Empirical Study based on Machine Learning Approach to Assess the QoS/QoE Correlation

簡介:   
在現在多媒體快速發展的時代，有良好的影音體驗也是一件很重要的事情，因此大部分的服務商都會想辦法去優化客戶的影音體驗，以達到最佳的客戶滿意度。對於影片而言，因為相對來講變因更多，所以也就更難以確認那些變因會影響到使用者的觀看體驗。本篇論文在探討的就是如何利用機器學習的方法，結合多種關於網路底層參數的輸入指標(即QoS)，如delay, jitter等、及外部環境因素，如影片的幀數、終端裝置等，對應到輸出的QoE，也就是使用者的主觀感受的狀況，並再細分為五個等第，希望藉此能夠在顧客反應問題前及時做出改善。本篇的使用者資料是收集於可控制的網路環境內，介於18~30歲之間，性別涵蓋男女，並有分常在網路上看影片串流與否等資訊。結合以上作為訓練的X值後，作者使用了Naïve bayes, SVM, 4-NN, Decision tree, Random forest, Neural network等六種分類器進行分析，並發現其中Random forest及decision tree的表現最佳，進一步分析過後，也發現在這兩種之中，Random forest的表現又相對更好一些，最後達到約0.75的F1 score。作者最後總結，在可控環境底下，確實可以用機器學習的方法來達到預測QoE的效果，而其中，Random forest又是對這個資料集來說效果最好的分類器。

優點:

相較於大部分的機器學習會有一個固定不變的事實，作者選用了一個相對主觀的事實(使用者評分)來做為機器學習的目標值，而且做出來的成效也不差，代表對於使用者直接標記的資料，機器學習也可以有相當不錯的表現。

缺點:

對於普遍機器學習模型來說，0.75的F1 score算是比較低的值。每個人對影片品質的感受差異可能很大，1~5分可能讓受試者難以準確評估。

建議改善:

收集更多資料集，達成較高的F1 score。採用更精確的標準去量化受試者對於影片的感受。

Digging into HTTPS: Flow-Based Classification

of Webmail Traffic

簡介:

相較於以往，越來越多人喜歡用網頁的客戶端來登入email系統查看郵件，雖然這樣的行為對於使用者來說是一種方便，但對於流量分類來說就會困難上許多。因為大部分的網路郵件系統為了安全都會使用HTTPS加密流量，沒辦法針對裡面的內文進行分析，因此對於分類來說就會更加困難。而flow-based feature也不足以分析HTTPS流量是否為webmail流量。本篇論文希望能夠提出比較新穎的幾個特徵來進行HTTPS webmail分類，不需要很準確但要確認這些特徵容易取得。因為webmail server通常都會在傳統的SMTP, IMAP, POP等server附近，且webmail因為有一直去跟server取得資料的特性，會有固定的流量模式可循。作者利用原先的5-tuple(source ip and port, destination ip and port, protocol)，進一步去分析service proximity, daily and weekly profile, session duration和periodicity等特殊特徵。Service proximity是指在同個子網中，webmail server通常會在比較相近的IP上，所以跟大部分IP不同群的流量就比較不會是webmail service。Daily and weekly profile是指對於daily/weekly的流量來說，webmail server間的相關性一定會比non-mail server還要更高。Session duration是因為現在的session都是使用持久連線，加上使用者都會花比較多的時間去閱讀郵件，因此絕大部分的連線時間都會超過25秒。

優點:

缺點:

建議改善:

Failure Diagnosis Using Decision Trees

簡介:

優點:

缺點:

建議改善:

Detecting Fake Accounts on Social Media

簡介:

優點:

缺點:

建議改善:

Identifying and Discriminating Between Web and

PeertoPeer Traffic in the Network Core

簡介:

優點:

缺點:

建議改善:

Analyzing Load Profiles of Energy Consumption to Infer Household Characteristics Using Smart Meters

簡介:

優點:

缺點:

建議改善:

MACHINE LEARNING METHODS FOR SPAM E-MAIL CLASSIFICATION

簡介: 近年垃圾郵件越來越難以用單一規則，如email address或主旨內的特定字詞來防禦，因此現在主要使用ML的方式來進行垃圾郵件分類。這篇論文的作者提出了naïve bayes, KNN, SVM, NN, Artificial Immune System, rough set等方法來達成垃圾郵件分類的目的。

Support Vector Machine meets Software Defined

Networking in IDS domain

簡介:   
利用統計資料及SVM演算法做異常偵測。作者使用OpenFlow switch及Controller來實作此演算法，擷取可被SDN處理的封包資訊(如packet length, duration)等，並比較如果使用全資料及只使用SDN可擷取的資料的效能差異。比較後發現，雖然使用SDN的資料會造成Ture Positive比率微幅上升，但也還在可接受的範圍內。作者同時比較

Learning-based and Data-driven TCP Design for Memory-constrained IoT

簡介:   
IoT裝置需要有可靠的傳輸及極低的延遲，因此常使用TCP來傳輸，但TCP Reno是使用AIMD的方法，即便bottleneck的throughput有所改變也無法跟著一起調整。作者提出使用Q-Learning來調整CWND的方法，但考量到Q-Learning有極大的記憶體消耗，因此作者也提出了一些function approximation的方法，試圖以近似的結果來處理以節省記憶體開銷。在每800秒切換7.5Mbps及2.5Mbps的情況下，標準的TCPLearning提升了throughput 33%，降低了12%的延遲，並在function approximation的情況下，大幅節省了記憶體並還是能維持不錯的結果。

Depthwise Separable Convolution based Passive

Indoor Localization using CSI Fingerprint

簡介:  
利用CSI的頻譜影像作為CNN的Input，來預測/學習室內定位的資訊。使用CSI的原因是不會因為路徑而影響，並且可實行被動定位，即不需攜帶電子設備也可以達成定位目的。並透過depthwise separable convolution的技術，針對每層影像取得其單層及其深度的資料，結合本篇作者使用震幅差異產生的強化資料，取得比起以往CSI定位的方式還要更高的accuracy及更少的epoch數，也代表對於無論2D、3D矩陣，使用CNN都可以取得很好的效果。

優點:

缺點:

建議改善:

UAV Path Planning for Wireless Data Harvesting: A Deep Reinforcement Learning Approach

簡介:  
使用DDQN來做無人機路徑規劃以取得IoT的資訊，並針對不同的scenario parameter得到最快及最安全的飛行路徑。無人機在飛行時會遇到電量問題、障礙物、不可飛行區域及訊號被阻礙的問題，如何找出能讓throughput最大及讓飛行時間最少即是本篇論文要解決的問題。本篇的state有比較centered map及non-centered map的差異，發現在centered map的情況下(即永遠將無人機置於地圖中央)會有較好的效果，並將資訊使用CNN轉換成6個action的Q value。本篇的貢獻在發現centered map與DDQN結合，也可以用於路徑規劃等問題。

優點:

缺點:   
只在模擬環境中模擬，現實世界的街道及狀況一定不可能與模擬狀況一樣

建議改善:

An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks

簡介:

優點:

缺點:

建議改善: