UNIVERSITE BADJI MOKHTAR-ANNABA

DEPARTEMENT D'INFORMATIQUE
M1 IATI

Apprendre à Ordonner 'Learn to Rank'

par

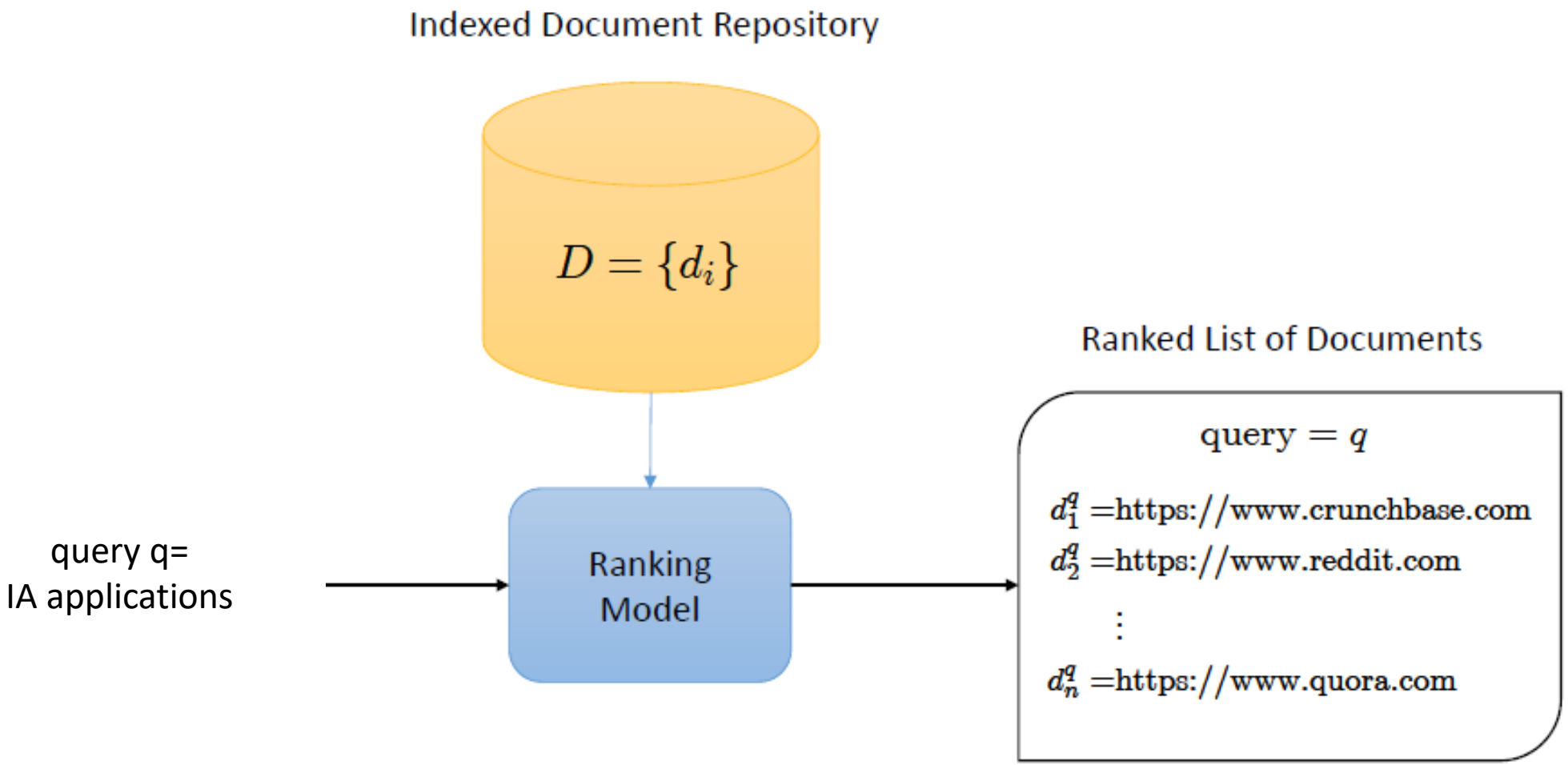
Dr. Samira LAGRINI

Année universitaire:2024/2025

Introduction

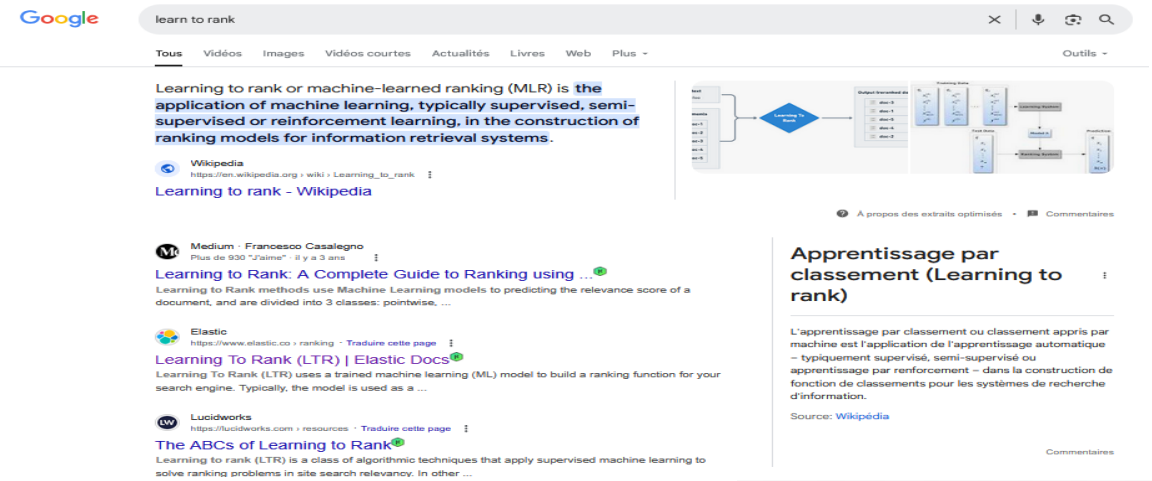
- Apprendre à ordonner (Learning to Rank, LTR) est un problème de l'apprentissage supervisé qui a pour objectif d'apprendre à un modèle à attribuer un classement (ou un score de pertinence) à des éléments (documents, produits..) dans une liste.
- Le but est de trier des éléments en fonction de leur pertinence pour une requête donnée.
- Plutôt que de prédire une simple étiquette binaire (pertinent ou non pertinent), le modèle LTR doit apprendre à ordonner les éléments selon un ordre de pertinence.

