

DÉVELOPPEMENT ET APPLICATION DE MÉTHODES DE TRAITEMENT AUTOMATIQUE DES LANGUES SUR LES CAUSES MÉDICALES DE DÉCÈS POUR LA SANTÉ PUBLIQUE

// DEVELOPMENT AND APPLICATION OF NATURAL LANGUAGE PROCESSING METHODS TO MEDICAL CAUSES OF DEATH FOR PUBLIC HEALTH

Aude Robert¹, Yasmine Baghdadi², Pierre Zweigenbaum³, Claire Morgand¹, Cyril Grouin³, Thomas Lavergne³, Aurélie Névéal³, Anne Fouillet², Grégoire Rey¹ (gregoire.rey@inserm.fr)

¹ Inserm-CépiDc, Le Kremlin-Bicêtre, France

² Santé publique France, Saint-Maurice, France

³ LIMSI, CNRS, Université Paris-Saclay, Orsay, France

Soumis le 20.06.2019 // Date of submission: 06.20.2019

Résumé // Abstract

Introduction – Les causes médicales sont renseignées par les médecins sur les certificats de décès en texte libre avec une grande variété d'expressions. Les méthodes de traitement automatique des langues (TAL) permettent d'envisager leur exploitation dans des délais courts. Cet article décrit la démarche retenue pour développer ces méthodes et illustre leur utilisation pour l'alerte sanitaire.

Méthodes – L'identification des méthodes performantes s'inscrit dans le cadre d'un challenge international. Celui-ci consiste à transmettre aux équipes participantes un corpus de données, comprenant le texte libre et les codes CIM-10 considérés comme le *gold standard*, pour développer leur outil de prédiction des codes CIM-10, et à évaluer indépendamment les performances des outils sur un jeu de test.

Certaines méthodes ont été utilisées pour la classification des causes en texte libre dans des regroupements pertinents pour la surveillance réactive de la mortalité.

Résultats – Les meilleurs résultats ont été obtenus à l'aide de réseaux de neurones sur le corpus américain et avec des méthodes à base de règles sur le corpus français. Une méthode mixte, incluant des règles et de la classification par *support vector machine* (SVM), a produit des résultats meilleurs ou proches sur les deux corpus.

L'analyse de l'évolution temporelle de quatre regroupements de causes pour la surveillance réactive de la mortalité a mis en évidence des événements attendus (épidémies) et inhabituels.

Discussion – L'expérience du challenge et l'application pour la surveillance à visée d'alerte montrent l'apport et la performance des méthodes de TAL pour appuyer l'exploitation réactive des données de mortalité pour la santé publique.

Introduction – Medical causes are filled in by doctors on free-text death certificates with a wide variety of expressions. The use of natural language processing methods (NLP) is necessary to set a reactive data treatment process. This article describes the approach adopted to develop these methods and illustrates their use for health alert.

Methods – The identification of successful methods falls within an international challenge. This consists of transmitting a data package, including free text and ICD-10 codes considered as the gold standard, to the participating teams to develop their ICD-10 predictions tool, and to independently evaluate the performance of the tools on a test set.

Some methods were used for the classification of free-text causes in groups of causes specifically relevant to reactive mortality monitoring.

Results – The best results were obtained using neural networks on the American corpus and with rules-based methods on the French corpus. A mixed method including rules and classification by SVM, produced better or near results on both corpora.

The analysis of the temporal evolution of four groupings of causes for the reactive monitoring of mortality has highlighted expected events (epidemics) and unusual events.

Discussion – The international challenge experience and the application for alert surveillance show the contribution and performance of NLP methods to support the reactive use of mortality data for public health.

Mots clés : Mortalité, Traitement automatique des langues, Codage médical

// **Keywords**: Mortality, Automatic language processing, Medical coding

Introduction

Le volet médical du certificat de décès constitue une source d'information riche pour la recherche et la surveillance en santé dans tous les domaines¹. Cependant, pour identifier une situation sanitaire critique ou l'effet d'une intervention en santé publique, cette information doit être traitée de façon précise et constante, dans des délais courts.

