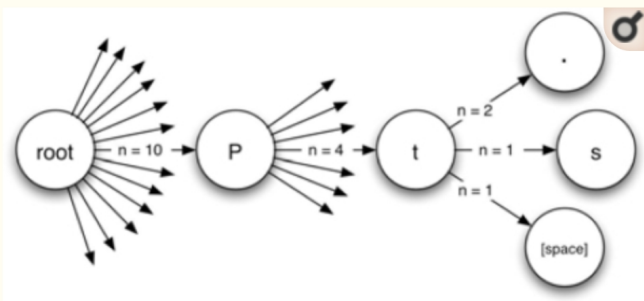


Figure 1

Trie data structure. The probability of observing an 's' given the preceding string "Pt" is $\frac{1}{4}$, or 25%. The freedom following "pt" is 3.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655800/>

<https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=6983&context=etd>

The **roses** in the vase by the door ? Competing answers: **is**, **are**

$P(\text{The roses in the vase by the door are})$

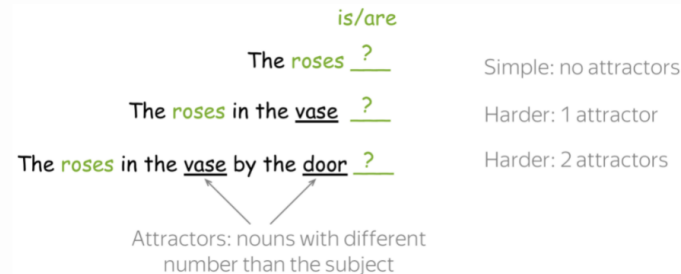
$P(\text{The roses in the vase by the door is})$

Is the correct answer ranked higher?

$P(\dots\text{are}) > P(\dots\text{is})$?

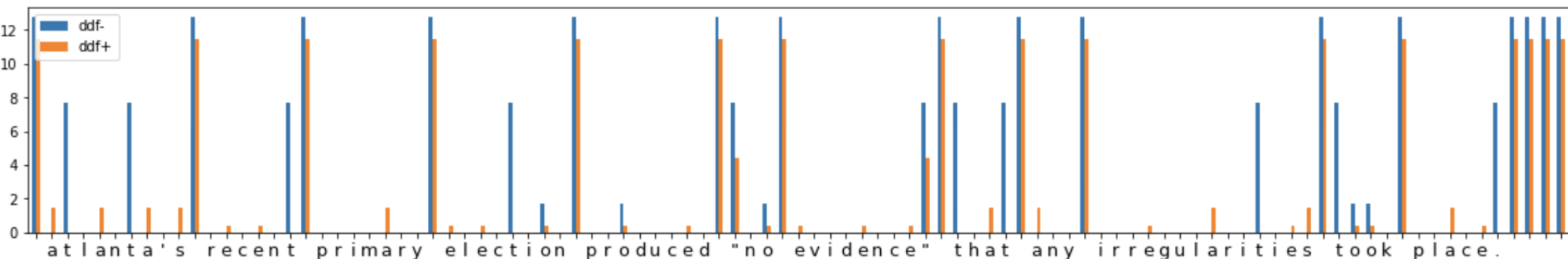
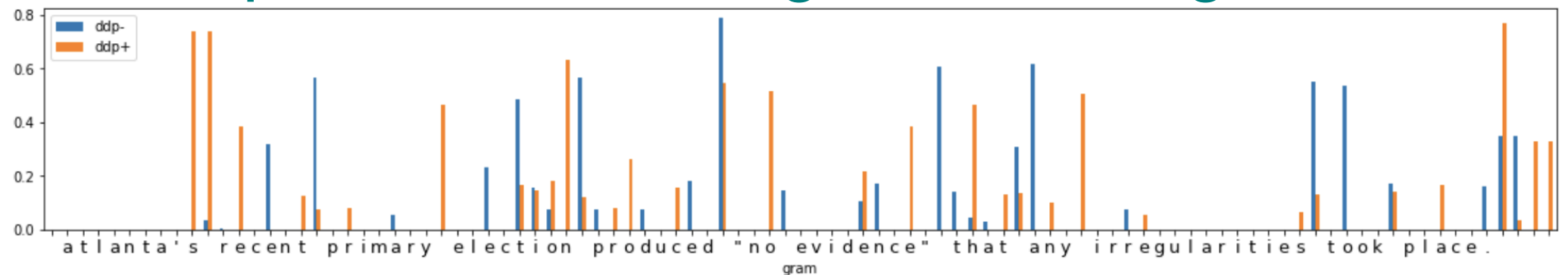
A very popular phenomenon to look at is subject-verb agreement, initially proposed in the [Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies](#) paper. In this task, contrastive examples consist of two sentences: one where the verb agrees in number with the subject, and another with the same verb, but incorrect inflection.

