

Exploitation de LLM pour l'extraction de graphes de connaissances causaux en assurance

Abdallah Arioua, CDO chez Relyens

Aurélien Couloumy, CEO chez Dylogy



1. Contexte

Quelles sont les limites actuelles ?

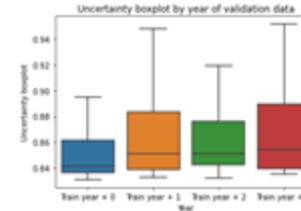
Format



Restitution



Inaltérabilité

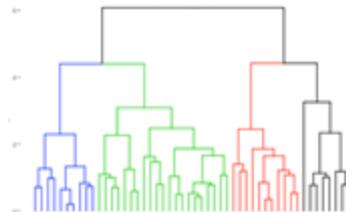


Source

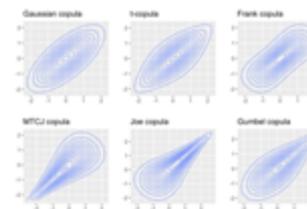


Connaissance

Agrégation



Liens



Multi-dimensionnalité



1. Contexte

Explorer la connaissance à l'aide de l'IA Gen

Objectif

- Disposer de plus de connaissances, et améliorer les applications associées via des techniques d'IA Gen.

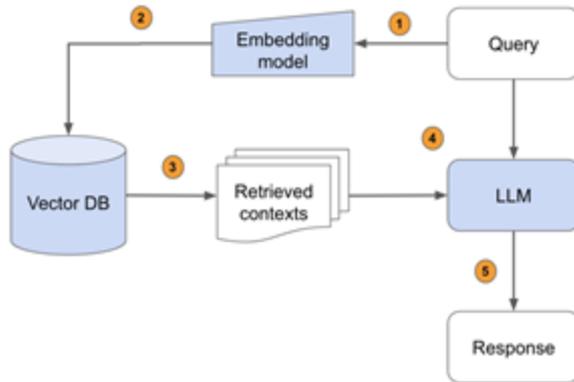


Figure: Exemple de système RAG

Périmètre

- Cas n°1 - assurance RC Médicale
- Cas n°2 - assurance Energie



Figure: site éolien et photovoltaïque

Cas d'usage

- Analyse de dossiers sinistres :

- ✓ Contexte client
- ✓ Déroulé des événements
- ✓ Impacts financiers
- ✓ Avis d'expert
- ✓ Interventions juridiques

[...]

2. Méthodes

Idées générales à mettre en place

Source et format

- Revenir aux données non structurées brutes
- Qualifier la notion d'évènement
- Atomiser l'information



Figure: exemple de rapport de sinistre

Restitution et agrégation

- Poser un modèle de données : contexte, caractéristiques d'évènement, relations.
- Tendre vers une représentation standard, idéalement codifiée.
- Agréger sous contrôle les résultats



```
{  
  "properties": {  
    "event": "Unusual wind event",  
    "event_category": "Windstorm",  
    "event_code": "GEAI",  
    "date_and_time": "2021-05-04T19:50:00",  
    "duration": "1 day",  
    "is_ambiguous": false,  
    "natcat": {  
      {  
        "natcat": true,  
        "properties": {  
          "name": "Lena",  
          "type": "Unusual wind event",  
          "characteristics": "Wind scattered damage  
across solar array"  
        }  
      }  
    }  
  },  
  ...  
}
```

Figure: exemple de résultats JSON

Inaltérabilité, liens, dimensionnalité

- Lier les évènements dans le temps et de manière causale
- Utiliser un Directed Acyclic Graph (DAG)

