

Vous êtes ici :

Accueil ▫ Actualités

▫ **Mesure de l'innovation technologique : tirer parti de l'analyse sémantique des brevets**

OBSERVATOIRE DES SCIENCES ET TECHNIQUES (OST)

Mesure de l'innovation technologique : tirer parti de l'analyse sémantique des brevets

Publié le 30 octobre 2019

Le département Observatoire des Sciences et Techniques (OST) lance un projet de développement de nouveaux indicateurs pour mesurer l'innovation technologique grâce à l'analyse sémantique des brevets. Dominique Guellec, conseiller scientifique de l'OST et pilote du projet, en précise les objectifs et les enjeux.



Dominique Guellec, conseiller scientifique (OST - Hcéres)

En septembre dernier, vous avez rejoint l'OST pour piloter le projet d'analyse sémantique des brevets qui permettra notamment de produire des indicateurs reflétant la nouveauté et l'impact des inventions. Quelles sont les limites rencontrées aujourd'hui dans la mesure de l'innovation technologique et quel serait l'apport de l'analyse sémantique ?

Afin de vous proposer des vidéos, des boutons de partage, des contenus remontés des réseaux sociaux et d'élaborer des statistiques de fréquentation, nous sommes susceptibles de déposer des cookies tiers sur votre machine. Cela ne peut se faire qu'en obtenant, au préalable, votre consentement pour chacun de ces cookies.

Cette approche a ses mérites mais elle peut être améliorée. En effet le brevet, s'il reflète effectivement l'innovation, fonctionne selon sa propre logique, juridique et stratégique, qui est source de bruit et de biais statistiques. Les techniques nouvelles de traitement automatique du langage (TAL) permettent d'analyser le texte des brevets, qui inclut une description de l'invention, et d'en déduire des indicateurs reflétant non plus le brevet en tant que tel, mais directement l'invention qu'il décrit, ce qui sera plus précis du point de vue de la technologie. Le TAL est une branche de l'intelligence artificielle ; il est utilisé dans nombre de domaines telle l'interprétation des questions (sur nos téléphones portables), la classification de textes etc., et il est l'objet de recherches très actives.

Quels sont les objectifs du projet ?

Le projet va développer et tester des indicateurs reflétant la dynamique de l'innovation en appliquant le TAL aux brevets. Différents indicateurs pourraient être compilés, reflétant respectivement la nouveauté, l'impact ou d'autres caractéristiques des inventions. L'agrégation de ces indicateurs au niveau d'une entreprise, d'un organisme de recherche ou d'un pays refléterait la dynamique d'innovation de l'entité concernée.

Quelles données et méthodes seront utilisées ?

Les données de brevets sont publiques. Il existe différentes bases accessibles et suffisamment propres pour des traitements statistiques (comme la base de l'OST). Les données textuelles devraient y être ajoutées pour ce projet, ce qui est possible car elles sont rendues disponibles par les offices de brevets. Ce projet se concentrerait sur un ou plusieurs des principaux offices de brevets : OEB, OMPI ou USPTO.

La méthode consistera, dans une première étape, à vectoriser les textes (brevets), ce qui permettra dans une seconde étape de les comparer quantitativement. La vectorisation vise à capter les dimensions importantes du texte, les principaux aspects de l'invention ; elle permet également de réduire la dimensionnalité des documents et le bruit qui lui est associé. Différentes méthodes de TAL pourraient être utilisées pour ce projet, par exemple le *word embedding*, ou des techniques très récentes utilisant les réseaux de neurones artificiels. Ces techniques sont en progrès constant, les efforts actuels visent à mieux capter le contexte et la syntaxe au sein des documents. Une seconde étape consistera à calculer les indicateurs de nouveauté (distance sémantique entre un brevet et les brevets antérieurs) et d'impact (distance entre un brevet et les brevets postérieurs) : différents types de distance seront calculées, puis validées notamment par des humains (experts).

Pour quelles applications et productions ?

La nouveauté et l'impact sont des caractéristiques fondamentales de toute invention, et disposer d'indicateurs fiables serait utile pour toutes les analyses de l'innovation : à des fins d'évaluation, à des fins analytiques ou à des fins stratégiques. Ces indicateurs intéresseraient donc les chercheurs, les entreprises, les institutions publiques et les responsables de politiques publiques.

Les méthodes issues de ce projet pourraient aussi être appliquées à d'autres données comme les publications scientifiques ou les descriptifs d'inventions publiés sur Internet.

Ce projet permettra la production de nouvelles méthodes quantitatives ; de bases de données d'accès ouvert ; d'indicateurs de l'innovation ; de rapports analytiques utilisant ces indicateurs.

Quelles sont les ressources nécessaires ?

La réalisation de ce projet, estimée sur 2 ans, requiert :

1. des bases de données adéquates (les données existent, il faut les assembler) ;
2. des capacités de calcul importantes ;
3. des qualifications avancées en TAL, data science, et programmation.

L'OST possède déjà des compétences approfondies en data science, qui seront au cœur de ce projet ; elles pourraient être complétées par un partenariat avec une équipe de recherche spécialisée dans le TAL (par exemple du CNRS).

Repères biographiques

Dominique Guellec est conseiller scientifique à l'OST. Il contribue aux activités de l'OST, notamment dans les domaines des statistiques de brevets et de l'évaluation des politiques publiques en matière de recherche et d'innovation. Il est en charge d'un projet consistant à utiliser les techniques d'analyse sémantique des brevets, pour produire, notamment, des indicateurs reflétant la nouveauté et l'impact des inventions.

Jusqu'en août 2019, Dominique Guellec était chef de la division des politiques de science et technologie à l'Organisation de Coopération et de Développement Économique (OCDE). À ce titre, il a par exemple dirigé l'étude de 2014 sur le système de recherche et d'innovation de la France. Il avait auparavant été responsable des statistiques de science et technologie à l'OCDE, pilotant la révision du Manuel de Frascati en 2001 et celle du manuel des statistiques de brevets en 2009. Il a été également chef économiste à l'Office Européen des Brevets, où il a mis en place la base de données Patstat. Il a publié ou co-publié de nombreux articles académiques et plusieurs ouvrages sur l'innovation et la croissance, en français et en anglais (entre autres : Économie de l'innovation, éditions la Découverte 2018 ; The Economics of the European Patent System, Oxford University press, 2007).

Dominique Guellec est administrateur de l'INSEE, ancien élève de l'ENSAE.