Examples can be of different complexity depending on the number of **attractors**: other nouns in a sentence that have different grammatical number and can "distract" a model from the subject.

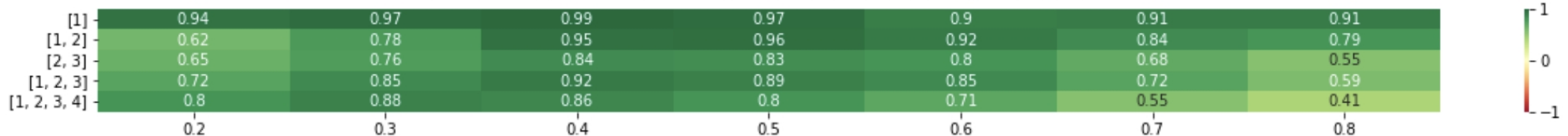


https://lena-voita.github.io/nlp_course/language_modeling.html

Unsupervised Learning for Text Segmentation



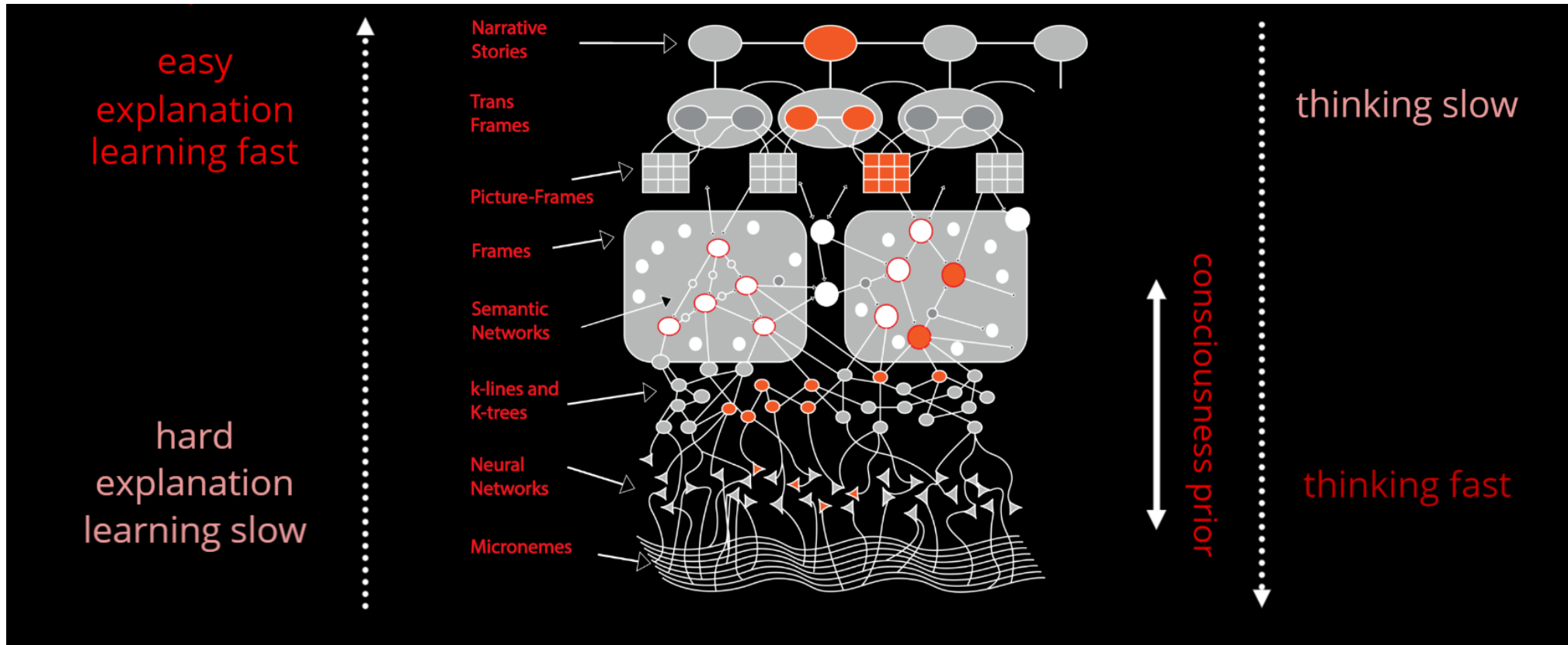
F1 - Brown filtered at 0.0001



<https://github.com/aigents/pygents/blob/main/tokenization.md>

Copyright © 2022 Anton Kolonin, Aigents®

Marvin Minsky's Narrative Story is a Scenario of a Process



<https://towardsdatascience.com/explainable-ai-vs-explaining-ai-part-1-d39ea5053347>

