

Taller 4: Redes bayesianas para pronosticar condiciones climáticas severas en Colorado

Franco Cornejo Gonzalez (4 horas)

Departamento de Ingeniería Informática

Universidad de Santiago de Chile, Santiago, Chile

franco.cornejo.g@usach.cl

Resumen—Este trabajo presenta un modelo de Redes Bayesianas (RB) para pronosticar condiciones climáticas severas en Colorado, utilizando el dataset Hailfinder. Se emplearon métodos como Hill-Climbing y Max-Min Hill-Climbing, integrando conocimiento experto mediante listas negras y blancas. El modelo ajustado incluye 54 nodos y 67 aristas, capturando relaciones clave entre viento, humedad, inestabilidad y nubes. Los resultados destacan que la inestabilidad (0.29) y la cobertura de nubes (0.27) son los factores que por separado son más determinantes en la predicción de eventos severos.

Palabras claves—Redes Bayesianas, Predicción Climática, Hailfinder.

1. INTRODUCCIÓN

El estado de Colorado juega un papel crucial en los fenómenos climáticos de los Estados Unidos ya que esta citado en el corazón del país. Colindando por el Oeste con montañas rocosas y dando inicio a las llanuras por el Este. Esta caracterización singular lo sitúa como el punto de inicio de múltiples eventos climáticos severos, tales como, tornados, granizos y ciclones. Que nacen en Colorado y se propagan hacia el Este del país.

La comprensión de las condiciones que dan lugar a los eventos climáticos severos es esencial para prevenir catástrofes y mitigar sus impactos en las comunidades. Los científicos han centrado sus esfuerzos en modelar y comprender los patrones climáticos de Colorado, aprovechando una rica base de datos meteorológicos recopilados mediante estaciones terrestres, radares y observaciones satelitales. Estos datos, permiten la aplicación de minería de datos para comprender mejor el clima e incluso lograr su predicción.

En este contexto, las redes bayesianas emergen como una herramienta poderosa para modelar y predecir condiciones meteorológicas severas. Estas redes permiten integrar datos observados con conocimiento experto, proporcionando una representación probabilística de las relaciones entre variables climáticas. Este enfoque no solo sirve para hacer predicciones como otros métodos, sino que también permite realizar análisis detallados sobre cómo las diferentes variables influyen entre sí desembocando en la probabilidad de eventos severos.

El presente trabajo tiene como objetivo construir y analizar un modelo basado en redes bayesianas utilizando el reconocido dataset de Hailfinder. El propósito es doble: por

un lado, explorar la capacidad predictiva del modelo para identificar eventos severos; por otro, realizar consultas sobre cómo variables específicas, como el viento, la humedad y la cobertura de nubes, afectan la probabilidad de tormentas severas.

2. METODO Y DATOS

2.1. Descripción del Dataset

El dataset utilizado en este trabajo corresponde a Hailfinder. Este conjunto de datos consta de 56 variables que representan distintos aspectos del clima, incluyendo viento, nubes, humedad, estabilidad, inestabilidad y convección, además de variables de salida que describen el pronóstico regional. Todas estas variables ya se encuentran discretizadas. A continuación, se presenta una descripción de las principales categorías de variables.

2.1.1. Viento: El viento, tanto en movimientos verticales como horizontales, es crucial para la formación de tormentas y convección. Las variables relacionadas y sus posibles estados son:

- **N07muVerMo:** Movimiento vertical (StrongUp, WeakUp, Neutral, Down).
- **SubjVertMo:** Juicio subjetivo del movimiento vertical (StrongUp, WeakUp, Neutral, Down).
- **QGVertMotion:** Movimiento cuasi-geostrófico (StrongUp, WeakUp, Neutral, Down).
- **CombVerMo:** Combinación de movimientos verticales (StrongUp, WeakUp, Neutral, Down).
- **WndHodograph:** Perfil de viento con altitud (DCVZFavor, StrongWest, Westerly, Other).

2.1.2. Nubes: La cobertura y las sombras de las nubes afectan significativamente los fenómenos climáticos al modificar la radiación solar y la convección. Las variables relacionadas son:

- **VISCloudCov:** Cobertura de nubes visibles (Cloudy, PC, Clear).
- **IRCloudCover:** Cobertura de nubes infrarrojas (Cloudy, PC, Clear).
- **CombClouds:** Cobertura combinada (Cloudy, PC, Clear).
- **CldShadeOth:** Sombras de nubes (Cloudy, PC, Clear).
- **CldShadeConv:** Sombras asociadas a la convección (None, Some, Marked).

