

Tabla de Contenido

1. Introducción	1
1.1. Motivación	1
1.2. Alcances	2
1.3. Objetivos	3
1.4. Hipótesis de Trabajo	4
1.5. Acuerdos de Confidencialidad	4
1.6. Estructura de la Tesis	5
2. Estado del Arte y Marco Teórico	6
2.1. Metodologías para Estimación de Riesgo Criminal	6
2.2. Introducción a Procesos Puntuales	11
2.3. Definición de Procesos Puntuales	12
2.4. Función de Intensidad	13
2.5. Estimación de Función de Intensidad	13
2.5.1. Definición de Kernels y Tipos	13
2.5.2. <i>Kernel Density Estimation</i>	14
2.5.3. Modelo por Mezcla de Gaussianas	16
2.5.4. Métricas para Análisis de Modelos Espaciales	16
2.5.5. Métricas de Validación Predictiva	21
3. Metodología de Caracterización de Riesgo Criminal Incluyendo Elementos Fenomenológicos	23
3.1. Procesamiento de Datos	24
3.1.1. Separación de Datos	24
3.2. Procesamiento Espacial	25
3.2.1. Elección Previa de Parámetros	25
3.2.2. Restricciones Paramétricas	28
3.3. Procesamiento Temporal	30
3.3.1. Preprocesamiento de Series Temporales	31
3.3.2. Modelo por Mezcla de Kernels Periódicos	33
3.3.3. Reconstrucción de Señales Mediante Wavelet	34
3.4. Procedimiento de predicción	37
3.4.1. Proyección de Señales Pretéritas	38
3.5. Resumen del Algoritmo a Utilizar	38
4. Caso de Estudio: Modelación de Riesgo Criminal Espacio-Temporal con	

Datos Reales de Comunas del Gran Santiago	40
4.1. Validación de Metodología Utilizando una Comuna	41
4.1.1. Descripción Zona de Estudio	41
4.1.2. Base de Datos	41
4.1.3. Resultados Preliminares	43
4.1.4. Estudio de Criterios de Término de Iteraciones en Modelos Espaciales	50
4.1.5. Estudio Cantidad de Componentes Iniciales en Modelos Espaciales . .	51
4.1.6. Estudio de Mecanismo sobre Valores Propios de Matrices de Covarianza	56
4.1.7. Estudio de Inclusión de Calles en la Caracterización de Riesgo	59
4.1.8. Resultados Predictivos Generales	62
4.2. Resultados de Metodología para Diferentes Comunas	69
4.2.1. Resultados AUC Promedio para Cada Comuna de Estudio	70
4.2.2. Discusión	71
4.3. Comparación de Resultados Predictivos con Otros Métodos	72
5. Conclusiones	76
5.1. Trabajo Futuro	78
5.1.1. Con Respecto a Métricas de Evaluación Espaciales	78
5.1.2. Propuesta de Método <i>Online</i> para Determinar Número de Componentes Óptimos para GMM	78
5.1.3. Con Respecto a Condiciones Iniciales para GMM	78
5.1.4. Con Respecto a Mecanismo sobre Valores Propios de Matrices de Covarianza	79
5.1.5. Con Respecto a Algoritmo EM	79
5.1.6. <i>Framework</i> de Entrenamiento de Parámetros	79
Bibliografía	81

Índice de Tablas

4.1. Muestra de nombre de delitos y código de delitos para agrupado alcoholes.	42
4.2. Cantidad de delitos según agrupados anual	42
4.3. Parámetros para caracterización <i>a priori</i> y <i>a posteriori</i> típicas, donde <i>n_components</i> es la cantidad de componentes de la GMM, <i>covariance_type</i> indica el tipo de covarianza siendo en este caso con vectores propios arbitrarios, <i>min_covar</i> es el mínimo valor de la covarianza diagonal, <i>params</i> corresponde a los parámetros a entrenar siendo <i>weights</i> , <i>means</i> y <i>covariances</i> y <i>niter</i> es la cantidad de iteraciones del algoritmo EM.	44
4.4. Resumen mejores λ análisis estático fechas 2015-01-01 a 2015-05-01.	64
4.5. Resumen mejores λ análisis estático fechas 2016-01-01 a 2016-05-01.	65
4.6. Rangos de AUC de predicción para años 2015 y 2016 del análisis estático	65
4.7. Resumen mejores λ análisis series de tiempo perfectas fechas 2015-01-01 a 2015-05-01.	66
4.8. Resumen mejores λ análisis series de tiempo perfectas fechas 2016-01-01 a 2016-05-01.	66
4.9. Rangos de AUC de predicción para años 2015 y 2016 del análisis con series de tiempo perfectas	66
4.10. Resumen mejores λ análisis predicciones estándar fechas 2015-01-01 a 2015-05-01.	67
4.11. Resumen mejores λ análisis series de tiempo estándar fechas 2016-01-01 a 2016-05-01.	68
4.12. Rangos de AUC de predicción para años 2015 y 2016 del análisis con series de tiempo estándar	68
4.13. Porcentaje de aciertos para un 10% de cobertura para las diferentes comunas trabajadas considerando el mejor resultado según el turno policial y versión del algoritmo (estático o dinámico). Basado en los resultados de Anexos: Resultados AUC promedio para cada comuna de estudio.	75
5.1. Nombre de delitos y sus agrupados según el tipo DMCS.	85