Evgeni Vityaev's Invariant Structures

ANNEX IN FINANCIAL FORECASTING

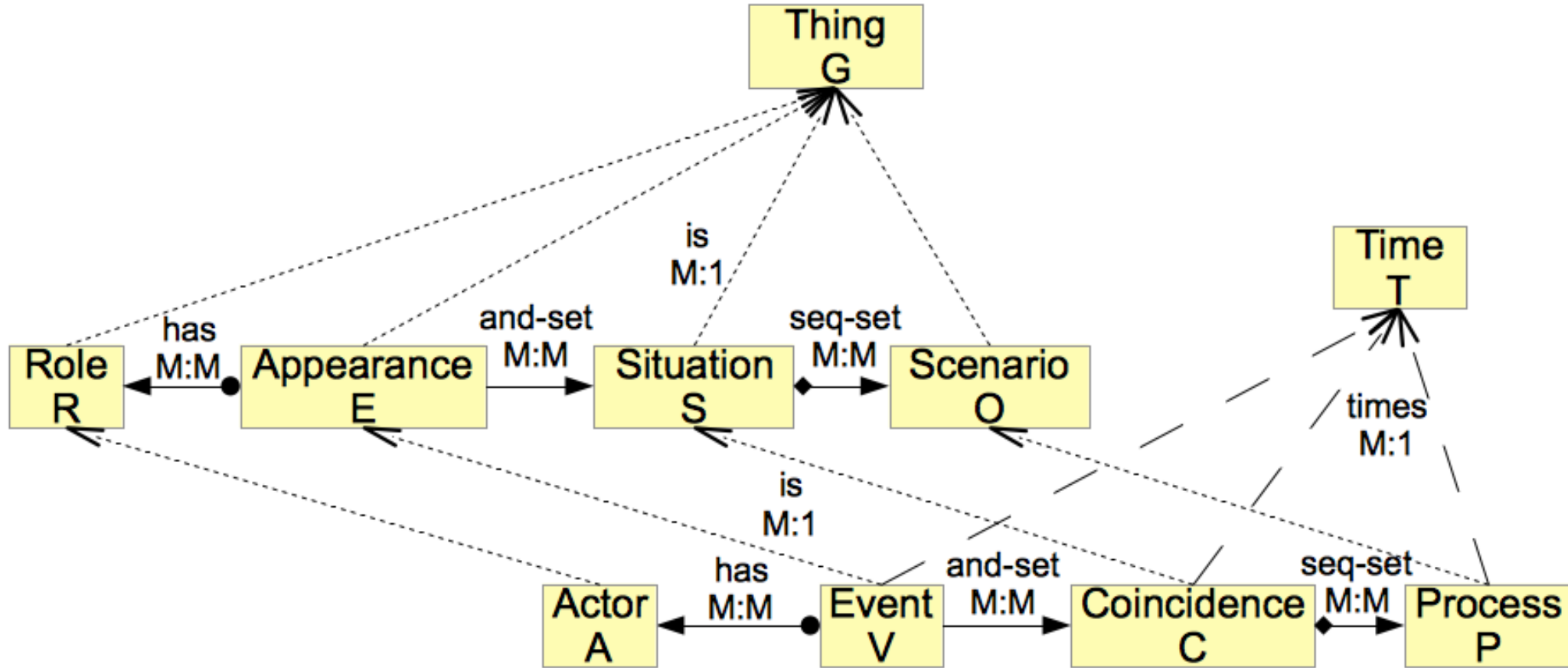
Mining temporally invariant structures (“scenarios”) in historical time series data recording financial “processes” with the **Discovery** system.

http://www.math.nsc.ru/AP/ScientificDiscovery/PDF/data_mining_for_financial_applications.pdf

