

機器學習應用在交通運輸

第一篇 Automated Transportation Mode Detection Using Smart Phone Application via Machine Learning: Case Study Mega City of Tehran:

在過去的幾十年中，旅行行為變得更加複雜，尤其是在大城市，例如伊朗首都德黑蘭。決策者需要更準確，更全面的信息來規劃城市交通。與傳統的基於紙質和電話的調查相反，最近已使用一種新的高效數據收集方法，該方法使用了信息技術，例如可跟踪乘客行程的基於 **GPS** 的數據收集方法。利用這種新方法，本研究的主要目的是分析收集的數據，以便使用一種稱為隨機森林的新型機器學習方法來區分乘客使用的運輸方式。該模型不僅以近 **96%** 的高精度對交通方式進行分類，例如汽車，公共汽車和步行，而且還基於兩個重要指標（均值降低準確性和基尼係數）確定了分類過程中最具影響力的屬性。結果表明，**GPS** 軌道的即時速度和準確性是交通方式分類中影響最大的屬性。運輸計劃人員可從此準確而全面的旅行行為數據（使用模式）中製定政策，從而從中受益匪淺。

第二篇 A Machine Learning System for Routing Decision-Making in Urban Vehicular Ad Hoc Networks:

在車輛自組織網絡（**VANET**）中，由於車輛的高移動性，網絡拓撲和通信鏈路經常發生變化。關鍵挑戰包括如何縮短傳輸延遲並提高傳輸穩定性。在建立路由路徑時，大多數研究都集中在檢測交通並選擇具有較高車輛密度的道路以傳輸數據包，從而避免了前轉情況和減少傳輸延遲。但是，由於車輛密度變化非常快，因此此類方法可能無法通過定期監視每條道路來獲得準確的實時交通密度。在本文中，我們提出了一種新穎的路由信息系統，稱為機器學習輔助路由選擇（**MARS**）系統，以估計路由協議所需的信息。在 **MARS** 中，借助機器學習，可在路邊單位維護道路信息。我們使用機器學習來預測車輛的行駛，然後選擇一些具有更好傳輸能力的合適路由路徑來傳輸數據包。此外，**MARS** 可以幫助根據目的地的預測位置和兩個轉發方向上的估計傳輸延遲來確定兩個 **RSU** 之間的轉發方向。我們提出的系統可以為 **VANET** 提供及時的路由信息，並大大提高網絡性能。

第三篇 Predicting Future Driving Risk of Crash-Involved Drivers Based on a Systematic Machine Learning Framework:

本文的目的是預測中國崑山市涉及撞車事故的駕駛員的未來駕駛風險。提出了系統的機器學習框架來處理三個關鍵技術問題：**1.定義駕駛風險； 2.發展風險驅動因素； 3.建立可靠且可解釋的機器學習模型。**高風險（**HR**）和低風險（**LR**）驅動程序由五個不同的場景定義。從七年的碰撞/違規記錄中提取了許多功能。駕駛員之前兩年的撞車/違規信息被用來預測其後兩年的駕駛風險。使用一年的滾動時間窗口，開發了四個連續時間段的預測模型：**2013 – 2014、2014 – 2015、2015 – 2016 和 2016 – 2017**。嘗試了四種基於樹的集成學習技術，包括隨機森林（**RF**），帶決策樹的 **Adaboost**，梯度增強決策樹（**GBDT**）和極端梯度增強決策樹（**XGboost**）。進行時間遷移性測試和後續研究以驗證訓練後的模型。定義駕駛風險的最佳方案是多維的，包括事故重現，嚴重性和錯誤承諾。**GBDT** 似乎是所有時間段的最佳模型選擇，在最新數據集（即 **2016-2017** 年）上，可接受的平均精度（**AP**）為 **0.68**。九個主要特徵中的七個與危險駕駛行為有關，這些行為與駕駛風險呈非線性關係。在相對較短的時間間隔（**1-2** 年）內保持模型的可移植性。對於風險預測任務，需要考慮適當的風險定義，複雜的違規/崩潰功能以及先進的機器學習技術。擬議的機器學習方法很有希望，因此可以更有效地啟動安全乾預措施。