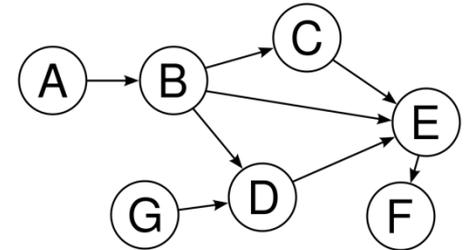
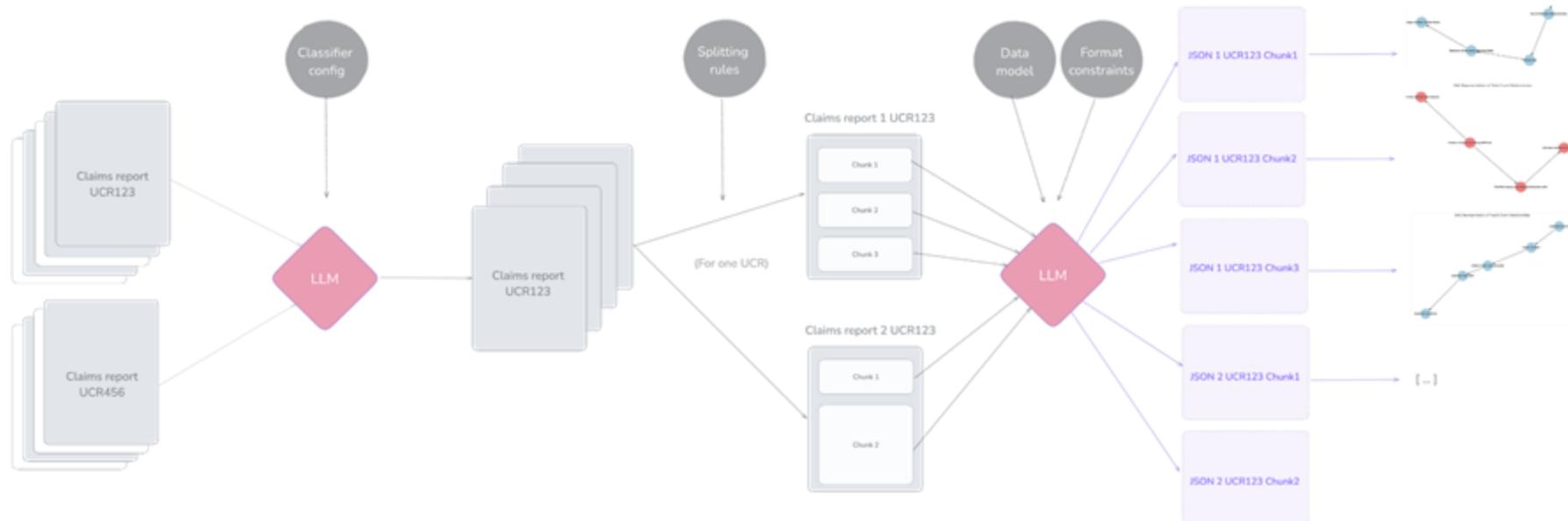


Figure: exemple de structure en DAG

2. Méthodes

Processus d'acquisition de la connaissance (1/2)



