

Cerebra, un système de recommandation de candidats pour l'e-recrutement

A. CASAGRANDE, F. GOTTI, G. LAPALME

RALI - Université de Montréal

Vendredis Quanti

4 octobre 2019



Plan

- 1 Introduction
- 2 Revue de la littérature
- 3 Cerebra
- 4 Améliorations
- 5 Conclusion

Contexte

Butterfly Predictive Project (BPP)

- ➔ Domaine des ressources humaines
- ➔ Identifier les candidats potentiels pour un poste donné
- ➔ 70 % des recruteurs reconnaissent avoir des difficultés à trouver les profils qui correspondent
- ➔ Collaboration avec le partenaire industriel LittleBigJob

Objectifs techniques et scientifiques

- ➔ Conception d'outils permettant d'effectuer une veille intelligente et dynamique du domaine de l'e-recrutement ;
- ➔ Prédiction de succès d'un candidat dans un poste dans une entreprise ;
- ➔ Constitution automatique de ressources linguistiques ;
- ➔ Nouvelle approche concernant l'appariement entre les candidatures et l'offre d'emploi ;
- ➔ Génération assistée de la marque personnelle pour les candidats.

Système de recommandation Cerebra

Objectifs de Cerebra

- ➔ Améliorer la récolte et l'agrégation d'information sur des candidats provenant de multiples sources (dont les données publiques sur les médias sociaux) ;
- ➔ Améliorer la mise en correspondance entre des candidats et des offres d'emploi ;
- ➔ Prédire le succès d'un candidat dans un nouveau poste.

Appariement offres d'emploi/candidatures

- ➔ Approche sémantique avec le Système IMPAKT (Colucci *et al.*, 2003, 2007, 2013)
- ➔ Approche par des mesures de similarité (Kessler *et al.*, 2008)
- ➔ Approche par apprentissage machine (Faliagka *et al.*, 2013)
- ➔ Approche par méthode vectorielle et utilisation des réseaux sociaux (Diaby et Vianet, 2014)

Tâche difficile à cause de la structure libre des candidatures et des offres d'emploi

Ressources : Offres d'emploi

Titre : LEAD DEVELOPER**Descriptif :**

[description de la société S qui cherche un candidat]

Pour accompagner la transformation digitale de ses clients, *S* a créé le pôle *P*. Ce pôle d'expertise rassemble équipe IT & des consultants métier, qui ensemble [...]

Nous déployons aujourd'hui ces solutions pour des clients de toute taille, en France et à l'international (plus de 60 pays couverts aujourd'hui).[...] Compétences techniques recherchées : PHP, Bootstrap, Javascript, idéalement ExtJS/SQL (MySQL) et Talend et DOT NET VB/C#, Sont un plus. Diplômé(e) d'une école d'ingénieur ou équivalent, vous justifier d'au moins 4 ans d'expérience dans le développement de solutions web.

Travaillant en étroite collaboration avec le manager de l'équipe IT de *P*, vous intervenez sur la conception, le développement et le maintien de nos solutions web. Dans ce cadre, vous prenez en charge les sujets suivants :

Réalisation des spécifications techniques de nos solutions, sur la base des cahiers des charges et spécifications fonctionnelles fournies par les consultants métiers,

Elaboration de plans de tests, accompagnement de leur réalisation,

Développement de nos applications,

Maintenance et support technique des solutions développées, en collaboration avec le reste de l'équipe de développement.

Résolution de problèmes technique de l'équipe, notamment ceux remontés par les collaborateurs moins expérimentés,

Contrôle de la qualité et amélioration continue de la qualité des livrables de l'équipe,

Localisation : Lyon

Langue : Français

Ressources : Profils

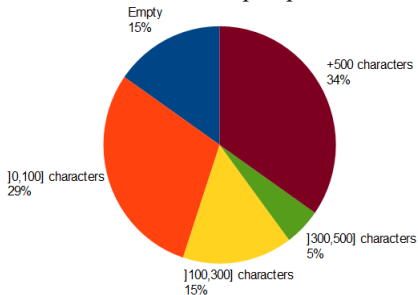
- ➔ 25 millions de profils issus de LinkedIn, Viadeo, Indeed, ...
- ➔ Profils canadiens ou français
- ➔ Profils "anonymisés" (sans nom et prénom)

Exemple de profil :

Formation	2015–2015 2003–2005	ESGI Infosup
Expérience	2008–maintenant	Responsable informatique : <i>société</i>
Compétences	Microsoft Office, Anglais, Microsoft Windows, Cloud Computing, Virtualisation, Cisco Technologies, ITIL, Linux, Microsoft Exchange, VoIP	

Difficultés liées aux profils

Nombre de caractères par profil



Statistiques

	Effectif	%
Nombres de profils	25 345 054	
Profils doublons	983 200	
Nombre de profils sans doublons	24 361 854	100%
Avec au moins une expérience	17 815 286	73.1%
Avec au moins une formation	11 368 534	46.7%
Avec un titre-fonction	22 589 868	92.7%
Avec un pitch	4 538 639	18.6%
Avec au moins une compétence	7 953 709	32.6%
Avec au moins une relation	21 093 111	86.6%

