
Vers un système de recommandation de profils experts dans l'industrie des procédés

Yann Duperis^{a, b} — Adrian-Gabriel Chifu^a — Bernard Espinasse^a — Sébastien Fournier^a — Arthur Kuehn^b

^a LIS ^b Moben & Rooster

RÉSUMÉ. La dématérialisation des processus de recrutement n'a pas fait disparaître toutes les frictions inhérentes à cette activité. La recherche automatisée d'un candidat idéal se heurte toujours à la difficulté à modéliser correctement les besoins exprimés en langage naturel dans une offre d'emploi. Le recrutement d'experts, notamment, est particulièrement difficile. En effet, ces profils concernent une proportion réduite des recrutements et leur prise en charge informatisée nécessite une connaissance précise du secteur d'activité concerné. Dans cet article, nous proposons l'architecture d'un système de recommandation de profils experts dans l'industrie des procédés afin d'assister ce type de recrutements.

ABSTRACT. The digitalization of recruitment processes did not smooth out all frictions related to this activity. Automatically searching for an ideal candidate is still hindered by one major issue: adequately modeling the needs expressed in natural language in a job offer. This problem is even more obvious when one considers the use case of recruiting expert profiles. Indeed, these profiles represent a minority of the recruitments. To assist a recruitment process concerning these profiles, an automated system requires an accurate knowledge of their activity sector. In this article we introduce the architecture of an expert profile recommender system applied to the process industry sector.

MOTS-CLÉS : Système de recommandation de poste, Recherche d'expert, Web sémantique.

KEYWORDS: Job recommender system, Expert search, Semantic web.

1. Introduction

L'émergence et la démocratisation progressive du Web ont conduit, en ce début de XXIème siècle à une dématérialisation croissante de nombreuses activités. L'activité du recrutement, et sa symétrie la recherche d'emploi, ont radicalement changé en moins de trois décennies. Aujourd'hui, les organisations cherchant à recruter du personnel utilisent massivement le Web pour publier leurs offres d'emploi. De leur côté, les individus utilisent également le Web pour numériser et structurer leurs CV sous forme de profils de réseaux sociaux professionnels (tels que *LinkedIn.com* ou le français *Viadeo.com*). Ces derniers peuvent également candidater en ligne.

La facilité offerte par la candidature numérique a tendance à submerger les recruteurs qui en sont rapidement venus à utiliser des systèmes permettant de filtrer les candidatures. Ces systèmes ont été initialement construits sur des méthodes classiques de recherche d'information à base de mots-clés. La connaissance de ce fonctionnement a conduit les candidats à adapter leurs CV, allant, pour certains, jusqu'à les transformer en « sacs de mots-clés ». Ce problème, commun au domaine de la recherche d'information, a un impact non négligeable sur l'activité du recrutement car il standardise les profils et décourage leurs auteurs de mentionner certaines informations (voire de les développer, tout simplement).

Nous présentons dans cet article l'architecture d'un système de recommandation de candidats visant à adresser ce problème. La spécificité principale de nos travaux vient de la nature des besoins en compétences adressés. En effet, nous souhaitons proposer un système de recommandation spécialisé dans le recrutement d'experts dans le secteur de l'industrie des procédés. Cette industrie étudie les processus de la matière pour obtenir des produits en optimisant les ressources en matières premières et en énergie¹.

Comme tous les secteurs d'activité économique concernés par une spécialisation, ce dernier est sujet à l'émergence d'un vaste réseau de connaissances implicites et d'un sous-langage partagé. Contrairement à d'autres secteurs d'activité, ces connaissances sont moins accessibles au « grand public » et font donc plus difficilement l'objet d'un apprentissage par transfert à partir de ressources linguistiques génériques et publiques. Les méthodes s'appliquant à des secteurs moins ésotériques, tels que l'informatique², ou s'appliquant à tout type de besoin ne parviennent pas à adresser le problème de la recommandation de profils rares. Le système de recommandation que nous proposons devra donc être capable d'obtenir un rappel important pour ce type de profils. Ces travaux sont menés en collaboration avec un cabinet de conseil en ingénierie spécialisé dans ce secteur d'activité.

Dans cet article, nous introduirons premièrement les concepts fondamentaux du domaine du recrutement dans la section 2. Une fois le domaine présenté, nous pro-

1. Y sont donc fréquemment inclus les secteurs de la pétrochimie, du nucléaire et de la chimie.

2. Le secteur de l'informatique agrège une proportion non négligeable des contributions présentes dans la littérature

poserons une formalisation du problème et l'architecture d'un système de recommandation y répondant en section 3. Les modules composant ce système ainsi que les sous-problèmes qu'ils adressent seront présentés dans la section 4. Enfin, nous concluons cet article par la section 5 en présentant nos pistes de recherche et les principaux verrous identifiés.