Comme dans la quasi-totalité des pays industrialisés, la structuration du volet médical suit les recommandations de l'Organisation mondiale de la santé (OMS)², celles-ci visant une meilleure comparabilité internationale des données produites. Lorsque la saisie des certificats se fait par voie électronique, l'OMS recommande également que les certificateurs, obligatoirement des médecins pour des décès survenus en France, ne soient pas guidés par des outils d'aide au remplissage courants sur internet tels que des listes déroulantes ou de l'auto-complétion. Ces recommandations visent à ce que les données dépendent au minimum du mode de recueil, mais également à maximiser la richesse du langage utilisé.

Le Centre d'épidémiologie sur les causes médicales de décès (CépiDc) de l'Inserm remplit une mission légale de production de la statistique nationale sur les causes médicales de décès. Cela nécessite de coder les entités nosologiques déclarées (diagnostics, états morbides, traumatismes, actes médicaux) à partir du texte initialement rédigé. Le codage est effectué en utilisant la Classification internationale des maladies (CIM) pour la statistique officielle. Pour l'alerte sanitaire, les entités nosologiques déclarées peuvent être directement classées dans des indicateurs syndromiques construits pour répondre à un objectif de détection réactive de variations inhabituelles de la mortalité ou de mesure de l'impact d'évènements connus sur la mortalité, tels qu'une épidémie saisonnière ou une vague de chaleur. En l'absence de

traitement automatisé efficace du texte médical, la surveillance syndromique repose sur des données administratives sans information sur les causes médicales de décès³.

Or le texte libre donne inévitablement lieu à une grande variété d'expressions différentes pouvant être codées ou classées de façon identique, à la présence d'abréviations, et de fautes d'orthographe ou de frappe, qui rendent le travail de codage fastidieux et long. Le tableau 1 montre différents exemples de standardisation uniformisant le codage à partir d'entrées multiples lors de la certification. C'est pourquoi seul un traitement automatique du texte médical permet d'obtenir une information suffisamment normalisée pour une exploitation en santé publique dans des délais courts.

Parmi les évolutions méthodologiques apparues ces dernières années, les méthodes de traitement automatique des langues (TAL) ouvrent des perspectives d'exploitation facilitée des données textuelles.

Cet article vise à décrire quelles méthodologies, organisationnelle et technique, ont été retenues pour développer ces méthodes et donne un exemple d'utilisation possible pour l'alerte sanitaire.

Matériel et méthodes

Challenge international

Les méthodes de TAL se décomposent essentiellement en trois grandes familles : les méthodes fondées sur des règles de décision définies *a priori*, les méthodes faisant appel à des techniques d'apprentissage statistique et les méthodes hybrides combinant les deux précédentes. Chacune de ces familles comprend un très grand nombre d'approches possibles, pouvant dépendre du jeu de données et de la tâche à réaliser. Compte tenu du peu de connaissances préexistantes sur la tâche spécifique de codage des causes médicales de

Tableau 1

Type de standardisation nécessaire pour le codage des causes médicales de décès

Type de standardisation	Texte brut	Texte standardisé	Code CIM retenu par entité
Uniformisation de termes	Kc, néo, adénoK, cancer, K	Cancer	CXXX
Uniformisation de termes (orthographe)	Anévrysme	Anévrisme	I72X
	Broncho-pneumopathie	Bronchopneumopathie	J18X
Développement d'abréviations	IDM	Infarctus du myocarde	I25X
	IRA	Insuffisance respiratoire aiguë	J96X
	PNP	Pneumopathie	J1X
Développement de termes pluriquivalifiés	Métastases cérébrales, hépatiques, osseuses	Métastases cérébrales, métastases hépatiques, métastases osseuses	C7XX
	Brûlures torse et jambe	Brûlures torse, brûlures jambe	T2XX
Développement d'abréviations polysémiques, fonction du contexte	IRC	Insuffisance rénale chronique ou insuffisance respiratoire chronique	NXXX ou JXXX
Suppression de termes non pertinents	Probable, possible, plausible, très...	Ignorés pour le codage mais gardés pour les recherches	NC

