



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

PREDICCIÓN DE VELOCIDAD EN UNA AUTOPISTA URBANA UTILIZANDO
ESTADOS DE TRÁFICO Y MODELOS DE INTELIGENCIA ARTIFICIAL

TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN GESTIÓN DE OPERACIONES

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

BASTIÁN ALONSO VEJAR TEMER

PROFESOR GUÍA:
FERNANDO ORDOÑEZ PIZARRO

MIEMBROS DE LA COMISIÓN:
FRANCO BASSO SOTZ
SEBASTIÁN RÍOS PÉREZ

SANTIAGO DE CHILE
2019

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE MAGÍSTER EN GESTIÓN DE OPERACIONES; INGENIERO CIVIL INDUSTRIAL
POR: BASTIÁN ALONSO VEJAR TEMER
FECHA: 2019
PROF. GUÍA: FERNANDO ORDOÑEZ PIZARRO

PREDICCIÓN DE VELOCIDAD EN UNA AUTOPISTA URBANA UTILIZANDO ESTADOS DE TRÁFICO Y MODELOS DE INTELIGENCIA ARTIFICIAL

Cada día millones de Santiaguinos se desplazan a sus respectivos trabajos y estudios, en donde miles de ellos utilizan la autopista central, la autopista concesionada mas grande de la región como la vía de transporte para sus recorridos diarios. Durante gran parte del día, las velocidades de esta autopista fluctúan entre 90 y 110 km/h, a excepción de eventos de congestión, en donde dado que la cantidad de vehículos que intenta pasar por esta vía es superior a su capacidad máxima, la velocidad disminuye, no superando los 50 km/h.

El objetivo de esta tesis es lograr predecir de la manera más certera posible la velocidad promedio de un segmento de la autopista en un horizonte de corto plazo, utilizando información sobre las condiciones operativas de este segmento y de los que se encuentran anteriores a este. Estudios anteriores han propuesto una gran variedad de modelos de tráfico para predecir velocidad, flujo y volumen. Sin embargo, el desempeño de estos modelos y metodologías ha sido ampliamente criticado, ya que gran parte de los estudios se centra en el simple procesamiento de datos y aplicaciones de los modelos. Esta tesis plantea considerar esta crítica y utilizar dos modificaciones sobre la arquitectura de estos modelos de inteligencia artificial, ambas apoyadas en la teoría de las fases de tráfico, la existencia de estados sin congestión y estados congestionados, donde en este último los vehículos no pueden transitar a la velocidad que les gustaría. La forma en que se plantea discernir entre estos estados es mediante una regla sobre la velocidad, apoyado por un análisis gráfico del comportamiento de los datos.

El desempeño de las metodologías adoptadas es comparada en contraste a una variedad de modelos. Al revisar en la literatura los modelos con mejores resultados durante la última década, es posible escoger los siguientes modelos para la comparación: una Red Neuronal Artificial (ANN), un Bosque Aleatorio (RF), un Support Vector Regressor (SVR) y una Long Short-Term Memory Recurrent Neural Network (LSTM). Adicionalmente, estos modelos fueron comparados con una regresión lineal debido a lo común que es su implementación en los estudios. Estos resultados son evaluados con dos métricas: el error absoluto medio (MAE) y el error porcentual absoluto medio (MAPE).

De acuerdo con los experimentos realizados, se considera que los modelos previamente utilizados predicen de manera satisfactoria, particularmente el caso de la Artificial Neural Network (ANN) y la Long Short-Term Memory Recurrent Neural Network (LSTM) donde estos predicen con errores absolutos medios menores a 4.74 km/h para horizontes temporales de hasta 30 minutos. Adicionalmente, se considera que una de las dos modificaciones realizadas es exitosa, debido a que la segunda de estas logra predecir velocidades con un MAE de 4.69 km/h para cualquier horizonte temporal considerado, siendo esta una disminución del error de aproximadamente un 1 %-6 % dependiendo del horizonte temporal seleccionado respecto a los resultados de la ANN y la LSTM.

A mi gata

Tabla de Contenido

1. Introducción	1
1.1. Objetivos	1
1.2. Estructura de la tesis	2
2. Revisión de literatura	3
2.1. Según tipo de variable	3
2.2. Según enfoque metodológico	4
2.2.1. Métodos paramétricos	4
2.2.2. Métodos no-paramétricos	5
2.2.3. Métodos basados en modelos de tráfico	6
3. Metodología y desarrollo	8
3.1. Modelos de machine learning analizados	8
3.1.1. Random Forest	9
3.1.2. Support Vector Regressor	9
3.1.3. Artificial Neural Network	10
3.1.4. Long Short Term Memory Recurrent Neural Network	12
3.2. Modificaciones realizadas	14
3.2.1. Clasificador de estados de congestión (CEC)	16
3.2.2. Probabilidad de estado de congestión (PEC)	17
3.3. Indicadores de error	18
4. Descripción del caso de estudio	20
4.1. Imputación de datos	21
4.2. Transformación de datos	22
4.3. Clasificación de congestión	22
5. Resultados	25
5.1. Comparación modelos	25
5.2. Modelos propuestos	26
6. Conclusión	28
6.1. Alcances y limitaciones	28
6.2. Recomendaciones para trabajo futuro	29
7. Bibliografía	30

Índice de Tablas

4.1. Análisis exploratorio de datos	21
5.1. Comparación de modelos considerados	25
5.2. Comparación de modificaciones con modelos de mejor desempeño	26
5.3. Matriz de confusión según horizonte temporal	27
5.4. Evaluación de cada una de las ANN del modelo PEC	27
6.1. Cantidad de valores considerados según velocidad	29

Índice de Ilustraciones

3.1. Función sigmoide	10
3.2. Arquitectura básica de una red neuronal artificial	11
3.3. Iteraciones sobre una RNN	12
3.4. Nodo de una LSTM	14
3.5. Cambio de estado a congestión sincronizada. Fuente: Kerner (2009)	15
3.6. Ejemplo arquitectura del método CEC	17
3.7. Ejemplo arquitectura del método PEC	18
4.1. Flujo vehicular diario por hora	21
4.2. Velocidad-flujo mes de Octubre 2016 (Izquierda) y Noviembre 2016 (Derecha)	23
4.3. Velocidad-flujo año 2016 (Izquierda) y 2017 (Derecha)	23
4.4. Velocidades instantáneas minutales del día 06/10/2016	23
4.5. Flujo-velocidad Octubre 2016 - Junio 2017	24

Capítulo 1

Introducción

Las predicciones de corto plazo en tráfico tratan de estimar la velocidad o cantidad de vehículos que pasan a través de un arco o una serie de arcos en un horizonte temporal determinado, generalmente menor a una hora. Estas predicciones de tráfico, en particular de tiempo y velocidad, cada vez han tenido mayor relevancia para investigadores a medida que los avances en tecnologías han permitido obtener una mayor y mejor cantidad de datos y mejores y más complejos algoritmos, que ha permitido a organismos tanto públicos como privados generar iniciativas más informadas que ayudan a la población general a desplazarse de manera más eficiente.

La presente tesis se centra en la aplicación y comparación de los principales modelos de inteligencia artificial utilizados en la literatura reciente, en relación a dos modificaciones en la arquitectura de los modelos, con la intención de aprovechar fenómenos particulares de tránsito para obtener predicciones más certeras.

Los principales modelos a comparar con las modificaciones son un Bosque Aleatorio o Random Forest (RF por su nombre en inglés), Regresor de Soporte de Vectores (SVM), una Red Neuronal Artificial (ANN) y una Long Short-Term Memory Recurrent Neural Network (LSTM) mediante el uso del lenguaje de programación Python. Estos modelos fueron utilizados para estimar velocidades a corto plazo -cinco, diez, quince y treinta minutos- en una sección de tres pistas en un único sentido de la Autopista Central, ubicadas a altura cercana del paso superior Lo Espejo. Las predicciones de velocidad fueron realizadas utilizando datos relacionados con velocidad y flujo vehicular, obtenidos mediante cuatro pórticos de cobro electrónico de peaje de la misma autopista.

1.1. Objetivos

El objetivo general del estudio es proponer una nueva metodología que entregue una mejor predicción de velocidades a corto plazo y analizar las virtudes de esta en comparación a otros métodos clásicos de predicciones de tráfico a corto plazo.

Los resultados se comparan principalmente en términos de su capacidad predictiva.

Para lograr este objetivo general, se plantean los siguientes objetivos específicos:

- Levantamiento de variables relevantes para la predicción de tráfico a corto plazo y obtención de datos y variables necesarias mediante técnicas de preprocesamiento y transformación de datos.
- Investigación y desarrollo de principales modelos de predicción de variables de tráfico a corto plazo.
- Determinar actuales falencias de los métodos actuales y desarrollar modificaciones y mejoras sobre estos para obtener mejores predicciones.
- Evaluar el desempeño de las predicciones de velocidad de los modelos y comparar con las modificaciones propuestas.

1.2. Estructura de la tesis

Esta tesis ha sido estructurada en un total de cinco capítulos. El primero de estos es la presente introducción, el segundo corresponde a una revisión bibliográfica de estudios relacionados a predicciones de tráfico a corto plazo, separado en dos categorías: según variable predicha y según enfoque metodológico, incluyendo un análisis de cómo ha cambiado la elección de estos en las últimas décadas y cuáles son las falencias actuales. El tercer capítulo presenta una descripción detallada de los algoritmos utilizados, las falencias y críticas que se les realiza y las modificaciones en las arquitecturas planteadas. El cuarto capítulo corresponde a una descripción del caso aplicado, donde se presentan los datos utilizados, las imputaciones, transformaciones y el por qué de estas. En el quinto capítulo se presentan los resultados de la predicciones de velocidad a corto plazo en autopista, para cada tipo de modelo utilizado. Finalmente, se presentan las principales conclusiones del trabajo realizado, aprendizajes, problemas adicionales encontrados y líneas generales para trabajo futuro. Por último, se puede encontrar la bibliografía utilizada en este estudio, en donde es posible encontrar las investigaciones referenciadas.

Capítulo 2

Revisión de literatura

Desde principios de la década de los ochentas, las predicciones de tráfico a corto plazo han sido una parte fundamental de los sistemas integrados de transporte (SIT) e investigación relacionada. Más aún, durante esta última década, un especial énfasis se ha puesto por parte de los investigadores en el desarrollo y aplicación de modelos matemáticos para la más certera predicción de tráfico. Esto se puede apreciar en la revisión de literatura realizada por Ermagun y Levinson (2018) que muestra que la cantidad de publicaciones relacionadas ha aumentado en un 1000 % en un período de 16 años. Se cree que este aumento ocurre principalmente debido a los grandes avances en tecnología, lo que ha permitido mejorar las capacidades de recopilar información y manejar grandes cantidades de datos.