2.1.3. *Humedad*: La humedad es un factor determinante en la formación de nubes y tormentas. Las variables incluidas son:

- **SatContMoist**: Contenido de humedad saturada (VeryWet, Wet, Neutral, Dry).
- **RaoContMoist**: Humedad según el método de Rao (VeryWet, Wet, Neutral, Dry).
- **CombMoisture**: Humedad combinada (VeryWet, Wet, Neutral, Dry).
- **AreaMoDryAir**: Presencia de aire seco (VeryWet, Wet, Neutral, Dry).

2.1.4. *Estabilidad e Inestabilidad*: Estas variables describen la resistencia de la atmósfera al movimiento vertical del aire:

- **AMInstabMt**: Inestabilidad matutina en montañas (None, Weak, Strong).
- **InsInMt**: Inestabilidad general en montañas (None, Weak, Strong).
- **CapChange**: Cambios en la estabilidad (Decreasing, LittleChange, Increasing).
- **CapInScen**: Estabilidad en escenarios específicos (LessThanAve, Average, MoreThanAve).

2.1.5. *Convección*: La convección, que describe el movimiento ascendente de aire caliente, es fundamental para la formación de tormentas. Las variables relacionadas incluyen:

- **OutflowFrMt**: Flujo descendente desde las montañas (None, Weak, Strong).
- **MorningBound**: Presencia de límites atmosféricos matutinos (None, Weak, Strong).
- **Boundaries**: Límites significativos en la atmósfera (None, Weak, Strong).

2.1.6. *Variables de Salida*: Las variables de salida representan los pronósticos de condiciones severas en diferentes regiones de Colorado:

- **MountainFest**: Pronóstico para la region 1 (XNIL, SIG, SVR).
- **PlainsFest**: Pronóstico la region 2 (XNIL, SIG, SVR).
- **N34StarFest**: Pronóstico combinado para regiones 2, 3 y 4 (XNIL, SIG, SVR).
- **R5Fest**: Pronóstico para la Región 5 (XNIL, SIG, SVR).

Estas 4 variables pueden tomar los valores XNIL, SIG y SVR que respectivamente hacen referencia nulo, significativo y severo. Un pronostico de eventos climáticos como granizo, huracanes y tornados.

2.2. Métodos usado

2.2.1. *Redes Bayesianas*: Las Redes Bayesianas (RB) son modelos Las Redes Bayesianas (RB) son herramientas poderosas para modelar dependencias probabilísticas entre variables de interés, representadas en forma de Grafos Dirigidos Acíclicos (DAG). En estos grafos, los nodos corresponden a variables y las aristas reflejan influencias causales directas entre ellas. Estas relaciones se cuantifican mediante probabilidades condicionales, lo que permite

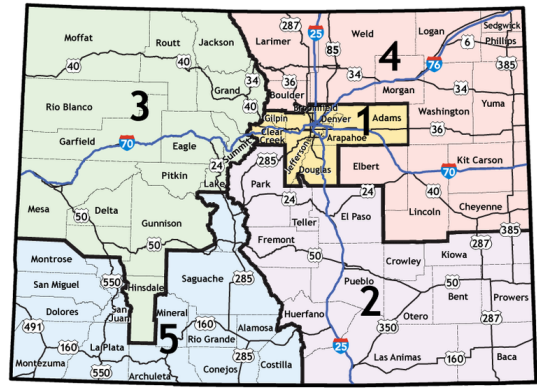


Figura 1: Mapa

integrar conceptos fundamentales de teoría de grafos, estadística bayesiana y optimización.

Una característica destacada de las RB es su capacidad para incorporar el conocimiento experto a través de listas negras (blacklists) y listas blancas (whitelists). Las listas negras definen restricciones que prohíben conexiones específicas entre nodos, mientras que las listas blancas aseguran la existencia de aristas consideradas esenciales por expertos en el dominio. Este enfoque permite construir modelos más coherentes y ajustados a la realidad observada, combinando datos con conocimiento previo.

El propósito de una RB va más allá de modelar dependencias: también incluye resolver problemas de aprendizaje estructural, manejar restricciones de listas negras y blancas, y propagar evidencia a lo largo de la red. Esto se logra mediante algoritmos especializados que permiten tanto la construcción de la estructura como la inferencia, facilitando una comprensión formal e intuitiva de las dependencias entre variables y su impacto en distintos contextos (Chacón and R., 2024).