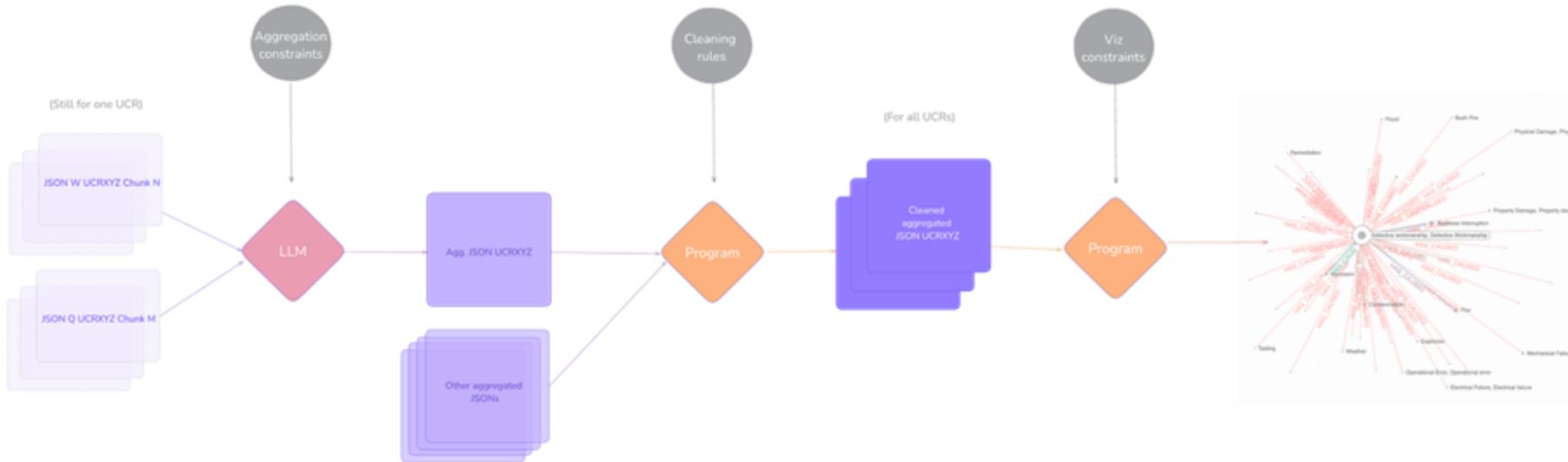
Étape 1 - Pour tous les dossiers, nous identifions les données textuelles pertinentes : fichiers PDF et rapports de sinistres.

Étape 2 - Diviser les rapports en sous-morceaux de texte (chunk) et transmettre tous les chunks associés à une UCR.

Étape 3 - Pour tous les chunks de toutes les déclarations d'une UCR, générer le JSON qui représente le graphe des événements causaux.

2. Méthodes

Processus d'acquisition de la connaissance (2/2)



Étape 4 - Pour chaque UCR, nous rassemblons les JSON pour créer un fichier agrégé unique afin d'illustrer l'ensemble du graphe causal de l'événement.

Étape 5 - En raison de la diversité des labels, nous nettoyons le contenu des agrégats avec des approches floues et sémantiques pour obtenir une information normalisée.

Étape 6 – Nous agrégeons et visualisons tous les graphes en un méta-graphe qui permet de représenter les relations entre tous les événements

2. Méthodes

Eléments techniques associés



Cas 1 - RC Médicale

Données et modèle :

Rapports de sinistres + revues scientifiques
Modèle cible + terminologies médicales pour les entités.

Pré-processing :

OCR, Deduplication, Semantic chunking, Vectorisation

Modèles :

Modèles locaux et GPTs.

Evaluation :

"LLM-as-Judge (Self-Critique)" et "HIL (human-in-the-loop)"

Cas 2 - Energie

Données et modèle :

1030 rapports de sinistres traduits en 800 graphes représentant plus de 400 sinistres uniques. Un modèle de données contenant 96 entités.

Pré-processing :

Textextract OCR, Deduplication, Fixed chunking, Vectorisation

Modèles :

Embedding (gte-large-en-v1.5). LLM (Sagemaker + Bedrock = Claude 3.5 + Llama 3 8B/70B). Cleaning (fuzzy via levenshtein & clustering dbscan)

Evaluation :

Via metadata interne et partiellement data synthetic

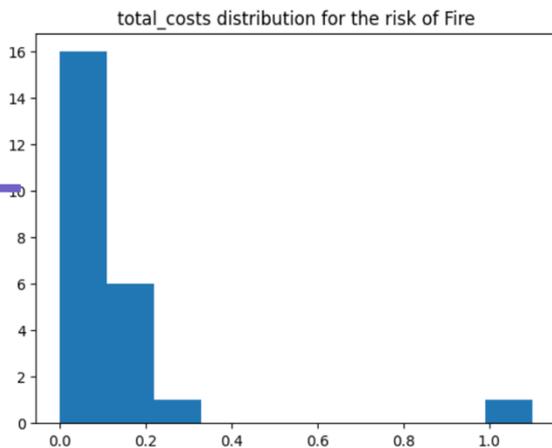
3. Résultats

Lecture du JSON agrégé (méta-graphe causal)