Ressources : Univers

Id	Nom	Univers apparentés
2	Aéronautique et aérospatiale	9 Industries automobiles 18 Digital, e-commerce, big data, jeux électroniques 17 Défense et armement, police, sécurité, transport de fonds 20 Énergie, eau, nucléaire, pétrole, gaz 42 Télécoms, hébergement, internet
13	Conseil et services informatiques, édition de logiciels	18 Digital, e-commerce, big data, jeux électroniques 42 Télécoms, hébergement, internet 10 Banque, finance, capital risque, fonds privés 22 Équipements électriques et électroniques, composants, matériel informatique 8 Audiovisuel, cinéma, spectacles, média, publicité, événementiel, divertissement, communication

Interface

Entrez une offre d'emploi

Offre entrée manuellement ▾

Titre

Avionic Methods Agent

Description

Nous assistons présentement notre client, situé sur la rive nord de Montreal, afin de combler une position temporaire d'une durée d'environ 2 mois comme Agent Méthodes Avionic. Responsabilités: Développer et élaborer des instructions de travail pour la réparation, la modification et la maintenance Assurer un soutien aux techniciens Travailler avec le service d'ingénierie Requis:

Langue français ▾ Pays France ▾

2 - Identifiez l'univers associé à cette offre

Trouver l'univers

0 : tous les univers ▾

3 - Combien de candidats désirez-vous évaluer (au maximum)?

5 ▾ Chercher

Appariement

Deux algorithmes

- ➔ WordMatcher
- ➔ SkillFinder

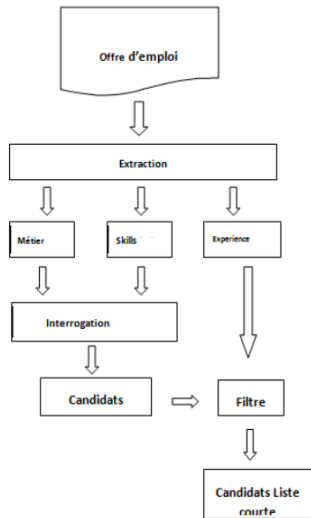
WordMatcher

- ➔ Approche vectorielle
- ➔ Création de vecteur à partir de l'offre et des profils
- ➔ Intersection des vecteurs

SkillFinder

3 étapes

- Extraction des informations utiles
- Recherche de profils
- Classement des résultats



Extraction des informations utiles

Titre d'emploi

- ➔ Suppression des chaînes inutiles
- ➔ Normalisation

Compétences

- ➔ Extraction à l'aide des compétences extraites des profils

Années d'expérience

- ➔ Recherche de *experience*, *expérience*, *années*, *years*, *related*, etc.

Exemple :

Titre extrait	lead developer
Compétences extraites	gestion de projets, résolution de problèmes
Expérience extraite	4 ans

Recherche de profils

- ➔ Utilisation de Lucene
- ➔ Création d'une requête booléenne
- ➔ Ajout de la langue et du pays recherchés
- ➔ Exemple : `expfunc="lead developer" AND skillname:(gestion de projets, résolution de problèmes) AND univid=13.`

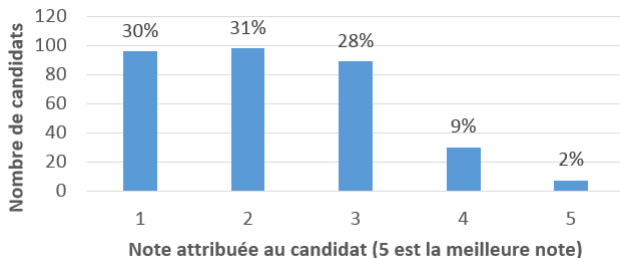
Classement des résultats

Composition du score :

- ➡ sous-score titre d'emploi
- ➡ sous-score compétences
- ➡ sous-score expérience
- ➡ sous-score lucene

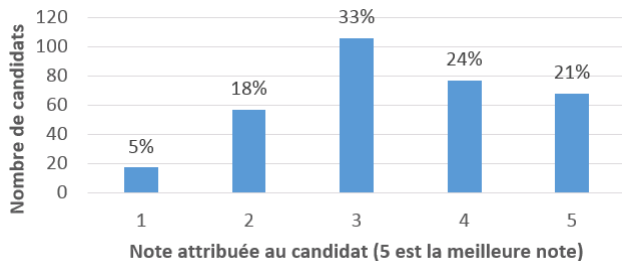
Évaluations

- Réalisées par notre partenaire à partir d'offres réelles
- 86 offres évaluées (320 candidats)
- Notation de 1 à 5 des profils proposés