Se plantea que el modelamiento puede ser categorizado de manera general basado en dos factores básicos: el tipo de variable y el enfoque metodológico. Estas distintas categorías y su respectiva literatura serán presentadas en este capítulo.

2.1. Según tipo de variable

Los investigadores tienen que definir claramente sus metas y alcances para así identificar qué es lo que desean predecir. Investigadores no solo han realizado esfuerzos en analizar y predecir tiempos de viaje, sino que también velocidades instantáneas, velocidades espaciales, volumen y flujo. Dadas las relaciones básicas y fundamentales existentes entre estas variables, en oportunidades es posible obtener los valores buscados mediante cálculos de los otros. Los principales determinantes del tipo de la variable buscada depende de los objetivos del estudio y la información que se encuentra disponible.

Adicionalmente, los investigadores necesitan definir correctamente el horizonte predictivo que necesitan para cumplir sus objetivos. El horizonte de predicción define la resolución temporal de la información usada como dato de entrada en términos de intervalos de tiempos. En la literatura, generalmente se define (y aplica) como un intervalo de corto plazo valores entre 0-30 minutos. Es posible encontrar posturas disidentes de cuánto sería un tiempo apropiado para las predicciones; mientras que por una parte Kwon et al (2000) y Ishak y Al-Deek (2002)

concluyeron que la exactitud predictiva de los modelos disminuye considerablemente a medida que el horizonte predictivo aumenta, y por otra, Park et al (1998) y Van Lint et al (2002) argumentan que existe una considerable disminución en la exactitud de sus modelos al ser examinados en intervalos de tiempo menores a un minuto, debido a las grandes fluctuaciones de los parámetros de tráfico, en donde estas variaciones disminuyen a medida que el intervalo de agregación de la data aumenta. Dada la existencia de este trade-off entre cuán grande o pequeño tiene que ser el horizonte de predicción, definir apropiadamente este intervalo de agregación es un factor clave que define la calidad de la información utilizada en los modelos predictivos.

2.2. Según enfoque metodológico

Vlahogianni et al (2004) y Vlahogianni et al (2014) compilaron una revisión bibliográfica exhaustiva de las técnicas de predicción a corto plazo, categorizando las predicciones de velocidades y tiempos de viaje en dos principales enfoques: métodos paramétricos y no paramétricos. De manera similar, Guo et al (2013) y Meng et al (2015) consideran además de modelos estadísticos y data-driven (de manera análoga a los de tipo paramétrico y no-paramétrico, respectivamente) la existencia de una tercera categoría: métodos basados en modelos de tráfico. Estas definiciones serán utilizadas y descritas a continuación:

2.2.1. Métodos paramétricos

Los métodos estadísticos suelen predecir las condiciones de tráfico futuras aplicando conocimientos teóricos sobre una base de datos, para así estimar los parámetros óptimos que logren explicar las relaciones entre variables dependientes e independientes o la teoría de estados espaciales.

El modelo más básico utilizado por investigadores es el promedio histórico. Dada la sencillez de estos modelos, generalmente no son investigados de por sí, sino que son comparados frente a otros con mayor poder predictivo, tal como lo ha realizado Smith y Demetsky (1997), entre otros.

Basándose en la familia de modelos lineales autoregresivos de promedios históricos (ARIMA) utilizados inicialmente por Ahmed y Cook (1979), una variedad de investigadores han realizado estudios similares, entre ellos Hamed et al (1995) utilizando un ARIMA (0,1,1) y Williams (2001) añadiendo la covarianza logra utilizar un modelo ARIMAX. De forma similar, Williams et al (1998) y Williams y Hoel (2003) desarrollaron y compararon un modelo lineal autoregresivo de promedios históricos con modificación estacional (SARIMA) con mejores resultados que los obtenidos por anteriores modelos ARIMA. Estos estudios mostraron un claro problema en esta familia de modelos, ya que estos tendían a concentrarse en valores cercanos a la media y tienen una gran dificultad en predecir fluctuaciones extremas; como las que efectivamente ocurren en el tráfico.

D'Angelo (1999) y Ishak y Al-Deek (2002) utilizaron series de tiempo no lineales para

predecir tiempos de viaje en autopista, en donde el último de estos destaca el hecho de que la capacidad predictiva del modelo aumenta considerablemente cuando se encuentra en un estado no congestionado a diferencia de uno congestionado.

Por su parte, tanto Kwon et al (2000) como Sun et al (2003) utilizaron regresiones lineales para predecir flujo, en donde se concluye que este tipo de modelos funciona de manera bastante certera para tiempos pequeños, mientras que utilizar información histórica entrega una mejor predicción para horizontes de predicción mas largos, conclusiones que también fueron obtenidas en el análisis de Ishak y Al-Deek (2002) para las series de tiempo no lineales. Tanto Kwon como Sun consideran sus estudios exitosos al realizar comparaciones con una feed-forward neural network y un modelo K-nearest neighbor, respectivamente. A pesar de lo anterior, ambos investigadores destacan las oportunidades futuras de estos últimos, donde concluyen que mediante desarrollo e investigación podrían superar sus métodos.

Finalmente, también se pueden considerar como series de tiempo los filtros de Kalman utilizados por Chen y Chien (2001), Xie et al (2007) Guo y Williams (2010) y Guo et al (2014), en donde de manera similar al resto de los métodos estadísticos, destacan como principal obstáculo la baja capacidad predictiva cuando existen fluctuaciones considerables, como el paso de un estado descongestionado a uno con congestión.

2.2.2. Métodos no-paramétricos

Los modelos data-driven o no-paramétricos intentan encontrar relaciones y patrones ocultos a partir de los datos sin que estos sean explícitamente entregados al modelo. Dentro de esta clasificación se incluyen support vector machine, redes neuronales, algoritmos genéticos, regresiones no paramétricas y enfoques híbridos que combinan distintos algoritmos para reducir dimensionalidad y realizar trabajo de forma más eficiente.

Máquina de Soporte de Vectores o Support Vector Machine (SVM) es un algoritmo de aprendizaje supervisado que funciona bajo la idea de encontrar hiperplanos que separen de forma óptima a los puntos de una clase de otra. Wu et al (2004) comparó exitosamente un modelo de SVM en contraste a un promedio histórico y al algoritmo que era utilizado por la central de la autopista estudiada. Vanajakshi y Rilett (2004) realizando un estudio similar, concluyeron que en un escenario con una cantidad limitada de datos el SVM predice mejor que una red neuronal. Castro-Neto et al (2009) presenta que su modelo SVM predice de forma más certera en los eventos atípicos en comparación a una red neuronal. Por otra parte, Asif et al (2014) utilizaron métodos como k-means clustering, principal component analysis, y self-organizing maps para encontrar relaciones espacio-temporales y así pasarlas a un modelo SVM, con resultados exitosos. Theja (2010), Yao et al (2014) y Yao (2016) han demostrado que SVM es un método comparable a las redes neuronales en cuanto capacidades predictivas.

De forma similar, Bosques aleatorios o Random Forest (RF) también han sido utilizados por investigadores para predecir condiciones operativas de tránsito. Hamner (2010) exitosamente desarrolló e implementó una metodología con Random Forest que mejoró en más de un 60 % los métodos utilizados anteriormente en la autopista de estudio.

El método K-Nearest Neighbor (KNN) es un método de clasificación supervisado, que se basa en la estimación del valor de la función de densidad de probabilidad o directamente la probabilidad a posteriori de que un elemento pertenezca a una determinada clase. En las últimas tres décadas, grandes avances se han realizado en esta área desde los primeros intentos de Davis et al (1991). Entre ellos se destaca Clark (2003) que predijo tráfico de la autopista M25 de Londres utilizando un método KNN mediante observaciones de las condiciones operativas de la vía (principalmente ocupancia y flujo), lo que le permitió promediar estos valores para obtener una predicción certera de la velocidad futura de cada estado. Sin embargo, ya que el método fue solamente testeado durante tres semanas, no deja claro cuán generalizable puede ser la aplicación de KNN dada la pequeña cantidad de información entrenada. Por su parte, Yu et al (2016) compara los resultados obtenidos por el método KNN con un Support Vector Machine y una Artificial Neural Network, concluyendo que aunque no es estrictamente mejor que estos métodos, es una alternativa efectiva y viable para las predicciones de tráfico a corto plazo.

Las aplicaciones de las redes neuronales artificiales o artificial neural networks (ANN) en predicciones de tráfico a corto plazo se extienden desde la configuración de un Multilayer Perceptron utilizado por Clark (1993), Smith and Demetsky (1994, 1997) e Innamaa (2005), hasta estructuras más complejas como Modular Neural Network (Park and Rilett 1998) Time-Delayed Neural Networks (TDNN) (Abdulhai et al. 1999, Dia 2001, Ishak et al 2003); Recurrent Neural Network (RNN) (Lingras et al 2002); Dynamic Neural Networks (Shen 2008) y Bayesian Combined Neural Network Approach (Zheng et al 2006).

Recientemente, avances en la investigación relacionada a redes neuronales y Deep Learning ha vuelto un poco más difusa la clasificación de estos modelos, donde Huang (2014) propone una arquitectura que depende de dos partes: una Deep Belief Network (de forma como una pila de Restricted Boltzmann Machines) y una Multitask Regression. Ma (2015), Tian (2015), Fu et al (2016), Yu et al (2017) y Zhao et al (2017) realizaron predicciones utilizando Long-Short-Term Memory Recurrent Neural Network (LSTM) con resultados prometedores en comparación a modelos con redes neuronales menos complejas, donde Mackenzie et al (2018) comparó una LSTM con una Hierarchical Temporal Memory Neural Network. Polson y Solokov (2017) utilizaron una arquitectura basada en Feed-Foward Neural Network para predecir comportamientos en dos momentos fuera de lo usual: a la salida de un juego de football americano y en una tormenta de nieve. Por último, Ma (2017) probó una Deep Convolutional Neural Network con la idea de convertir las dinámicas espacio-temporales del tráfico en imágenes que describirían las relaciones de tiempo y espacio mediante una matriz bi-dimensional, comparando su método con cuatro métodos clásicos: mínimos cuadrados, KNN, ANN, Random Forest y tres de deep learning: stacked autoencoder, RNN y LSTNN, con resultados exitosos.

2.2.3. Métodos basados en modelos de tráfico

Estos modelos se apoyan fundamentalmente en conocimientos de teoría de tráfico, intentando describir y representar funcionalmente las interacciones entre variables físicas, que describen los fenómenos de tránsito de forma macroscópica y microscópica. Los modelos macroscópicos se concentran en la predicción de comportamientos de tráfico basado en analogías

del tráfico vehicular con dinámicas de gas y fluidos, mientras que los modelos microscópicos considera el comportamiento individual con gran detalle para analizar las interacciones entre vehículos en la red. Conocimientos obtenidos mediante teoría, simulaciones o práctica en la vida real.