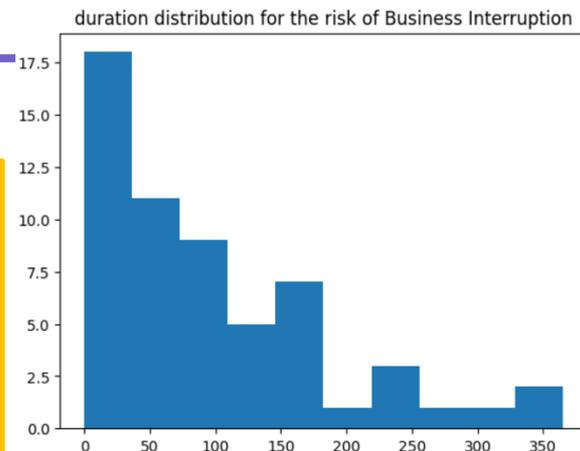
JSON

ID: 33
EVENT: FIRE

```
"event_category": [
  "Fire",
  "Fire/Smoke",
  "Firefighting",
  "Fire/Explosion",
  "Fire Control"
],
"event_code": [
  "GEAF"
],
"total_costs": [
  "2000000",
  "4500000",
  "6500000",
  "",
  "0",
  "7600000",
  "3000000",
  "8000000",
  "27187485",
  "6585000"
],
"country": [
  "US",
  "Thailand",
  "Australia",
  "Belgium",
  "Saudi Arabia",
  "Poland",
  "USA",
  "Colombia",
  "Czech Republic",
  "Spain",
  "Qatar",
  "Norway",
  "NZ",
  "Vietnam"
],
```



```
"duration": [
  "85",
  "40",
  "60",
  "2",
  "219",
  "123",
  "162",
  "240"
],
"industry_type": [
  "Geothermal Energy",
  "Electricity",
  "Oil",
  "Petrochemical",
  "Chemicals",
  "Plastics",
  "LNG",
  "Gas",
  "Manufacture of ethylene and its derivatives",
  "Oil and Gas",
  "Ethanol Production"
],
"site_size": [
  "",
  "190 acres",
  "800 hectares",
  "300 acres",
  "31449 acres",
  "500km2",
  "282 acres"
],
"site_employees": [
  "",
  "4",
  "4",
  "500",
  "0",
  "180",
  "750",
  "300",
  "320"
],
```

