

Détermination de la classe de QoS des connexions IP du réseau mobile par l'apprentissage non supervisé

Alassane Samba^{*,**} Philippe Dooze^{*}

^{*}Orange Labs, 2 avenue Pierre Marzin, 22300 Lannion
alassane.samba@orange.com,
philippe.dooze@orange.com,
<http://www.orange.com>

^{**}Télécom Bretagne, 2 rue de la Châtaigneraie, 35510 Cesson-Sévigné
alassane.samba@telecom-bretagne.com
<http://www.telecom-bretagne.com>

1 Introduction

Pour la mesure de la Qualité de Service réseau (QoS), les opérateurs utilisent plusieurs indicateurs, tels que le débit, la latence, le taux de perte de paquets, etc. Traditionnellement, la technique utilisée pour évaluer cette QoS consistait à déterminer sur avis d'expert du réseau des seuils sur chacun de ces indicateurs. Ces indicateurs, ainsi que leurs interactions, entrent en jeu dans la QoS réseau et in fine dans la qualité ressentie par le client. Ainsi, les méthodes d'apprentissage non supervisé ou Clustering permettant une analyse multidimensionnelle, sont une piste d'amélioration de l'évaluation de la QoS. De plus, de par leur nature, elles permettent d'intégrer facilement de nouvelles dimensions, ouvrant la voie à d'autres études.

Dans ces travaux, la méthode appelée K-means a été utilisée pour déterminer des classes de qualité sur un service data mobile. Les travaux ont été menés sur des données de trafic data d'un réseau mobile 3G, notamment, au niveau des sessions TCP (Transmission Control Protocol).

2 Données et mise en forme

Les données initiales sont issues d'une capture de 2 semaines de sondes d'un réseau mobile collectant des informations relatives au service de vidéo streaming sur le trafic IP (Internet Protocol). Ces données agrégées au niveau de la session TCP ont permis de calculer les indicateurs suivants : le débit moyen de téléchargement, le taux de perte de paquets, le taux de retransmission de paquets, le délai de réponse du serveur distant à la première requête du terminal et le délai d'établissement de la session TCP. L'analyse a été menée sur un échantillon représentatif d'environ 300 000 sessions.

3 Utilisation de la méthode K-means et optimisation de ses paramètres

L'algorithme K-means a été appliqué sur les indicateurs cités ci-dessus. Le K-means est une méthode d'apprentissage non supervisé. C'est un algorithme simple d'agrégation des observations autour de centres mobiles. Les principaux paramètres du K-means sont le nombre de classes et les points choisis comme centres. Un inconvénient de l'algorithme est qu'il faut

Détermination de la classe de QoS par Clustering

définir au préalable le nombre de classes. Ce nombre peut être fondé sur une expertise métier. Mais dans une démarche exploratoire comme la nôtre, une procédure itérative, telle que décrite ci-après peut être utilisée pour choisir le nombre de classes et améliorer les résultats

Les résultats du K-means peuvent être comparés grâce à l'inertie intra-classe totale et l'inertie inter-classe obtenues à la fin. La première mesure l'homogénéité dans les classes et le second l'hétérogénéité entre elles. L'inertie ou variance constitue la somme des carrés des distances. Plus l'inertie intra-classe totale est petite, plus chaque classe est homogène en moyenne, et meilleur est le Clustering. Cependant, l'inertie intra-classe totale est une fonction décroissante du nombre de centres K. Elle est minimale quand K est égal au nombre d'observations. Pour optimiser les résultats nous avons déroulé le K-means plusieurs fois, en faisant varier K (ici, entre 2 et 50). Du fait de la sensibilité des résultats du K-means aux centres aléatoirement choisis à l'initialisation, pour chaque K nous avons déroulé plusieurs K-means (en l'occurrence 20), afin d'en choisir le meilleur, c'est-à-dire celui qui a la plus petite inertie intra-classe totale. Ensuite, on a observé l'évolution de l'inertie intra-classe totale en fonction de K. Et on a choisi le K adéquat à partir duquel, la diminution de l'inertie intra-classe totale observée n'est plus très grande, et ne vaut pas la peine d'avoir un plus grand nombre de classes. Ainsi, K=12 semble un bon compromis.

4 Résultats du clustering

Une analyse plus fine des tendances à l'intérieur de chacune des 12 classes et les résultats d'une ACP (Analyse en composantes principales) sur les données, qui offre une meilleure visualisation du nuage de points, ont permis un regroupement en quatre superclasses facilement dénommables comme suit :

- Les sessions TCP sans problème particulier
- Les sessions TCP rencontrant un problème de débit
- Les sessions TCP connaissant une latence réseau
- Et les sessions TCP ayant rencontré à la fois des problèmes de débit et de latence réseau.

On dispose dès lors, après cet apprentissage non supervisé suivi d'un regroupement en superclasses plus parlantes, d'une nouvelle information de QoS générale basée sur l'observation de tous les indicateurs. Cette information peut être utilisée pour établir un modèle prédictif pour qualifier toute nouvelle session TCP du même service.

5 Conclusion

La méthode K-means, appliquée à des indicateurs de QoS de sessions TCP, a permis, dans cette étude, de déterminer des classes de QoS. L'évaluation est fondée, non plus sur des seuils, mais sur une analyse multidimensionnelle permettant d'isoler les sessions TCP selon leur niveau de QoS générale. Cette démarche appliquée à un service du réseau mobile 3G peut s'étendre à d'autres services IP.

Des agrégations de l'indicateur de QoS générale, qui découle de cette apprentissage, au niveau cellule-période ou client-période permettent d'avoir une vue de la QoS à ces différentes échelles. Ces travaux se poursuivent par la construction d'un moteur dynamique d'apprentissage de flux de données de QoS, intégrant des dimensions supplémentaires, afin de fournir des classes de QoS dynamiques plus intelligentes s'approchant du ressenti client.

Summary

This document summarizes a study held on IP QoS of mobile network. A K-means clustering method helped to better characterize QoS classes.