Generalmente, es necesario relacionar estos conocimientos con los métodos anteriormente mencionados, ya que la naturaleza de estas investigaciones obligan a evaluar con datos. En el modelo macroscópico de Van Lint et al (2003, 2005) se asocia la existencia de estados operativos a una State-Space Neural Network para realizar predicciones. Stathopoulos y Karlaftis (2003) utilizando series de tiempo capturaron una gran variedad de dinámicas del flujo del tránsito mediante las interacciones de sus variables.

Dada la complejidad de establecer modelos que expliquen y comprendan la totalidad de interacciones en el tráfico (Yu et al 2016), las investigaciones recientes han requerido que estos estudios no se realicen de forma pura, es decir, que funcionen de forma híbrida con alguno de los dos enfoques nombrados.

Capítulo 3

Metodología y desarrollo

Los tres tipos de enfoques mencionados en el capítulo anterior (paramétricos, no-paramétricos y basados en tráfico) se han utilizado en varias oportunidades para estimar tanto tiempos de viaje como velocidades. En esta última década, es posible apreciar un cambio en el tipo de método usado, donde se ha dejado de lado la utilización de métodos paramétricos y ha ocurrido un constante aumento en el uso de métodos no-paramétricos, particularmente de Machine Learning (Ermagun et al. (2018)). Se plantea que esto se debe a una variedad de factores, entre ellos a la mejora en la capacidad de obtener y almacenar una gran cantidad de datos y a los considerables avances de estos últimos, ya que en una gran mayoría de los estudios recientes estos métodos superan con creces a sus contrapartes estadísticas.

Mientras que se entiende que los modelos paramétricos y basados en modelos de tráfico tienen una mayor capacidad de interpretación en comparación a los no-paramétricos, grandes esfuerzos se han realizado para calibrar y mejorar estos últimos. Además, considerando los grandes avances que han ocurrido en las tecnologías de información y las capacidades de recopilación y procesamiento de datos, han hecho que estos modelos implícitos se hayan vuelto una alternativa bastante atractiva.

Para los modelos basados en datos, tanto paramétricos como no-paramétricos, el tipo de método usado, la calidad de la información y los parámetros usados en el desarrollo e implementación son vitales para la calidad predictiva.

3.1. Modelos de machine learning analizados

Uno de los puntos relevantes a considerar en esta tesis, es la comparación de cuatro modelos utilizados recurrentemente por investigadores, los que son una Artificial Neural Network (ANN), un Support Vector Machine (SVM), un Random Forest (RF) y una Long Short Term Memory Recurrent Neural Network (LSTM). Se utilizan estos modelos para la comparación con las arquitecturas planteadas debido a sus capacidades de obtener resultados certeros incluso en bases de gran tamaño.

Adicionalmente, se plantean dos modificaciones sobre la arquitectura de los modelos, en donde se aprovecha conocimiento sobre fenómenos del tráfico aplicado sobre los modelos. La explicación y desarrollo de estos algoritmos será presentado más adelante en el presente capítulo.

3.1.1. Random Forest

El modelo Random Forest es un modelo de aprendizaje supervisado planteado por L. Brieman (2001), que se basa en la idea de utilizar el procedimiento de ‘Bootstrap Aggregation’ o ‘Bagging’ sobre una gran cantidad de árboles de decisión. Este procedimiento plantea que la combinación de varios modelos de aprendizaje incrementa la exactitud de las predicciones, reduciendo la varianza al promediar los modelos que son ruidosos o sesgados. Por lo que no presenta problemas con bases de datos de gran tamaño.

Cada uno de los árboles de un Random Forest dependen de un vector aleatorio sampleado independientemente cuya distribución es completamente idéntica para los vectores de cada uno de los árboles. Los árboles estiman características de los valores ingresados al identificar zonas de datos con parámetros o características similares. Al analizar el desempeño de la data de entrenamiento de cada árbol por separado, es posible detectar los árboles que son más eficientes al predecir los valores requeridos y en consecuencia darle más relevancia a estos ponderándolos por valores mayores en el resultado final, mientras que a la vez se descartan los que consiguen los peores resultados, así obteniendo las mejores estimaciones posibles.

3.1.2. Support Vector Regressor

El Support Vector Regression es un modelo de aprendizaje no supervisado propuesto por V. Vapnik y H. Drucker (1997). Es una modificación sobre el modelo Support Vector Machine en que se plantea representar las observaciones como puntos en el espacio, puestas de tal manera que es posible separar con un hiperplano separador que permite categorizar las variables de un tipo u otro.

El algoritmo Support Vector Regressor pretende resolver el siguiente problema:

$$\begin{aligned} \min_w \quad & \frac{1}{2} \|w\| \\ \text{s.a} \quad & y_i - \langle w, x_i \rangle - b \leq \varepsilon, \\ \text{s.a} \quad & \langle w, x_i \rangle + b - y_i \leq \varepsilon, \end{aligned}$$

donde x_i corresponde a la observación asociada a y_i . El producto interno más el intercepto $\langle w, x_i \rangle + b$ es la predicción de la muestra, finalmente ε corresponde a un parámetro libre que sirve como límite: todas las predicciones tienen que estar dentro del rango ε de las observaciones. Variables de holgura son agregadas para permitir errores y aproximaciones en caso que el problema sea infactible.

3.1.3. Artificial Neural Network

La Artificial Neural Networks (ANN) o redes neuronales artificiales corresponden a sistemas computacionales inspiradas en la interconectividad de las redes neuronales biológicas de los animales (Zurada 1992). Estas redes neuronales tienen la capacidad de aprender características o comportamientos similares al considerar ejemplos sin que explícitamente se les diga cuáles son sus objetivos. Los elementos básicos de las redes neuronales y la regla de aprendizaje Back-propagation serán presentados a continuación:

- **Nodo:** El elemento más básico corresponde a la neurona, también conocida como nodo. Un nodo toma una serie de inputs y computa un output normalizado acorde a una función de transferencia.
- **Pesos de conexión:** Una red neuronal está compuesta de varios nodos unidos entre ellos mediante conexiones, en donde algunos de los outputs de estos nodos corresponden a inputs de otros, todas estas conexiones tienen fuerza distinta, con un peso (entre cero y uno) asociado a cada conexión.
- **Función de transferencia:** Generalmente el estado de output de una neurona puede ser caracterizada como prendida o apagada. El cambio de un estado a otro es activado cuando la suma ponderada de los inputs y los pesos pasa cierto límite generado por esta función. A modo de ejemplo, un tipo de función de activación o transferencia corresponde a la función sigmoide, la cual está definida de la siguiente manera:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

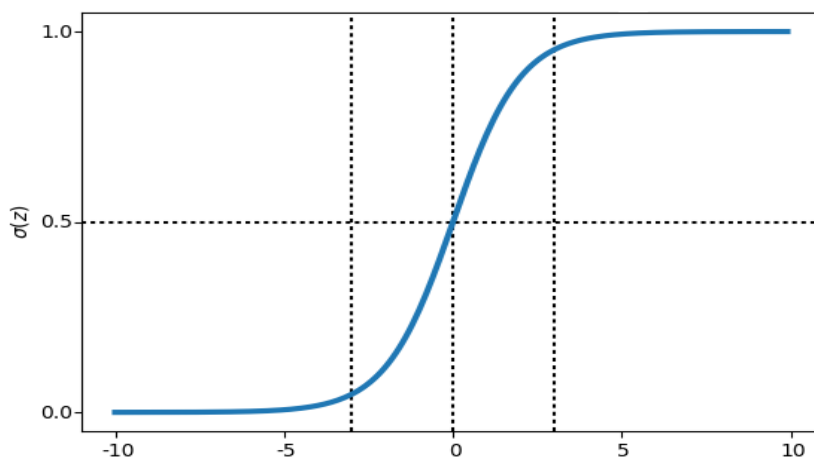


Figura 3.1: Función sigmoide

- Estructura topologica: Realmente cualquier tipo de estructura topologica puede ser utilizada, en donde para facilitar los problemas de analisis y visualización se organiza las neuronas en capas, con los nodos en capas adyacentes las unas a las otras. Una red neuronal tiene una capa de input y una o mas capas ocultas, llamadas así porque todas sus conexiones son internas de la red. Cada nodo perteneciente a una capa oculta toma como input todos los nodos de la capa anterior y propaga su valor la capa siguiente. La figura 3.2 a continuación presenta la arquitectura básica de una red neuronal.

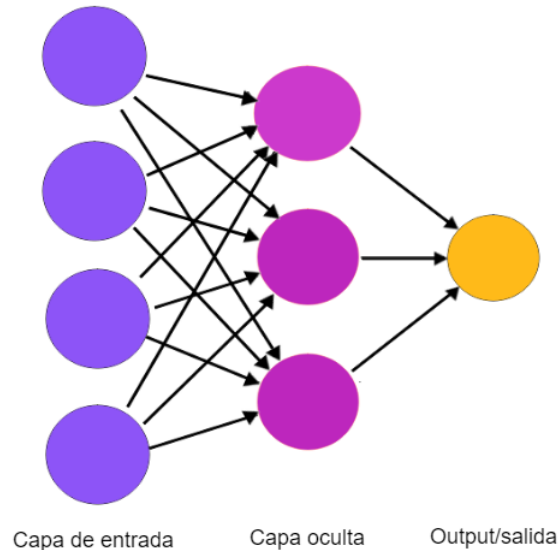


Figura 3.2: Arquitectura básica de una red neuronal artificial

- Back Propagation El paradigma de Back Propagation es un método utilizado para calcular el gradiente necesario para calcular los pesos usados en la red. La idea básica es ajustar el peso de las neuronas calculando el gradiente de la función de error y este error es propagado hacia el resto de las capas de la red. Para llevar a cabo el entrenamiento de forma efectiva se necesita un gran número de ejemplos del comportamiento deseado. Un resumen de los pasos iterativos involucrados con el entrenamiento es el siguiente:
 1. *Presentar una observación de la data utilizada normalizando entre 0 y 1 y entregarla a la capa de entrada de la red.*
 2. *Utilizando los pesos de conexión actuales de la red, computar el output de la observación ingresada.*
 3. *Comparar el output obtenido con el output deseado.*
 4. *Calcular una medida de error como función del vector de pesos w de la siguiente forma $E(w) = \frac{1}{2} \sum_{d \in D} (y^{(d)} - o^{(d)})^2$, este error define una superficie sobre el espacio de pesos que permite identificar la dirección en la cual los pesos deberían ser modificados para disminuir dicho error.*
 5. *Modificar los pesos de la conexión en una cantidad pequeña con la idea de acercarse al output deseado.*

Dada la naturaleza del algoritmo, que impide saber a los investigadores cuáles son las características que considera cada una de sus neuronas, la cantidad de capas y neuronas

por capa, así como la inicialización de los pesos y la función de transferencia escogidas dependen de cada caso de estudio en particular, siendo estos parámetros escogidos mediante experimentación.