<https://arxiv.org/pdf/cs/0208022.pdf>

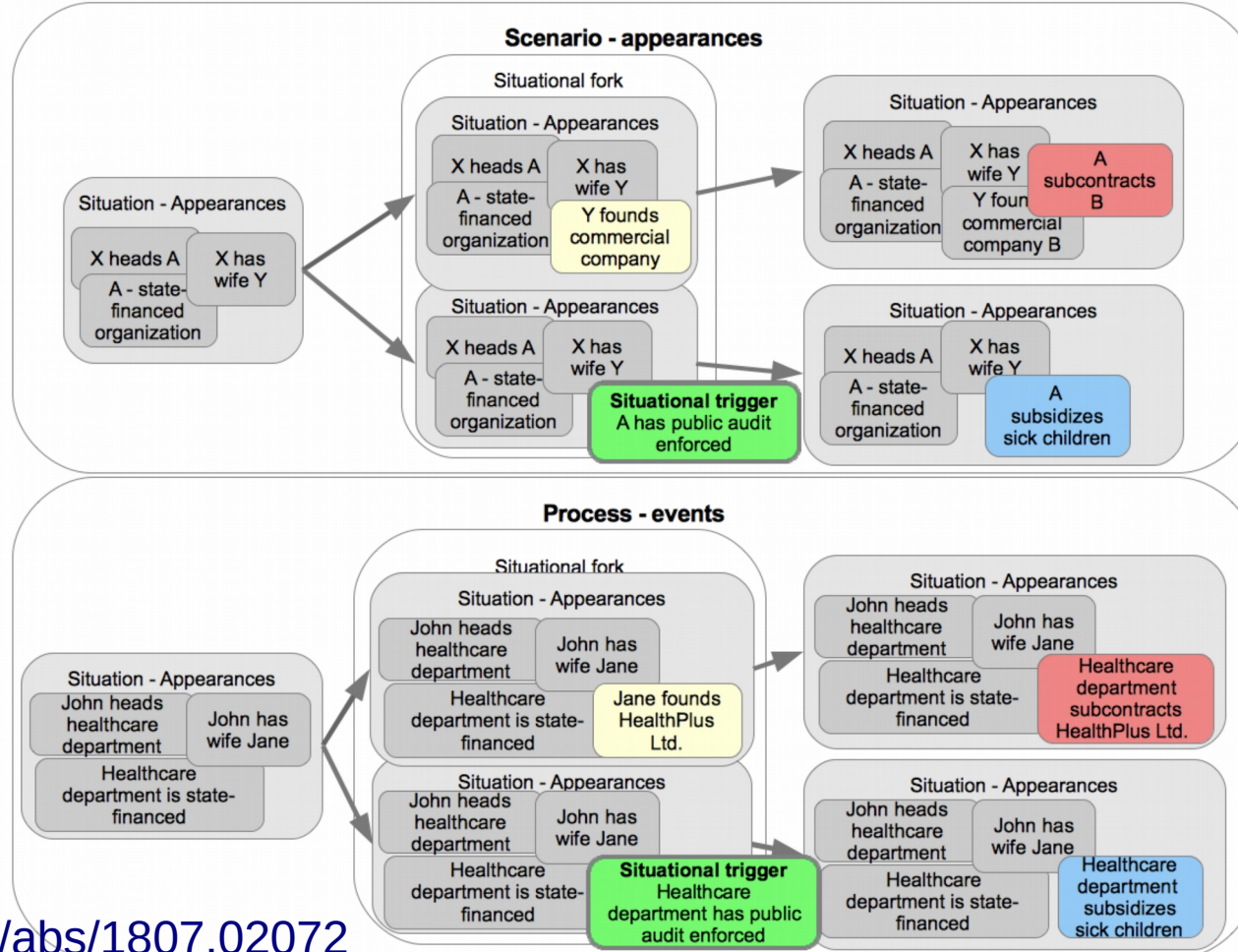
structure1	structure2	structure3	structure4	weekday	week
			forecast for	Friday	forecast week
		forecast for		Thursday	forecast week
		up	up	Wednesday	forecast week
	forecast for			Tuesday	forecast week
forecast for	up			Monday	forecast week
up		current day	current day	Friday	current week
	current day	down	down	Thursday	current week
current day	down	anchor2	anchor2	Wednesday	current week
	down anchor1	down		Tuesday	current week
down		anchor1		Monday	current week
			up	Friday	one week ago
				Thursday	one week ago
				Wednesday	one week ago
				Tuesday	one week ago
				Monday	one week ago
			up	Friday	two weeks ago
				Thursday	two weeks ago
				Wednesday	two weeks ago
				Tuesday	two weeks ago
			anchor1	Monday	two weeks ago
				Friday	three weeks ago
				Thursday	three weeks ago
				Wednesday	three weeks ago
				Tuesday	three weeks ago
				Monday	three weeks ago
training 0.74	training 0.72	training 0.7	training 0.7		
testing 0.78	testing 0.73	testing 0.71	testing 0.82		

