

Predicción de Múltiples Puntos de Series de Tiempo Utilizando Support Vector Machines

E. Guzmán – Ramírez¹, E. Bautista – Thompson², J. Figueroa – Nazuno³

Centro de Investigación en Computación
Instituto Politécnico Nacional
Unidad Profesional Adolfo López Mateos.
Zacatenco. Col. Lindavista, CP 07738
México D.F.

erick_igr@yahoo.com¹, ebautista@correo.cic.ipn.mx², jfn@cic.ipn.mx³.

RESUMEN

Se presenta la evaluación de la predicción de múltiples puntos de series de tiempo, mediante un corrimiento de ventana para Support Vector Machines (SVM) con dos funciones de kernel distintas (lineal y con base radial). Para la evaluación se utilizó un conjunto de treinta series de diferente origen. Se observa un buen desempeño de las SVM para la predicción de los primeros puntos utilizando la función de kernel radial, a pesar de la expansión del error de predicción.

Palabras clave:

Funciones de Kernel para SVM, Predicción de series de tiempo, Support Vector Machines.

I. INTRODUCCIÓN.

Una serie de tiempo es un conjunto de observaciones que expresan la dinámica de la variable de un sistema. Así, una serie de tiempo s se puede definir como un conjunto de valores reales consecutivos, $S = \{x_i \in R \mid x_1, x_2, \dots, x_N\}$. Uno de los principales objetivos del estudio de series de tiempo es la predicción. Dicho problema puede ser representado separando la serie $s = \{x_1, \dots, x_N\}$ en ventanas $w = (x_i, \dots, x_{i+p-1})$ de tamaño p , entonces, hay que encontrar una función $f: R^k \rightarrow R$ tal que $f(x_i, \dots, x_{i+p-1}) = x_{i+p}$ para cada $i \in \{0, N-p\}$ [1]. Utilizando una representación como ésta es posible abordar el problema de predicción de series de tiempo con Support Vector Machines (SVM), las cuales son máquinas de aprendizaje capaces de realizar clasificación binaria (reconocimiento de patrones) así como tareas de aproximación de funciones (cálculo de regresión).

En los últimos años las SVM se han convertido en un tema de investigación de gran interés, dando lugar a una amplia gama de aplicaciones reportadas en la literatura [2, 3, 4]. Sin embargo, en el campo de la predicción de series de tiempo, se encuentran pocos reportes de trabajos que hagan uso de las SVM, ya que debido a las características

del algoritmo, no es posible obtener directamente una predicción de más de un punto de la serie a la vez, como es común en muchas otras técnicas. Por ello que en este trabajo, se presenta el empleo de una representación de ventana deslizante en el conjunto de entrenamiento del algoritmo SVM para realizar predicciones de múltiples puntos. Se reportan los resultados obtenidos del proceso de predicción de treinta series de tiempo, clasificadas como series sintéticas y series experimentales, siendo las primeras obtenidas mediante selección de los parámetros de una función generadora y las segundas representan fenómenos económicos, físicos, fisiológicos, etc. Los experimentos fueron realizados utilizando dos tipos distintos de funciones kernel para la SVM: (dot kernel) y con base radial (radial basis). La evaluación del desempeño de la técnica se realizó mediante la comparación gráfica de los puntos predichos y los puntos originales de la serie, además del cálculo de dos tipos de error de predicción. El artículo está estructurado de la siguiente manera: en la Sección II se describe la estructura de las SVM, en la Sección III se presentan las funciones de kernel utilizadas en este trabajo, la Sección IV describe la metodología utilizada para la realización de los experimentos, la Sección V se presentan los resultados obtenidos, finalmente, en la Sección VI se presentan las conclusiones del trabajo.

II. SUPPORT VECTOR MACHINES.

Las SVM son sistemas de aprendizaje introducidos por Vladimir Vapnik [5] que utilizan un espacio hipotético de funciones lineales en un espacio de características, las cuales entrenadas con un algoritmo de optimización implementa una tendencia de aprendizaje estadístico [6, 7]. Dicha teoría de aprendizaje resuelve el problema de encontrar una función f de una clase de funciones $(f_\alpha)_\alpha \in \Lambda$ del tal forma que

minimice el riesgo esperado $R[f]$ respecto a una función de pérdida L , cuando la distribución de probabilidad de los ejemplos $P(x)$ y sus clasificaciones $P(y|x)$ no son conocidos, y tienen que ser estimados de un número finito de ejemplos $(x_i, y_i)_i \in I$.