2. Définition des concepts fondamentaux du recrutement

Le système de recommandation présenté dans cet article est spécifique au domaine du recrutement. Avant d'aborder la conception de ce système, cette section présentera un tour d'horizon des concepts fondamentaux du domaine en question. Ces concepts semblent accessibles au sens commun et sont par conséquent peu définis formellement dans les travaux du domaine de la recherche d'information appliquée au recrutement. Une étude de cette littérature fait cependant apparaître, explicitement ou implicitement, des sens différents associés à un même terme.

2.1. Compétences

Dans une grande majorité des travaux, dont les nôtres, le concept de compétence est central. Dans cette section, nous analysons les utilisations de ce concept dans la littérature du domaine de la recherche d'information appliquée au recrutement. Comme pour d'autres concepts manipulés ici, celui de compétence est bien plus rigoureusement étudié par des auteurs d'autres domaines, citons par exemple les travaux présentés dans (Coulet, 2011). Nous n'étudions ici que l'utilisation de ce concept dans le domaine de la recherche d'information. Nous n'avons pas pour ambition de proposer une synthèse de ces définitions.

Nous pouvons observer deux paradigmes différents dans les travaux étudiés. Nous qualifierons le premier d'*extensif* car il ne repose pas sur des définitions formelles mais plutôt sur des ensembles *ad hoc*. Nous qualifierons le second, plus rare, d'*intensif* car il s'inscrit dans un contexte plus expressif sémantiquement. Cette section sera conclue par la définition que nous avons retenue pour nos travaux.

2.1.1. Paradigme extensif

Ce paradigme est majoritaire dans la littérature et il n'est que rarement exprimé explicitement. Nous en faisons donc une synthèse à partir de l'analyse des différentes approches rencontrées. Dans ce paradigme, une compétence est une propriété d'un individu. Cette propriété peut être binaire ou associée à un niveau.

Lors de la rédaction d'une offre d'emploi, un recruteur va mentionner les compétences que le candidat doit posséder pour réaliser les tâches qui lui seront confiées. Les compétences ainsi mentionnées dans un profil ou une offre d'emploi sont des formes de surface propres au secteur d'activité. Les concepts sous-jacents ne disposent pas de définitions intensives et leurs extensions sont souvent floues.

Pour cette raison, les compétences sont restées modélisées comme des mots-clés. Dans le cas de réseaux sociaux professionnels tels que *LinkedIn.com* ou *CareerBuilder.com*, les utilisateurs éditant leur profil disposent d'une section dédiée à l'indication de leurs compétences. Cette saisie est guidée par l'usage de taxonomies ou de dictionnaires de formes de surface normalisées. Les chercheurs collaborant avec *LinkedIn.com* ont très tôt adopté une définition floue du concept de compétence lors de la conception de leur taxonomie (Bastian *et al.*, 2014). Dans (Zhao *et al.*, 2015), les auteurs collaborent avec *CareerBuilder.com* pour produire la taxonomie *SKILL* à partir de l'exploitation de 60 millions de profils professionnels et 1.6 millions d'offres d'emploi.

2.1.2. *Paradigme intensif*

À l'autre extrémité du spectre, certains auteurs adoptent des approches bien plus proches des sciences cognitives et proposent des méthodes reposant sur des définitions intensives du concept de compétence. Dans les travaux de (Paquette, 2020), les auteurs proposent une modélisation sémantique reposant sur la considération suivante : une compétence est la mise en application procédurale de méta-connaissances propres à des aptitudes génériques (telles que diagnostiquer, réparer, évaluer) à des situations et connaissances concrètes (l'automobile par exemple) et dans un méta-domaine d'aptitude spécifique (Cognitif, Affectif, Social ou Psychomoteur). D'après les auteurs, ce modèle permet de modéliser à différents niveaux de granularité la plupart des compétences. D'un point de vue pratique, le modèle OWL peut être connecté à des ontologies de domaine en liant une compétence à des connaissances spécifiques du domaine cible.

Ce paradigme est plus rare car la construction manuelle d'ontologies expressives est une démarche coûteuse (voir 2.1.1). Le coût en calcul résultant de l'application de raisonnements à un graphe de connaissances riche est également prohibitif pour l'intégration à un système de recommandation. Pour finir, l'usage d'un formalisme sémantiquement riche requiert une étape d'extraction d'information non triviale depuis les documents textuels, ou faiblement structurés, originaux.