Introduction



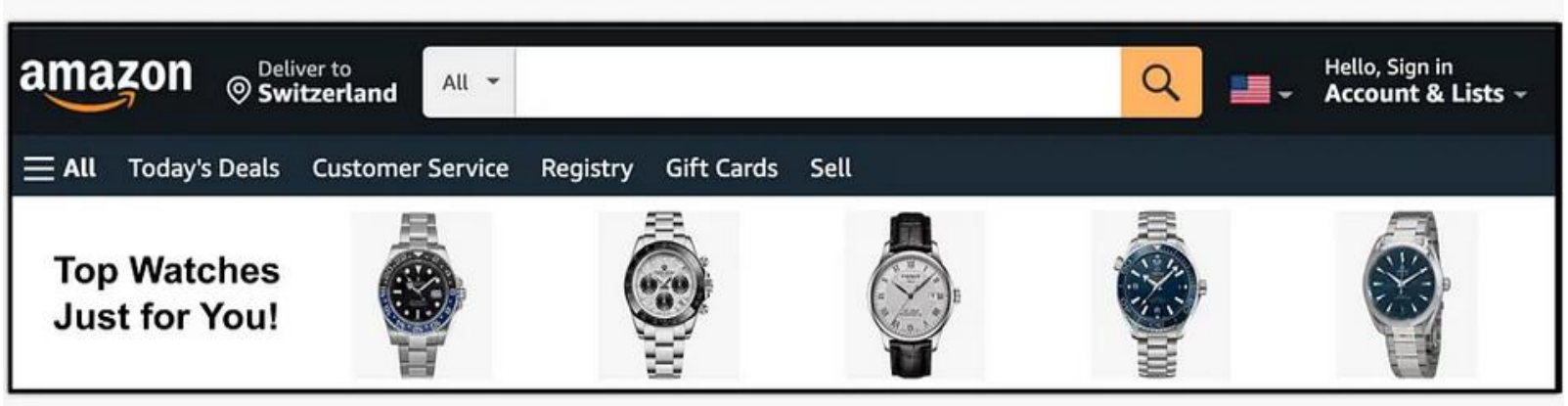
Applications Pratiques de LTR

Optimisation des résultats des moteurs de recherche

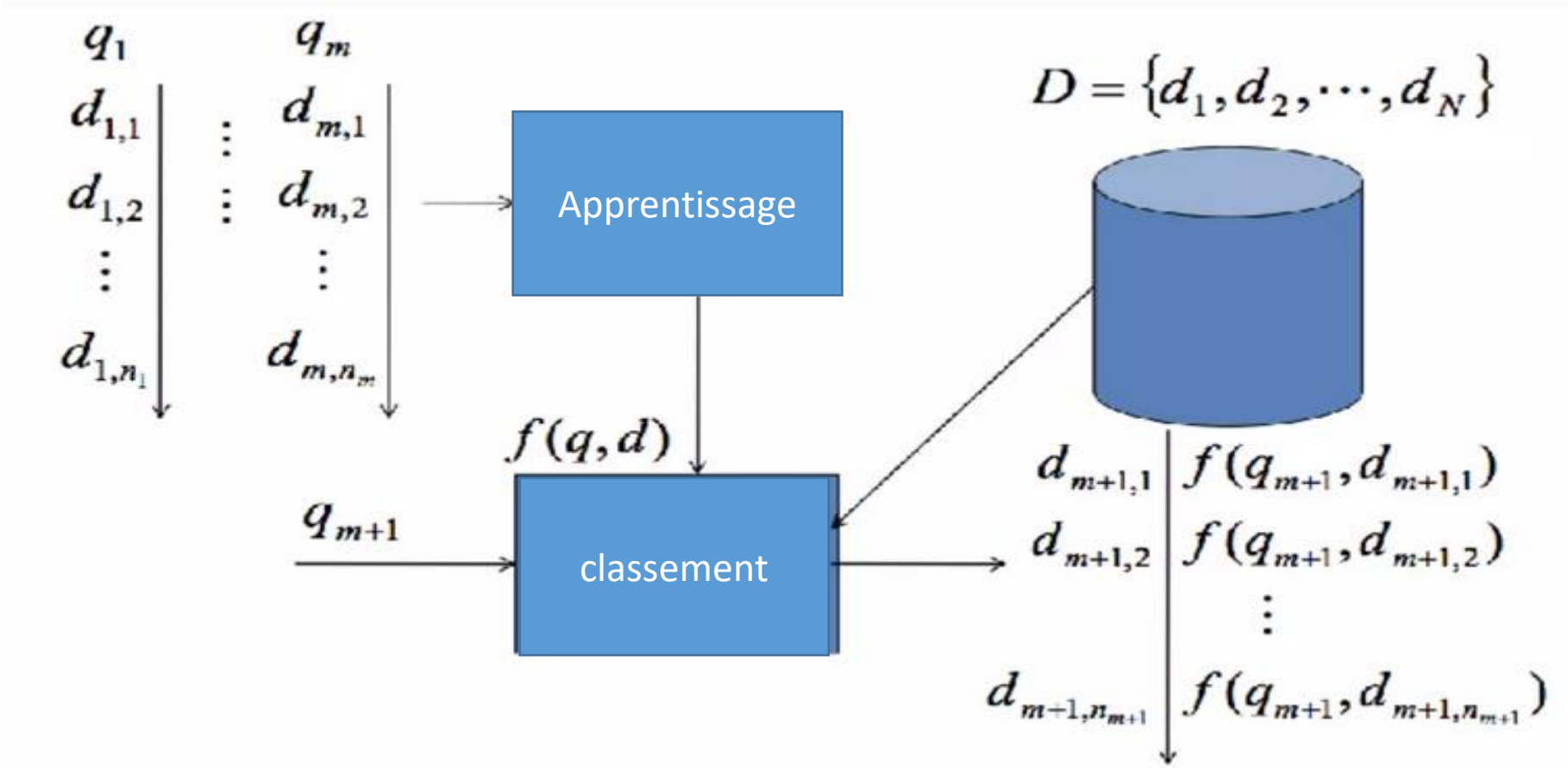


E-commerce et Publicité en ligne

Systèmes de Recommandation