décès à l'aide de méthodes de TAL, la première approche adoptée par le CépiDc-Inserm, en partenariat avec le Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (LIMSI-CNRS), a consisté à participer à l'organisation d'un challenge international ouvert à toutes les équipes souhaitant participer, dans le cadre de la plateforme d'évaluation CLEF e-Health, le principe ayant fait ses preuves qu'en définissant bien les règles d'une tâche partagée⁴, la multiplicité des acteurs, donc des approches, permet d'identifier rapidement par essai-erreur les méthodes les plus performantes⁵. Trois campagnes successives ont été créées, mais nous décrivons ci-dessous les grandes étapes de la tâche organisée dans le cadre du challenge de 2017, par ailleurs décrite de façon plus technique dans l'article de Névéol A. *et al*⁶.

L'organisation de la tâche s'est décomposée en trois grandes étapes.

La première étape était la construction du corpus de données. Ce jeu de données comprenait l'ensemble des lignes des certificats électroniques renseignés en texte libre par voie électronique sur chaque ligne des certificats. Un sous-ensemble contenait également la liste des codes CIM 10^e révision (CIM-10) obtenus après codage manuel des données, considéré comme le *gold standard*. Cette liste de codes était à déduire à partir du texte libre par des outils de TAL pour chacun des certificats du jeu d'évaluation. Chaque certificat de décès reçoit en moyenne quatre codes, appartenant le plus souvent à des chapitres différents. La complexité de la tâche réside dans le traitement du texte libre mais aussi dans la possibilité qu'une entité nosologique, selon son contexte (âge, autres causes présentes sur le certificat ou lieu de décès), peut être codée différemment. De plus, le nombre de codes à prédire sur une ligne est inconnu *a priori*. Les données de la France et des États-Unis ont été utilisées pour cette tâche. Le fait de proposer plusieurs langues permet de susciter davantage de participations de la communauté internationale des experts en TAL, souvent à la recherche de corpus de données riches, et de tester la souplesse et la capacité des outils par rapport à différentes langues.

Les jeux ont été décomposés en un jeu d'apprentissage et de développement et un jeu de test. Afin de se mettre dans une situation de vie réelle, avec des

nouvelles données à coder rapidement et qui ne sont pas forcément représentatives des données historiques (nouvelles expressions rencontrées, nouveaux codes ajoutés ou supprimés et évolution du codage), le jeu de test a été construit à partir des données les plus récentes, 2014 pour les données françaises et les mois les plus récents de 2015 pour les données américaines.

Au final, les jeux français et américains comprenaient respectivement 509 103 et 58 262 codes CIM-10 dont 74% et 68% dans le jeu d'entraînement et développement (tableau 2).

À chaque certificat était associé le sexe, l'âge quinquennal et le type de lieu de décès (ex. voie publique) du sujet décédé, des variables susceptibles d'influencer le codage des causes, mais ne permettant pas de ré-identifier les individus décédés.

La seconde étape a consisté à transmettre aux équipes participantes le jeu d'entraînement et de développement, incluant le texte brut et les codes CIM-10 à prédire. Cet envoi était accompagné des descriptifs précis des corpus et des versions du dictionnaire faisant la correspondance entre les expressions standardisées rencontrées et le(s) code(s) CIM-10 correspondant(s), mis à jour annuellement par les producteurs de données. Les équipes avaient alors quatre mois pour tester et affiner leur outil de prédiction des codes CIM-10.