2.1.3. *Définition retenue pour nos travaux*

Au vu des forces et faiblesses des deux paradigmes présentés précédemment, nous optons pour le paradigme extensif. En effet, nous pensons qu'une modélisation sémantiquement expressive des compétences ne parviendra pas à saisir la diversité des usages et l'extension floue des concepts en pratique.

Dans le cadre de nos travaux, nous manipulerons par conséquent les compétences comme des symboles dénués de définition intensive et en accord avec la définition suivante³:

3. Cette définition présente l'avantage d'être en accord avec l'étymologie de *Skill* (compétence en anglais) qui vient du vieux norrois « Skil » qui signifie différence ou distinction.

Une compétence est la capacité d'un individu à mettre en oeuvre un ensemble de savoirs et de savoir-faires dans le cadre de la réalisation d'une tâche plus ou moins stéréotypée. Cette capacité est telle qu'il existe une différence qualitative et quantitative de résultat entre le travail fourni par une personne possédant la compétence en question et celui d'une personne ne la possédant pas.

La possession d'une compétence est ici définie comme binaire, nous ne lui associons pas de niveau. Ce choix peut être contestable et certains travaux adoptent d'ailleurs une approche différente. Nous pensons cependant que l'évaluation du niveau de compétence d'un individu n'est pas dénuée d'arbitraire et implique des questionnements supplémentaires tels que l'implication concrète de l'échelle choisie. Si la connaissance du niveau d'un individu est essentielle, nous considérons que nous avons affaire à deux compétences distinctes (bien que sémantiquement liées). Dans notre acceptation, une compétence doit rester un indice quant à la capacité d'un individu à mener à bien une tâche.

2.2. *Tâche*

Le concept de *Tâche* est généralement une abstraction regroupant un ensemble d'actions, de comportements et d'interactions effectués par un individu et s'intégrant dans un processus économique. Dans le cas de secteurs réglementés ou de métiers particulièrement normalisés, les tâches sont suffisamment stéréotypées pour permettre leur conceptualisation (les travaux présentés dans (Khobreh, 2017) sont particulièrement intéressants à cet égard).

Pour d'autres secteurs d'activité, tel que celui que nous étudions, la réalité de l'activité économique entreprise est bien trop complexe et mouvante pour permettre la conceptualisation d'une tâche exprimée en langage naturel. La plupart des travaux vont par conséquent plutôt considérer les tâches comme des sections du document d'où extraire des mots-clés ou comme une entrée textuelle brute pour un modèle d'apprentissage profond (Qin *et al.*, 2018).

Dans le cadre de nos travaux, nous ne comptons pas procéder à une modélisation des tâches comme propriétés d'une offre d'emploi ou d'un profil. Nous ne les utiliserons pas dans le mécanisme d'appariement au coeur de notre système de recommandation. Cependant, nous pensons qu'elles peuvent être particulièrement utiles lors de la phase d'extraction d'informations à partir des documents textes.

2.3. *Métier*

Le dernier concept abordé lors de ce rapide tour d'horizon est celui de *Métier*. Ce concept est également une abstraction regroupant un ensemble de tâches plus ou moins fixes selon le secteur, l'organisation ou même l'individu l'exerçant. Certains

travaux (Jiechieu et Tsopze, 2020) sont d'ailleurs dédiés à la prédiction d'un métier à partir d'un profil de compétences.

L'exemple le plus stéréotypé est celui du métier de « Chef de projet ». En effet, ce nom de métier est fréquent dans de nombreux secteurs et agrège des réalités très différentes. Un chef de projet travaillant dans le secteur de l'industrie logicielle n'aura que peu de points communs avec un chef de projet opérant dans les travaux neufs.

De plus, il est communément admis que l'évolution rapide de l'économie et la fréquente survenue de « disruptions » accélèrent l'apparition et la disparition de métiers, imposant ainsi un re-centrement de la gestion de la main d'oeuvre sur les compétences des individus. Cette évolution est abordée par (De Smedt *et al.*, 2015) dans les motivations de la création de l'ontologie ESCO et la présentation de l'écosystème européen des référentiels professionnels. Nous comptons donc inscrire notre approche dans cette évolution de paradigme tout en prenant un compte les noms de métiers présents dans nos documents car ils sont, malgré tout, porteurs d'information.

3. Système de recommandation de profils proposé

Cette section introduira le système de recommandation de candidats que nous proposons. Elle commencera par une formalisation du problème et aboutira à une proposition d'architecture permettant de produire les recommandations définies. La section suivante (4), présentera ensuite chacun des modules envisagés. Pour chacun des modules, adressant un sous-problème spécifique, nous présenterons les approches rencontrées lors de l'analyse de la littérature spécialisée.