3.1.4. Long Short Term Memory Recurrent Neural Network

Las redes neuronales recurrentes (RNN) son una clase de red neuronal cuyas conexiones forman un ciclo, lo que permite trabajar con data temporal, en donde el adjetivo recurrente refiere a la existencia de conexiones que permiten a un nodo a volver a si mismo. La expresión más sencilla para una red neuronal recurrente se denota de la siguiente manera:

$$h_t = \sigma(W_h X_t + W_r h_{t-1})$$

$$y_t = \sigma(W_y h_t)$$

En donde X_t corresponde a la data de entrada en el tiempo t , σ la función de activación utilizada, W_h corresponde a la matriz de pesos de la capa de entrada a la capa oculta, W_r denota los pesos de la capa oculta a la capa de salida, h_t corresponde a los pesos de los nodos ocultos en el tiempo t e y corresponde al valor de nodo de salida en el tiempo t . Figura 3.3 ilustra la arquitectura característica de las redes neuronales recurrentes:

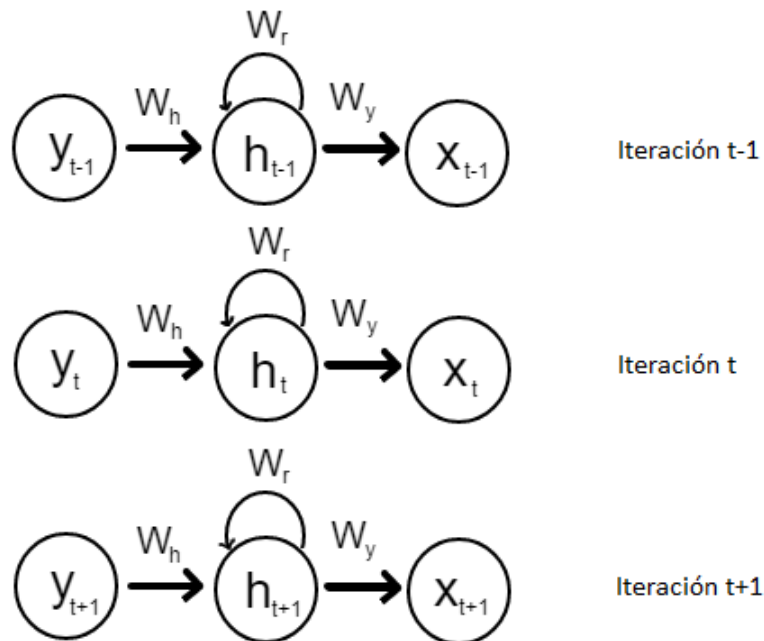


Figura 3.3: Iteraciones sobre una RNN

Debido a que esta arquitectura recurrente hace que el algoritmo de Backpropagation in-factible debido al problema del gradiente desvaneciente (Hochreiter (1991)) y que este no termine nunca de iterar, se introduce Backpropagation Through Time (BPTT) para expandir la arquitectura actual, de forma que se optimizen los pesos de la iteración actual en conjunto con la iteración anterior. La función de error es implementada como el cuadrado del error dada por la siguiente ecuación:

$$E_t = \sum (y_t - y)^2$$

Donde E_t es el error recursivo, y_t representa el valor predicho del tiempo en t e y denota el valor buscado.

La arquitectura de la Long Short-Term Memory Recurrent Neural Network (Hochreiter y Schmidhuber (1997)) es similar a la utilizada en la arquitectura de las redes neuronales artificiales, pero las capas ocultas de la LSTM contiene conexiones adicionales, que utilizan la arquitectura clásica de las redes neuronales recurrentes, que les permite además tener la capacidad de aprender o olvidar patrones.

Cada nodo de la LSTM esta compuesta de cuatro conexiones o puertas: llamadas generalmente input gate, output gate, forget gate y memory cell las cuales son utilizadas para controlar el flujo de información.

El primer paso para la LSTM es decidir que tipo de información será retenida o olvidada por cada nodo. Esta decisión es hecha por la ‘forget gate’ utilizando la función sigmoide, esta analiza los valores de h_{t-1} y x_t y entrega un numero f_t entre 0 y 1, donde 1 representa ‘retener completamente’ mientras que 0 significa ‘olvidar completamente’, mientras que valores intermedios representan aprender cantidades limitadas de información. El siguiente paso es decidir que nueva información se considera necesaria para ser almacenada en el estado de la celda. Esto tiene dos partes, primero una capa llamada input gate que utiliza la función sigmoide para entregar un valor i_t que decide que valores serán actualizados en la actual iteración y segundo una capa que mediante la función tangente hiperbólica crea un nuevo vector de posibles valores C'_t que son candidatos a ser agregados al estado del nodo. Estos valores son combinados al multiplicar el estado anterior de la celda por f_t y sumar el valor obtenido de i_t por los posibles valores a actualizar C'_t para actualizar el valor del estado de la celda. Finalmente se escoge el hecho de si es que el estado final será propagado a la próxima celda o no, esta decisión es realizada por la output gate. Tanto las fórmulas correspondientes a las distintas conexiones así como una ilustración de un nodo típico de la LSTM son presentadas a continuación.

Forget gate:

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f)$$

Input gate:

$$\begin{aligned} i_t &= \sigma(W_i * [h_{t-1}, x_t] + b_i) \\ C'_t &= \tanh(W_c * [h_{t-1}, x_t] + b_c) \\ C_t &= f_t * C_{t-1} + i_t * C'_t \end{aligned}$$

Output gate:

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

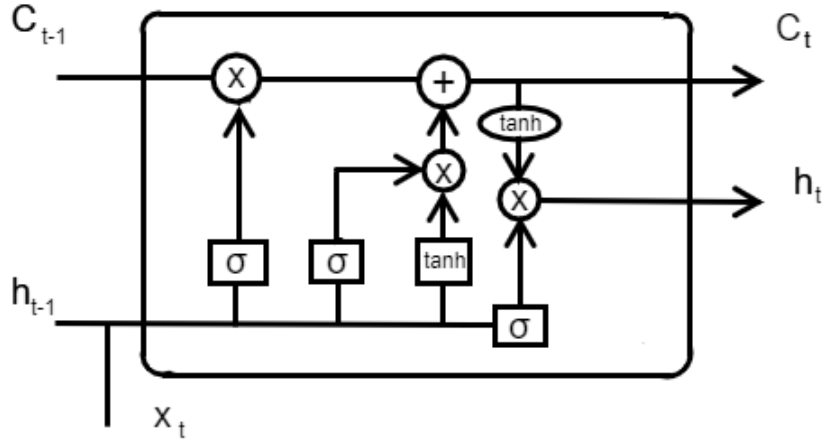


Figura 3.4: Nodo de una LSTM

Donde x_t representa a la observación ingresada en el tiempo t. Los valores de i_t , f_t , o_t , C_t , h_t representan los valores de la input gate, forget gate, el vector de activación de la output gate, el estado de la celda y el vector de output del nodo en el tiempo t, respectivamente. Los valores de W_i , W_f , W_o corresponden a las matrices de pesos de la input gate, forget gate y output gate, respectivamente. b_i , b_f , b_c y b_o representan el sesgo de su respectiva capa a minimizar y σ corresponde a la función sigmoide.

3.2. Modificaciones realizadas

Se plantea que existe una brecha entre los estudios realizados por investigadores de tráfico e investigadores que desarrollan algoritmos de ciencia de datos, donde los métodos generados por los primeros de estos no han sido desarrollados hasta su máximo potencial y los segundos se podrían ver beneficiados en aprovechar los conocimientos ya existentes y aplicarlos en sus estudios.

Por este motivo, esta tesis plantea dos modificaciones alternativas a la arquitectura de los modelos utilizados, apoyándose en la separación de fases de tránsito planteada en Kerner (1998) y Kerner (1999). Kerner plantea la existencia de distintos estados posible de tráfico: flujo libre, flujo sincronizado y congestión ampliada móvil. Kerner argumenta que las observaciones de estos estados son muy similares entre sí, pero difieren considerablemente unas de las otras.

Kerner plantea que estos puntos de inflexión entre un estado y otro aparecen a medida que la autopista llega a su densidad crítica, produciendo que los vehículos se vean forzados a disminuir su velocidad y pasar a un estado de congestión de menor velocidad y flujo.

Esto se puede apreciar en la figura 3.5, donde en niveles de flujo moderado se observan condiciones de flujo libre (F). Una vez que se llega a la saturación de la vía (maximum capacity), se produce un fenómeno de inestabilidad que lleva consigo un cambio de comportamiento, el que finalmente, se traduce a un estado de flujo sincronizado (S), de menor velocidad.

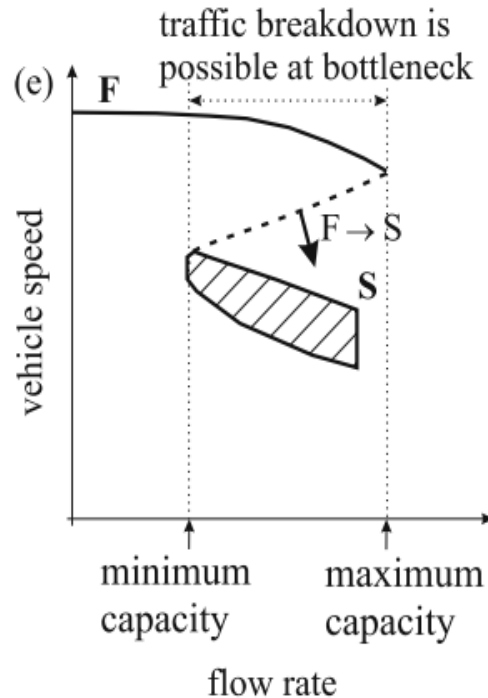


Figura 3.5: Cambio de estado a congestión sincronizada. Fuente: Kerner (2009)

Dado que el comportamiento del tráfico en una vía en que transitan vehículos depende de una infinidad de factores físicos, meteorológicos, culturales, entre otros, para determinar los distintos estados de congestión se plantea graficar y analizar dos interacciones de los datos y a partir de esto la generación de un meta algoritmo para delimitar dos estados.