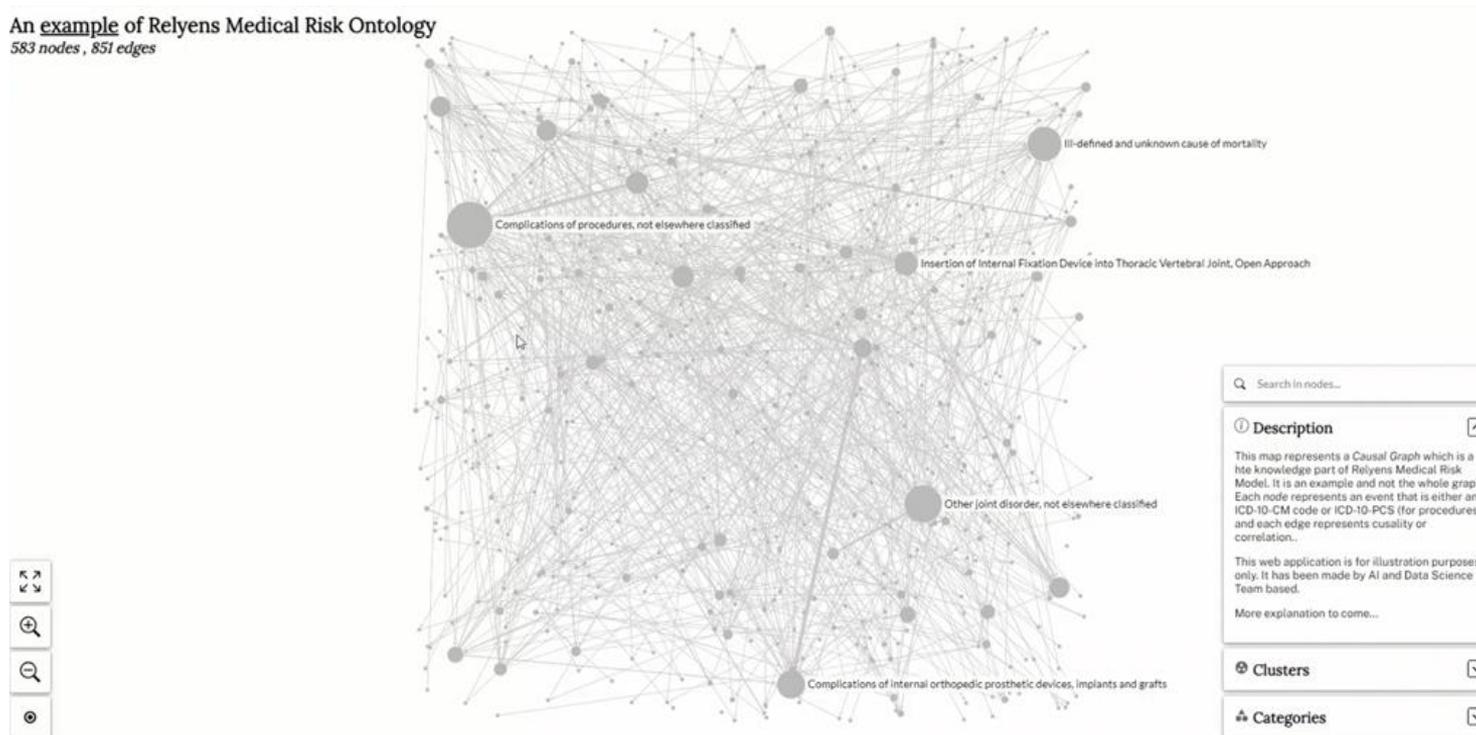


3. Résultats

Visualisation et explication du graphe (1/2)

Graphe

An example of Relyens Medical Risk Ontology
583 nodes , 851 edges



3. Résultats

Visualisation et explication du graphe (2/2)

Visualisation complémentaires

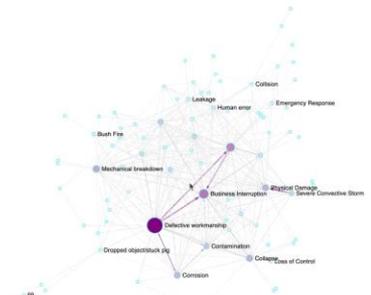


Figure: exemple de meta graph causal par catégorie d'évènement en assurance énergie