3.1. Formalisation de la recommandation de profil

Posons P , l'ensemble des profils professionnels des individus et O , l'ensemble des offres d'emploi, p_j un profil spécifique, o_i une offre d'emploi spécifique et (o_i, p_j) un appariement entre une offre d'emploi et un candidat. L'ensemble des recommandations possibles est défini par

$$R = \{(o_i, p_j, s_{ij})\} \forall o_i \in O, \forall p_j \in P \quad [1]$$

$$s_{ij} \in [0, 1]$$

Avec R l'ensemble des recommandations possibles et s_{ij} le score de pertinence de l'appariement (o_i, p_j) . Ce concept de *pertinence* sera abordé dans (4.4.1).

Le coeur du problème est ce que nous appellerons la *fonction d'évaluation de pertinence d'un appariement*, f_s , qui, à une offre d'emploi o_i et à un profil p_j va associer un score de pertinence s_{ij} .

$$f_s : O \times P \rightarrow [0, 1] \quad [2]$$

C'est l'application de cette fonction à un *appariement candidat* qui va permettre de déterminer si un profil est qualifié pour une offre.

3.2. Sous-problèmes à adresser

Afin de procéder à la recommandation de profils, tel que décrite dans (3.1), le système proposé devra adresser les sous-problèmes suivants:

- Extraction de caractéristiques depuis les documents textuels et modélisation des offres d'emploi et des profils dans un espace commun ;
- Stockage et indexation des profils à recommander ;
- Filtrage des profils dont les appariements avec une offre d'emploi requête devront être évalués ;
- Évaluation des appariements candidats à une recommandation.

La figure 1 représente l'architecture simplifiée du système envisagé. Chaque sous-problème est adressé par un composant. Les flux de données sont représentés par des flèches de couleurs différentes en fonction du moment où le flux se produit. La section suivante présentera, pour chaque composant, le sous-problème à adresser, les approches rencontrées dans la littérature et nos pistes de recherche.

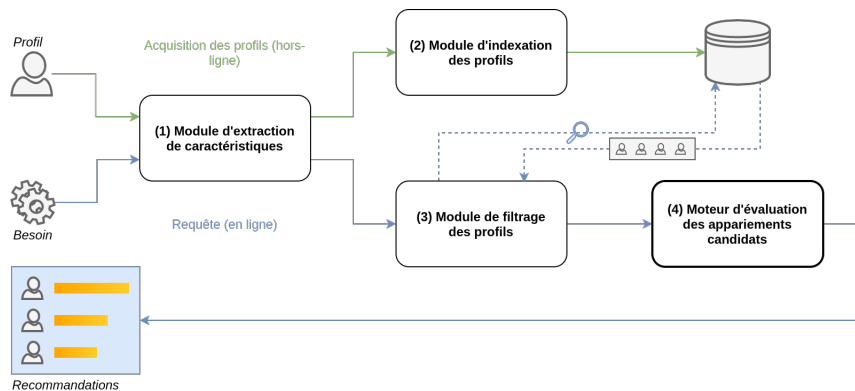


Figure 1. Architecture du système de recommandation de candidats proposé

4. Composants du système proposé

4.1. Module d'extraction de caractéristiques

Pour appliquer la fonction d'évaluation de pertinence d'un appariement, il convient de produire une représentation de cet appariement. Nous évoquerons dans cette

section les caractéristiques essentielles d’une requête à laquelle le système doit répondre en analysant un exemple d’offre d’emploi.

4.1.1. Exemple d’offre d’emploi

La figure 2 présente un exemple d’offre d’emploi que nous utiliserons comme support de réflexion. Cette offre d’emploi est relativement standard. Elle contient en effet les éléments suivants: nom du métier à exercer, pré-requis (comme dans (Bastian *et al.*, 2014) nous considérons tous ces éléments comme des compétences), tâches et contexte de travail.

Un oeil exercé au métier du recrutement et au secteur du nucléaire remarquera que le profil recherché est particulièrement atypique. Certains des mots-clés de cette offre d’emploi sont essentiels, en effet « site nucléaire » est ici une forme de surface du concept de *CNPE* (Centre Nucléaire de Production d’Électricité) et va déclencher une cascade de raisonnements intériorisés chez un recruteur spécialisé et chez un candidat du secteur qui postule à l’offre. Il est également pertinent de relever qu’un ensemble non continu des mots-clés de cette offre peut être lié à un concept pertinent du domaine. Il s’agit des mots clés « END », « CND », « ultrasons », « COFREND niveau 3 UT », « TOFD » et « CIFM ». En effet, dans l’activité du contrôle non destructif (CND), une certification COFREND est une certification attestant de la capacité d’un individu à opérer des dispositifs de contrôle non destructifs dans différentes industries.