$$R[f] = \iint L(y, f(x)) dP(y|x) dP(x) . \quad (1)$$

El algoritmo de las SVM resuelve éste problema minimizando el riesgo regularizado $R_{reg}[f]$ el cual es la suma ponderada del riesgo empírico $R_{emp}[f]$ con respecto a los datos $(x_i, y_i)_i = 1 \dots n$ y a un término de complejidad $\|w\|^2$

$$R_{reg}[f] = R_{emp}[f] + \lambda \|w\|^2 . \quad (2)$$

En su formulación básica, la SVM encuentra una función de decisión $f(x) = \text{sign}(\omega \cdot x + b)$ que minimiza el error de predicción en el conjunto de entrenamiento y procura el mejor desempeño en la generalización.

III. FUNCIONES DE KERNEL.

Una de las principales características de las SVM es el uso de funciones de kernel para extender la clase de funciones de decisión al caso no lineal [6]. Esto se hace mapeando los datos desde el espacio de entrada X a un amplio espacio de características χ mediante una función Φ , y resolviendo el problema de aprendizaje lineal en χ .

$$\Phi : X \rightarrow \chi . \quad (3)$$

La función real Φ no necesita ser conocida, es suficiente tener una función de kernel k que calcule el producto interno en el espacio de características.

$$k(x, y) = \Phi(x) \cdot \Phi(y) . \quad (4)$$

A. Kernel Lineal.

El kernel lineal $k(x, y) = x \cdot y$ es la función de kernel más simple. La función de decisión toma la forma $f(x) = \omega \cdot x + b$. Cuando se usa éste kernel en el análisis de series de tiempo, el modelo resultante es un modelo auto regresivo de orden p

($AR[p]$). Si las series son generadas por el mismo modelo AR se les considera similares.

B. Kernel con base radial.

Las funciones de kernel con base radial (FBR) toman la forma $k_\gamma(x, y) = \exp(-\gamma \|x - y\|^2)$, siendo γ una constante de proporcionalidad cuyo rango de valores útiles debe ser estimado para cada aplicación en particular. Para las series de tiempo se ha mostrado, que la información dentro de una ventana de una serie, puede ser obtenida de otras ventanas que son similares en términos de la distancia euclidiana entre ellas, lo que hace útil al kernel con base radial para la predicción [1].

IV. METODOLOGÍA.

A. Representación del conjunto de entrenamiento

La representación de ventana deslizante para el conjunto de entrenamiento de la SVM, permite utilizar los últimos p valores de una ventana para predecir el siguiente valor de una serie de tiempo (véase la Sección I). Esto es posible debido a que el corrimiento de la ventana a través de la serie, permite a la SVM (mediante la función de kernel utilizada) estimar la dependencia existente entre cada uno de sus valores, de esta forma se obtiene la información necesaria, para realizar la predicción del siguiente valor a partir de los valores anteriores de la serie. Entonces, para realizar una predicción de múltiples puntos es necesario crear un nuevo conjunto de entrenamiento cada vez que se realiza una predicción simple (de un solo punto). El nuevo conjunto se obtiene agregando al conjunto de entrenamiento original la ventana $w' = (x_{i+p-2}, \dots, x_{i+p})$, siendo x_{i+p} el valor correspondiente a la predicción anterior.

La evaluación de la técnica de predicción con corrimiento de ventana, se llevó a cabo modificando la herramienta mySVM [8]. En los experimentos se utilizaron las funciones de kernel con base radial y lineal para establecer un punto de comparación, entre dos diferentes funciones de kernel de SVM en la predicción de series de tiempo.

B. Series de tiempo utilizadas

Los experimentos se realizaron utilizando treinta series de tiempo de las cuales se tomaron los primeros 950 datos como conjunto de entrenamiento, siendo organizados en ventanas de tamaño 3 y realizándose la predicción de los 50

puntos siguientes de cada serie. Experimentalmente se observó que la SVM presenta dificultades para detectar tanto variaciones grandes (del orden de 1×10^3) como pequeñas (del orden de 1×10^{-4}) en los datos de entrenamiento, razón por lo cual las series de tiempo fueron normalizadas para un rango de valores $[-1, 1]$.

La selección de las series utilizadas se hizo bajo el criterio de tener un conjunto representativo de series de tiempo con diferente dinámica y dificultad para ser predichas [9, 10]. A continuación se presentan las series de tiempo utilizadas en los experimentos así como la descripción de su naturaleza.