- Velocidad - Flujo: Analizar qué límite corresponde al punto de inflexión entre un estado donde transitan vehículos a velocidades altas y uno que transitan a velocidades más bajas de lo que desearían los usuarios.
- Velocidad en el tiempo: Verificar el comportamiento de la velocidad durante intervalos de tiempo menores a un día, analizando la existencia de dos (o más) estados y estableciendo límites entre estos estados a partir de características de los datos.

Meta-algoritmo para la elección del punto de inflexión:

1. *Analizar las observaciones con flujo correspondiente al percentil superior (definido según el problema a evaluar) de todos los datos disponibles*
2. *Calcular el promedio de la velocidad de estas observaciones*
3. *Utilizar este valor calculado como el límite que separa los estados con o sin congestión*

Si bien este es un método heurístico basado en el enfoque teórico de Kerner, los resultados

mostraron que logra introducir mejoras significativas, lo que podría indicar la necesidad de investigar esta y otras sistematizaciones de este enfoque en futuras investigaciones.

3.2.1. Clasificador de estados de congestión (CEC)

La primera de estas modificaciones, aprovecha la información anterior para plantear un cambio en la manera en que la información de entrada es introducida al modelo. Los pasos a seguir son los siguientes:

1. Establecer patrones y límites que establezcan si una observación se encuentra en estado de congestión i .
2. Asignar a cada observación de la base de datos completa un valor categórico que determine el estado de la observación, a partir de los límites creados en el punto anterior.
3. Para cada estado i establecido, se ajusta un modelo de predicción i que determine la velocidad futura de las observaciones i , utilizando un porcentaje de los datos disponibles. En el caso de modelos de inteligencia artificial esto se conoce como la fase de entrenamiento.
4. Predicción de todas las observaciones a partir de su modelo respectivo y comparación completa con la velocidad efectivamente ocurrida.

Así, cada modelo será entrenado individualmente para responder al comportamiento para cada uno de los estados específicos, bajo el razonamiento de que si una observación obtenida en el tiempo t se encuentra en un estado i , la observación $t + 1$ tiene una alta probabilidad de encontrarse en el mismo estado i . La figura 3.6, a modo de ejemplo con dos estados, muestra que dependiendo del estado en que cada observación se encuentre si esta será asignada a la ANN 1 ‘congestionada’ o a la ANN 2 ‘descongestionada’.

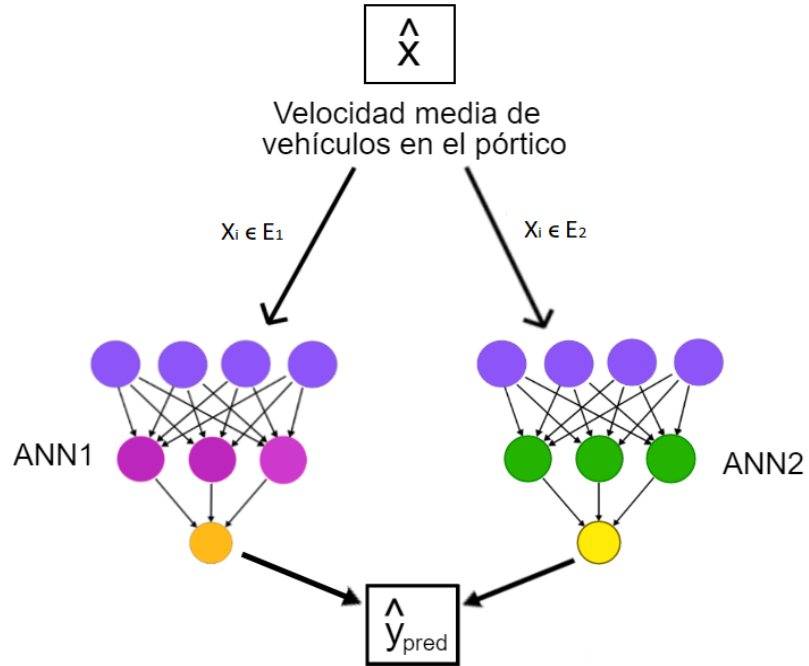


Figura 3.6: Ejemplo arquitectura del m todo CEC

3.2.2. Probabilidad de estado de congesti n (PEC)

La segunda metodolog a propuesta tiene grandes similitudes con la anterior, aunque lleva la asignaci n de estado un paso m s adelante. Se plantea la inclusi n de valores asociados a la probabilidad de cambiar o de mantenerse en su respectivo estado, de forma que esta informaci n sea utilizada como nuevo l mite para asignar a qu  modelo debe ser entregada cada observaci n.

Esta probabilidad fue determinada utilizando una red neuronal clasificadora, que pretende predecir cu l es la probabilidad de que la siguiente observaci n (seg n el horizonte predictivo escogido) se encuentre en estado de congesti n i . A continuaci n en la figura 3.7, se presenta un ejemplo del comportamiento de la arquitectura para un caso con dos estados, estado congestionado y uno sin congesti n.

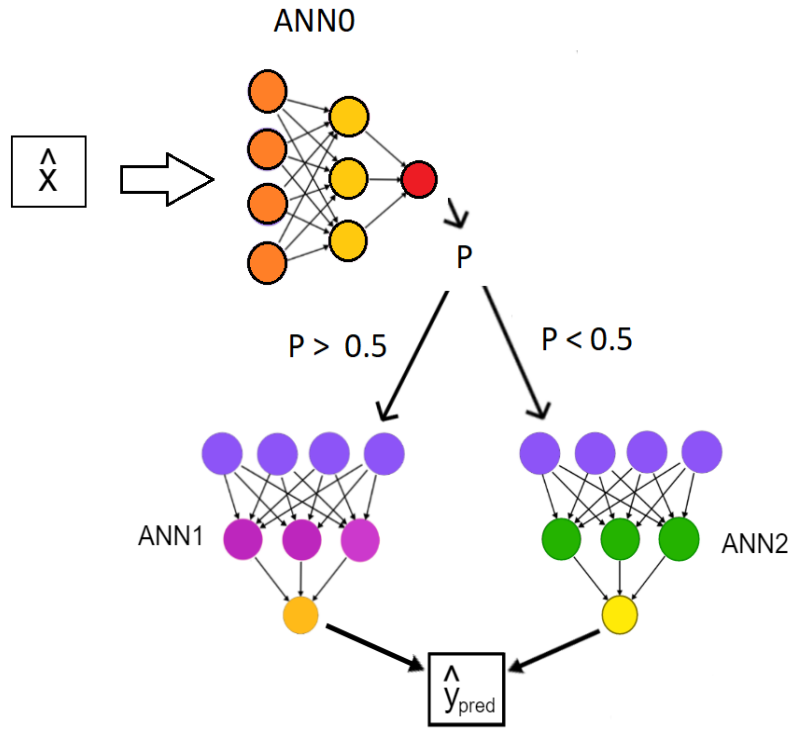


Figura 3.7: Ejemplo arquitectura del método PEC

Las variables explicativas son ingresadas en el sistema a través de la ANN 0, cuya misión es entregar un valor de la probabilidad P de que la siguiente observación se encuentre en estado de congestión. Al entrenar y ajustar esta red inicial a los datos, cada observación queda con una probabilidad asociada de mantenerse (o cambiar) de estado. Así, es posible pasar cada observación disponible por un filtro: si el valor de la probabilidad entregada por la red anterior es superior a 0.5, la observación es utilizada como dato de entrenamiento para la ANN 1, y en caso contrario, cuando P es menor a 0.5, es utilizada como dato de entrenamiento de la ANN 2. Para evitar que la red utilice como información el resultado que estamos buscando, se utiliza un set de datos de entrenamiento (correspondiente al 80 % de los datos) y un set de prueba (correspondiente al 20 % restante).

3.3. Indicadores de error

Las métricas de error en establecer el desempeño de los estimadores de velocidades fueron el error absoluto medio o Mean Absolute Error (MAE) y el error porcentual absoluto medio o Mean Absolute Percentage Error (MAPE), que se calculan de la siguiente manera:

$$MAE = \frac{1}{n} \sum |y_i - \bar{y}_i|$$

$$MAPE = \frac{1}{n} \sum \frac{|y_i - \bar{y}_i|}{y_i}$$

Donde para ambas ecuaciones n es el tamaño de la muestra, \bar{y}_i corresponde al valor predicho para la observación i e y_i corresponde al valor real.

Capítulo 4

Descripción del caso de estudio

Para realizar las estimaciones de las velocidades futuras de un autopista urbana, se utilizó información disponible de cuatro pódicos consecutivos (en orden: PA26, PA24, PA22 y PA20) de la Autopista Central dirección hacia el Sur. Esta autopista se encuentra ubicada en Santiago de Chile, siendo la autopista urbana concesionada más extensa del país, extendiéndose por 60,5 km de largo y conectando a más de 15 comunas de la ciudad. La predicción fue realizada únicamente para el pódico PA20 considerando la información de los pódicos anteriores a este.

Se dispone de información sobre las condiciones operativas de la autopista agrupadas en periodos de un minuto, desde octubre del 2016 a marzo del 2018. La data corresponde a 781.063 observaciones minutales (542 días) en cada uno de los pódicos, donde estas observaciones contienen las siguientes variables:

- Velocidad instantánea media (i, j, t) : Calculada como el promedio de la velocidad instantánea de los vehículos de tipo i que cruzan el pódico j en el minuto t .
- Desviación estándar de la velocidad instantánea (i, j, t) : La desviación estándar de la velocidad instantánea de los vehículos i que cruzan el pódico j en el minuto t .
- Flujo vehicular (i, j, t) : Corresponde a la cantidad de vehículos de tipo i que cruzan el pódico j en el minuto t .
- Fecha: Corresponde al día en que fue tomada la observación.
- Hora: Corresponde a la hora y minuto en que fue tomada la observación.

Los tipos de vehículo i corresponden a vehículos livianos, vehículos pesados (clasificados como camiones de dos o más ejes) o motocicletas. Mientras que los pódicos j corresponden a los pódicos PA20, PA22, PA24 y PA26 representados gráficamente en la figura 4.1 a continuación.



Figura 4.1: Flujo vehicular diario por hora

4.1. Imputación de datos

Para poder trabajar con datos nulos o faltantes sobre las velocidades (cuya ocurrencia se debe por falta de vehículos en ciertos momentos del día) se planteó la idea de promediar el dato anterior con el dato futuro, bajo el supuesto que cuando existe un bajo flujo de vehículos estos andarían a velocidades similares. Siguiendo bajo esta misma línea, se decidió trabajar con el intervalo horario de 6:30 a 22:59, debido a que en este intervalo horario el flujo vehicular es considerablemente inferior. Esta decisión fue realizada para evitar ruido y overfitting de los modelos, ya que estos se verían alterados al estar cerca de un tercio del día con flujo casi inexistente. De la tabla 4.1 es posible ver que la composición del tráfico en la sección y periodo analizado esta dada por un alto porcentaje de vehículos ligeros, seguido de un considerable porcentaje de vehículos pesados.