Ces mots-clés disséminés dans le texte n’ont pas nécessairement de sens lorsqu’ils sont pris individuellement. Ce sont des symboles dont l’association dans le contexte linguistique du domaine va activer le concept *Certification COFREND Ultrasons de Niveau 3* dans l’esprit d’un expert du domaine. Le module d’extraction de caractéristiques proposé devra donc être capable de détecter le concept de la compétence sous-jacente à l’aide de ces mots clés.

4.1.2. Conceptualisation des compétences

Posons w un n -gramme du texte et c un concept, il n’existe pas toujours un lien direct $w \rightarrow c$. Dans le cas présenté ici, c’est un sous-ensemble non continu des n -grammes du texte W , que nous nommerons W_c , qui permet d’estimer la probabilité de mention du concept c dans l’offre d’emploi o_i .

Formellement;

$$P(c|o_i) \sim P(c|W_c) \quad [3]$$

$$W_c \subset W$$

Cette probabilité est directement liée à $P(c|W_c)$, mais peut aussi être conditionnée par des éléments spécifiques de o_i tels que le secteur d’activité renseigné, le titre du poste, le nom de l’employeur, etc. Nous pensons que la proposition d’une méthode de conceptualisation du texte des offres d’emploi (mais aussi des profils) est es-

sentielle pour identifier correctement les éléments d'importance primordiale pour le succès d'un recrutement.

4.1.3. *Approche envisagée*

Au vu de la spécificité des besoins considérés dans notre étude et du faible volume d'exemples disponibles, nous faisons l'hypothèse qu'une approche basée sur une forme d'apprentissage risque de ne pas parvenir à capturer suffisamment de régularités statistiques pour offrir des résultats satisfaisants. Nous envisageons par conséquent l'application d'une méthode reposant, au moins partiellement, sur des connaissances du domaine explicites. Suite à l'étude de la littérature concernant la construction de taxonomies de compétences (Kessler et Lapalme, 2017 ; Dhoubi *et al.*, 2018 ; De Smedt *et al.*, 2015), nous envisageons d'exploiter un corpus d'offres d'emploi et de profils pour construire automatiquement un dictionnaire de compétences propre à notre domaine. Les ressources à utiliser pour guider l'étape de conceptualisation feront également l'objet de recherches.

DESCRIPTION DU POSTE

Nous recherchons activement pour un de nos clients – Acteur important du secteur nucléaire :

1 Responsable END / CND – ultrasons

vos principales missions sont de :

- Réalisation des chantiers END/CND sur sites nucléaires : pilotage des équipes, relation client, organisation du travail, respect des consignes d'hygiène et de sécurité, vérification de la conformité du matériel, réalisation des contrôles qualité, etc.
- Réaliser des travaux de contrôle dans les temps et le budget imparti, application des dossiers de réalisation et des procédures internes
- Participer aux essais de mise au point et de qualification de nouveaux outillages et procédés selon l'actualité du service
- Représenter la société sur chantier et être référent auprès du client

PROFIL RECHERCHÉ

De formation technique bac+ 2 à bac + 5 en mécanique/matériaux, avec une expérience confirmée en conduite de chantiers CND/END, sur des sites industriels en exploitation dans le nucléaire.

COFREND niveau 3 UT demandé, TOFD secteur CIFM

Mobilité France entière (grand déplacement).

Véhicule de service

AUTRES INFORMATIONS

Flamanville est la première affectation, vous serez amené à intervenir sur d'autres sites sur le territoire national en grand déplacement.

Figure 2. *Exemple d'offre d'emploi spécifique fournie par notre partenaire*

4.2. Module d'indexation des profils

Les profils collectés et pré-traités doivent être stockés et indexés pour permettre leur récupération à l'arrivée d'une requête. Les solutions de stockage retenues vont fortement dépendre du mécanisme de filtrage des candidats à évaluer pour un besoin (4.3). En fonction de la modélisation retenue (4.1), plusieurs solutions telles que l'indexation des documents pré-traités par Lucene (Casagrande *et al.*, 2017) ou le stockage dans une base de données relationnelle (Tinelli *et al.*, 2009; Stencel *et al.*, 2018) sont possibles. Nous attirons l'attention du lecteur sur la discrétion de la plupart des auteurs quant à la solution retenue pour le stockage des profils.

4.3. Module de filtrage des profils

Habituellement, les systèmes de recommandation sont utilisés en ligne et intégrés dans d'autres applications (Ramanath *et al.*, 2018). Le temps de réponse d'un tel système se doit donc d'être inférieur à quelques secondes. La fonction d'évaluation de pertinence, f_s , ne peut par conséquent pas être appliquée à tous les appariements candidats. Il est donc essentiel de s'appuyer sur un mécanisme de filtrage (Tinelli *et al.*, 2009) ou une méthode de pré-scoring (Casagrande *et al.*, 2017; Ramanath *et al.*, 2018) permettant de filtrer des candidats de l'ensemble des profils P .