Construire un Modèle LTR



Collecte des données d'entraînement

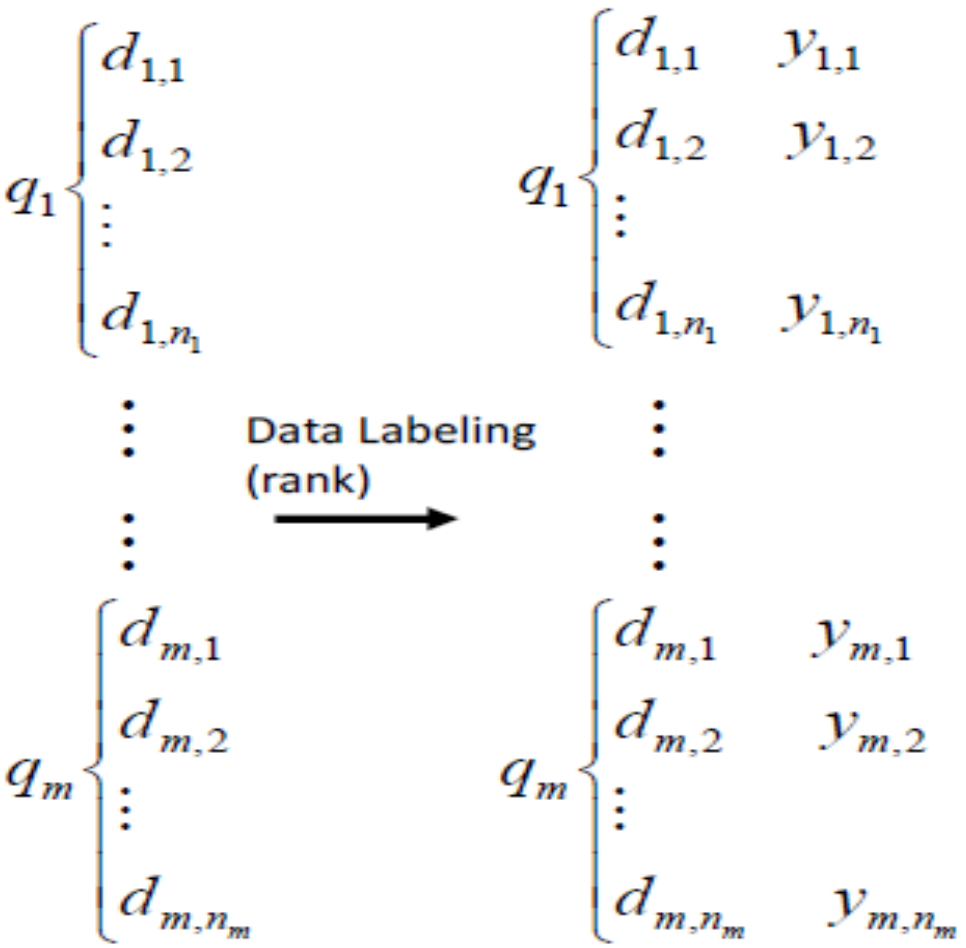
- Les données d'entrainement sont des exemples d'éléments classés par pertinence :

➢ **Requêtes et résultats** :

Liste de requêtes et de documents associés.