Foundation (Generic) Activity Ontology



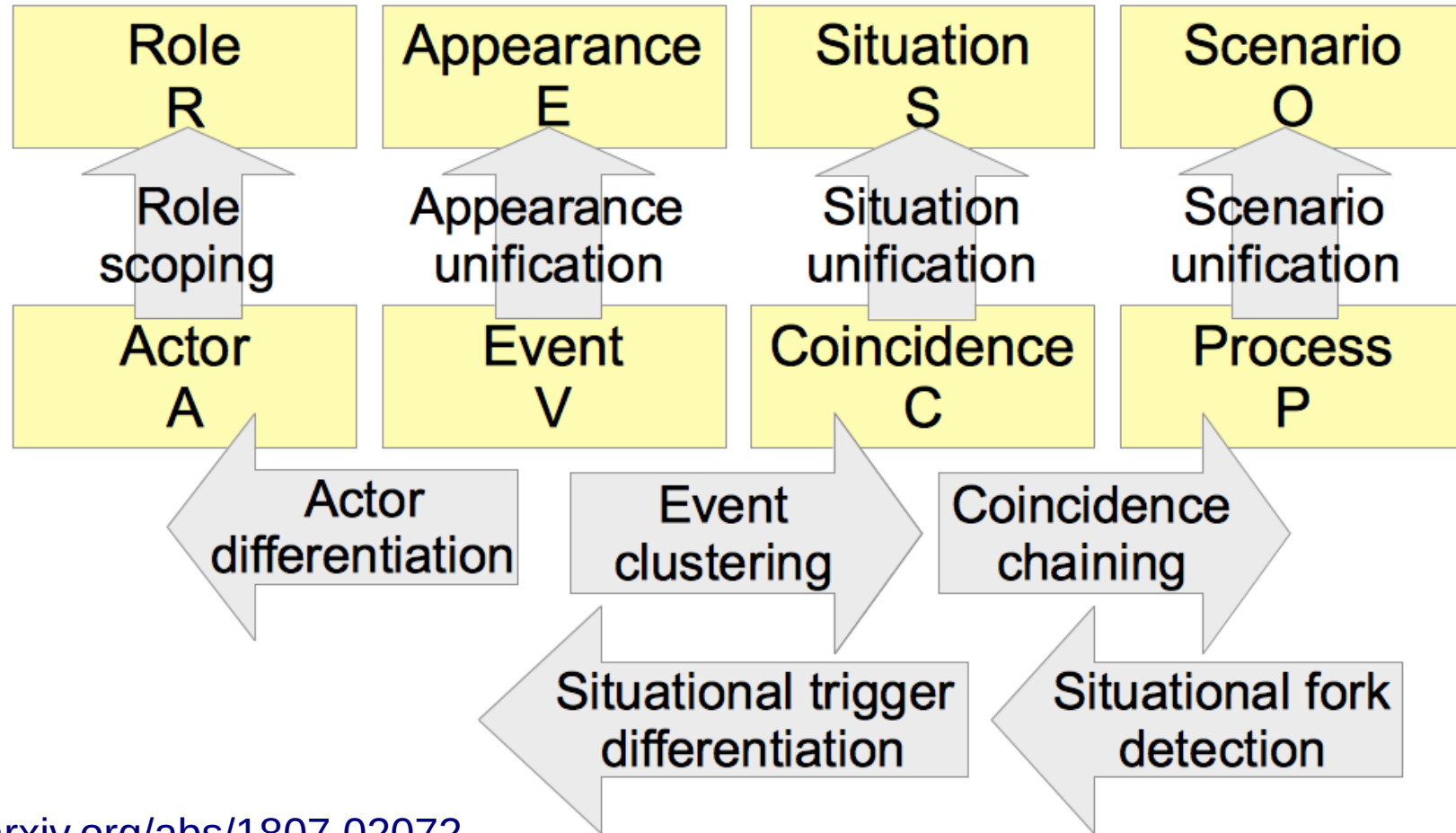
<https://arxiv.org/abs/1807.02072>

Upper (Domain) Activity Ontology Example



<https://arxiv.org/abs/1807.02072>

Scenario Mining in Process Data



<https://arxiv.org/abs/1807.02072>

Thank you and welcome!

Anton Kolonin

akolonin@aigents.com

Facebook: [akolonin](#)

Telegram: [akolonin](#)



<https://aigents.com>