Formellement,

$$\begin{aligned} f_f : O &\rightarrow P^n \\ P_i^f &= f_f(o_i) \end{aligned} \tag{4}$$

Avec f_f la fonction de filtrage, n le nombre de profils à retourner pour produire des appariements candidats et P_i^f l'ensemble des profils filtrés par rapport à la requête o_i tel que

$$\begin{aligned} P_i^f &\subset P \\ |P_i^f| &\ll |P| \end{aligned} \tag{5}$$

Et, idéalement,

$$\forall r_{ij} \in R_i^f \mid \forall r_{ik} \notin R_i^f \mid s_{ij} \geq s_{ik} \tag{6}$$

Avec R_i^f l'ensemble des recommandations produites pour les appariements candidats.

La conception de ce module n'est pas l'objet principal de nos recherches. Nous évaluerons cependant expérimentalement la fonction de filtrage qui sera retenue. Pour

ce faire, nous proposons de comparer, à f_s constante, une métrique telle que le Rappel@k avec et sans l'application de la fonction de filtrage.

4.4. Moteur d'évaluation des appariements candidats

La fonction d'évaluation d'appariement introduite en 3.1 est le coeur d'un système de recommandation tel que celui que nous proposons dans cet article. Elle permet d'associer à un appariement candidat (o_i, p_j) un score de pertinence. Dans les faits, ce score de pertinence peut être assimilé à la probabilité qu'un expert du domaine juge le profil p_j apte à remplir les tâches définies, ou non, dans o_i .

4.4.1. Définition de la pertinence d'un appariement

Nous avons, jusqu'ici, utilisé le terme de *pertinent* sans en donner la définition. Selon la portion du processus de recrutement à couvrir, sa définition va varier:

- (a) Filtrer les candidatures (*Screening*) ou détecter des candidats passifs potentiels (*Sourcing*). La pertinence d'un appariement est, ici, une prédiction de la capacité d'un individu à exercer une activité professionnelle donnée ;
- (b) Identifier les appariements susceptibles d'être acceptés par les deux parties et de conduire à un recrutement. Dans ce cas, la pertinence recouvre la définition précédente et contient également la prédiction de la contractualisation entre les deux parties ;
- (c) Identifier les appariements susceptibles d'être acceptés et conduisant à une collaboration fructueuse entre les deux parties.

La pertinence définie en (a) peut être considérée comme factuelle et s'estime à l'aide d'informations explicites (prise de contact, réponse à une candidature, etc.). L'ajout de la prédiction d'acceptation d'une offre d'emploi (b) implique la prise en compte de variables cachées telles que la situation familiale d'un individu, ses souhaits, ses exigences salariales, sa satisfaction dans son activité actuelle, etc. Les auteurs de (Stencel *et al.*, 2018) essayent de prédire cette issue du processus de recrutement et avancent une précision de 16.3% contre 4.1% pour un expert humain. La prédiction du bon déroulement d'une collaboration (c) repose, elle aussi, sur des variables cachées. La définition même du « bon déroulement d'une collaboration » ne va pas sans choix arbitraires. Quelle que soit cette définition, sa mesure et son intégration dans un jeu de données d'entraînement ou d'évaluation présuppose une couverture numérique bien plus profonde et complète qui s'étend potentiellement sur de multiples applications. Cette ambition semble, en 2020, être l'apanage de grandes organisations dont les systèmes d'information et les procédures internes sont suffisamment matures (Guo *et al.*, 2016).

Par conséquent, nous focaliserons nos travaux sur la conception d'une fonction d'évaluation de pertinence d'un appariement basée sur la définition (a) de la pertinence. Nous construirons un jeu de données d'évaluation avec l'aide de notre parte-

naire qui évaluera la pertinence des exemples d'appariements à l'aide de son expertise dans le secteur d'activité étudié.

4.4.2. *Approches adoptées dans la littérature*

La conception d'un module implémentant cette fonction f_s est intrinsèquement liée aux choix retenus lors de l'étape de modélisation des éléments de O et de P . Deux aspects de cette modélisation sont particulièrement importants :

- Le formalisme utilisé: documents textuels bruts (Maheshwary et Misra, 2018 ; Qin *et al.*, 2018), sacs de symboles scorés⁴ (Casagrande *et al.*, 2017 ; Dhouib *et al.*, 2020) ou encore assertions sémantiques intégrées un graphe de connaissances en lien avec une ontologie du domaine (Tinelli *et al.*, 2009) ;
- Les caractéristiques retenues: compétences, nom de métier, tâches, expériences professionnelles d'un individu, secteur d'activité, entreprise, contraintes (telles que la mobilité, la langue ou des certifications), etc.

