

Réponse graduée : De l'utilité d'un modèle stochastique

Application au cas français

Modèle est ici entendu au sens au sens d'une formalisation mathématique supposée reproduire de manière approchée la réalité d'un phénomène dans le but d'en reproduire le fonctionnement pour permettre de comprendre, et éventuellement d'agir (définition reprise du «Notions de base utiles en modélisation stochastique de phénomènes complexes», Jean-Baptiste DENIS, INRA, Département de Mathématiques et Informatique Appliquées^o et Magali RITZ-BRICAUD).

Se fondant sur l'observation de la mise à disposition sur les réseaux P2P d'oeuvres sous droits, la réponse graduée est un phénomène complexe où l'aléatoire intervient à de nombreuses étapes du processus mis en oeuvre :

- Le nombre des "incidents détectés" est de cinq à dix fois supérieur à celui transmis en aval aux fins de traitement.
- Les "incidents transmis" ne sont pas tous traités.
- La réitération à un même accès de faits de mise à disposition n'est constatée qu'à l'issue d'une chaîne de traitements, en raison du caractère dynamique de la plupart des adresses IP.

Cette complexité se trouve encore accrue par la pluralité des comportements des internautes, par leurs changements de comportement à réception d'une première recommandation, par l'introduction d'un délai de "viduité" dans la prise en compte des "incidents transmis", etc. Si la théorie des probabilités permet de conduire analytiquement certains calculs, le processus est trop complexe pour qu'il soit possible de mener à bien ainsi l'ensemble de ceux qui sont nécessaires à une compréhension de la réalité. D'où l'utilité de recourir à une simulation sur ordinateur de l'aspect stochastique du phénomène, pour tenter de répondre à des questions comme :

- Pourquoi le taux de réitération observé est-il de 10% en France et de 30% aux Etats-Unis?
- Comment varie le taux de réitération observé en fonction du pourcentage d'internautes qui cessent toute mise à disposition sur les réseaux P2P à réception d'une première recommandation?

La simulation présentée utilise comme variables d'entrée les volumes observés au cours de l'année 2012 sur la plateforme "audiovisuelle", opérée par Trident Media Guard pour le compte de l'ALPA (Association de Lutte contre la Piraterie Audiovisuelle). Cette année a été choisie car c'est la première :

- où se trouve écartée tout risque de divulgation de données directement ou indirectement personnelles, puisque les opérateurs de télécommunications ne peuvent conserver au delà d'un an l'historique des adresses IP dynamiques affectées sur Internet à un titulaire d'abonnement ;
- où s'arrête la décroissance et s'amorce un plateau dans le nombre de mises à disposition, sur les réseaux P2P, des dix films qui arrivent en tête de ce mode d'appropriation illicite.

Le modèle reproduit les caractéristiques moyennes d'un mois de 2012 :

- nombre de mises à disposition quotidiennes d'une oeuvre audiovisuelle dont les ayants droit ont décidé que leur prestataire suivrait la diffusion sur Internet : "incidents détectés" ;
- pourcentage quotidien des adresses IP où sont mises à disposition plus d'une de ces oeuvres, tel que le fait apparaître l'examen des "incidents validés", c'est-à-dire des incidents où la réalité de la mise à disposition a été constaté par téléchargement d'un extrait et comparaison à une empreinte caractéristique de l'oeuvre : IP "plurielles" ;
- pourcentage quotidien de ces oeuvres sur IP "plurielles" dans le nombre total des oeuvres mises à disposition, toutes IP, IP "plurielles" et IP "simples" confondues.

Réponse graduée : De l'utilité d'un modèle stochastique

Application au cas français

Il introduit la notion d'IP "uniques", analogue à celle de "visiteurs uniques" : une même IP, repérée plusieurs fois au cours d'un même mois comme mettant des oeuvres à disposition n'est comptée qu'une fois. Le nombre d'IP "uniques" qu'il retient est en accord avec la statistique publiée par Médiamétrie/Netratings sur l'évolution du nombre de visiteurs uniques sur les réseaux P2P de 2009 à 2013, lorsqu'on fait application du ratio de 1,88 en 2012 entre le nombre d'internautes et le nombre d'abonnements haut et très haut débit et que l'on suppose que 4 visites sur 5 d'un site P2P "dédié à la contrefaçon audiovisuelle" donnent lieu à mise à disposition d'une oeuvre audiovisuelle dont les ayants droit ont décidé que leur prestataire suivrait la diffusion sur Internet.

En une première application, l'exploitation du modèle permet d'introduire un pourcentage d'abandon complet de la pratique du P2P à réception d'une première recommandation qui rend compte du taux de répétition observé.

En une deuxième application, l'exploitation du modèle permet de simuler à titre exploratoire l'effet que pourrait avoir une modification des paramètres actuels de la "réponse graduée phase I". En effet une décroissance forte et continue du nombre de mises à disposition sur les réseaux P2P des dix films qui arrivent en tête de ce mode d'appropriation illicite a été enregistrée tout au long de l'année 2011, mois après mois ; elle a abouti à une division par 3 en 12 mois. Le fait qu'après cette décroissance, - qu'il y a toutes raisons d'attribuer à l'impact de la réponse graduée -, une stabilisation intervenue au début de l'année 2012 se prolonge sur toute l'année 2013, conduit à s'interroger sur l'intérêt que pourrait présenter un traitement spécifique des IP "plurielles", voire la modification de certains seuils comme le nombre maximal d'adresses IP qui peuvent être collectées par jour par le prestataire des ayants droit, ou le nombre de lettres remises contre signature que la Commission de Protection des Droits de l'HADOPI est en mesure d'envoyer.

Si ces deux applications suffisent, leur semble-t-il, à établir la faisabilité d'une élaboration d'un modèle stochastique de la réponse graduée, et l'utilité de disposer d'un tel outil, les auteurs tiennent à souligner le caractère préliminaire de leurs travaux ; ils souhaitent que des statistiques enrichies sur les données observées viennent améliorer la fiabilité du modèle, des sondages sur le comportement des internautes destinataires des recommandations l'affiner, des modélisations additionnelles le rapprocher encore plus de la réalité des processus mis en oeuvre, des expérimentations localisées et temporaires en confronter les prédictions aux résultats de l'expérience.

Le 10 septembre 2014

Jean BERBINAU
Ancien maître de conférences
à TelecomParisTech
en Théorie de l'Information
Ingénieur Général des Mines
(en retraite - origine Telecom)

Patrick WAELBROECK
Maître de conférences
à TelecomParisTech
Membre fondateur de la chaire
"Valeurs et politiques des
informations personnelles"