Tipo	Composición [%]	Velocidad [Km/h]			
		Promedio	Desv. Est.	Máxima	Mínima
Livianos	72,2 %	94,3	14,6	197	2,5
Pesados	26,8 %	82,6	11,2	176	2
Motos	1,1 %	89,6	13,3	222	2

Tabla 4.1: Análisis exploratorio de datos

4.2. Transformación de datos

Se consideró la necesidad de generar variables adicionales que explican condiciones operativas de la vía, siendo estas variables descritas a continuación:

- Densidad vehicular livianos: Calculada como el flujo de vehículos livianos dividido por la velocidad instantánea media de los vehículos livianos.
- Composición para el vehiculo (i): Calculada como el flujo del tipo del vehiculo i dividida por el flujo total. (i = autos, camiones, motocicletas)
- Delta velocidad livianos (a): Calculada como la diferencia entre la velocidad obtenida en la observación menos la velocidad obtenida en la observación a minutos atrás. (a = 1, 2, 3, 5, 10, 15)
- Día: Variable cualitativa ordinal según el día de la semana que es: lunes, martes, miércoles, etc.
- Tipo de día: Variable cualitativa ordinal sobre si el día es hábil o feriado. De los 542 días considerados 26 de ellos son feriados.
- Hora categorizada: La hora fue agrupada en intervalos de 15 minutos, bajo el planteamiento de que existen comportamientos similares en dichos intervalos. Esta variable categórica asigna el mismo valor a observaciones que se encuentren dentro del mismo intervalo, es decir, las observaciones con tiempos 6:08 y 6:09 obtienen el mismo valor asociado al intervalo horario 6:00-6:15, el cual sería distinto a cualquiera que no se encuentre en ese intervalo horario.

Además, fue generada la variable a predecir $VelocidadFutura(j,t,t_1)$ correspondiente a la velocidad obtenida en el mismo pórtico j en el periodo $t + t_1$ siendo $t_1 = 5, 10, 15, 30$ minutos.

Como última medida, las variables continuas, es decir, las no categóricas, fueron normalizadas sacando la media de su variable y escalando según la desviación estándar, así calculando el valor normalizado de la variable x como:

$$z = \frac{(x - u)}{s}$$

Donde u representa la media de la muestra y s corresponde a la desviación estandar de la misma. La decisión de normalizar fue tomada posterior a la implementación de los modelos debido a que estos predicen considerablemente peor y en oportunidades no son ni capaces de predecir cuando las variables no se encuentran normalizadas.

4.3. Clasificación de congestión

Como fue mencionado en el capítulo anterior, dos análisis y un meta algoritmo fueron utilizados para determinar los límites que asignan si una observación se encuentra en congestión o no: relación flujo y velocidad para encontrar el punto de inflexión y estudiar las fluctua-

ciones de velocidad durante periodos cortos de tiempo para discretizar el comportamiento de los distintos estados. En la figura 4.2, 4.3 y 4.4 se presentan gráficamente los distintos comportamientos de la autopista:

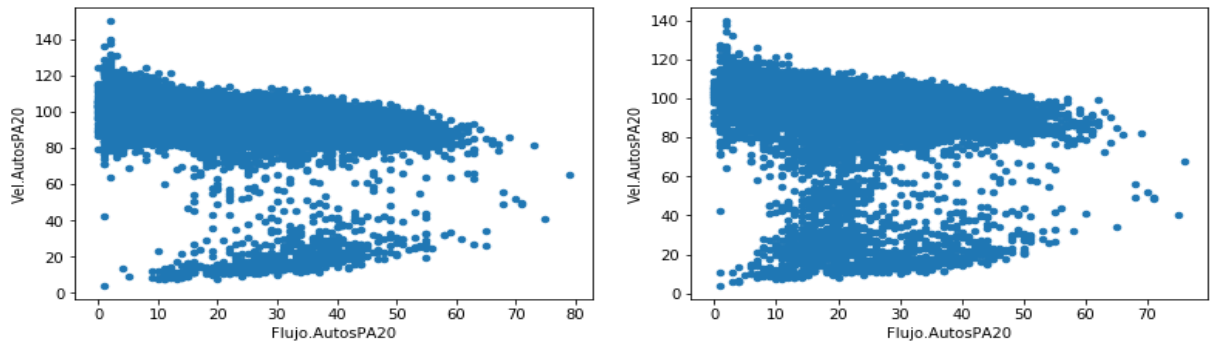


Figura 4.2: Velocidad-flujo mes de Octubre 2016 (Izquierda) y Noviembre 2016 (Derecha)

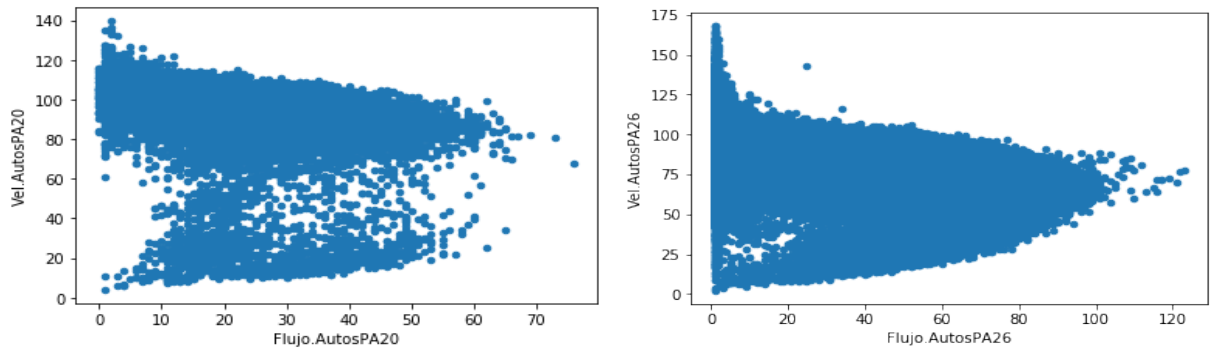


Figura 4.3: Velocidad-flujo año 2016 (Izquierda) y 2017 (Derecha)

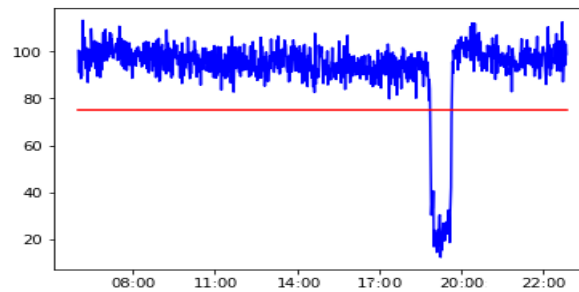


Figura 4.4: Velocidades instantáneas minutales del día 06/10/2016

A partir de estos análisis de los datos y el meta-algoritmo presentado en el capítulo anterior, se decide asignar dos estados, uno con congestión y uno descongestionado. Este valor es calculado mediante el promedio de la velocidad de las observaciones cuyo flujo que superen el percentil 99,5 %.

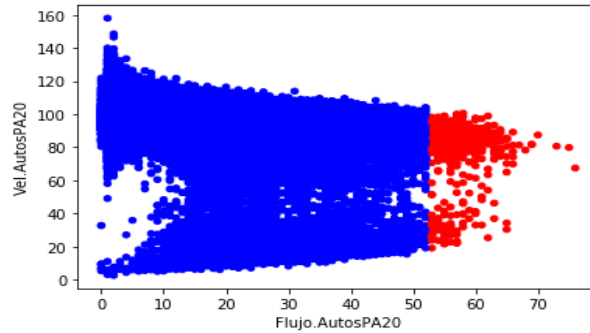


Figura 4.5: Flujo-velocidad Octubre 2016 - Junio 2017

En la figura 4.5 se presentan en rojo las observaciones cuyos valores del flujo superan el percentil 99.5 %. El promedio de la velocidad de estas observaciones resulta ser 79,88 km/h. Por lo tanto, los estados de congestión corresponden a las observaciones en que la velocidad registrada por el p rtico sean menores a este valor, mientras que se consideran como observaciones en estado sin congesti n para el caso contrario. En las figuras anteriores claramente se puede discernir entre los distintos comportamientos.

Este l mite de 80 km/h, establece a un estado en el cual los veh culos comienzan a generar una alta interferencia entre ellos, lo que corresponde a un nivel de servicio de congesti n F (Level of service, HCM 2010). Este l mite adem s permite contar con una cantidad de datos adecuada para la estimaci n. Mientras que para velocidades menores no exist an la cantidad de datos adecuados para realizar los experimentos, para l mites de velocidades mayores se volv a m s dif cil discernir entre qu  observaciones corresponden a congesti n y cu les no.

Capítulo 5

Resultados

La variable a predecir corresponde a la $VelocidadFutura(t)$ -a horizontes temporales de 5, 10, 15 y 30 minutos- para el pórtico PA20 definida en el capítulo anterior, utilizando toda las variables categóricas y continuas descritas en el capítulo anterior para el momento de la observación, que incluyen velocidades instantáneas actuales y anteriores, flujos, composiciones, densidades y variables temporales.

5.1. Comparación modelos

Primero, se evaluó el desempeño de los cuatro modelos de inteligencia artificial, para los cuatro horizontes temporales (HT) (5 minutos, 10 minutos, 15 minutos y 30 minutos) realizando una comparación con las métricas propuestas de cada uno de ellos: red neuronal artificial (ANN), bosque aleatorio (RF), support vector regressor (SVR) y long short-term memory recurrent neural network (LSTM) en contraste además a una regresión lineal (LR) para las dos métricas de error consideradas. Los resultados de esta comparación se presentan a continuación en la tabla 5.1:

	ANN		RF		LR		SVR		LSTM	
HT	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
5 min	4.15	5.12	4.71	5.97	4.23	5.85	5.44	8.95	4.15	5.28
10 min	4.30	5.77	4.88	6.65	4.43	6.09	5.52	9.29	4.34	5.81
15 min	4.42	6.21	4.95	7.72	4.61	6.73	5.97	9.75	4.40	6.12
30 min	4.74	7.32	5.33	8.82	5.06	8.31	6.30	11.02	4.70	7.10

Tabla 5.1: Comparación de modelos considerados

Los resultados muestran que utilizando los dos modelos con mejor desempeño (ANN y LSTM) es posible obtener errores absolutos medios menores a 4.74 km/h al comparar la predicción realizada con respecto a la velocidad efectivamente ocurrida, independiente del horizonte predictivo. Además, es posible apreciar que el desempeño de la ANN y la LSTM superan el de la regresión lineal y el bosque aleatorio para cualquier horizonte temporal,

mientras que estos últimos tienen un mejor desempeño al del support vector regressor.