Índice de Ilustraciones

2.1.	Métodos matemáticos para predecir ocurrencia de crímenes de acuerdo al problema que se desea solucionar. Fuente: Libro de W. Perry <i>et al.</i> [8].	7
2.2.	Diferentes métodos para la construcción de <i>Hot-Spots</i> . a) ocurrencia de eventos criminales, b) estimación de elipses mediante la desviación estándar de los delitos a su centro, c) designación de riesgo según zonas administrativas policiales, d) mapeo del territorio a través de grillas y finalmente en e) la estimación del riesgo mediante KDE (<i>Kernel Density Estimation</i>). Fuente: S. Chainey <i>et al.</i> [9].	8
2.3.	Estimación de las intensidades de fondo para el caso temporal y espacial. Fuente: [10].	9
2.4.	Kernel periódico con $x \in [0, 2]$, $x' = 0.5$, $\sigma = 1.0$, $p = 0.7$ y $l = [0.25, 0.35, 0.5]$	15
2.5.	Coefficiente de Silhouette para cada dato de entrenamiento considerando datos artificiales generados por 4 clusters versus un clustering mediante 10 clusters.	18
2.6.	Promedio del coeficiente de Silhouette para diferente cantidad de clusters o componentes del algoritmo GMM.	18
2.7.	Métrica de desviaciones para diferente número de clusters o componentes del algoritmo GMM.	19
3.1.	Metodología general para caracterización y predicción de riesgo criminal. . .	23
3.2.	Esquema general para procesamiento espacial.	26
3.3.	Ejemplificación de mecanismo	30
3.4.	Esquema general para procesamiento temporal.	30
3.5.	Método de correlaciones: para un conjunto de centros \mathbf{c} y periodos \mathbf{p} finitos se calcula la correlación entre la serie temporal suavizada $\bar{R}(t)$ con el kernel periódico para un determinado centro y periodo (construcción del conjunto $\Omega_{\text{corrj}}^{\text{pr}}$). Finalmente se ordenan según su grado de correlación y se escogen los elementos más correlacionados (construcción del conjunto $\mathcal{PC}_{\text{corrj}}^{\text{pr}}$). Fuente: tesis del Sr. Pablo Fuentes [6].	35
3.6.	Ecuaciones (izquierda) que definen funciones bases típicas de Transformada Wavelet Continua con su respectiva forma (derecha). Fuente: Trabajo de C. Torrence <i>et al.</i> [11]	36
3.7.	Señal temporal típica suavizada asociada a un kernel espacial con su respectivo EEW para periodos de tiempo de 0 a 120 días, y su histograma de energía promedio para bandas de ancho de 3 días. Fuente: tesis del Sr. Pablo Fuentes [6].	36
3.8.	Resumen de metodología	39

