Search trends symptoms dataset, un jeu de données anonymisées sur les recherches sur Google liées aux symptômes associés au Covid-19.

La recherche académique est aujourd'hui très active et de nombreux chantiers sont ouverts. Certaines pistes de recherche sont particulièrement prometteuses, notamment pour répondre à des questions importantes. D'abord, comment combiner équité et vie privée ? En effet, si la confidentialité différentielle est aujourd'hui une référence pour la protection de la vie privée lors de la publication de données, son impact sur l'équité n'est pas anodin. Son adoption par le Census a par exemple posé clairement ce problème. En effet, le bruit ajouté aux requêtes pour anonymiser les données du recensement peut impacter différemment certains groupes vis-à-vis d'autres. C'est notamment le cas des statistiques concernant les langues minoritaires, utilisées afin de décider dans quelles langues le matériel de vote doit être distribué. Le taux d'erreur, c'est-à-dire l'impact du bruit ajouté, varie entre les juridictions en fonction du nombre de locuteurs minoritaires recensés.

DANS LA TÊTE DE L'ATTAQUANT

Autre problématique, comment compléter les techniques de protection formelles comme la confidentialité différentielle avec des approches empiriques d'attaque ? L'objectif est d'évaluer l'efficacité d'une anonymisation au-delà des garanties théoriques du modèle. La tâche est cruciale, car un modèle de vie privée ne capture pas tout. Il est ici nécessaire de se mettre à la place d'un attaquant qui observerait des données anonymisées et tenterait d'en extraire une information sensible. De quelles connaissances auxiliaires dispose-t-il (partagées sur les réseaux sociaux, obtenues de façon illicite…) ? Avec quelles méthodes va-t-il exploiter les données anonymisées et les connaissances auxiliaires (réidentification, reconstruction ou inférence d'appartenance) ? Le niveau de protection assuré par une méthode d'anonymisation peut être estimé en comparant le succès d'attaques sur les données brutes et sur les données anonymisées. Cette approche empirique, jamais exhaustive, de gestion du risque est complémentaire des approches formelles, car elle guide les choix complexes de modèle de vie privée, d'algorithme d'anonymisation et de paramétrage (par exemple, le budget).

Enfin, dernier exemple : comment combiner chiffrement et perturbations aléatoires liées à la confidentialité différentielle, par exemple dans un contexte de calcul sécurisé (donc hors du domaine d'application initial de la confidentialité différentielle, c'est-à-dire l'anonymisation) ? Les techniques de protection habituelles sont parfois trop coûteuses ou trop rigides quand les données sont massives ou distribuées. En autorisant un dévoilement contrôlé d'informations au cours du calcul, les techniques de perturbation aléatoire satisfaisant la confidentialité différentielle peuvent compléter les techniques cryptographiques en accélérant les performances tout en conservant des garanties formelles. Le prix à payer est la perte en précision des calculs due aux perturbations aléatoires.

En pleine ébullition, les travaux sur la confidentialité différentielle et son amélioration nous promettent un avenir où notre vie privée sera garantie de façon plus sûre qu'elle ne l'était jusqu'à présent. De la sorte, nous aurons moins de réticences à confier nos données. À l'heure de l'open data, c'est une bonne nouvelle.

— **Les auteurs** —

> Tristan Allard
est maître de conférences à l'université de Rennes 1 et rattaché au laboratoire Irisa.

> Louis Béziaud
est doctorant en informatique à l'université Rennes 1 et à l'université du Québec à Montréal.

> Sébastien Gambs
est titulaire de la chaire de recherche du Canada en analyse respectueuse de la vie privée et éthique des données massives, et professeur au département d'informatique de l'université du Québec à Montréal.

— **À lire** —

S. Wagh *et al.*, DP-cryptography : Marrying differential privacy and cryptography in emerging applications, *Commun. ACM*, vol. 64 (2), pp. 84-93. 2021.