Géolocalisation

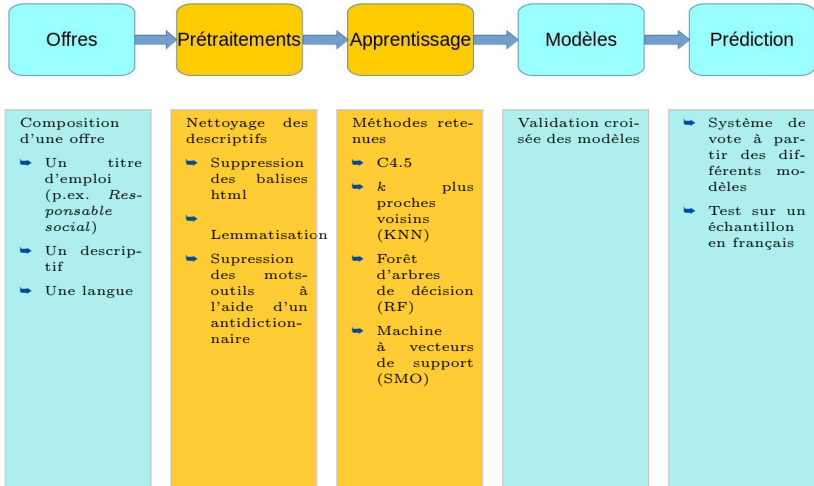
- Recherche de la ville à partir de fichiers ressources
- Recherche "à l'envers"
- Améliorations des évaluations



Détection de l'univers

- **But** : détection automatique du secteur d'activité d'une offre d'emploi
- **Motivation** : trouver des candidats issus du même secteur que l'offre
- **Méthode** : apprentissage machine, entraînement sur un corpus de
 - 56k en anglais
 - 31k en français

Processus



Méthodes d'apprentissage

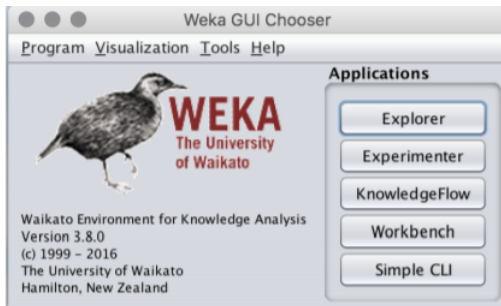
- ➔ **C4.5** (Quilan,2014) est une méthode d'apprentissage qui produit un arbre de décision et qui se base sur la mesure d'entropie.
- ➔ **K-plus-proches voisins (KNN)** classe une instance de test à l'aide des instances qui lui sont le plus similaires. L'algorithme calcule pour un individu sa distance à tous les autres individus de la base et identifie la classe majoritaire parmi les k individus ayant la plus petite distance. Nous avons choisi de prendre $k = 1$.
- ➔ **Forêt d'arbres de décision (RF)** (Breiman,2001) : plusieurs arbres de décision sont entraînés sur des sous-ensembles de données légèrement différents. On procède ensuite à un vote majoritaire pour obtenir la classe prédite. Nous avons choisi de considérer 100 arbres de décision.
- ➔ **SMO (sequential minimal optimization)** (Platt,1999) est un algorithme pour résoudre le problème d'optimisation quadratique qui se produit lors de l'entraînement des machines à vecteurs de support (SVM). Les SVM cherchent à déterminer l'*hyperplan de marge optimale* qui sépare correctement les données de classes distinctes.

Création des corpus d'apprentissage et de test

- **Recruteurs** : deux stratégies
 - univers selon l'entreprise
 - univers selon le titre d'emploi
- **Corpus d'apprentissage et de test** :
 - 2 corpus d'apprentissage et 2 corpus de test par langue
 - création automatique de ces corpus à l'aide de :
 - concordancier univers/entreprises
 - concordancier univers/emplois
 - création 2 corpus de test étiquetés manuellement pour le français

Outil utilisé : Weka

- ➔ **Weka** : suite de logiciels d'apprentissage automatique
- ➔ **Utilisation** via une interface ou comme librairie Java



Résultats

Validation croisée

Corpus anglais	Précision	Rappel	F-mesure
C4.5	56%	55%	55%
KNN	73%	71%	72%
RF	77%	69%	69%
SMO	70%	70%	70%
Corpus français			
C4.5	49%	49%	49%
KNN	66%	63%	64%
RF	74%	63%	63%
SMO	61%	61%	61%

Échantillon de test en français

	Sect. correct
C4.5	54%
KNN	68%
RF	65%
SMO	67%

Système de vote

Secteur correct : 74 %

Corpus étiqueté manuellement (compagnie)

	1000 mots		1500 mots	
	Univ. correct	Univ. apparenté	Univ. correct	Univ. apparenté
C4.5	14%	36%	10%	34%
KNN	7%	24%	7%	24%
RF	15%	48%	15%	49%
SMO	18%	38%	23%	40%

- ➡ concordancier incomplet
- ➡ expertise de l'étiqueteur

Bilan

- ➔ Résultats encourageants
- ➔ Première version prometteuse

Limites

- ➔ Difficultés de comparaison
- ➔ Expériences trop anciennes
- ➔ Orthographe

Pistes de recherche futures

- ➔ Améliorations de la fonction score
- ➔ Etude approfondie du vocabulaire des offres et des profils
- ➔ Intégration de l'ontologie développée par l'équipe
- ➔ Amélioration de la localisation (élargissement aux villes voisines)

Merci de votre attention