La troisième étape consistait à transmettre le jeu de test aux équipes participantes, uniquement avec le texte brut et sans les codes CIM-10, et à demander en retour rapide le résultat fourni par leur outil de TAL. Des mesures de performance des outils étaient alors utilisées pour les évaluer en comparaison des codes CIM de référence (*gold standard*). Ces mesures étaient :

$$\text{Précision } P = \frac{\text{vrais positifs}}{\text{vrais positifs} + \text{faux positifs}}$$

$$\text{Rappel } R = \frac{\text{vrais positifs}}{\text{vrais positifs} + \text{faux négatifs}}$$

$$F\text{-mesure} = \frac{2 \times P \times R}{P + R}$$

Tableau 2

Statistiques descriptives des corpus de données utilisées dans le cadre de la tâche de codage automatique des causes de décès CLEF e-Health 2017

Nombre	France			États-Unis		
	Entraînement et développement (2006-2013)	Test (2014)	Total	Entraînement et développement (2015)	Test (2015)	Total
Certificats	93 694	31 690	125 384	13 330	6 665	19 995
Lignes	276 103	91 962	368 065	32 714	14 834	47 548
Mots	1 673 643	599 127	2 272 770	90 442	42 819	133 261
Codes CIM10	377 677	131 426	509 103	39 334	18 928	58 262

Les mesures ont été calculées sur l'ensemble des codes CIM-10, puis uniquement sur l'ensemble des codes de causes externes ou correspondant à des morts violentes (codes V01 à Y98). En effet, ces causes sont particulièrement suivies pour évaluer les politiques de prévention, et sont susceptibles d'être décrites de façon assez différente par les certificateurs.

Application à la surveillance réactive de la mortalité à visée d'alerte

La classification des données pour la surveillance syndromique répond à la même logique de classement automatique que le codage médical en CIM-10, avec une moindre exigence de précision du codage. C'est pourquoi l'expérience des deux premiers challenges internationaux, sans être directement répliquable, a été utile pour dégager deux principales méthodes de classification automatique des entités nosologiques (exprimées en texte libre) dans les indicateurs syndromiques. Il s'agit de : 1/ une méthode à base de règles et 2/ une méthode de classification SVM, intégrant les résultats de la méthode par règles. Ces méthodes ont été choisies parmi celles ayant démontré les meilleures performances pour le codage selon la CIM-10 sur le corpus français et pour leur simplicité pour une utilisation quotidienne et réactive. Les règles et paramètres de la classification SVM utilisés dans le cadre du challenge ont été adaptés pour répondre aux spécificités de cette tâche. L'entraînement et l'évaluation de la méthode de classification par apprentissage supervisé ont été effectués à partir d'un échantillon de 4 500 certificats de décès annotés manuellement (attribution d'un indicateur syndromique à chaque entité nosologique).

Les deux méthodes de classification ont ensuite été appliquées pour classer l'ensemble des entités nosologiques des certificats électroniques rédigés entre 2012 et 2016 (203 797 certificats)⁷. Le nombre hebdomadaire de décès, pour chacun des indicateurs syndromiques construits pour les besoins de la surveillance réactive de la mortalité, a été calculé. Les fluctuations de la proportion de décès pour ces indicateurs parmi les décès toutes causes confondues ont été analysées rétrospectivement, afin d'évaluer si des événements saisonniers ou inhabituels connus et validés dans le passé auraient été détectés à partir de ces indicateurs.

Les résultats pour quatre indicateurs (grippe, infection respiratoire aiguë basse (Irab), décès en lien avec la chaleur, brûlure/corrosion) sont présentés dans cet article. Ils permettent 1/ pour les deux premiers, de suivre l'évolution des épidémies saisonnières respiratoires hivernales⁸, 2/ pour le troisième, de suivre l'impact direct des épisodes de fortes chaleurs⁸, 3/ pour le dernier, d'évaluer la capacité de détection de l'impact d'un événement inhabituel, ponctuel et local (brûlure/corrosion en rapport avec la survenue d'un incendie mortel en Normandie en 2016).