Mientras que la ANN tiene el mejor desempeño de todos los modelos para horizontes temporales menores a 10 minutos considerando tanto MAE como MAPE; el desempeño de la LSTM para horizontes de tiempos mayores es más certera y tiene una mayor robustez al evaluar sobre distintas iteraciones. Esta diferencia sería aún mayor si es que la variable 'DeltaVel(a)', que indica las diferencias de velocidades no fuera incluida. La ANN aprovecha de mejor manera esta información adicional ya que la LSTM en su caso base tiene la capacidad de retener y utilizar información de observaciones pasadas, mejorando el desempeño de la primera de estas en un 3%. Se argumenta que estos modelos son exitosos, bajo la consideración de que los vehículos se desplazan a velocidades promedio de 100 km/h, lo que significaría que los valores de error entre 4 a 6 km/h serían bastante pequeños.

Llama la atención el mal desempeño del SVR en comparación al resto de los modelos, al obtener errores superiores a 1 - 1.5 km/h en comparación a la ANN y LSTM equivalente a más de un 30% - 35% de error sobre estos. Este mediocre desempeño del SVR no es solamente respecto a su capacidad predictiva sino que también en la cantidad recursos utilizados. Mientras que cualquiera de los otros modelos logra ser entrenado y predecir en un tiempo menor a una hora, el tiempo de entrenamiento generalmente supera las 30 horas, en oportunidades incluso superando las 50 para la predicción para un horizonte temporal único. Diversas modificaciones en el preprocesamiento de datos fueron realizadas para disminuir el tiempo de entrenamiento así como los errores en la predicción, tales como la no normalización de los datos, el uso de distintos intervalos horarios para la predicción y la remoción de los datos temporales, sin mejoras significativas. Independiente del modelo escogido, una vez entrenados estos pueden predecir de forma instantánea.

En todos los casos antes presentados, los modelos presentan la menor capacidad predictiva al tratar de estimar valores extremos de velocidad, sobreestimando considerablemente las velocidades mínimas/máximas y cuando existen cambios abruptos de las mismas, que en este estudio se terminan considerando como cambios de estado.

5.2. Modelos propuestos

Seleccionado los mejores enfoques, se realiza la comparación de sus capacidades predictivas con las arquitecturas planteadas, para todos los horizontes de predicción propuestos. Se presentan los resultados en la tabla 5.2 a continuación:

HT	ANN		LSTM		CEC		PEC	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
5 min	4.15	5.12	4.15	5.28	4.27	5.46	3.97	4.60
10 min	4.30	5.77	4.34	5.81	4.42	5.75	4.17	4.87
15 min	4.42	6.21	4.40	6.12	4.52	6.61	4.27	5.12
30 min	4.74	7.32	4.70	6.70	4.71	7.26	4.69	5.52

Tabla 5.2: Comparación de modificaciones con modelos de mejor desempeño

Los resultados son satisfactorios, mientras que no es posible concluir una mejora significativa por parte de la arquitectura CEC, se concluye que la arquitectura PEC es una clara mejora sobre todos los modelos utilizados, incluyendo el desempeño de la red neuronal y la LSTM.

Dos puntos son importantes a analizar sobre la arquitectura PEC: cuán precisas son las capacidades predictivas de la red neuronal clasificadora acerca de la detección de cambios de estados y segundo, cómo cada una de las redes de la arquitectura predice individualmente. En la tabla 5.3 a continuación se presentan matrices de confusión para cada intervalo horario, mientras que en la tabla 5.4 se presenta individualmente el desempeño de las dos redes neuronales regresoras entrenadas a partir de los estados considerados, que mediante la unión de estos resultados se obtiene desempeño final del método PEC:

Horizonte Temporal	Valor predicho	Valor actual		Precisión
		Sin congestión	Congestión	
5 min	Sin congestión	96,7 %	0,2 %	98,6 %
	Congestión	1,1 %	1,9 %	
10 min	Sin congestión	96,7 %	0,2 %	98,4 %
	Congestión	1,3 %	1,7 %	
15 min	Sin congestión	96,6 %	0,4 %	98,2 %
	Congestión	1,5 %	1,6 %	
30 min	Sin congestión	96,6 %	0,3 %	97,8 %
	Congestión	1,9 %	1,2 %	

Tabla 5.3: Matriz de confusión según horizonte temporal

HT	Modelo sin congestión		Modelo con congestión		PEC	
	MAE	MAPE	MAE	MAPE	MAE	MAPE
5 min	3.90	4.17	7.69	27.23	3.97	4.60
10 min	4.05	4.35	10.34	33.24	4.17	4.87
15 min	4.15	4.55	11.39	30.41	4.27	5.12
30 min	4.49	4.91	11.21	24.50	4.69	5.52

Tabla 5.4: Evaluación de cada una de las ANN del modelo PEC

La precisión de la predicción de estados futuros disminuye a medida que se extiende el horizonte temporal, mediante un aumento en la cantidad de oportunidades en que se clasifica que el próximo estado como congestión aún cuando esta no lo este. Se perjudica principalmente la cantidad de predicciones correctas de estado de congestión.

Por otra parte, se aprecia en la tabla 5.4 que el desempeño del modelo sin congestión es considerablemente superior a su contraparte y que la disminución en capacidades predictivas a medida que aumentan los horizontes temporales del primero de estos, es el principal causante de la pérdida de precisión del modelo final, el cuál obtiene resultados similares a los modelos clásicos a medida que el horizonte temporal alcanza valores cercanos a la media hora. Se plantea que mediante mejoras al modelo de asignación de estados a mediano plazo este podría predecir con considerables ventajas que los otros modelos utilizados.

Capítulo 6

Conclusión

La mayor contribución de esta tesis es la introducción de una metodología alternativa para la predicción de velocidades en corto plazo, obteniendo mejores resultados que modelos recurrentemente utilizados en la literatura. Existe una variedad de alternativas para utilizar esta metodología en casos aplicados, tanto como información para los usuarios de la autopista, así como para el ruteo y planificación de los servicios de emergencia.

A partir de los experimentos realizados, se plantea que las predicciones obtenidas por los modelos ANN, LSTM, PEC y CEC son bastante buenas debido a los bajos valores obtenidos del MAE, donde en las modificaciones fue posible caracterizar el tráfico en dos estados previamente mencionados. En circunstancias en las que no hay congestión, las velocidades fluctúan entre un rango entre 90-110 km/h, mientras que en un segundo estado con situaciones de congestión, las velocidades disminuyen considerablemente en intervalos de tiempo muy cortos. Se establece que la principal dificultad encontrada en los modelos utilizados en la literatura, sigue siendo la capacidad de predecir la discontinuidad en la velocidad, es decir, el momento en el cual se producirá un cambio de estado, el fenómeno de caída abrupta de la velocidad, los motivos por los cuales fueron planteadas los cambios en las arquitecturas.

Apoyándose en el trabajo realizado, se concluye que la etapa de preprocesamiento y transformación de datos es una de la más significativa para una predicción más certera y obtención de resultados finales, ya que gran parte de las mejoras en el desarrollo de los modelos ocurrieron debido a una mayor y mejor cantidad de información.

6.1. Alcances y limitaciones

Los resultados de las metodologías son bastante satisfactorias, se presupone que es posible aminorar aún más los errores obtenidos y lograr un mejor ajuste de los modelo al agregar información adicional, por ejemplo, datos relacionados con accidentes o eventos extraordinarios de tráfico, como fallas en los accesos o días previos a festivos.

Sin embargo, se comenta a partir de esto, que errores tan bajos en estados de descongestión

se encuentran dentro de una dispersión aceptable y esperable para los usuarios de tránsito, principalmente considerando los resultados obtenidos en la literatura, como se muestra en la tabla 6.1, donde el desempeño (capacidades predictivas) de la arquitectura PEC es superior a una variedad de investigaciones con distintos enfoques y modelos. Se argumenta a partir de los resultados de este estudio que el principal problema a resolver es determinar cuándo aparecerá congestión y cuál será la velocidad de dicha congestión cuando esta aparezca.

Autor	Metodo de predicción	MAPE 10 min (%)
Tang et al. (2016)	Fuzzy neural network	6.83
Yao et al. (2016)	SVM with spatiotemporal parameters	7.05
Ma et al. (2015)	LSTM	4.52
Jia et al. (2016)	Deep belief network	7.33
Jia et al. (2016)	ARIMA	7.83
Véjar (2019)	PEC	4.87

Tabla 6.1: Cantidad de valores considerados según velocidad

El análisis mostró que la introducción de la probabilidad de que las observaciones se encontraran en un estado de congestión, mejora la capacidad predictiva del modelo. El análisis indica que el método de estimación basado en inteligencia artificial, logra introducir de manera adecuada el concepto de probabilidad. Sin embargo, éste depende de la definición de estado congestión y cómo se asigna cada variable a dicho estado.

Si bien esta tesis ha tratado de incorporar conceptos teóricos a partir de análisis visuales relacionados con la teoría de los estados de tráfico de Kerner, se considera que podría ser un avance relevante sistematizar y objetivizar la forma de introducir el concepto de congestión o estados de tráfico al modelo.

6.2. Recomendaciones para trabajo futuro

Se plantean las siguientes líneas como avance:

- Analizar cómo la información temporal agregada afecta el modelo. Se explica que el desempeño de la red neuronal artificial es tan similar a la de la LSTM debido a la inclusión de variables temporales, tales como la diferencia en velocidad. Por lo tanto, se deberían analizar qué otros factores temporales serían relevantes para la formulación de este tipo de modelos.
- Establecer de forma más certera y sistemática como se define un estado de congestión. Además, se plantea la necesidad de utilizar otros métodos de clusterización e introducir análisis de los conceptos tradicionales de la ingeniería de tránsito como relaciones flujo-velocidad y densidad vehicular.

Capítulo 7

Bibliografía

Abdulhai, B., Porwal, H., & Recker, W., “Short-Term Freeway Traffic Flow Prediction Using Genetically-Optimized Time-Delay-Based Neural Networks”, TRB 1999 Annual Meeting CD-ROM, 1999.

Ahmed, M. S., & Cook, A. R., “Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques”, Transportation Research Record, 722, Transportation Research Board, Washington, DC, pp. 1-9, 1979.

Asif, M. T., Dauwels, J., Chong Yang Goh, Oran, A., Fathi, E., Muye Xu, & Jaillet, P. (2014). Spatiotemporal Patterns in Large-Scale Traffic Speed Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 15(2), 794–804.