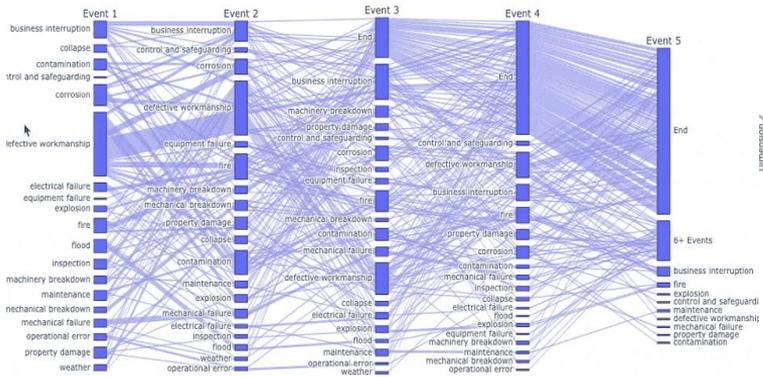


Figure: exemple de visualisation par chronologie de catégorie d'évènement

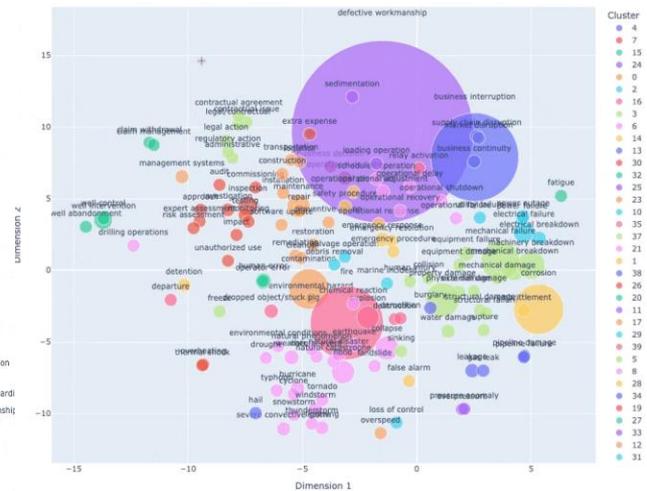


Figure: exemple de représentation de l'étape de standardisation des propriétés de noeuds (catégorie d'évènement)

4. Usages et discussions

Quelques applications en assurance

Cas 1 - RC Médicale

- KG-Grounded Q&A/Chat (Dev)
- Liability Detection (R&D)
- Evidence-based Prevention (R&D)

Cas 2 - Energie

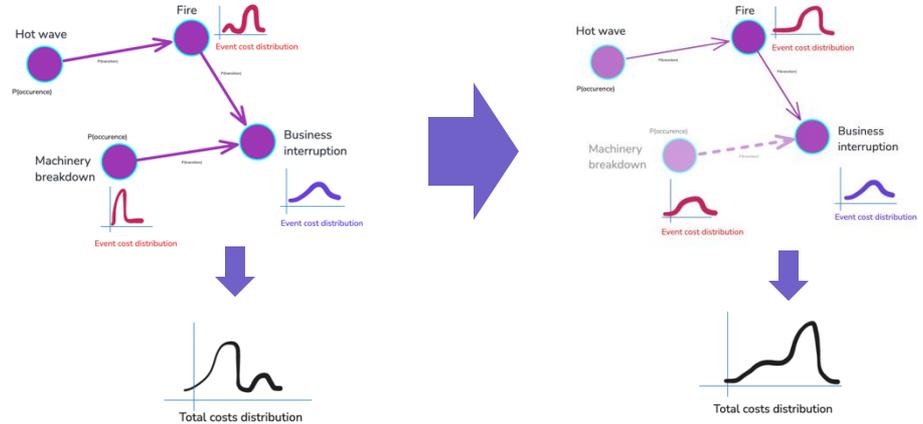
- KG-Grounded Q&A/Chat (Dev)
- Recommendation engine (Dev)
- Pricing and other actuarial services (R&D)

What are the main event category associated with a photo voltaic panel site / solar panel ?

The main event categories associated with photovoltaic panel sites (solar panel sites) are:

1. Defective workmanship
2. Bush Fire
3. Hail
4. Property Damage
5. Business Interruption
6. Windstorm
7. Physical Damage
8. Burglary
9. Control and safeguarding
10. Electrical breakdown
11. Operational disruption
12. Inspection
13. Contamination
14. Fire
15. Electrical Breakdown

Figure: exemple d'interface et de résultats d'un outil agentique pour requêter la base de connaissance



4. Usages et discussions

Conclusions et perspectives



Cas 1 - RC Médicale

- Induction d'ontologie & mise à jour.
- Human-in-the-loop et validation de données.
- Business et process (adoption et intégration).