- Seno. Serie periódica generada por la función $f(x) = \text{sen}(x)$, para 10 ciclos.
- Vanderpol. Serie periódica generada por una ecuación diferencial que es modelo de un circuito oscilador de un tubo de vacío [11].
- Qperiodic2. Serie cuasi periódica obtenida de la medición de una variable de velocidad (cm/s) en un experimento anular para reproducir un Flujo de Couette (Reología).
- Qperiodic3. Serie cuasi periódica obtenida de forma análoga a la serie Qperiodic2.
- Mackey-Glass. Serie caótica generada por una ecuación diferencial de retardo temporal: Modelo de formación de células sanguíneas blancas (linfocitos) [12].
- Logistic. Serie caótica generada por un mapa interpretado como un modelo ecológico de las variaciones anuales de poblaciones de insectos [11].
- Lorenz. Serie caótica generada por un sistema de ecuaciones diferenciales: modelo de convección de fluidos (convección de Rayleigh-Benard) la cual se presenta en la atmósfera terrestre [11].
- Rossler. Serie caótica generada por un sistema de ecuaciones diferenciales: modelo simplificado de Lorenz [12].
- Ikeda. Serie caótica generada por un mapa: modelo de la dinámica de pulsos de luz que viajan a través de un medio no lineal [11].
- Henon. Serie caótica generada por un mapa: modelo simplificado del mapa de Poincaré para el modelo de Lorenz [11].
- Tent. Serie caótica generada por un mapa de tipo lineal por partes [11].
- A1(Concurso Santa Fe). Serie compleja obtenida a partir de mediciones experimentales de la intensidad de un láser NH3 Infrarrojo Lejano [13].
- D1 (Concurso Santa Fe). Serie compleja generada por un modelo de la dinámica de una partícula amortiguada en un potencial interacción [13].
- Laser (Concurso Santa Fe). Serie compleja obtenida a partir de mediciones experimentales de la intensidad de pulsos de láser Infrarrojo Lejano, condiciones de la frecuencia ≥ 3 frecuencia del sistema que genera A1 [13].
- Dow Jones. Serie compleja obtenida a partir Índice Dow Jones del NYSE (New York Exchange).
- Kobe. Serie compleja obtenida a partir acelerograma del sismo de Kobe del enero de 1995.
- ECG. Serie compleja obtenida a partir electrocardiograma humano estándar extremidad (Limb Lead II), voltaje amplificado digitalizado a 1000 Hz [14].
- EEG. Serie compleja obtenida a partir electroencefalograma humano [14].
- ASCII. Serie compleja generada a partir código de texto ASCII.
- El niño. Serie compleja obtenida a partir medición experimental de la dinámica variable del fenómeno climático el niño.
- HIV DNA. Serie compleja obtenida a partir código del DNA del Virus de Inmunodeficiencia Humana HIV (1 = A, 2 = C, 3 = G, 4 = T).
- Human DNA. Serie compleja obtenida del código del DNA Humano.
- Lovaina (Concurso Universidad de Lovaina). Serie compleja generada a partir de datos formato ASCII [15].
- Plasma. Serie compleja obtenida a partir medición de una variable de un experimento con plasma.
- Primos. Serie compleja generada a partir números primos.
- S&P500. Serie compleja obtenida a partir Índice Financiero de Standard & Pool 500 empresas más importantes de los valores de Nueva York.
- Star. Serie compleja obtenida a partir medición de la intensidad luminosa de una estrella variable.
- Brownian Motion. Serie estocástica generada a partir del modelado del movimiento browniano (proceso de ruido blanco integrado).

- White Noise. Serie estocástica generada a partir del modelado de proceso de ruido blanco (ruido aleatorio uniforme).

C. Evaluación de Predicciones

La representación de ventana con corrimiento, realizada en el conjunto de entrenamiento de la SVM, requiere que la misma se retroalimente con los resultados previamente predichos, por lo anterior es de esperarse que el desempeño de la predicción se vea afectado, debido al error acumulado conforme se avanza en los puntos predichos. Para determinar si una predicción fue exitosa o no, se realizó la comparación gráfica de los puntos originales de la serie y los puntos predichos por la técnica, en donde se consideraron dos factores: el seguimiento de la tendencia de la serie original y el número de puntos predichos. Para cuantificar el error de la predicción se emplearon las siguientes fórmulas de error utilizadas en la literatura relacionada con la predicción de series de tiempo [9]:

