

Deux approches pour catégoriser le risque

Natalia Grabar, Niña Kerry

STL UMR 8163 CNRS, Université Lille 3 et Lille 1

natalia.grabar@univ-lille3.fr, ninia@boujut.com

<http://natalia.grabar.perso.sfr.fr/>

Résumé. Le risque chimique ou alimentaire couvre les situations où les produits chimiques sont dangereux pour la santé et consommation humaine ou animale, et pour l'environnement. Les experts qui assurent le contrôle et la gestion de ces substances se retrouvent face à de gros volumes de littérature scientifique, qui doit être analysée pour appuyer la prise de décisions. Nous proposons une aide automatique pour l'analyse de cette littérature. Nous abordons la tâche comme une problématique de catégorisation: il s'agit de catégoriser les phrases des textes dans les classes du risque lié aux substances. Nous utilisons deux approches: par apprentissage supervisé et la recherche d'information. Les résultats obtenus avec l'apprentissage supervisé (toute classe confondue, F-mesure autour de 0,8 pour le risque alimentaire, entre 0,61 et 0,64 pour le risque chimique) sont meilleurs que ceux obtenus avec par recherche d'information (toute classe confondue, F-mesure entre 0,18 et 0,226 pour le risque alimentaire, entre 0,20 et 0,32 pour le risque chimique). Le rappel est compétitif avec les deux approches.

1 Introduction

Le risque chimique ou alimentaire se manifeste lorsque les produits chimiques sont dangereux pour la santé et consommation humaine ou animale, et pour l'environnement. Si certains produits et substances sont maintenant clairement identifiés comme dangereux (*e.g.* l'amiante, l'arsenic, le plomb), nos connaissances actuelles sur d'autres substances sont moins complètes. Nous nous intéressons en particulier au risque alimentaire (*e.g.* l'arsenic, les nitrates, la listeria, la dioxine) et au risque chimique (*e.g.* le bisphénol A, les phtalates). Ces substances entrent souvent dans la composition de produits courants et peuvent avoir l'effet nuisible sur l'organisme humain. Le contrôle sur la commercialisation de ces substances est effectué par des organismes sanitaires dédiés, comme EFSA (European Food Safety Authority) ou ANSES (Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail). Les experts se retrouvent face à une littérature scientifique abondante et doivent l'étudier pour avoir une base solide pour la prise de décisions. L'objectif de notre travail consiste à proposer une aide automatique pour l'analyse de la littérature scientifique afin de détecter les phrases indicatives du risque induit par ces substances. Nous abordons cette tâche comme une problématique de catégorisation : les phrases des textes doivent être catégorisées dans les classes du risque. Nous présentons les données (section 2) et approches utilisées (sections 3 et 4). Nous discutons ensuite les résultats obtenus et concluons avec les pistes pour les travaux futurs (section 5).