Résultats

Challenge international

Les équipes participantes en 2017 provenaient d'Australie, de France, d'Allemagne, d'Italie et de Russie. Les équipes comprenaient souvent des membres issus de différentes disciplines : science des données, informatique, statistique, archivage et clinique. Le LIMSI, en tant que membre organisateur, n'a participé que de façon non officielle. Quinze résultats, provenant de 9 équipes ont été reçus pour le corpus américain et 9 résultats provenant de 6 équipes pour le corpus français. L'ensemble des résultats est décrit plus amplement ailleurs⁶.

Les outils développés par les équipes faisaient appel à des techniques très diverses. La plupart utilisaient des ressources terminologiques, dont les dictionnaires fournis par les organisateurs. Les participants se sont servis d'outils déjà existants en les adaptant à la problématique du codage en CIM-10. Parmi ceux-ci on retrouve des bibliothèques Python servant pour faire de l'apprentissage avec un SVM multi-étiquettes, la plateforme de recherche SOLR permettant l'utilisation de méthode de recherche d'information ou encore les réseaux de neurones. Certaines équipes ont aussi testé la combinaison de deux de ces méthodes afin de proposer un modèle hybride.

Les meilleurs résultats sur le corpus américain étaient obtenus par l'équipe russe à l'aide de méthodes utilisant des réseaux de neurones, avec une précision de 0,89, un rappel de 0,81 et une F-Mesure de 0,85 sur l'ensemble des causes⁷. Bien que nettement inférieurs, leurs résultats sur les causes externes étaient également les meilleurs (P=0,58, R=0,36 et F=0,44).

Sur le corpus français, les meilleurs résultats étaient obtenus par une équipe française, essentiellement à base de règles utilisant des ressources terminologiques propres et un correcteur orthographique, aussi bien sur les données réalignées (P=0,84, R=0,78, F=0,80) que sur les données brutes (P=0,86, R=0,69, F=0,76)⁸. Là encore, les meilleurs résultats étaient obtenus par la même équipe sur les causes externes réalignées (P=0,53, R=0,47, F=0,50) et brutes (P=0,57, R=0,43, F=0,49).

L'équipe organisatrice du LIMSI obtenait, avec une méthode mixte faisant appel à des règles et de la classification par SVM, des résultats meilleurs ou proches sur le corpus américain (P=0,90, R=0,80, F=0,84 sur l'ensemble des causes et P=0,72, R=0,37, F=0,49 sur les causes externes) et sur le corpus français (P=0,85, R=0,88, F=0,87 sur l'ensemble des causes et P=0,63, R=0,67, F=0,65 sur les causes externes).

Application à la surveillance réactive de la mortalité à visée d'alerte

Au regard des méthodes ayant obtenu les meilleurs résultats dans le challenge sur le corpus français, les entités nosologiques contenues dans les certificats

électroniques de 2012 à 2016 ont été classées dans les indicateurs syndromiques à partir de la méthode à base de règles et du modèle SVM. Les indicateurs « grippe » et « Irab » constituent traditionnellement deux traceurs de l'impact des épidémies saisonnières hivernales, le premier étant spécifique à l'épidémie de grippe, le second permettant une surveillance des pathologies respiratoires au sens large. La dynamique de ces deux indicateurs montrait une augmentation marquée sur les périodes hivernales (figure 1a). La proportion de décès de l'indicateur « grippe » parmi les décès toutes causes confondues augmentait spécifiquement sur les périodes épidémiques de grippe et était quasi nulle en dehors (figure 1a). L'indicateur « Irab » présentait également des augmentations en dehors des périodes épidémiques de grippe, comme c'est en particulier le cas sur les deux semaines de janvier 2016 ou sur l'été (S29-2015) et l'automne 2015 (S44-2015) (figure 1b).

La proportion de certificats mentionnant un effet de la chaleur (coup de chaleur/hyperthermie) parmi l'ensemble des décès était plus faible, mais montrait une hausse nette, notamment à travers la méthode par règles, sur la semaine 28 de 2015, au cours de laquelle une vague de chaleur majeure était survenue sur une large part du territoire (figure 1c).