Error Medio Raíz Cuadrada (RMSE). Representa la medida típica del error de predicción, siendo particularmente sensible a las discrepancias grandes entre los valores original y predicho de la serie de tiempo.

$$RMSE(x) = \left[\frac{1}{N} \sum_{i=1}^N (x_i^f - x_i^o)^2 \right]^{\frac{1}{2}} \quad (5)$$

Error Bias (BE). Es una medida de la tendencia de un modelo a sobre o sub predecir un valor. Un BE positivo indica una tendencia a sobre predecir, es decir, que el algoritmo se ajustó demasiado a los datos de entrenamiento por lo que, de haber cambios considerables en la dinámica de la serie, se puede tener dificultad al predecir; por el contrario, un BE negativo indica una tendencia a sub predecir, lo cual significa que el algoritmo no fue capaz de captar totalmente la dinámica de la serie. Un BE cercano a cero indica que el algoritmo se ajustó correctamente a la dinámica de la serie predicha.

$$BE(x) = \frac{1}{N} \sum_{i=1}^N (x_i^f - x_i^o) \quad (6)$$

Para cada una de las fórmulas anteriores N representa al número de puntos de la serie y x_i^f y x_i^o representan respectivamente al i -ésimo punto predicho por la técnica y al i -ésimo punto de la serie original.

V. RESULTADOS.

A continuación se muestran dos ejemplos de los resultados obtenidos que ilustran el desempeño de las SVM en la predicción de series de tiempo.

La Fig. I muestra la predicción de 50 puntos de la serie Tent utilizando una función de kernel radial con $\gamma = 0.01$. Nótese el seguimiento de la tendencia de la serie a pesar de las discrepancias entre las amplitudes de la señal original y la predicha.

La Fig. II corresponde a 50 puntos predichos de la serie Seno con una función de kernel lineal. Obsérvese como la exactitud de la predicción disminuye conforme aumenta el número de puntos predichos.

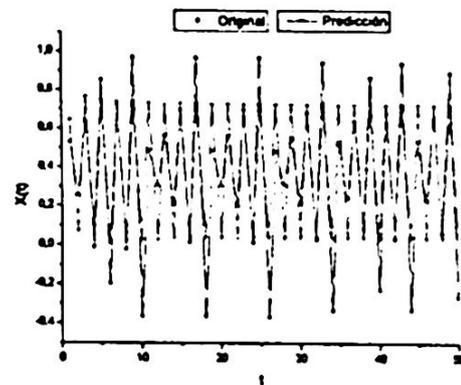


FIGURA I. Comparación entre la predicción y puntos originales de la serie Tent con una función de kernel radial ($\gamma = 0.01$).

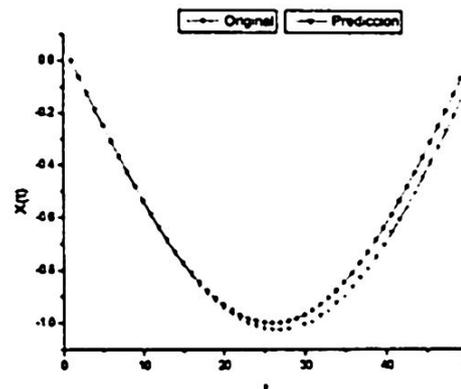


FIGURA II. Comparación entre la predicción con kernel lineal y los puntos originales de un segmento de la serie Seno.

Se encontró que la función de kernel de tipo radial es la que obtiene mejores resultados en la predicción de series de tiempo, ya que fue capaz de

predecir 13 de las series utilizadas, mientras que la función de kernel lineal únicamente fue capaz de predecir tres series: Seno, Vanderpol y Qperiodic3. Asimismo, se observó que el algoritmo SVM es capaz de predecir series con diferente comportamiento dinámico: periódico, cuasi periódico, caótico y complejo [10]. Finalmente, en la Tabla I se presentan los errores RMSE y BE calculados para los 50 puntos predichos en los experimentos exitosos con la función de kernel radial (FBR), además de su correspondiente Alcance de Predicción (AP), definido como el número de puntos predichos, para los cuales se satisfacen los factores de evaluación de las predicciones descritos en la Sección IV.

