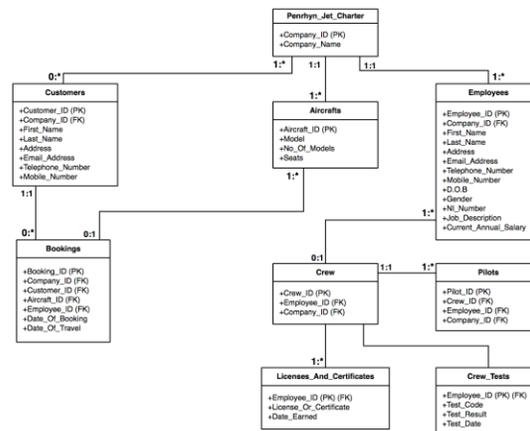


Cross-fertilization between DB and ML : Application to imprecise questions and classification

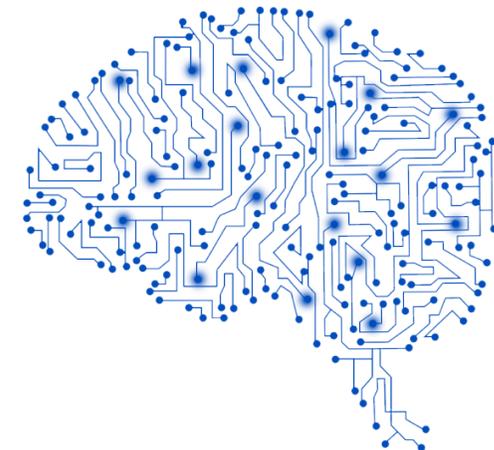
Marie Le Guilly
Jean-Marc Petit
Marian Scuturici

Bases de données relationnelles



- Stocker et interroger des données
- Robuste et pratique
- Vastement utilisé

Apprentissage automatique



- Extraire des connaissances
- Valoriser les données
- Comprendre

Deux applications:

1. ML vers DB: extensions de requêtes SQL pour les questions imprecises
2. DB vers ML: évaluation de la faisabilité de classification en utilisant les dépendances fonctionnelles

EXTENSIONS DE REQUÊTES SQL

■ Motivations:

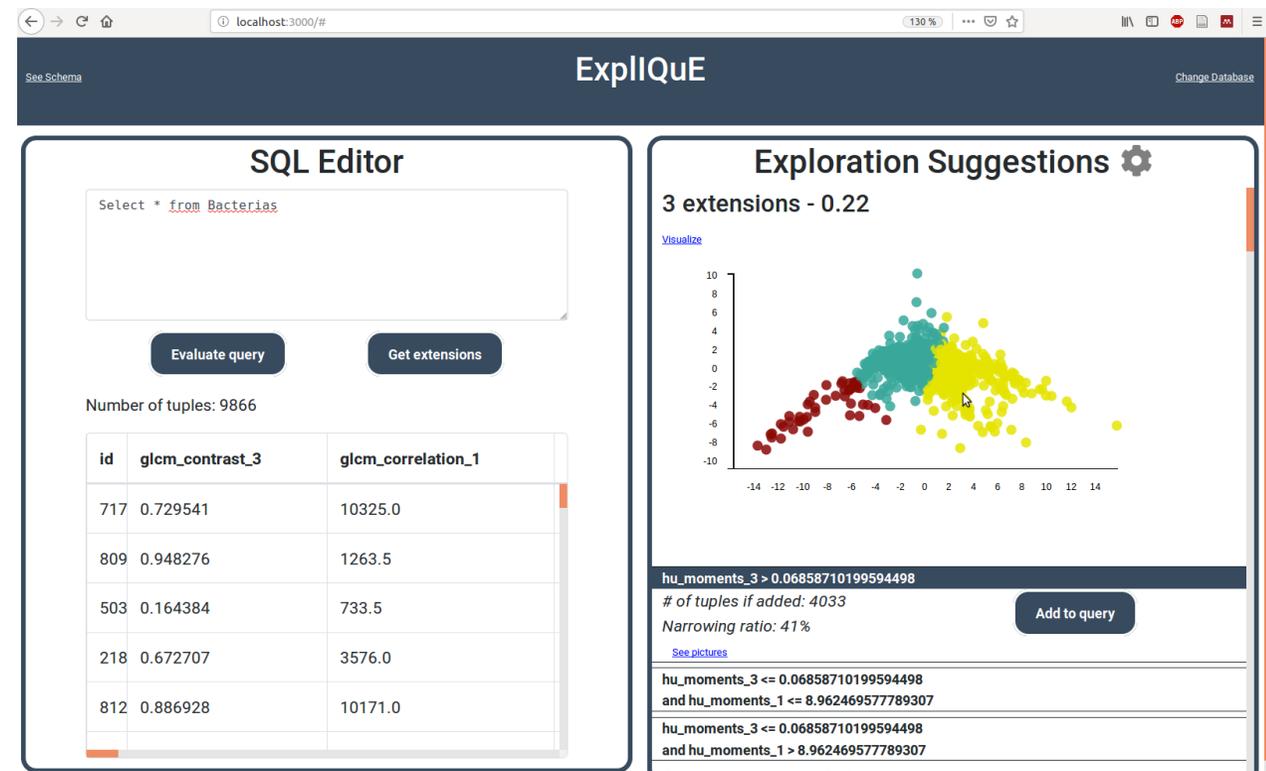
- Parfois difficile de trouver la requête SQL désirée
- Pas mal de conditions imprécises difficiles à traduire en SQL
Ex: "plus petit que la moyenne", "assez grand", etc.
- Processus itératif nécessaire, beaucoup de tuples à trier

■ Proposition:

- Suggérer des ensemble d'extension d'une requête initiale
- Résumer les tuples avec les extensions
- Et proposer des clauses de sélection supplémentaires pour tendre vers le résultat désiré

■ Solution:

- Algo: basé sur clustering et arbre de décision
- Interface web avec visualisations



DÉPENDANCES FONCTIONNELLES ET CLASSIFICATION

- Motivations:
 - Beaucoup de problèmes de classification
 - DF: valide l'existence d'une fonction dans une relation
 - Modèle de classification: recherche d'une fonction
- Proposition: valider l'existence (ou non) d'une fonction avant de la déterminer
- Contribution:
 - Estimation d'une borne supérieure de la précision d'un modèle
 - Recherche de contre-exemples
 - Analyse des contre-exemples avec les experts du domaine
 - Génération de jeu synthétiques complexes pour la classification

FD Validation

Enter the FD to validate:

Ticket, Age, Gender → Survived

Validate FD

The FD is not valid. Counterexamples:

id	t2.id	Ticket	Age	Gender	t1.Survived	t2.Su
	5	2nd	Child	Male	yes	no
	6	2nd	Child	Male	no	yes
	8	1st	Ad...	Male	no	yes
	8	1st	Ad...	Male	no	yes

G1(X→Y) = 17%
G2(X→Y) = 90%
G3(X→Y) = 70%

[See Heatmap](#)