Le nombre de décès avec une mention de « brûlure/corrosion » dans le certificat était faible et variait de 0 à 3 décès par semaine, à l'exception de la semaine 22 de 2015 (avec 7 décès) et de la semaine 31 de 2016, où 17 décès ont été identifiés (figure 1d). L'investigation de ce second pic a confirmé l'identification de l'impact de l'événement que nous cherchions à mettre en évidence, celui de l'incendie mortel en Normandie responsable de 13 décès.

Discussion

Les résultats du challenge ont montré le fort potentiel du traitement automatique des langues et ont permis d'identifier les stratégies prometteuses pour accélérer le codage des causes de décès et l'utilisation réactive de ces causes pour la veille et l'alerte sanitaire.

Avant d'entrer dans les chaînes de production du codage pour la statistique nationale, fiable et comparable à l'international, ces différents développements doivent encore faire l'objet de tests et de validations, notamment afin d'identifier les cas pour lesquels le traitement manuel reste indispensable. C'est possiblement le cas des causes externes, mêmes si ces résultats pourraient évoluer avec l'apparition d'une case à cocher sur les circonstances apparentes du décès en cas de mort violente dans les nouveaux certificats de décès. Le choix de méthodes fondées sur des règles ou sur de l'apprentissage statistique doit également tenir compte des enjeux d'évolution de terminologies et de règles de codage, comme l'implémentation de la onzième révision de la CIM prévue dans les années à venir. Dans tous les cas,

un effort conséquent d'annotation manuelle devra être fait pour valider l'application de règles ou pour construire de nouveaux corpus d'apprentissages.

Certaines méthodes semblent atteindre des résultats performants sur plusieurs langues. C'est un point important dans la perspective d'implémenter ces méthodes dans un outil de codage automatique générique tel que le logiciel international Iris⁹. En effet, de nombreux pays manifestent aujourd'hui leur manque de ressource pour traiter de façon rigoureuse et reproductible l'importante quantité d'informations présentes sur les certificats de décès. Il s'agit-là d'un enjeu de santé publique majeur¹⁰.

Par ailleurs, le cadre des recommandations de l'OMS sur l'aide à la certification est relativement contraignant, pour des motifs de comparabilité entre les différents pays et entre les différents modes de certification². Sous condition que la certification électronique soit suffisamment déployée, et que des outils de TAL soient accessibles de façon générique à l'international, il pourrait être envisagé à terme de les intégrer dans des outils de recueil plus interactifs, vérifiant la cohérence de la saisie ou proposant des corrections au bénéfice du certificateur, tout en respectant les objectifs poursuivis par l'OMS.

