Review/summary of article:
X. Chen, H. Pao, and Y. Lee (2014). Efficient traffic speed forecasting based on massive heterogeneous historical data. In IEEE International Conference on Big Data, Washington, DC, USA, pages 10-17.

Chen et al. (2014) developed a model to predict traffic speed on highways, by applying a computational intelligence approach using historical data. The model was conceived to provide traffic predictions with high confidence level, to be used by both traffic administrators and drivers.

The authors consider using data from several sources, including Intelligent Transportation System (ITS), weather conditions, and other non-ITS data. In the proposed model, a hybrid approach is applied: the k-Nearest-Neighbors algorithm is applied to select those subsets of data that provide the most useful information for the predicted scenario. In turn, a small-scale Gaussian Process Regression built from that information is applied for traffic forecasting.

The prediction model needs to be re-trained frequently to incorporate up-to-date data. Therefore, the results are computed using a Map-Reduce model for data processing, implemented over the Hadoop framework. The performance is further enhanced by applying a Support Vector Registration technique to determine the task-to-machine assignments.

The experimental evaluation of the proposed forecasting method was performed over a real scenario on the I5N road in San Diego, California, U.S. Specific information from the Research Data Exchange, a platform for transportation data sharing (U.S. Department of Transportation, https://www.its-rde.net) is used. ITS data includes the speed, flow, and occupancy for vehicles, collected using sensors and loop-detectors on the road. The weather data include visibility reports from weather stations near San Diego, available from the National Oceanic and Atmospheric Administration. Several experiments were conducted to evaluate the prediction accuracy of the computational intelligence methods by analyzing different number of sensors and time periods, different data sizes for training, and different sets of historical data.

The main results of the experimental analysis indicate that the proposed method accurately predicts the traffic speed of vehicles in the studied scenario, having an average forecasting error smaller than 2 miles per hour. The robustness of the method is also highlighted, as using larger datasets allows improving the forecasting results. Furthermore, a reduction of up to 69% on the execution time is reported for the parallel/distributed approach implemented over the Hadoop framework, when compared against the efficiency of the same algorithm running on a single machine.

This article presents an interesting approach for using big data analysis and computational intelligence to predict traffic speed. The proposed implementation using Map-Reduce over Hadoop demonstrates that distributed computing is a valuable tool for reducing the prediction times.

Comentarios

La reseña es extensa, está escrita para incluirse en un trabajo de extensión considerable (por ejemplo, un artículo de revista de más de 15 páginas o una tesis/informe de proyecto). En un artículo científico la reseña debe tener una menor extensión, escribirse en un único párrafo y debe estar bien enfocada en el comentario del trabajo y sus principales contribuciones en el contexto de la investigación que se reporta en el artículo.

La reseña se redacta en **pasado**, al corresponder a un trabajo previamente realizado.

Todas las frases y párrafos son autocontenidos y pueden comprenderse sin necesidad de otras frases o párrafos.

La reseña se redacta organizando su contenido desde los conceptos más generales a los conceptos más específicos.

Las frases y párrafos están vinculadas apropiadamente por conectores.

Análisis del contenido

El primer párrafo presenta la referencia bibliográfica, el principal aporte, la metodología utilizada y la motivación, contextualizando el trabajo relevado.

Chen et al. (2014) developed a model to predict traffic speed on highways, by applying a computational intelligence approach using historical data. The model was conceived to provide traffic predictions with high confidence level, to be used by both traffic administrators and drivers.

El segundo párrafo presenta detalles sobre los datos utilizados y se profundiza en la descripción del enfoque algorítmico propuesto.

The authors consider using data from several sources, including Intelligent Transportation System (ITS), weather conditions, and other non-ITS data. In the proposed model, a hybrid approach is applied: the k-Nearest-Neighbors algorithm is applied to select those subsets of data that provide the most useful information for the predicted scenario. In turn, a small-scale Gaussian Process Regression built from that information is applied for traffic forecasting.

El tercer párrafo presenta comentarios específicos sobre las características del modelo propuesto, justifica la elección del modelo algorítmico y presenta la técnica utilizada para mejorar el desempeño del método propuesto.

The prediction model needs to be re-trained frequently to incorporate up-to-date data. Therefore, the results are computed using a Map-Reduce model for data processing, implemented over the Hadoop framework. The performance is further enhanced by applying a Support Vector Registration technique to determine the task-to-machine assignments.

A continuación se describen los detalles del análisis experimental realizado, incluyendo el escenario geográfico, los repositorios de datos considerados en el estudio, los datos específicos y el objetivo de los experimentos computacionales que se llevaron a cabo.

The experimental evaluation of the proposed forecasting method was performed over a real scenario on the I5N road in San Diego, California, U.S. Specific information from the Research Data Exchange, a platform for transportation data sharing (U.S. Department of Transportation, https://www.its-rde.net) is used. ITS data includes the speed, flow, and occupancy for vehicles, collected using sensors and loop-detectors on the road. The weather data include visibility reports from weather stations near San Diego, available from the National Oceanic and Atmospheric Administration. Several experiments were conducted to evaluate the prediction accuracy of the computational intelligence methods by analyzing different number of sensors and time periods, different data sizes for training, and different sets of historical data.

Posteriormente, los resultados experimentales son presentados y comentados, tomando en cuenta la calidad de resultados, la robustez y la eficiencia computacional del método propuesto.

The main results of the experimental analysis indicate that the proposed method accurately predicts the traffic speed of vehicles in the studied scenario, having an average forecasting error smaller than 2 miles per hour. The robustness of the method is also highlighted, as using larger datasets allows improving the forecasting results. Furthermore, a reduction of up to 69% on the execution time is reported for the parallel/distributed approach implemented over the Hadoop framework, when compared against the efficiency of the same algorithm running on a single machine.

Finalmente, se presenta un análisis valorativo de la propuesta en el contexto de la investigación que está desarrollando por parte de quien realiza la reseña.

This article presents an interesting approach for using big data analysis and computational intelligence to predict traffic speed. The proposed implementation using Map-Reduce over Hadoop demonstrates that distributed computing is a valuable tool for reducing the prediction times.