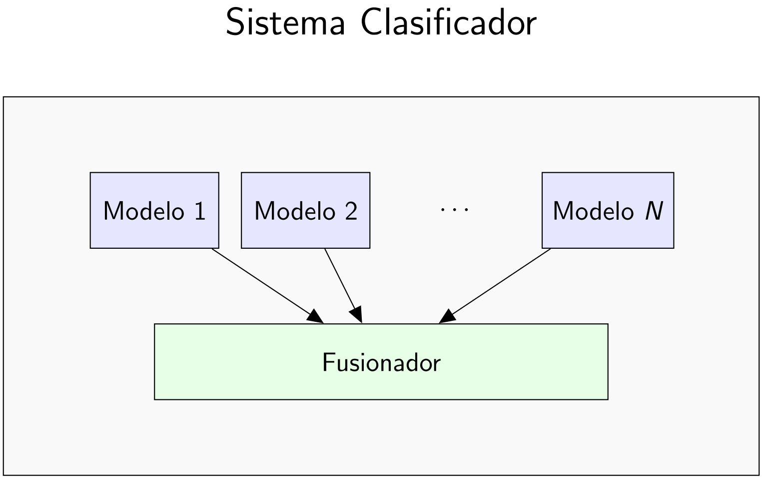
Método de evaluación para modelo integrante de un sistema clasificador (Análisis de peor caso)

Introducción

Se tiene un sistema clasificador con distintos modelos en paralelo que pueden ver features distintas de los datos. Luego se fusionan sus resultados mediante algún método, resultando en un único resultado del sistema. Está enfocado particularmente en clasificación binaria.

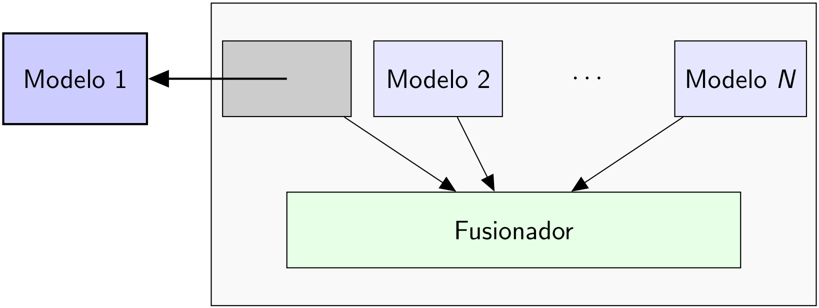
Introducción



Introducción

Se busca evaluar un modelo para ver como este impacta en el rendimiento del sistema. En otras palabras, se quiere poder evaluar que tan bien se espera que funcione el sistema dado que cierto modelo sea usado.

Introducción



Información disponible

Datos de los que se dispone para el problema:

- El sistema incompleto, sin el modelo
- Datos para la evaluación del sistema
- Una función que asocia las confusiones posibles del sistema a costos
- El modelo que se pretende evaluar
- Datos para la evaluación del modelo

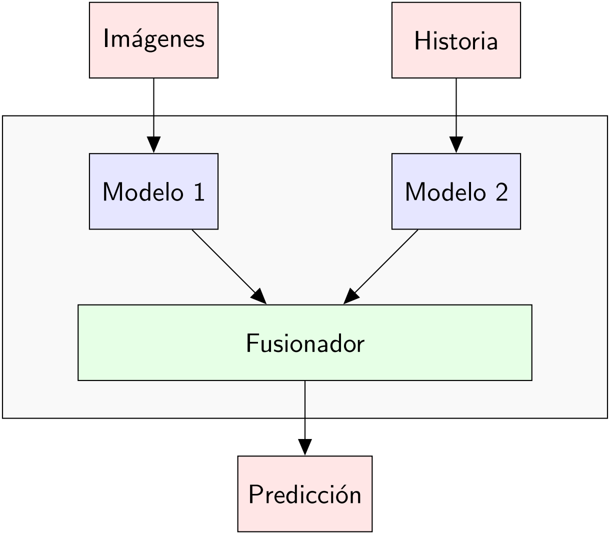
Problema

Para evaluar el modelo, una opción es correr el sistema con el modelo y ver que resultados se obtienen. Para esto es necesario que quien tenga el modelo, también tenga acceso al sistema. Incluso aunque eso fuera posible, se podría querer evitar la ejecución del sistema completo para cada modelo posible que se busque evaluar. El objetivo es proponer un método que sea de utilidad para comparar potenciales modelos sin la necesidad de evaluar el sistema nuevamente para cada uno.