3. RESULTADOS

Para construir las Redes Bayesianas (RB), se emplearon los métodos de Hill-Climbing (HC), Max-Min Hill-Climbing (mmHC) y Max-Min Parents and Children (mmPC). De los modelos generados, se seleccionó aquel que más se asimiló a las estructuras reportadas en la literatura. Posteriormente, el modelo seleccionado fue ajustado utilizando conocimiento experto, basado en la documentación del dataset.

3.1. Aplicación y comparación de HC, mmHC y mmPC

Método	HC	mmHC	mmPC
Nodos	56	56	56
Arcos	64	34	35
BIC	-990476	-1144947	—

Tabla 1: Comparación de Modelos Generados por Diferentes Métodos

De los tres modelos evaluados, se seleccionó el generado por el método HC debido a su mejor puntaje BIC, lo que refleja una mayor compatibilidad entre la estructura aprendida

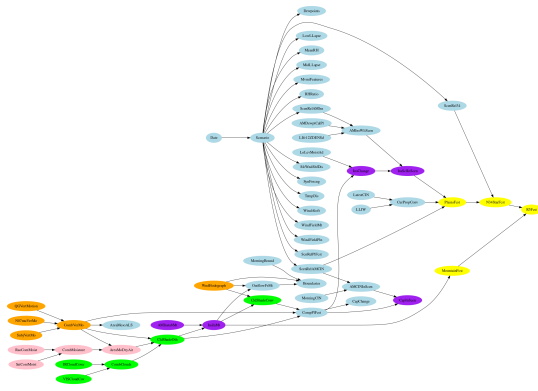


Figura 2: Modelo HC modificado con conocimiento experto

y los datos. Además, este modelo mostró un notable parecido con las estructuras propuestas en la literatura, destacándose como la base ideal para ajustes adicionales.

Posteriormente, se introdujeron modificaciones al modelo HC utilizando listas negras (*blacklists*) y listas blancas (*whitelists*), con el fin de incorporar conocimiento experto y alinear el modelo aún más con lo reportado en estudios previos.

La lista blanca incluyó las conexiones que reflejan dependencias fundamentales, como:

- De **AreaMesoALS** hacia **AreaMoDryAir**, **CldShadeOth** y **CompPIFest**.
- De **ScnRelPIFest**, **CapInScen** y **CurPropConv** hacia **PlainsFest**.

Por otro lado, la lista negra eliminó conexiones inconsistentes con la literatura, como las dependencias directas de las variables de salida (**PlainsFest**, **N34StarFest** y **R5Fest**) hacia variables de entrada como **LatestCIN** y **LLIW**.

Con estas modificaciones, el modelo resultante combina datos, conocimiento experto y teoría, logrando una estructura que refleja tanto la información del dataset como las relaciones propuestas en investigaciones previas. El modelo resultante consta un 54 nodos, 67 aristas y un BIC de -991544 que es ligeramente peor al del modelo sin modificar pero a nuestro entender es mejor ya que incorpora conocimiento de expertos y esto nos da mas confianza que un modelo generado por heurísticas que puede estar sesgado por los datos de entrenamiento.

En la Figura 2 se presenta el modelo final. Para facilitar su interpretación, las variables han sido categorizadas y coloreadas según su tipo: las variables de salida están marcadas en amarillo, las relacionadas con el viento en naranja, las de nubes en verde, las de humedad en rosa y las de inestabilidad en púrpura. Esta representación visual destaca las conexiones clave entre las categorías, ayudando a comprender cómo se relacionan estas variables para predecir eventos climáticos severos.

3.2. Propagación de la Evidencia

Para analizar el impacto de las distintas variables climáticas sobre la predicción de eventos severos, se

seleccionó la variable de salida **MountainFest** cuyo nodo es amarillo y esta citado en la parte más baja de la Figura 2. Esta elección se basó en su relevancia geográfica, ya que esta variable representa las condiciones en las montañas, específicamente en áreas como Denver, que es la región más habitada de Colorado.