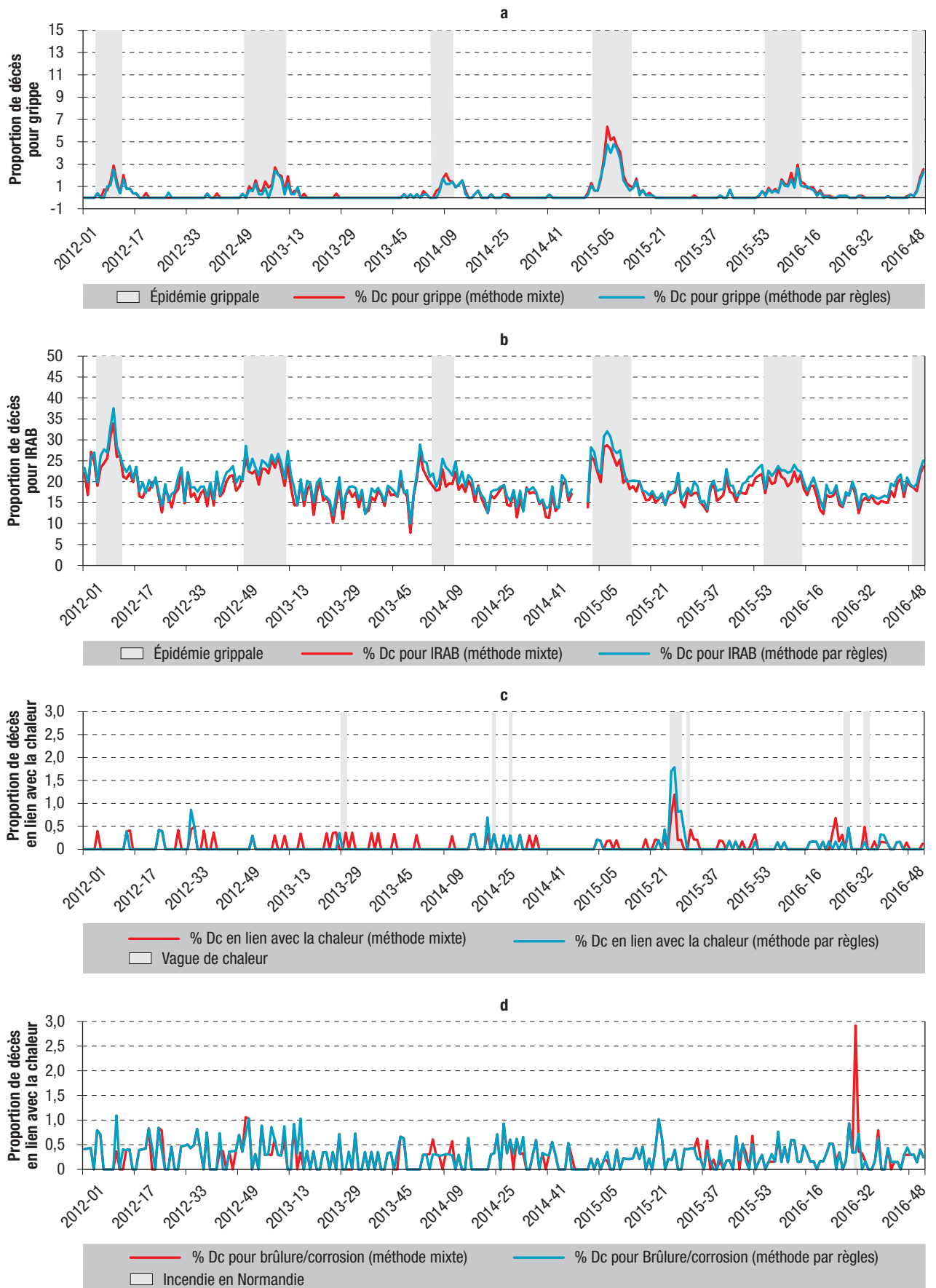
Ces méthodes de classification des causes médicales en texte libre pour la construction en routine des indicateurs syndromiques peuvent être moins spécifiques que le codage de la base finalisée des causes médicales de décès. C'est pourquoi, en attendant de nouvelles avancées, le TAL constitue d'ores et déjà une amélioration majeure pour la surveillance réactive de la mortalité à visée d'alerte¹¹. En effet, celle-ci repose actuellement sur la surveillance des évolutions des données administratives de mortalité (sans information sur les causes des décès), limitant l'interprétation des événements identifiés par ce système³.

Cependant, à l'image de la tâche de codage des causes de décès, les performances des méthodes de classification dans certains indicateurs syndromiques destinés à suivre des pathologies rares ou complexes nécessitent encore un travail d'amélioration¹².