➢ **Pertinence**

Scores ou labels de pertinence

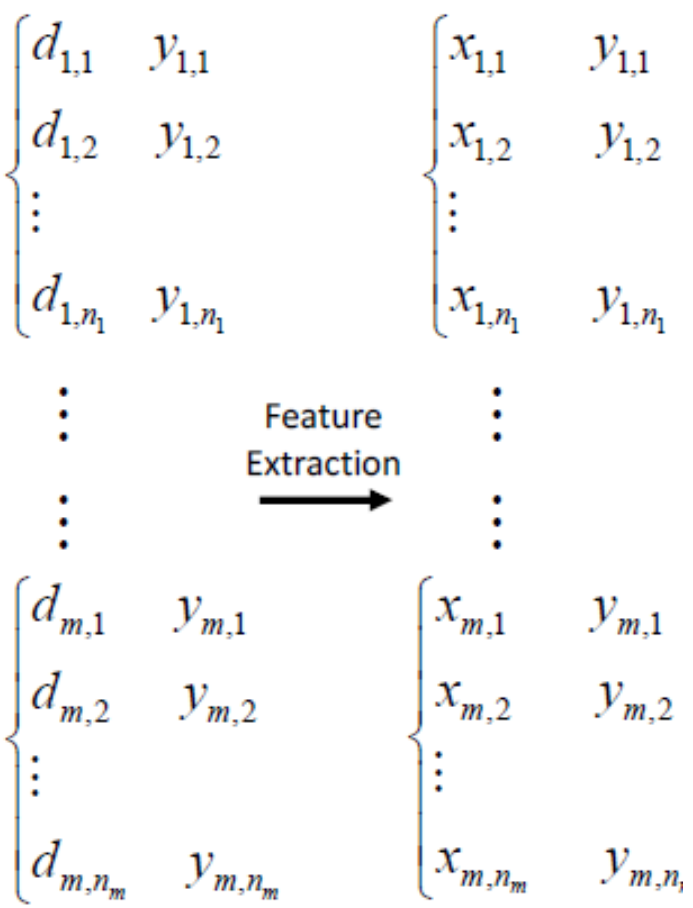


Extraction des caractéristiques

- Pour chaque élément à classer, extraire des caractéristiques (features) qui peuvent influencer son classement:

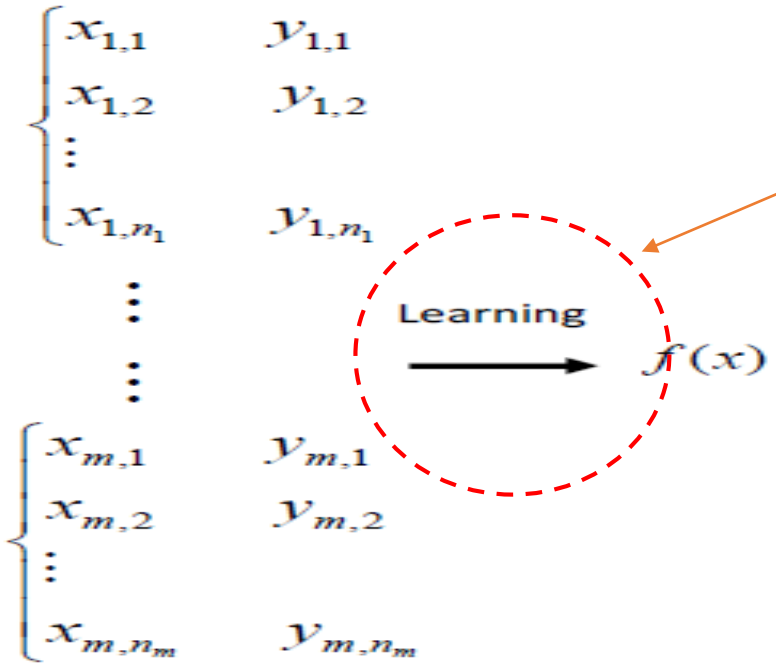
✓ Fréquences de mots de requete, longueur du texte, etc.
✓ Date de publication, auteur, etc.
✓ **Interactions utilisateur** : Clics,.. etc.
✓ …….

- Chaque élément est transformé en un vecteur de caractéristiques qui sera utilisé pour l'entraînement du modèle.