第四篇 A Machine-Learning Approach to Distinguish Passengers and Drivers Reading While Driving:

駕駛員分心是交通事故的主要原因之一。近年來，隨著連接性和社交網絡的發展，駕車時使用智能手機的頻率越來越高，安全問題也越來越嚴重。開車時發短信，打電話和讀書是由於使用智能手機而引起的干擾。在本文中，我們提出了一種非侵入性技術，該技術僅使用來自智能手機傳感器和機器學習的數據來在讀取車輛中的消息時自動區分駕駛員和乘客。我們在不同情況下建模和評估了七種最先進的機器學習技術。卷積神經網絡和梯度提升是我們實驗中效果最好的模型。結果顯示準確性，準確性，召回率，**F1** 得分和 **kappa** 指標均優於 **0.95**。

第五篇 Predicting Freeway Travel Time Using Multiple Source Heterogeneous Data Integration:

高速公路的行駛時間受許多因素的影響，包括交通量，惡劣天氣，事故，交通管制等。我們採用多源數據挖掘方法來分析高速公路的行駛時間。我們從中國湖南省的 **G5513** 高速公路收集了通行費數據，天氣數據，交通事故處理日誌以及其他歷史數據。使用支持向量機 (**SVM**)，我們提出了基於這些數據庫的旅行時間預測模型。新的 **SVM** 模型可以模擬旅行時間與那些因素之間的非線性關係。為了提高支持向量機模型的精度，我們應用人工魚群算法對支持向量機模型參數進行了優化，包括內核參數 σ ，非敏感損失函數參數 ε 和罰分參數 **C**。具有反向傳播 (**BP**) 神經網絡和通用 **SVM** 模型的 **SVM** 模型，使用從高速公路 **G5513** 收集的歷史數據。結果表明，優化後的 **SVM** 模型的準確性分別比 **BP** 神經網絡模型和普通 **SVM** 模型的準確性高 **17.27%**和 **16.44%**。

第六篇 Smartphone Transportation Mode Recognition Using a Hierarchical Machine Learning Classifier and Pooled Features From Time and Frequency Domains:

本文開發了一種新穎的兩層分層分類器，該分類器提高了傳統運輸方式分類算法的準確性。本文還通過提取新的頻域特徵來提高分類精度。許多研究人員已經從全球定位系統數據中獲得了這些功能。但是，此數據未在本文中排除，因為系統使用可能會耗盡智能手機的電池，並且某些區域可能會丟失信號。我們提出的兩層框架與以前的分類嘗試在三個方面不同：**1**) 使用貝葉斯規則將兩層的輸出組合起來，以選擇具有最大後驗概率的運輸方式；**2**) 提出的框架將新提取的特徵與傳統上使用的時域特徵相結合以創建特徵池；**3**) 基於分類模式，在每一層中使用不同的提取特徵子集。使用了幾種機器學習技術，包括 **k** 最近鄰，分類和回歸樹，支持向量機，隨機森林以及隨機森林和支持向量機的異構框架。結果表明，提出的框架的分類精度優於傳統方法。將時域特徵轉換為頻域還可以在新空間中添加新特徵，並提供對信息丟失的更多控制。因此，在一個大型池中組合時域和頻域特徵，然後選擇最佳子集會比單獨使用任一域產生更高的準確性。提出的兩層分類器的最大分類精度為 **97.02%**。

第七篇 Travel-Time Prediction using Gaussian Process Regression: A Trajectory-Based Approach:

本文涉及地圖上任意起點-終點對的行進時間預測任務。與大多數現有研究僅關注交通繁忙的特定路段（路段）不同，如果將路徑之間的相似性定義為核心函數，我們的方法允許我們概率性地預測沿未知路徑（路段序列）的行駛時間。我們的第一個創新是使用 **astring** 內核來表示路徑之間的相似性。我們的第二個新想法是將高斯過程回歸應用於概率旅行時間預測。我們用真實的流量數據測試了我們的方法。