



Détection et analyse des signaux faibles pour un service caché Lanceurs d'alerte

Julien Maitre

► To cite this version:

Julien Maitre. Détection et analyse des signaux faibles pour un service caché Lanceurs d'alerte. APVP 2019 – l'Atelier sur la Protection de la Vie Privée, Jul 2019, Saint-Valery-sur-Somme, France. hal-02552561

HAL Id: hal-02552561

<https://hal.science/hal-02552561>

Submitted on 23 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Détection et analyse des signaux faibles pour un service caché Lanceurs d'alerte

Julien Maitre

L3i University of La Rochelle La Rochelle, France julien.maitre@univ-lr.fr

I. INTRODUCTION

Pour les décideurs, l'objectif principal est de prendre des décisions éclairées face à l'augmentation drastique des signaux transmis par des systèmes d'information toujours plus nombreux. Des phénomènes de saturation des capacités de nos systèmes conduisent à des difficultés d'interprétation ou même à refuser les signaux précurseurs de faits ou d'événements. La prise de décision est limitée par les nécessités temporelles et nécessite donc un traitement rapide de la masse d'informations. Être capable de détecter en un temps fixe, les bons signaux porteurs d'informations utiles dans un contexte de stratégie d'anticipation, est un défi devenu permanent pour de nombreux acteurs économiques. Il est donc nécessaire de développer, sous la forme de plateformes d'investigation (cf. Figure 1), de nouveaux services d'aide à la décision pour les politiques et les organisations chargées de ces activités. Les prises de décision, qui doivent porter à la fois sur la crédibilité de la source d'information et sur la pertinence des informations révélées dans un événement, nécessitent des algorithmes robustes pour détecter les *signaux faibles*, extraire, analyser les informations fournies par ce dernier et s'ouvrir à un contexte d'information plus large. [1]

II. OBJECTIFS

Nous proposons de nous concentrer sur les deux points essentiels suivants: (1) la détection des *signaux faibles* et (2) l'extraction des informations qu'ils véhiculent. La Figure 1 donne notre vision d'une plateforme d'investigation où les acteurs interagissent entre eux par le biais d'outils et de processus sécurisés. Notre but est donc la détection de signaux précurseurs dont la présence contiguë dans un espace de temps et de lieux donné anticipe l'avènement d'un fait observable. Cette détection est facilitée par les premières informations fournies par un lanceur d'alerte sous forme de documents. Ils exposent des faits prouvés, unitaires et ciblés mais aussi partiels et relatifs à un événement déclencheur. Le lanceur d'alerte fournit des informations qui ne sont pas encore détectables / apparentes sur les réseaux sociaux spécialisés. Elles permettent de dessiner le contour des signaux à venir sur les réseaux, facilitant ainsi leur détection et l'extraction des informations qu'ils véhiculent.

La procédure d'investigation proposée repose donc sur la détection de *signaux faibles* à partir des documents reçus. Trois actions sont donc entreprises :

- Action 1: Analyse automatique de contenus avec un minimum *d'a priori*. Identification des informations per-

tinentes. Indicateur de cohérence des thèmes obtenus. Détection des *signaux faibles*;

- Action 2: Agrégation de connaissances. Enrichissement de l'information.
- Action 3: Visualisation analytique. Mise en perspective de l'information par la création de représentations visuelles et de tableaux de bord dynamique.

Notre étude porte sur la recherche des mots-clés relevant d'un thème "signal faible" éventuellement présent dans un corpus de documents transmis par un lanceur d'alerte. Nos premiers résultats portent sur la première phase de la chaîne de traitements de la plateforme, autrement dit, l'analyse des documents reçus doit simultanément permettre de : (1) découvrir les thèmes ou l'arborescence des thèmes portés par les documents; (2) classer les documents relativement aux thèmes; (3) détecter les mots-clés pertinents composant les thèmes, (4) et enfin, découvrir les mots-clés du thème "signal faible". Nous prôtons l'utilisation d'une approche conjointe, modèle thématique (LDA [2]) et plongement lexical (Word2Vec [3]). La première vise principalement à décrire des documents et des collections de documents en leur assignant des distributions de thèmes, qui à leur tour ont des distributions de mots assignés. Elle capture ainsi des associations au niveau des documents. La seconde cherche à positionner des mots dans un espace vectoriel latent. Elle n'est pas vraiment conçue pour décrire des documents mais permet la capture des associations très locales. Dans le cas par exemple de Word2vec, les mots sont représentés par un vecteur de longueur fixe et attribue une signification sémantique aux distances entre les représentations des mots, ce qui est une des caractéristiques recherchées des signaux faibles (mots sémantiquement reliés). Les deux approches s'avèrent donc complémentaires car le modèle s'applique pour la première sur la représentation d'un document par un vecteur de longueur fixe, le second s'attache à décrire un mot par un vecteur de longueur fixe.

Nous retenons pour notre étude les qualificatifs suivant des signaux faibles : rareté, unitaire, non relié à des paradigmes existants, nouveauté, anormalité, mots-clés sémantiquement reliés. Cette définition du signal faible et de ses caractéristiques est essentielle en effet, car elle conditionne le critère d'évaluation qui permet de déterminer quel cluster final a capturé les mots clés relevant du signal faible. Elle justifie l'approche conjointe modèle thématique / plongement lexical, et contraint le choix des méthodes utilisées : LDA et Word2vec (cf. Figure 2). Cette dernière permet notamment de construire