TABLA I.
RMSE, BE Y AP DE LAS PREDICCIONES CON FBR

Serie	RMSE	BE	AP
Seno	0.0682	0.00258	50
Vanderpol	0.70549	-0.02588	50
Qperiodic2	0.17429	-0.14772	50
Qperiodic3	0.46989	0.09800	50
Mackey-Glass	0.5007	0.06878	6
Logistic	1.08828	-0.05962	50
Lorenz	0.17389	0.15915	50
Rosler	0.4131	-0.07741	50
Tent	0.20708	0.02593	50
A1	7.37899	5.05720	7
Laser	0.32574	0.01397	25
Kobe	0.08582	-0.04860	50
Lovaina	0.16296	-0.13157	50

Nótese en la Tabla I que todos los valores del RMSE están por debajo de la unidad, excepto en las series Logistic y A1, en las cuales dicho valor se ve aumentado debido al error acumulado para los últimos puntos de la serie. Asimismo, se observa que en general, el algoritmo SVM no es particularmente susceptible a presentar sub o sobre predicciones, ya que solo en 2 casos (A1 y Lorenz) se tiene un BE mayor a 1.5×10^{-1} y solo en un caso (Qperiodic2) se presentó un BE menor a -1.4×10^{-1} . El Alcance de Predicción tiene un valor promedio de 37, lo cual significa que a pesar de la expansión del error se logra al menos mantener la tendencia de la serie en la predicción de los primeros 37 puntos.

VI. CONCLUSIONES.

El algoritmo presentado para la generación del conjunto de entrenamiento de las SVM, permitió

cuantificar el desempeño de esta técnica predicción de múltiples puntos de series de Del análisis de los resultados obtenidos trabajo se concluye, que las SVM con una de kernel con base radial constituyen una herramienta, para abordar el problema predicción de series de tiempo de múltiples La técnica muestra una buena capacidad ajustarse a la dinámica de la serie de durante el proceso de entrenamiento, demuestran los resultados del BE cercanos en la mayoría de las series predichas. Por las SVM alcanzan su mejor desempeño cuando requiere hacer predicciones a corto plazo, promedio en los primeros 37 puntos de predicción.

REFERENCIAS.

- [1] S. Rüping, "SVM Kernels for Time Series Analysis". Department AI Unit University of Dortmund.
- [2] H. Lodi, C. Saunders, J. Shawe-Taylor, N. Cristianini, Watkins, "Text Classification using String Kernels". Machine Learning Research 2, 2002.
- [3] C. Yeang, S. Ramaswamy, P. Tamayo, S. Mukherjee Rifkin, M. Angelo, M. Reich, E. Lander, J. Mesirov, "Molecular classification of multiple tumor types". Genome Research, MIT Whitehead Institute, One Square, Cambridge, 2001.
- [4] K. R. Müller, A. Samola, G. Rätsh, B. Schölkopf Kohlmorgen, V. Vapnik, "Predicting Time Series with S Vector Machines". Proceedings of ICANN'97, Springer 1327, p. 999-1004.
- [5] V. Vapnik, "Statistical Learning Theory". Wiley, Chichester, 1998.
- [6] N. Cristianini, J. Shawe-Taylor, "An introduction Vector Machines and other kernel based methods". University Press, UK, 2000.
- [7] S. Rüping, "Support Vector Machines and Learning Time". IEEE International Conference on Accoustics, Speech and Signal Processing, 2003.
- [8] S. Rüping, "mySVM Manual". CS Department University of Dortmund.
- [9] E. Bautista, J. Figueroa, "Matriz de Conocimiento Dificultad de Predicción de Series de Tiempo". VII Iberoamericano de Reconocimiento de Patrones, México, 2002.
- [10] A. Espinosa, J. Figueroa, "Análisis Caótico y Predicción la Contaminación Ambiental en la Ciudad de México" Congreso Internacional de Electrónica, Comunicaciones y Computadoras, CONTELECOMP'98, México, 1998.
- [11] E. Ott, "Chaos in Dynamical Systems". Cambridge Press, 2000.
- [12] E. Ott, "Coping with Chaos. Analysis of Chaotic Data and Exploitation of Chaotic Systems". Tim Sauer, James John Wiley & Sons. 1994.
- [13] A. S. Weingend, N. A. Gershenfeld, "Time Series Forecasting the Future and Understanding the Past". Wesley-Longman, 1994.
- [14] H. Kantz and T. Schreiber, "Nonlinear Analysis", Cambridge University Press. 2000.
- [15] J. A. K. Suykens and J. Vandewalle, "Nonlinear Model Advanced Black-Box Techniques", Kluwer Academic Press 1998.