En conclusion, cette expérience du challenge et l'application pour la surveillance à visée d'alerte présentée dans cet article montrent l'apport et la performance des méthodes de traitement automatique des langues pour appuyer de façon réactive l'exploitation des données de mortalité pour la santé publique. De plus, la forte participation de plusieurs équipes internationales au challenge, malgré l'absence de financement, montre que les données des causes de décès contribuent au développement de méthodes innovantes. En effet, ces données de santé, parmi les plus standardisées au niveau international, permettent à ces équipes de tester leurs outils sur de nouveaux jeux de données, avec des problématiques originales, multilingues et homogènes. ■

Figure 1

Proportion hebdomadaire de décès certifiés électroniquement entre 2012 et 2016 parmi l'ensemble des décès toutes causes confondues pour les indicateurs grippe, infection respiratoire aiguë basse (IRAB), brûlure/corrosion, et décès en lien avec la chaleur, France



Remerciements

Les auteurs remercient l'initiative de *Conference and Labs of the Evaluation Forum* (CLEF), pour leur soutien dans l'organisation du challenge de codage CIM-10 dans le cadre de CLEF eHealth de 2016 à 2018.

Ils remercient également Alix Bourrée pour son appui dans la mise en œuvre et l'évaluation des méthodes de classification pour la surveillance réactive de la mortalité.

Liens d'intérêt

Les auteurs déclarent ne pas avoir de liens d'intérêt au regard du contenu de l'article.

Références

- [1] Rey G. Les données des certificats de décès en France : processus de production et principaux types d'analyse. *Rev Médecine Interne*. 2016;37(10):685-93.
- [2] World Health Organization. ICD-10: International statistical classification of diseases and related health problems. 10th revision. Vol. 2. Geneva: WHO. 1992.
- [3] Baghdadi Y, Gallay A, Caserio-Schönemann C, Fouillet A. Evaluation of the French reactive mortality surveillance system supporting decision making. *Eur J Public Health*. 2018;29(4):601-7.
- [4] Brazier F, Treur J, Wijngaards N. Modelling interaction with experts: The role of a shared task model. In: *ECAI'96*. 12th European conference on artificial intelligence. John Wiley and Sons publisher; 1996. 241-5.
- [5] Huang CC, Lu Z. Community challenges in biomedical text mining over 10 years: Success, failure and the future. *Brief Bioinform*. 2016;17(1):132-44.
- [6] Névéal A, Anderson RN, Cohen KB, Grouin C, Lavergne T, Rey G, *et al*. CLEF eHealth 2017 multilingual information

extraction task overview: ICD10 coding of death certificates in English and French. In: *CLEF Evaluation Labs and Workshop: Online Working Notes*, CEUR-WS. 2017.

[7] Miftahutdinov Z, Tutubalina E. KFU at CLEF eHealth 2017 task 1: ICD-10 coding of English death certificates with recurrent neural networks. In: *CEUR Workshop Proceedings 2017 vol.1866*. 2017.

[8] Cabot C, Soualmia LF, Darmoni SJ. SIBM at CLEF eHealth Evaluation Lab 2017: Multilingual information extraction with CIM-IND. In: *CLEF Online Working Notes*. 2017.

[9] German Institute of Medical Documentation and Information (DIMDI). About Iris Institute [Internet]. <https://www.dimdi.de/dynamic/en/classifications/iris-institute/#about-iris>

[10] AbouZahr C, Gollogly L, Stevens G. Better data needed: Everyone agrees, but no one wants to pay. *The Lancet*. 2010;375(9715):619-21.

[11] Baghdadi Y, Bourrée A, Robert A, Rey G, Gallay A, Zweigenbaum P, *et al*. Automatic classification of free-text medical causes from death certificates for reactive mortality surveillance in France. *Int J Med Inform*. 2019;131:103915.

[12] Baghdadi Y, Bourrée A, Robert A, Rey G, Gallay A, Zweigenbaum P, *et al*. A New approach to compare the performance of two classification methods of causes of death for timely surveillance in France. *Stud Health Technol Inform*. 2019;264:925-9.

Citer cet article

Robert A, Baghdadi Y, Zweigenbaum P, Morgand C, Grouin C, Lavergne T, *et al*. Développement et application de méthodes de traitement automatique des langues sur les causes médicales de décès pour la santé publique. *Bull Epidémiol Hebd*. 2019;(29-30):603-9. http://beh.santepubliquefrance.fr/beh/2019/29-30/2019_29-30_5.html