Se realizaron consultas compuestas para evaluar la importancia de diferentes categorías de variables sobre la predicción de eventos severos en **MountainFest**. Estas consultas se enfocaron en cuatro grupos en los cuales se agrupan las variables sobre el viento (nodos naranja), la humedad (nodos rosado), la inestabilidad (nodos púrpura) y la cobertura de nubes (nodos verde). Los resultados se resumen en la Tabla 2.

Pregunta	"MountainFest"="Severo"
Vientos Fuertes	0.16
Muy Alta Humedad	0.16
Alta Inestabilidad	0.29
Alta Cobertura de Nubes	0.27

Tabla 2: Impacto de las Variables sobre la Predicción de Eventos Severos en MountainFest

Los resultados muestran que ninguna categoría por sí sola supera una probabilidad de 0.3 para predecir eventos severos. Sin embargo, se observa que las variables relacionadas con la inestabilidad y la cobertura de nubes tienen un mayor impacto (0.29 y 0.27, respectivamente) en comparación con las variables de viento y humedad (ambas con 0.16). Esto es consistente con lo reportado en la literatura, donde se señala que la inestabilidad atmosférica y la formación de nubes profundas son factores clave para desencadenar fenómenos severos como tormentas eléctricas y granizo (Abramson, Brown, Edwards, Murphy, and Winkler, 1996).

Estos resultados muestran la importancia de evaluar las interacciones entre categorías de variables y estudiar sus posibles sinergias, ya que al evaluar solo el impacto de un tipo de variables no se logra ver un gran impacto sobre la variable MountainFest. Además en el estudio igual se hicieron pruebas probando configuraciones mixtas de variables sin lograr mayor impacto sobre las variables de salida.

Finalmente se evaluó el modelo como clasificador contra un conjunto de test y para medir la precisión en la predicción de eventos climáticos severos en MountainFest dando como resultado un 60 % de aciertos.

4. CONCLUSIONES

En este trabajo se logró construir un modelo de RB bastante similar a los reportadas en la literatura, logrando una estructura cercana al modelo propuesto originalmente para el dataset de Hailfinder. Este enfoque permitió identificar y comprender la relevancia de ciertas variables sobre la predicción de eventos climáticos severos, como la inestabilidad atmosférica, la cobertura de nubes, el viento y la humedad.

El modelo ajustado fue evaluado como clasificador, obteniendo una precisión del 60 %. Este resultado refleja el

desafío inherente al modelado de fenómenos climáticos complejos, donde múltiples factores interactúan sinérgicamente.

La incorporación de listas negras y blancas permitió mejorar la interpretabilidad del modelo, destacando la importancia de integrar conocimiento experto para capturar relaciones que podrían no ser evidentes en los datos.

Finalmente, este trabajo muestra la dificultades de modelar fenómenos climáticos y hacer predicciones sobre esto.

REFERENCIAS

- [1] M. Chacón and F. A. B. R., “Capítulo iv: Redes bayesianas,” in *Taller de minería de datos avanzada PPT*, 2024.
- [2] B. Abramson, J. Brown, W. Edwards, A. Murphy, and R. L. Winkler, “Hailfinder: A bayesian system for forecasting severe weather,” *International Journal of Forecasting*, vol. 12, no. 1, pp. 57–71, 1996, probability Judgmental Forecasting. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0169207095006648>

ANEXO

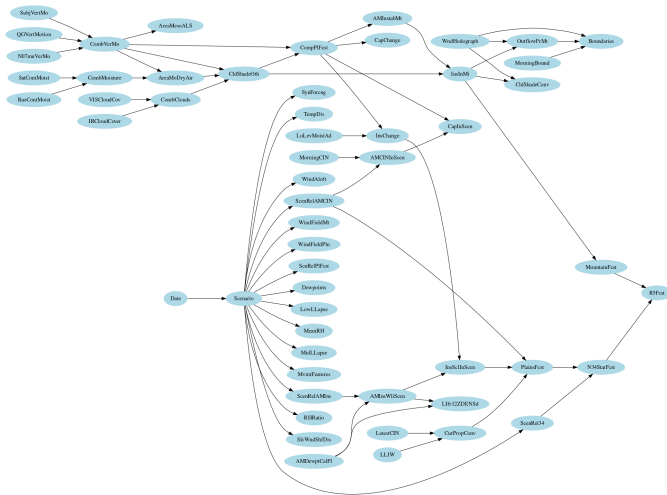


Figura 3: Modelo HC sin modificar



Figura 4: Modelo mmHC



Figura 5: Modelo mmPC