Sistema ejemplo

Se usará como ejemplo el siguiente sistema. Se tienen datos de tipo imágenes médicas y história clínica. El sistema consiste de dos modelos distintos, cada uno con su veredicto que son luego fusionados con algún método. El modelo a evaluar en este ejemplo es el modelo que recibe historias clínicas.

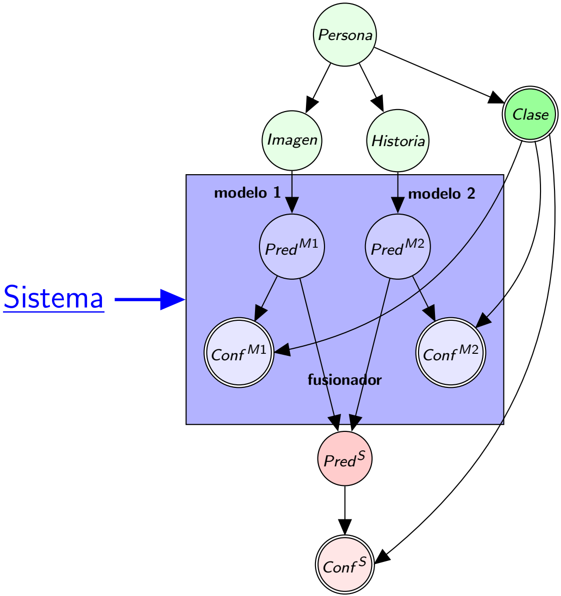
Sistema ejemplo



Redes Bayesianas

La forma que se va a utilizar para modelar el problema es mediante una red bayesiana causal. Una red bayesiana causal es un DAG que se usa para representar distribuciónes probabilisticas sobre variables no independientes. Se tienen nodos para representar entidades, y aristas que representan causalidad entre esas entidades. Los nodos con doble círculo son determinísticos dados los padres.

Redes Bayesianas



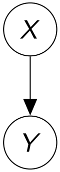
Intervención

En las redes bayesianas existe una operación interesante llamada intervención. Esta operación le asigna a la variable de cierto nodo un valor particular, con probabilidad 1 en todos los casos. Esta operación se representa en una consulta probabilistica con el operador **do**.

Intervención

Un ejemplo del operador **do**:

- $P(Y = y|X = x)$
- $P(Y = y|do(X = x))$



La consulta sin **do** es un condicional normal, donde miramos la probabilidad de $Y = y$ solo para los casos en los que $X = x$.

La consulta con el operador **do** se realiza sobre toda la distribución, asumiendo que X vale x siempre.

Estimación del costo del sistema

Para poder comparar entre modelos se separa el proceso en dos partes:

Primero es evaluado el sistema sin el modelo. Se calculan las probabilidades de error y acierto si el modelo hubiera fallado o acertado siempre, usando la operación de intervención.

Luego al tener esos valores, se ejecuta un modelo en particular y se intenta estimar alguna métrica del sistema asociada a ese modelo.

Estimación del costo del sistema

- Idealmente, se buscaría estimar el resultado de la evaluación del sistema si se hubiera corrido con el modelo. Esto no es posible ya que dados distintos modelos con los mismos resultados en su matriz de confusión, es posible que interactúen de distintas formas con el sistema y llevar a resultados finales muy distintos.

- Es por esto que la métrica de interés va a ser que dada la matriz de confusión de un modelo candidato para un conjunto de instancias, estimar el peor resultado posible de la evaluación del sistema sobre esas instancias.

