

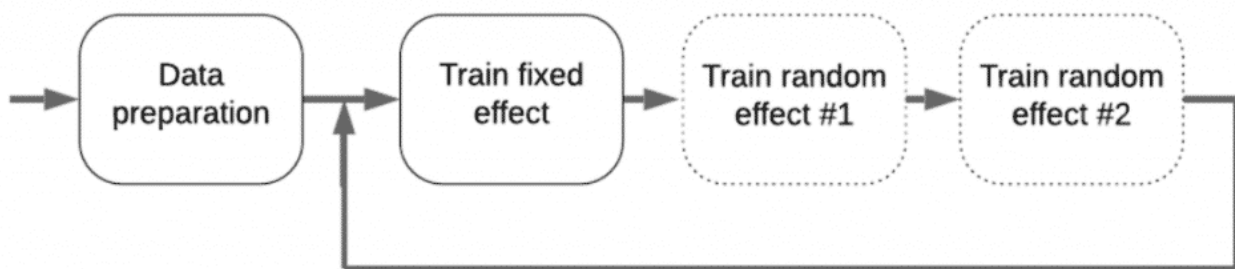
LinkedIn : l'autre vitrine open source de Microsoft

Dagli, DeText, GDMix... Autant de projets passés cette année en *open source*. Tous trois visent à favoriser, dans différents contextes, l'utilisation de l'apprentissage machine. Ils ont une source commune : LinkedIn.

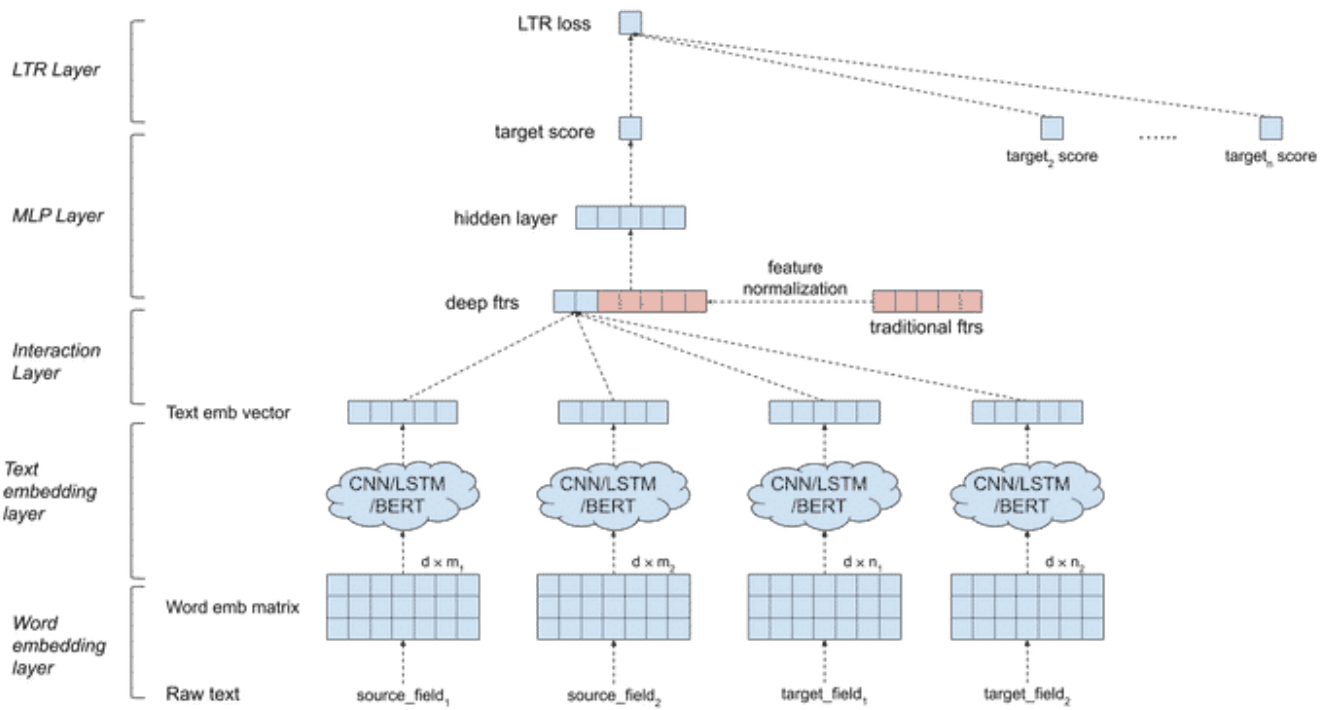
On les retrouve ainsi en action sur le réseau social. Par exemple au niveau des outils de recherche et de recommandation, avec [GDMix](#) (Generalized Deep Mixed).

Ce *framework* est censé favoriser l'entraînement des modèles qui portent les outils en question. Il les scinde pour cela en un modèle « global » (dont les paramètres sont fixes ou déterminés) additionné d'une multitude de « petits » modèles aux variables aléatoires.

Dans le cas d'une suggestion d'offre d'emploi, le modèle global pourra correspondre à un diplôme ou à une qualification. Chaque « petit » modèle représentera un membre de LinkedIn ou un groupe de membres. L'approche offre davantage de souplesse que le croisement d'entités (membres) et de propriétés (diplômes, qualifications...) au sein d'un même modèle. Lequel regrouperait, dans le cas présent, des dizaines de milliards d'éléments (LinkedIn revendique 706 millions de membres et prend en compte plus d'une centaine de paramètres pour l'affichage des offres d'emploi).



Autre *framework*, plus générique : [DeText](#). Son rôle : simplifier l'exploitation « transversale » des modèles de traitement du langage naturel. Il gère pour le moment le classement, la catégorisation et la complétion automatique de séquences. L'idée est de pouvoir enchaîner les tâches. Par exemple, une fois interprété le sens d'une phrase, déterminer celles qui en sont proches et lesquelles pourraient répondre à une requête précédemment formulée.



[Dagli](#) permet pour sa part de définir des modèles en Java – et autres langages JVM – sous la forme de graphes orientés acycliques.

LinkedIn a ajouté à cet ensemble un [Fairness Toolkit](#). Il s'agit d'une bibliothèque Spark pour évaluer l'« éthique » des modèles et des données d'entraînement.

Illustration principale © Michael J Berlin