C. Sahin *et al.*, A differentially private index for range query processing in clouds, 2018 IEEE 34th international conference on data engineering (ICDE), pp. 857-868, 2018.

C. Dwork et A. Roth, The algorithmic foundations of differential privacy, *Found. Trends Theor. Comput. Sci.*, vol. 9, pp. 211-407, 2014.

M. Alaggan *et al.*, BLIP : Non-interactive differentially-private similarity computation on bloom filters, *SSS*, 2012.

Pourquoi l'anonymat favorise-t-il la violence des internautes ?

Se masquer pour faire le mal

Gérald Bronner

En février 2015, en France, la Commission nationale consultative des droits de l'homme (CNCDH) émettait un avis sur les discours malveillants sur internet. Dès les premières lignes de son texte, elle dénonçait la « prolifération des contenus haineux sur la Toile ».

L'idée qu'internet représente un risque de radicalisation des points de vue n'est pas neuve. Elle a bien souvent été confirmée, tant à travers le constat de la diffusion de propos racistes que de la visibilité publique de déclarations tombant sous le coup de l'apologie d'actes de terrorisme. Plus généralement, et beaucoup moins grave, reconnaissons-le, quiconque a fréquenté les réseaux sociaux sait qu'une simple conversation virtuelle peut facilement, et parfois sur un simple malentendu, dériver vers des fâcheries, voire des insultes ou des menaces. Le pire est probablement atteint sur les forums des divers sites d'information où ceux qui échangent des commentaires ne sont même pas censés être « amis ».

La violence des discussions et la façon dont s'apostrophent, souvent par des noms d'oiseaux, ceux qui participent à ces échanges doivent sans doute beaucoup au fait qu'ils ne sont pas physiquement en contact les uns avec les autres : cette distance peut inspirer des audaces que la vie réelle n'autoriserait pas. Dans la vie ordinaire, la proximité spatiale entre les individus les enjoint souvent à éviter d'utiliser l'insulte ou l'invective. Ce n'est pas que la chose soit impossible, mais elle est plus rare, car en faire usage c'est prendre le risque de payer instantanément le prix de son agressivité. Au contraire, dès qu'une certaine distance physique rend un peu irréel le danger qu'implique toute interaction physique, la situation change. Un peu comme l'automobiliste, protégé par l'habitacle de son véhicule, hurle plus volontiers contre ses congénères que s'il les croisait sur le trottoir.

La garantie pacificatrice de la proximité physique est encore moins garantie sur les réseaux sociaux, et qu'ils facilitent un climat agonistique n'est donc pas surprenant. Les échanges sur internet incitent à ce que John Suler, de l'université Rider, à Lawrenceville, appelle la « désinhibition numérique ». Il a identifié six facteurs qui la favorisent, et l'un d'eux est l'anonymat derrière lequel se cachent certains internautes.

LES DÉRIVES DE L'ANONYMAT

De fait, beaucoup profitent de l'anonymat que confère un pseudonyme protecteur pour s'abandonner à l'outrance. Cet anonymat fait partie de l'idéologie fondatrice du web qui se voulait un projet d'émancipation par les communautés virtuelles, ce sur quoi insistait, dès les années 1980, l'écrivain et enseignant américain Howard Rheingold, l'un des gourous d'internet alors naissant. Dès lors, ces communautés vous offraient la possibilité d'être « un autre » afin de vous extraire des contraintes de votre identité géographique. Personne ne songeait alors que, protégés par des avatars virtuels, les individus en profiteraient pour laisser libre cours aux manifestations les plus obscures de leur personnalité.

Pourtant, dans un tout autre domaine, l'anthropologie avait montré les risques de

En bref

> Sur internet, invectives, insultes et fâcheries sont monnaie courante.

> Parmi les facteurs qui expliquent cette tendance, la dissimulation derrière l'anonymat est un des plus importants.

> On retrouve ce phénomène dans le comportement des guerriers tribaux : masqués ou grimés, ils sont plus violents.

> L'anonymat sur internet est en outre très répandu, surtout chez les jeunes.