Notación

Para definir el problema formalmente se introduce la notación $m_{ij}^{f,D}$ para denotar un valor de una matriz de confusión:

$$m_{ij}^{f,D} = |\{d \in D \mid f(d_1) = i \wedge d_2 = j\}|$$

- i es la clase real
- j es la clase que predice el modelo f sobre el dato
- D es el dataset utilizado, conteniendo un conjunto de pares d_1 instancia, d_2 clase real asociada, y su valor por defecto que representa todos los datos disponibles es D
- f es el modelo usado y puede valer:
  - M como el modelo a analizar
  - S como el sistema completo con el modelo
  - $S_{*\to x}$ como el sistema con un modelo intervenido que predice de forma constante la clase x para cualquier instancia

Especificación del problema

Se espera lo siguiente sobre el comportamiento del sistema para cualquier conjunto de instancias D:

- $\{d \in D \mid d_2 = 0 \wedge S(d_1) = 0\} \subseteq$
  $\{d \in D \mid d_2 = 0 \wedge S_{* \to p}(d_1) = 0\}$

$\backslash\backslash$ Agregar evidencia solo puede mejorar la predicción

Especificación del problema

Se espera lo siguiente sobre el comportamiento del sistema para cualquier conjunto de instancias D:

- $\{d \in D \mid d_2 = 0 \wedge S(d_1) = 1\} \subseteq \{d \in D \mid d_2 = 0 \wedge S_{*\to n}(d_1) = 1\}$

\\\\ Restar evidencia solo puede empeorar la predicción

Especificación del problema

Se espera lo siguiente sobre el comportamiento del sistema para cualquier conjunto de instancias D:

- $\{d \in D \mid d_2 = 0 \wedge S(d_1) = 0\} \subseteq$
  $\{d \in D \mid d_2 = 0 \wedge S_{*\to p}(d_1) = 0\}$
- $\{d \in D \mid d_2 = 0 \wedge S(d_1) = 1\} \subseteq$
  $\{d \in D \mid d_2 = 0 \wedge S_{*\to n}(d_1) = 1\}$
- $\{d \in D \mid d_2 = 1 \wedge S(d_1) = 1\} \subseteq$
  $\{d \in D \mid d_2 = 1 \wedge S_{*\to n}(d_1) = 1\}$
- $\{d \in D \mid d_2 = 1 \wedge S(d_1) = 0\} \subseteq$
  $\{d \in D \mid d_2 = 1 \wedge S_{*\to p}(d_1) = 0\}$

\\ Lo mismso para Clase Negative

Especificación del problema

Se espera lo siguiente sobre la función de costo asociada a las confusiones del sistema:

- $Costo^S_{00} \leq Costo^S_{01}$
- $Costo^S_{11} \leq Costo^S_{10}$

Los fallos tienen mayor costo que los aciertos.

Para este trabajo se va a asumir que los aciertos tienen costo 1 y los fallos costo 0.

Especificación del problema

Datos disponibles:

	Pred Pos	Pred Neg
Clase Pos	$m_{0,0}^{S_{*\to p},D}$	$m_{0,1}^{S_{*\to p},D}$
Clase Neg	$m_{1,0}^{S_{*\to p},D}$	$m_{1,1}^{S_{*\to p},D}$

Table: Matriz de confusión del sistema si Modelo = Positive

	Pred Pos	Pred Neg
Clase Pos	$m_{0,0}^{S_{*\to n},D}$	$m_{0,1}^{S_{*\to n},D}$
Clase Neg	$m_{1,0}^{S_{*\to n},D}$	$m_{1,1}^{S_{*\to n},D}$

Table: Matriz de confusión del sistema si Modelo = Negative

Especificación del problema

Otros datos disponibles:

	Positive	Negative
#Instancias	P	N

Table: Cantidad de instancias por clase

	Pred Pos	Pred Neg
Clase Pos	$m_{0,0}^{M,D}$	$m_{0,1}^{M,D}$
Clase Neg	$m_{1,0}^{M,D}$	$m_{1,1}^{M,D}$

Table: Matriz de confusión del modelo

Especificación del problema

Dados esos datos, se busca generar la peor matriz de confusión posible del sistema que sea válida (peor caso).