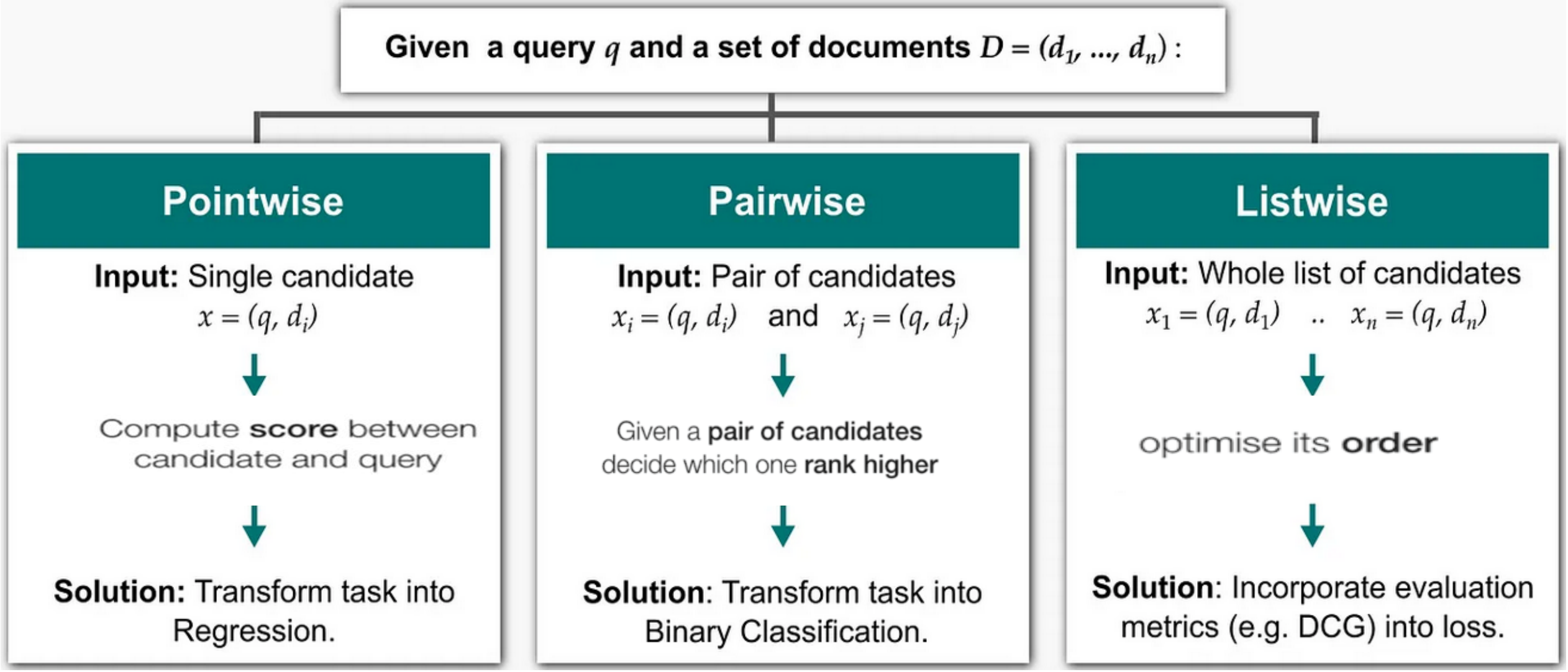
Entraîner le modèle

- Sélectionner un algorithme d'apprentissage supervisé pour entraîner le modèle ➔ Cela dépond de l'approche LTR



Quelle approche ????

Approches d'apprentissage LTR



Approche point par point (Pointwise)

- Apprendre à classer chaque élément de la liste indépendamment des autres.➔ input: $x = (q, d_i)$
- On attribue un score de pertinence entre chaque élément et la requête.

➔ Modéliser la tache comme un problème de régression

Algourithms courants

OC-SVM (One-Class Support Vector Machine), Régression linéaire, régression logistique.

Avantages : Simple à implémenter.

Limites : Ignore les relations entre les documents.

Approche Pointwise

Exemple

Supposons que vous avez les résultats suivants pour une requête "chatons" :

- Page A : Note 5
- Page B : Note 3
- Page C : Note 1
- L’objectif de l’approche pointwise est de prédire cette note pour chaque page en fonction de ses caractéristiques (par exemple, la présence du mot "chaton", le nombre de vues, etc.).

Approche par paire (Pairwise)

- Au lieu de classer directement chaque élément, on compare des paires d'éléments pour déterminer lequel des deux est plus pertinent. ➔ input : $x_i=(q,d_i)$ $x_k=(q,d_j)$

- Le modèle apprend à prédire quel élément parmi une paire est le plus pertinent et ensuite à en déduire un classement global. ➔ Modéliser la tache comme un problème de classification binaire.

Modèles courants

RankNet (réseaux de neurones), Ranking SVM, RankBoost, LambdaRank, and LambdaMART

Approche Listwise

- L’approche Listwise optimise directement l’ordre complet de tous les éléments dans la liste

Prendre en compte la relation entre les éléments d'une liste plutôt que de les traiter individuellement.

Modèles courants

ListNet, ListMLE, AdaRank, SVM MAP, Soft Rank, and AppRank.

Évaluation des performance

Évaluer la qualité du modèle à l'aide de métriques tel que:

Métrique	Description
NDCG (Normalized Discounted Cumulative Gain)	Mesure l'efficacité du classement en tenant compte de la pertinence et de la position des éléments dans la liste.
MAP (Mean Average Precision)	Moyenne de la précision pour plusieurs requêtes, prenant en compte la position des éléments pertinents.
P@k (Precision at k)	Mesure la proportion d'éléments pertinents parmi les kkk premiers résultats du classement.
R@k (Recall at k)	Mesure la proportion des éléments pertinents trouvés dans les kkk premiers résultats par rapport au total des éléments pertinents.


Merci de
votre
attention