Les approches reposant sur l'usage de textes bruts utilisent généralement des méthodes d'apprentissage profond afin de se libérer de l'étape de *Feature engineering*. Des réseaux profonds récurrents intégrant des mécanismes d'attention (Vaswani *et al.*, 2017) ont été proposés dans (Maheshwary et Misra, 2018 ; Nigam *et al.*, 2019). L'utilisation de ce type de réseaux profonds implique cependant les limitations suivantes : nombre d'exemples requis important, coût en temps de calcul important et taille de documents à ingérer limitée (bien qu'il existe des méthodes permettant de produire, par apprentissage profond, des représentations de documents plus longs (Zhang *et al.*, 2020), appliqué dans (Bian *et al.*, 2020)).

Bien que les approches à base d'apprentissage profond puissent se révéler intéressantes pour nos travaux (notamment pour des étapes de *Feature Engineering*), nous pensons que les particularités de notre sujet et des corpus à notre disposition ne nous permettront pas d'apprendre correctement les connaissances du domaine. Par conséquent, nous n'envisageons pas d'utiliser l'apprentissage profond pour la conception de f_s .

Des méthodes plus classiques d'apprentissage automatique sont également appliquées dans la littérature (Stencel *et al.*, 2018), avec une performance notable de l'algorithme *XGBoost* (Chen et Guestrin, 2016). Les auteurs de (Ramanath *et al.*, 2018) concluent d'ailleurs leur article en présentant l'arbitrage réalisé par *LinkedIn.com* pour l'implémentation de leur système de recommandation. En effet, le réseau profond présenté dans l'article n'améliore la précision obtenue que de 3% par rapport à *XGBoost*, ce qui ne justifie pas, pour eux, l'augmentation significative du temps de latence d'une recommandation.

4. Nous incluons dans cette large catégorie les modélisations reposant sur une extraction de symboles, associés à un quelconque score de pertinence, depuis un document. Y sont donc inclus les modélisations à base de *TF-IDF*, *Bag of Words* ou encore *Bag of Concepts*.

Notons que des méthodes basées sur des similarités entre représentations vectorielles sont également appliquées. Cette catégorie d'approches est hétérogène. S'y côtoient en effet des représentations basées sur des sacs de concepts (Dhouib *et al.*, 2020) et des plongements spécifiques au domaine du recrutement (Ramanath *et al.*, 2018 ; Dave *et al.*, 2018) construits à partir de graphes⁵.

4.4.3. Approche envisagée

À ce stade de nos travaux, nous envisageons de combler le manque d'exemples à notre disposition et la faiblesse des signaux présents dans le texte en nous appuyant sur des représentations vectorielles enrichies sémantiquement par une étape de conceptualisation (présentée dans 4.1.2) et reposant sur des connaissances explicites du domaine.

5. Conclusion

Dans cet article, nous avons proposé une analyse du problème de la recommandation appliquée au recrutement et avons également tissé un lien fort entre les considérations propres au domaine de la Recherche d'Information et celles propres au recrutement de profils spécifiques à un secteur d'activité. Nous avons fait apparaître, par l'étude d'un exemple représentatif, des verrous scientifiques que nous devons résoudre dans nos travaux afin de concevoir, avec notre partenaire industriel, un système de recommandation répondant aux contraintes exposées dans cet article. Nous avons également présenté l'état de l'art de la littérature spécialisée sur le sujet et l'avons mis en relation avec les contraintes de nos travaux et la spécificité de notre sujet de recherche. Nos prochaines contributions se focaliseront sur l'étape primordiale de la conception de la taxonomie des compétences utilisées et sur la conceptualisation des documents textuels vers des concepts de cette taxonomie.

Remerciements

Ces travaux ont été entrepris dans le cadre d'une thèse CIFRE réalisée au sein du LIS (Laboratoire d'Informatique et Systèmes) en partenariat avec Moben & Rooster⁶, un cabinet de recrutement spécialisé dans l'industrie des procédés et développant des solutions logicielles accompagnant la transformation digitale des entreprises industrielles.