	Prediccion Positive	Prediccion Negative
Clase Positive	$m_{0,0}^{S,D}$	$m_{0,1}^{S,D}$
Clase Negative	$m_{1,0}^{S,D}$	$m_{1,1}^{S,D}$

Table: Matriz de confusión del sistema

Especificación del problema

La peor es la que maximiza:

- $m_{0,0}^{S,D} * Costo_{00}^{S} + m_{0,1}^{S,D} * Costo_{01}^{S} + m_{1,0}^{S,D} * Costo_{10}^{S} + m_{1,1}^{S,D} * Costo_{11}^{S}$

Como los costos asociados a las confusiones de aciertos valen 1 y las de los fallos 0, es equivalente a maximizar:

- $m_{0,1}^{S,D} + m_{1,0}^{S,D}$

El análisis también vale para otros casos.

Planteo inicial

El problema se va a trabajar como un problema de búsqueda.

- El espacio de búsqueda van a ser todas las posibles combinaciones de resultados de clasificación en el sistema sobre las instancias disponibles.
- El objetivo va a ser hallar los conjuntos de resultados donde la cantidad de fallos sea máxima y se cumpla la especificación planteada

Comportamiento de instancias

Para lograr esto se va a analizar de forma más cercana los comportamientos posibles de las instancias.

La siguiente tabla tiene los posibles casos en los que se puede encontrar una instancia con respecto al resultado del sistema en ella en base al resultado del modelo. Se marca con una X los casos donde dado cierto resultado del modelo, el sistema falla para esa instancia.

Comportamiento de instancias

Recordando la assumption 2 de la clase Positive:

- $\{d \in D \mid d_2 = 0 \wedge S(d_1) = 0\} \subseteq \{d \in D \mid d_2 = 0 \wedge S_{*\to p}(d_1) = 0\}$

	$m_{01}^{S_{*\to p},i}$	$m_{01}^{S_{*\to n},i}$	Es Válido
Instancia 1			
Instancia 2	X		
Instancia 3		X	
Instancia 4	X	X	

Table: Resultados posibles en instancias de clase Positive

Comportamiento de instancias

Recordando la assumption 2 de la clase Positive:

- $\{d \in D \mid d_2 = 0 \wedge S(d_1) = 0\} \subseteq \{d \in D \mid d_2 = 0 \wedge S_{*\to p}(d_1) = 0\}$

	$m_{01}^{S_{*\to p},i}$	$m_{01}^{S_{*\to n},i}$	Es Válido
Instancia 1			SI
Instancia 2	X		
Instancia 3		X	
Instancia 4	X	X	

Table: Resultados posibles en instancias de clase Positive

Comportamiento de instancias

Recordando la assumption 2 de la clase Positive:

- $\{d \in D \mid d_2 = 0 \wedge S(d_1) = 0\} \subseteq \{d \in D \mid d_2 = 0 \wedge S_{*\to p}(d_1) = 0\}$

	$m_{01}^{S_{*\to p},i}$	$m_{01}^{S_{*\to n},i}$	Es Válido
Instancia 1			SI
Instancia 2	X		NO
Instancia 3		X	
Instancia 4	X	X	

Table: Resultados posibles en instancias de clase Positive

Comportamiento de instancias

Recordando la assumption 2 de la clase Positive:

- $\{d \in D \mid d_2 = 0 \wedge S(d_1) = 0\} \subseteq \{d \in D \mid d_2 = 0 \wedge S_{*\to p}(d_1) = 0\}$

	$m_{01}^{S_{*\to p},i}$	$m_{01}^{S_{*\to n},i}$	Es Válido
Instancia 1			SI
Instancia 2	X		NO
Instancia 3		X	SI
Instancia 4	X	X	