4.1.	División política del Gran Santiago y comuna de Puente Alto en color celeste.	41
4.2.	Caracterización espacial de modelos.	45
4.3.	Pertenencia de eventos \mathcal{E}_{pr} a $\mathcal{K}_j^{pr} \forall j = 1, \dots, 179$ (considerando centroides en color rojo) y construcción de histograma de cantidad de eventos de cada \mathcal{K}_j^{pr} .	45
4.4.	Serie de tiempo de ejemplo suavizada con KDE con su respectiva matriz de correlaciones destacando los primeros 10 pares de centros y periodos (en unidades de 8 horas)	46
4.5.	Serie de tiempo suavizada con KDE de un kernel de \mathcal{K}_j^{pr} y su versión reconstruida mediante Transformada Wavelet (ver Sección 3.3.3).	47
4.6.	Predicciones de 1 a 4 pasos considerando un 30 % del área de cobertura total de la comuna.	48
4.7.	Grillas de estudio para comuna de Puente Alto.	48
4.8.	Curva HRP v/s TIP para 4 pasos de predicción con el respectivo valor AUC y un ejemplo para 1 paso de predicción mostrando 164 grillas de mayor riesgo (3.9 %) con acierto de 3 de 7 delitos (42.8 %).	49
4.9.	Iteraciones de modelos espaciales <i>a priori</i> y <i>a posteriori</i>	51
4.10.	Criterios multi-objetivo (<i>Silhouette</i> con Aciertos y <i>Silhouette</i> con Desviaciones)	52
4.11.	Relación entre criterio de Desviaciones (Cohesión) y Aciertos para modelo <i>a priori</i> y <i>a posteriori</i> considerando 10 a 300 componentes espaciales correspondiente a los puntos que parten desde el borde inferior izquierdo y terminan en el borde superior derecho.	54
4.12.	Valores de AUC a un paso y promedio de tres pasos para diferentes componentes espaciales de modelo <i>a priori</i> y <i>a posteriori</i>	55
4.13.	Elipse de cada componente de modelo <i>a posteriori</i> graficando cada Gaussiana con 0.9 de probabilidad. El mecanismo se ejecuta considerando $r^2 = 6$ y $\gamma = 0.5$.	56
4.14.	Métricas espaciales para modelo <i>a posteriori</i> de 40 componentes iniciales, utilizando el mecanismo para diferentes γ y diferentes r^2 tal que N_{rest} (cantidad de componentes afectadas por el mecanismo) cambie uno en uno. Obs: Los criterios están normalizados entre 0 y 1.	57
4.15.	Valores de AUC a un paso y promedio de tres pasos para diferentes componentes espaciales de modelo <i>a priori</i> y <i>a posteriori</i>	59
4.16.	Gráfico de calles de la comuna de estudio diferenciando el valor de riesgo de cada una tanto para modelo <i>a priori</i> como <i>a posteriori</i> . Para este ejemplo las calles de importancia se grafican en escala de colores a partir del percentil 80 de riesgo, y las que están bajo este umbral son pintadas azul.	60
4.17.	Curva HRP vs TIP para modelo con geometría de calles para la fecha 2016-05-03.	61
4.18.	Resultados de algoritmo con segmentos de calles para tres pasos de predicción para la fecha 2016-05-03 mostrando el 20 % de los segmentos de mayor riesgo y en puntos rojos los delitos de predicción de la fecha.	62
4.19.	Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2015-01-01 a 2015-05-01.	64
4.20.	Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2016-01-01 a 2016-05-01.	64
4.21.	Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2015-01-01 a 2015-05-01.	65
4.22.	Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2016-01-01 a 2016-05-01.	66

4.23. Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2015-01-01 a 2015-05-01.	67
4.24. Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2016-01-01 a 2016-05-01.	68
4.25. Comunas de estudio de la Región Metropolitana.	69
4.26. AUC's promedio de 1 a 3 pasos de predicción considerando versión estándar , perfecta y estática del algoritmo en el intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio.	71
4.27. Resultados predictivos diarios usando celdas de 200x200[m ²]. En cuanto al método prospectivo los parámetros de kernels son 400 [m] espaciales y 8 [semanas] (figura de la izquierda) y parámetros óptimos (figura de la derecha) donde se considera 200 [m] espaciales y 39 [semanas] temporales. La intensidad de fondo para el método de procesos puntuales corresponde a un ancho de banda de 300 [metros] (figura de la izquierda) seleccionado mediante validación cruzada y 130 [m] (figura de la derecha) seleccionado para optimizar el número de eventos criminales a predecir. Fuente: Trabajo de Mohler <i>et al.</i> [10]	73
4.28. AUC para eventos criminales de Charlottesville desde Marzo 2014 a Febrero 2005 (izquierda). Curva HRP/TIP para mes de Julio 2004 (derecha). En estos resultados se compara los algoritmos propuestos por Wang <i>et al.</i> consistentes en modelos aditivos GAM (línea verde y azul) con el típico modelo mediante Hot-spots (línea negra) y otro llamado <i>Generalized Linear Models</i> (GLM) (línea roja). Fuente: Trabajo de Wang <i>et al.</i> [4]	74
5.1. <i>Framework</i> de entrenamiento sistemático de parámetros.	80
5.2. AUC's promedio de 1 a 3 pasos de predicción para intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio.	88
5.3. AUC's promedio de 1 a 3 pasos de predicción para intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio.	89
5.4. AUC's promedio de 1 a 3 pasos de predicción para intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio considerando las tres versiones del algoritmo.	90
5.5. AUC's promedio de 1 a 3 pasos de predicción para intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio considerando las tres versiones del algoritmo.	91