5. Dans le cas de *LinkedIn.com*, le graphe utilisé est l'*Economic Graph*, projet de long-terme de l'entreprise qui vise à modéliser l'économie mondiale sous forme de graphe des interactions entre les acteurs économiques. <https://www.linkedin.com/pulse/20121210053039-22330283-the-future-of-linkedin-and-the-economic-graph/>

6. <https://www.mobenrooster.com/>

6. Bibliographie

- Bastian M., Hayes M., Vaughan W., Shah S., Skomoroch P., Kim H., “Linked in skills: Large-scale topic extraction and inference”, *RecSys 2014 - Proceedings of the 8th ACM Conference on Recommender Systems*, p. 1-8, 2014.
- Bian S., Chen X., Zhao W. X., Zhou K., Hou Y., Song Y., Zhang T., Wen J.-R., “Learning to Match Jobs with Resumes from Sparse Interaction Data using Multi-View Co-Teaching Network”, p. 65-74, 2020.
- Casagrande A., Gotti F., Lapalme G., “Cerebra, un système de recommandation de candidats pour l’e-recrutement”, p. 6, 2017.
- Chen T., Guestrin C., “XGBoost: A scalable tree boosting system”, *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-August-2016, p. 785-794, 2016.
- Coulet J.-C., “La notion de compétence : un modèle pour décrire, évaluer et développer les compétences”, 2011.
- Dave V. S., Al Hasan M., Zhang B., AlJadda K., Korayem M., “A combined representation learning approach for better job and skill recommendation”, *International Conference on Information and Knowledge Management, Proceedings*, 1997-2006, 2018.
- De Smedt J., le Vrang M., Papantoniou A., “ESCO: Towards a Semantic Web for the European Labor Market”, *WWW2015 Workshop: Linked Data on the Web*, 2015.
- Dhouib M. T., Faron C., Tettamanzi A., Dhouib M. T., Faron C., “Injection of Knowledge in a Sourcing Recommender System To cite this version : HAL Id : hal-02996442 Injection of Knowledge in a Sourcing Recommender System”, 2020.
- Dhouib M. T., Zucker C. F., Tettamanzi A., Construction d’ontologie pour le domaine du sourcing, Technical report, 2018.
- Guo S., Alamadun F., Hammond T., “RésuméMatcher: A personalized résumé-job matching system”, *Expert Systems with Applications*, vol. 60, p. 169-182, 2016.
- Jiechieu K. F. F., Tsopze N., “Skills prediction based on multi-label resume classification using CNN with model predictions explanation”, *Neural Computing and Applications*, 2020.
- Kessler R., Lapalme G., “AGOHRA : Génération d’une ontologie dans le domaine des ressources humaines”, *Revue Traitement Automatique des Langues*, vol. 58, n° 1, p. 39-62, 2017.
- Khobreh M., “Ontology Enhanced Representing and Reasoning of Job Specific Knowledge to Identify Skill Balance”, 2017.
- Maheshwary S., Misra H., “Matching Resumes to Jobs via Deep Siamese Network”, *WWW ’18: Companion Proceedings of the The Web Conference 2018*, p. 87-88, 2018.
- Nigam A., Roy A., Singh H., Waila H., “Job recommendation through progression of job selection”, *Proceedings of 2019 6th IEEE International Conference on Cloud Computing and Intelligence Systems, CCIS 2019*, p. 212-216, 2019.
- Paquette G., *Learning, Design, and Technology*, 2020.
- Qin C., Zhu H., Xu T., Zhu C., Jiang L., Chen E., Xiong H., Chen E.-H., “Enhancing Person-Job Fit for Talent Recruitment: An Ability-aware Neural Network Approach”, p. 25-34, 2018.
- Ramanath R., Inan H., Polatkan G., Hu B., Guo Q., Ozcaglar C., Wu X., Kenthapadi K., Geyik S. C., “Towards deep and representation learning for talent search at LinkedIn”, *Internationa-*

- tional Conference on Information and Knowledge Management, Proceedings*, 2253-2262, 2018.
- Stencel K., Janusz A., Ciebiera K., Ślęzak D., Stawicki S., Drewniak M., “How to Match Jobs and Candidates - A Recruitment Support System Based on Feature Engineering and Advanced Analytics”, p. 503-514, 2018.
- Tinelli E., Cascone A., Ruta M., Di Noia T., Di Sciascio E., Donini F. M., “I.M.P.A.K.T.: An innovative semantic-based skill management system exploiting standard SQL”, *ICEIS 2009 - 11th International Conference on Enterprise Information Systems, Proceedings*, vol. AIDSS, p. 224-229, 2009.
- Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I., “Attention Is All You Need”, 2017.
- Zhang X., Wei F., Zhou M., “Hibert: Document level pre-training of hierarchical bidirectional transformers for document summarization”, *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, p. 5059-5069, 2020.
- Zhao M., Javed F., Jacob F., McNair M., “SKILL: A system for skill identification and normalization”, *Proceedings of the National Conference on Artificial Intelligence*, vol. 5, p. 4012-4017, 2015.