Table: Resultados posibles en instancias de clase Positive

Comportamiento de instancias

Recordando la assumption 2 de la clase Positive:

- $\{d \in D \mid d_2 = 0 \land S(d_1) = 0\} \subseteq \{d \in D \mid d_2 = 0 \land S_{* \to p}(d_1) = 0\}$

	$m_{01}^{S_{*\to p},i}$	$m_{01}^{S_{*\to n},i}$	Es Válido
Instancia 1			SI
Instancia 2	X		NO
Instancia 3		X	SI
Instancia 4	X	X	SI

Table: Resultados posibles en instancias de clase Positive

Ejemplos de resultados

El conjunto de resultados posibles sobre las instancias tiene las combinaciones de estos 3 resultados posibles para cada instancia (ninguna X, una X o dos X).

A su vez, la cantidad total de X (fallas) por columna son los valores calculados con la intervención ($m_{0,1}^{S_{*\to p},D}$ y $m_{0,1}^{S_{*\to n},D}$). Como estos valores fueron calculados y son conocidos, en el conjunto original se descartan todos los casos donde la cantidad de X por columna no sea la adecuada.

Ejemplos de resultados

Ahora sabiendo de que formas se pueden comportar las instancias, se observarán distintas posibles combinaciones de confusiones del sistema para las instancias de clase Positive, usando los resultados del modelo.

Al usar la confusión del modelo se fijan la cantidad de predicciónes Positive y Negative, pero hay muchas combinaciones de reusltados del modelo sobre las instancias que llevan a la misma matriz confusión. Estas variaciones pueden llevan a confusiones del sistema muy distintas entre sí.

Ejemplos de resultados

	$m_{01}^{S_{*\to p},i}$	$m_{01}^{S_{*\to n},i}$	$m_{01}^{S,i}$
Instancia 1			
Instancia 2		X	
Instancia 3			
Instancia 4	X	X	X
Instancia 5		X	X
Instancia 6		X	X
Instancia 7		X	X
Instancia 8		X	X

Table: Ejemplo Clase Positive 1

Cantidad de True Positives totales: 3
Cantidad de False Negatives totales: 5

Ejemplos de resultados

	$m_{01}^{S_{*\to p},i}$	$m_{01}^{S_{*\to n},i}$	$m_{01}^{S,i}$
Instancia 1		X	
Instancia 2		X	
Instancia 3			
Instancia 4	X	X	X
Instancia 5			
Instancia 6		X	X
Instancia 7		X	X
Instancia 8		X	X

Table: Ejemplo Clase Positive 2

Cantidad de True Positives totales: 4
Cantidad de False Negatives totales: 4

Ejemplos de resultados

	$m_{01}^{S_{*\to p},i}$	$m_{01}^{S_{*\to n},i}$	$m_{01}^{S,i}$
Instancia 1		X	
Instancia 2		X	
Instancia 3		X	
Instancia 4			
Instancia 5		X	X
Instancia 6		X	X
Instancia 7			
Instancia 8	X	X	X

Table: Ejemplo Clase Positive 3

Cantidad de True Positives totales: 5
Cantidad de False Negatives totales: 3

Ejemplos de resultados

	$m_{01}^{S_{*\rightarrow p},i}$	$m_{01}^{S_{*\rightarrow n},i}$	$m_{01}^{S,i}$
Instancia 1		X	
Instancia 2		X	
Instancia 3		X	
Instancia 4		X	
Instancia 5			
Instancia 6		X	X
Instancia 7			
Instancia 8	X	X	X

Table: Ejemplo Clase Positive 4

Cantidad de True Positives totales: 6
Cantidad de False Negatives totales: 2

Ejemplos de resultados

Se puede observar como para 4 modelos ejemplo con la misma matriz de confusión, se obtienen 4 resultados distintos en el sistema.