Cas 2 - Energie

- Appropriation métier et visualisation
- Evaluation (données synthétiques, benchmark causal)
- Stabilité des résultats (étude en cours XXX)
- Modèles de données (contexte statique, propriété des nœuds, propriétés des arêtes, id du cheminement)
- Définitions (événements, atomicité, etc.)
- Requêtage de graphes

Annexes

Références complémentaires (1/3)



Extract data into a causal graph

Long, Stephanie, Tibor Schuster, and Alexandre Piché. "Can large language models build causal graphs?." arXiv preprint arXiv:2303.05279 (2023).

Gopalakrishnan S, Chen VZ, Dou W, Hahn-Powell G, Nedunuri S, Zadrozny W. "Text to Causal Knowledge Graph: A Framework to Synthesize Knowledge from Unstructured Business Texts into Causal Graphs." Information. (2023).

Maisonnave, Mariano et al. "Causal graph extraction from news: a comparative study of time-series causality learning techniques." PeerJ. Computer science vol. 8 e1066. 3 Aug. 2022, doi:10.7717/peerj-cs.1066

Use causal graph with GNN

Sanchez-Lengeling, et al., "A Gentle Introduction to Graph Neural Networks", Distill, (2021).

Dallin Stewart, Dylan Skinner, Gwen Johnson, Jason Vasquez, "Causal Inference with Graph Neural Networks", Medium, (2024).

Behnam, Arman, and Binghui Wang. "Graph Neural Network Causal Explanation via Neural Causal Models." arXiv preprint arXiv:2407.09378 (2024).

Annexes

Références complémentaires (2/3)



Other use of causal graph

Ferreira, Simon, and Charles K. Assaad. "Identifiability of direct effects from summary causal graphs." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38. No. 18. 2024.

Shpitser, Ilya, and Judea Pearl. "Complete identification methods for the causal hierarchy." (2008).

Matheus Facure Alves, Github contributors. "Causal Inference for The Brave and True" (2022).

Benchmark

Jin, Zhijing, et al. "Can large language models infer causation from correlation?." arXiv preprint arXiv:2306.05836 (2023).

Annexes

Références complémentaires (3/3)



Other readings

Bareinboim, Elias, and Judea Pearl. "Causal inference by surrogate experiments: z-identifiability." arXiv preprint arXiv:1210.4842 (2012).

M., Karthika, J. Pearl, J. Tian. "Graphical models for inference with missing data." Advances in neural information processing systems 26 (2013).

Bareinboim, Elias, Jin Tian, and Judea Pearl. "Recovering from selection bias in causal and statistical inference." Probabilistic and causal inference: The works of Judea Pearl. 2022. 433-450.

Wikipedia contributors. "Causal graph." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 7 Nov. 2024. Web. 14 Nov. 2024.

Liu, Jiarui, et al. "Automatic Generation of Model and Data Cards: A Step Towards Responsible AI." arXiv preprint arXiv:2405.06258 (2024).

AlKhamissi, Badr, et al. "Opt-r: Exploring the role of explanations in finetuning and prompting for reasoning skills of large language models." arXiv preprint arXiv:2305.12001 (2023).

Ariffin, W. N. M., and S. Salleh. "Task Scheduling for Directed Cyclic Graph Using Matching Technique." Contemporary Engineering Sciences 8.17 (2015): 773-788.

Vaiva Vasiliauskaite, Tim S. Evans, Paul Expert, "Cycle analysis of Directed Acyclic Graphs", Physica A: Statistical Mechanics and its Applications, (2022).

Merci ! Avez-vous des questions ?



Abdallah ARIOUA
Email: abdallah.arioua@relyens.eu
<https://www.relyens.eu/fr>



Aurélien COULOUMY
Email: acouloumy@dylogy.com
www.dylogy.com