

# **SentiQ: Une approche logique-probabiliste pour améliorer la qualité de l'analyse des sentiments**

Wissam Mammam Kouadri\*\*\*, Salima Benbernou\*, Mourad Ouziri\*  
Themis Palpanas\*, Iheb Ben Amor\*\*

\*Université de Paris, France  
nom.prenom@u-paris.fr

\*\*IMBA Consulting, France  
prenom.nom@imbaconsulting.com

## **1 Introduction**

L'opinion exprimée dans les données issues des réseaux sociaux représente un facteur majeur dans le processus de prise de décision. Cependant, malgré les avancés de recherche dans le domaine de l'analyse du sentiment, elle reste une tâche difficile en raison de la richesse et la complexité du langage naturel qui permet d'exprimer le même sentiment de différentes manières. Pour illustrer cette complexité, considérons les deux phrases : (a) Donald Trump softens tone on Chinese investments et (b) Trump drops new restrictions on China investment. Bien que (a) et (b) soient structurées différemment, elles sont sémantiquement équivalentes et expriment la même idée. Les travaux de recherche ont consenti que les textes sémantiquement équivalents doivent avoir la même polarité (Positive, Négative, ou Neutre). Cependant, nous avons constaté à travers des expérimentations intensives, que les algorithmes d'analyse de sentiments ne détectent pas la similarité entre les documents et extraient des polarités différentes conduisant à des incohérences *intra et inter-algorithmes*. Les incohérences intra-algorithme se produisent lorsqu'un algorithme extrait des polarités différentes de documents sémantiquement équivalents. Les incohérences inter-algorithmes représentent le cas où deux algorithmes extraient des polarités différentes d'un document. Dans ce papier, nous présentons SentiQ, un framework basé sur les réseaux logiques probabilistes de Markov (MLN) qui identifie et résout les deux types d'incohérences tout en améliorant la précision des algorithmes.

## **2 SentiQ : Une approche logique probabiliste pour une analyse de sentiment de qualité**

Le vote majoritaire représente la solution triviale pour les problèmes d'incohérences intra et inter-algorithmes. Cependant, cette méthode attribue les mêmes pondérations à tous les algorithmes, ce qui implique une dégradation de la précision des algorithmes comme illustré dans Kouadri et al. (2020). Les limites du vote majoritaire ont été surmontées par des travaux de recherche comme celui de Ratner et al. (2020) qui attribue des pondérations aux algorithmes