La causa de esto son las instancias con una sola X. Sobre esas instancias, si el modelo predice Positive, el resultado del sistema es un True Positive y si el modelo predice Negative, el resultado del sistema es un False Negative.

Podar espacio

Ahora se busca descartar subconjuntos de resultados del sistema que tengan menos False Negatives que otros.

El objetivo es encontrar dentro del espacio de búsqueda, con combinaciónes de resultados del sistema para las instancias en una ejecución del modelo particular, la peor combinación o una dentro de las peores.

Podar espacio

Las confusiones del sistema se pueden particionar de la siguiente manera:

- $m_{0,0}^{S,D} = m_{0,0}^{S,D_{*\to p}} + m_{0,0}^{S,D_{*\to n}}$
- $m_{0,1}^{S,D} = m_{0,1}^{S,D_{*\to p}} + m_{0,1}^{S,D_{*\to n}}$

Con el objetivo de separar en casos por resultado del modelo y reducir el espacio de busqueda analizando casos donde el modelo acierta y el modelo falla por separado.

Podar espacio: Aciertos

Primero, analizando solo las instancias que fueron aciertos en el modelo, se descartan resultados donde la cantidad de errores no sea máxima. Las instancias que generan Falsos Negativos cuando el modelo acierta son las que en los ejemplos anteriores tienen doble X. Se quiere tener la cantidad máxima posible de esas como aciertos del modelo.

Podar espacio: Aciertos

Ejemplos de casos descartados:

	$m_{01}^{S_{*\rightarrow p},i}$	$m_{01}^{S_{*\rightarrow n},i}$	$m_{01}^{S,i}$
Instancia 1			
Instancia 2		X	
Instancia 3			
Instancia 4	X	X	X
Instancia 5		X	X
Instancia 6	X	X	X
Instancia 7			
Instancia 8		X	X

Table: Ejemplo Clase Positive descartado 1

Podar espacio: Aciertos

Ejemplos de casos descartados:

	$m_{01}^{S_{*\to p},i}$	$m_{01}^{S_{*\to n},i}$	$m_{01}^{S,i}$
Instancia 1			
Instancia 2		X	
Instancia 3			
Instancia 4		X	
Instancia 5		X	X
Instancia 6	X	X	X
Instancia 7			
Instancia 8	X	X	X

Table: Ejemplo Clase Positive descartado 2

Podar espacio: Aciertos

Ejemplos de casos NO descartados:

	$m_{01}^{S_{*\rightarrow p},i}$	$m_{01}^{S_{*\rightarrow n},i}$	$m_{01}^{S,i}$
Instancia 1		X	
Instancia 2	X	X	X
Instancia 3		X	
Instancia 4	X	X	X
Instancia 5			
Instancia 6			
Instancia 7			
Instancia 8		X	X

Table: Ejemplo Clase Positive no descartado 1

Podar espacio: Aciertos

Ejemplos de casos NO descartados:

	$m_{01}^{S_{*\rightarrow p},i}$	$m_{01}^{S_{*\rightarrow n},i}$	$m_{01}^{S,i}$
Instancia 1			
Instancia 2	X	X	X
Instancia 3			
Instancia 4	X	X	X
Instancia 5		X	X
Instancia 6		X	X
Instancia 7			
Instancia 8		X	X

Table: Ejemplo Clase Positive no descartado 2

Podar espacio: Aciertos

Si una de estas instancias con doble X considerada como acierto fuera en cambio un fallo, tiene que cambiar de lugar con otra para conservar la cantidad de X adecuada. Esa otra puede ser otra doble X, una sola o ninguna. En todos los casos la cantidad de fallos del sistema es igual o menor a la original.

La confusión resultante de fallos en el sistema sobre la clase Positive para estos casos donde el modelo acierta es única y es la siguiente:

- $m_{0,1}^{S,D_{*\to p}} = \mathrm{Min}(m_{0,0}^{M,D}, m_{0,1}^{S_{*\to p},D})$

Podar espacio: Fallos

Luego queda analizar los fallos. Las instancias con una X provocan un fallo en el sistema **solo** si el modelo falla, por lo que se busca que sea máxima. La poda descarta los casos en los que la cantidad de instancias con X única no sea máxima para los fallos del modelo.