Breiman, Leo. Random forests. *Machine learning* 45.1 (2001): 5-32.

Castro-Neto, M., Jeong, Y.-S., Jeong, M.-K., & Han, L. D. (2009). Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert Systems with Applications*, 36(3), 6164–6173.

Chen, M., & Chien, S. (2001). Dynamic Freeway Travel-Time Prediction with Probe Vehicle Data: Link Based Versus Path Based. *Transportation Research Record: Journal of the Transportation Research Board*, 1768, 157–161.

Clark, S.D., Dougherty, M.S., & Kirby, H.R., 1993. The use of neural networks and time series models for short-term traffic forecasting: a comparative study. PTRC 21st Summer Annual Meeting.

Clark, S. (2003). Traffic Prediction Using Multivariate Nonparametric Regression. *Journal of Transportation Engineering*, 129(2), 161–168.

D’Angelo, M., Al-Deek, H., & Wang, M., “Travel Time Prediction for Freeway Corridors”, *Transportation Research Record 1676*, Transportation Research Board, Washington, DC, pp.184–191, 1999.

Davis, G. A., & Nihan, N. L. (1991). Nonparametric Regression and Short-Term Freeway Traffic Forecasting. *Journal of Transportation Engineering*, 117(2), 178–188.

Dia, H., “An Object-Oriented Neural Networks Approach to Short-Term Traffic Forecasting”, *European Journal of Operational Research*, 131(2), pp. 253–261, 2001.

Dougherty, M. S., & Cobbett, M. R. (1997). Short-term inter-urban traffic forecasts using neural networks. *International Journal of Forecasting*, 13(1), 21–31.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., Vapnik, V. (1997). Support vector regression machines. In *Advances in neural information processing systems* (pp. 155-161).

Ermagun, A., Levinson, D. (2018). Spatiotemporal traffic forecasting: review and proposed directions. *Transport Reviews*, 1–29.

Fu, R., Zhang, Z., & Li, L. (2016). Using LSTM and GRU neural network methods for traffic flow prediction. 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC).

Guo, J., & Williams, B. M. (2010). Real-Time Short-Term Traffic Speed Level Forecasting and Uncertainty Quantification Using Layered Kalman Filters. *Transportation Research Record: Journal of the Transportation Research Board*, 2175(1), 28–37.

Guo, F., Krishnan, R., & Polak, J. (2013). A computationally efficient two-stage method for short-term traffic prediction on urban roads. *Transportation Planning and Technology*, 36(1), 62–75.

Guo, J., Huang, W., & Williams, B. M. (2014). Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. *Transportation Research Part C: Emerging Technologies*, 43, 50–64.

Habtemichael, F. G., & Cetin, M. (2016). Short-term traffic flow rate forecasting based on identifying similar traffic patterns. *Transportation Research Part C: Emerging Technologies*, 66, 61–78.

Hamed, M. M., Al-Masaeid, H. R., & Said, Z. M. B. (1995). Short-Term Prediction of Traffic Volume in Urban Arterials. *Journal of Transportation Engineering*, 121(3), 249–254.

Hamner, B. (2010). Predicting Travel Times with Context-Dependent Random Forests by Modeling Local and Aggregate Traffic Flow. 2010 IEEE International Conference on Data Mining Workshops.

Huang, W., Song, G., Hong, H., & Xie, K. (2014). Deep Architecture for Traffic Flow Prediction: Deep Belief Networks With Multitask Learning. *IEEE Transactions on Intelligent Transportation Systems*, 15(5), 2191–2201.

Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.

Innamaa, S. (2005). "Short-Term Prediction of Travel Time Using Neural Networks on An Interurban Highway", *Transportation*, Volume 32, pp, 649-669.

Ishak, S. & Al-Deek, H. (2002) "Performance Evaluation of Short-Term Time-Series Traffic Prediction Model", *Journal of Transportation Engineering*, 128(6), pp. 490– 498.

Ishak, S., Kotha, P., & Alecsandru, C. (2003). Optimization of Dynamic Neural Network Performance for Short-Term Traffic Prediction. *Transportation Research Record: Journal of the Transportation Research Board*, 1836, 45–56.

Jia Y., Wu J., Yiman Du. (2016). Traffic speed prediction using deep learning method. 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC).

Kerner, Boris S. "Experimental features of self-organization in traffic flow." *Physical review letters* 81.17 (1998): 3797.

Kerner, Boris S. (1999). "The physics of traffic." *Physics World* 12.8: 25.

Kerner, Boris S. (2009) *Introduction to modern traffic flow theory and control: the long road to three-phase traffic theory*. Springer Science Business Media.

Kwon, J., Coifman, B., & Bickel, P. (2000). Day-to-Day Travel-Time Trends and Travel-Time Prediction from Loop-Detector Data. *Transportation Research Record: Journal of the Transportation Research Board*, 1717, 120–129.

Lingras, P., Sharma, S., & Zhong, M. (2002). Prediction of Recreational Travel Using Genetically Designed Regression and Time-Delay Neural Network Models. *Transportation Research Record: Journal of the Transportation Research Board*, 1805, 16–24.

Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, 54, 187–197.

Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction. *Sensors*, 17(4), 818.

Mackenzie J., Roddick F., Zito R. (2018). An Evaluation of HTM and LSTM for Short-Term Arterial Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 1–11.

Meng, M., Shao, C., Wong, Y., Wang, B., & Li, H. (2015). A two-stage short-term traffic flow prediction method based on AVL and AKNN techniques. *Journal of Central South University*, 22(2), 779–786.

Park, D., L. Rilett, & G. Han. (1998). Forecasting Multiple-Period Freeway Link Travel Times Using Neural Networks with Expanded Input Nodes. *Proc., ASCE (1998) 5th International Conference on Applications of Advanced Technologies in Transportation*, Newport Beach, CA, pp. 325–332

Polson, N. G., Sokolov, V. O. (2017). Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, 79, 1–17.

Smith, B.L., & Demetsky, M.J. (1994). Short-term traffic flow prediction models—a comparison of neural network and nonparametric regression approaches. *IEEE International Conference on*. 1707-1709..

Smith, B. L., & Demetsky, M. J. 1997!. “Traffic flow forecasting: comparison of modeling approaches.” *J. Transp. Eng.*, 123 4!, 261– 266.

Shen, L. (2008). "Freeway Travel Time Estimation and Prediction Using Dynamic Neural Networks," Ph. D. Thesis, Florida International University.

Stathopoulos, A., & Karlaftis, M. G. (2003). A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies*, 11(2), 121–135

Sun, H., Liu, H., Xiao, H., He, R., Ran, B. (2003). Use of Local Linear Regression Model for Short-Term Traffic Forecasting. *Transportation Research Record: Journal of the Transportation Research Board*, 1836, 143–150.

Tian, Y., Pan, L. (2015). Predicting Short-Term Traffic Flow by Long Short-Term Memory Recurrent Neural Network. 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity).

Theja, P. V. V. K., Vanajakshi, L. (2010). Short Term Prediction of Traffic Parameters Using Support Vector Machines Technique. 2010 3rd International Conference on Emerging Trends in Engineering and Technology.

Van Lint, J. W. C., Hoogendoorn, S. P., & Van Zuylen, H. J. (2002). "Freeway Travel Time Prediction with State-Space Neural Networks - Modeling State-Space Dynamics with Recurrent Neural Networks", *Transportation Research Record*, 1811, Transportation Research Board, Washington, D.C, pp. 30–39.

Van Lint, J. W. C., & Van der Zijpp, N. J., (2003) "Improving A Travel Time Estimation Algorithm by Using Dual Loop Detectors", *Transportation Research Record*, 1855, Transportation Research Board, Washington, D.C., pp. 41–48.

Van Lint, J. W. C. (2004), *Reliable Travel Time Prediction for Freeways*, Ph.D Thesis, Submitted to Civil Engineering for Delft University of Technology, Netherland.

Van Lint, J. W. C., Hoogendoorn, S. P., & Van Zuylen, H. J., "Accurate Freeway Travel Time Prediction with State-Space Neural Networks under Missing Data", *Transportation Research, Part C: Emerging Technology*, 13(5–6), pp.347–369, 2005.

Van Lint, J. W. C., "Reliable Real-Time Framework for Short-Term Freeway Travel Time Prediction", *Journal of Transportation Engineering*, 132(12), pp. 921-932, 2006.

Vanajakshi, L., Rilett, L. R. (2004). A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed. *IEEE Intelligent Vehicles Symposium*.

Vlahogianni, E.I., Golias, J.C., & Karlaftis, M.G., 2004. Short-term traffic forecasting: Overview of objectives and methods. *Transp. Rev.* 24, 533–557.

Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2005). Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. *Transportation Research Part C: Emerging Technologies*, 13(3), 211–234.

Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43, 3–19.

Xie, Y., Zhang, Y., & Ye, Z. (2007). Short-Term Traffic Volume Forecasting Using Kalman Filter with Discrete Wavelet Decomposition. *Computer-Aided Civil and Infrastructure Engineering*, 22(5), 326–334.

Williams, B. (2001). Multivariate Vehicular Traffic Flow Prediction: Evaluation of ARIMA Modeling. *Transportation Research Record: Journal of the Transportation Research Board*, 1776, 194–200.

Williams, B. M., & Hoel, L. A. (2003). Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. *Journal of Transportation Engineering*, 129(6), 664–672.

Wu, C.-H., Ho, J.-M., Lee, D. T. (2004). Travel-Time Prediction With Support Vector Regression. *IEEE Transactions on Intelligent Transportation Systems*, 5(4), 276–281.

Yao, B., Hu, P., Zhang, M., Jin, M. (2014). A support vector machine with the tabu search algorithm for freeway incident detection. *International Journal of Applied Mathematics and Computer Science*, 24(2), 397–404.

Yao, B., Chen, C., Cao, Q., Jin, L., Zhang, M., Zhu, H., Yu, B. (2016). Short-Term Traffic Speed Prediction for an Urban Corridor. *Computer-Aided Civil and Infrastructure Engineering*, 32(2), 154–169.

Yu, B., Song, X., Guan, F., Yang, Z., Yao, B. (2016). k-Nearest Neighbor Model for

Multiple-Time-Step Prediction of Short-Term Traffic Condition. *Journal of Transportation Engineering*, 142(6), 04016018.

Yu, H., Wu, Z., Wang, S., Wang, Y., Ma, X. (2017). Spatiotemporal Recurrent Convolutional Networks for Traffic Prediction in Transportation Networks. *Sensors*, 17(7), 1501.

Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y., Liu, J. (2017). LSTM network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, 11(2), 68–75.

Zheng, W., Lee, D.-H., Shi, Q. (2006). Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach. *Journal of Transportation Engineering*, 132(2), 114–121.