Podar espacio: Fallos

Ejemplos de casos descartados:

	$m_{01}^{S_{*\to p},i}$	$m_{01}^{S_{*\to n},i}$	$m_{01}^{S,i}$
Instancia 1			
Instancia 2	X	X	X
Instancia 3			
Instancia 4	X	X	X
Instancia 5		X	X
Instancia 6		X	X
Instancia 7			
Instancia 8		X	X

Table: Ejemplo Clase Positive descartado

Podar espacio: Fallos

Ejemplos de casos NO descartados:

	$m_{01}^{S_{*\rightarrow p},i}$	$m_{01}^{S_{*\rightarrow n},i}$	$m_{01}^{S,i}$
Instancia 1		X	
Instancia 2	X	X	X
Instancia 3		X	
Instancia 4	X	X	X
Instancia 5			
Instancia 6			
Instancia 7			
Instancia 8		X	X

Table: Ejemplo Clase Positive NO descartado

Podar espacio: Fallos

Si una de estas instancias con X única considerada como fallo fuera en cambio un acierto, tiene que cambiar de lugar con otra. Esa otra puede ser una doble X, una X única o ninguna X. En todos los casos, la cantidad de fallos del sistema es igual o menor a la original.

La confusion resultante de fallos en el sistema sobre la clase Positive para estos casos donde el modelo falla es única y es la siguiente:

- $m_{0,1}^{S,D_{*\to n}} = \text{Min}(m_{0,1}^{M,D},\ m_{0,1}^{S_{*\to n},D}\text{- Min}(m_{0,0}^{M,D},\ m_{0,1}^{S_{*\to p},D}))$

Podar espacio: Clase Negative

El mismo análisis y las mismas podas se pueden hacer para la clase negative. Las confusiones resultantes serían las siguientes:

- $m_{1,0}^{S,D_{*\to n}} = \mathrm{Min}(m_{1,1}^{M,D},\ m_{1,0}^{S_{*\to n},D})$
- $m_{1,0}^{S,D_{*\to p}} = \mathrm{Min}(m_{1,0}^{M,D},\ m_{1,0}^{S_{*\to p},D}$- $\mathrm{Min}(m_{1,1}^{M,D},\ m_{1,0}^{S_{*\to n},D}))$

Confusiones resultantes

Las confusiones finales del sistema sobre todos los datos son:

- $m_{0,1}^{S,D} = \text{Min}(m_{0,0}^{M,D},\ m_{0,1}^{S_{*\to p},D}) +$
  $\text{Min}(m_{0,1}^{M,D},\ m_{0,1}^{S_{*\to n},D}\text{- Min}(m_{0,0}^{M,D},\ m_{0,1}^{S_{*\to p},D}))$
- $m_{0,0}^{S,D} =$ Instancias restantes de clase Positive
- $m_{1,0}^{S,D} = \text{Min}(m_{1,1}^{M,D},\ m_{1,0}^{S_{*\to n},D}) +$
  $\text{Min}(m_{1,0}^{M,D},\ m_{1,0}^{S_{*\to p},D}\text{- Min}(m_{1,1}^{M,D},\ m_{1,0}^{S_{*\to n},D}))$
- $m_{1,1}^{S,D} =$ Instancias restantes de clase Negative

Datasets distintos para evaluación

Hasta ahora se trabajó asumiendo que los datos en los que se evalúan las intervenciónes sobre el sistema, y los datos con los que se evalúa el modelo son los mismos, y se logró encontrar una función de costo de peor caso para el sistema. Originalmente se buscaba conseguir una función de costo que fuera útil también para casos donde los datos en los que se evalúa el sistema intervenido y los datos con los que se evalúa el modelo fueran distintos.

Datasets distintos para evaluación

- La función de costo de peor caso solo utiliza el accuracy de los resultados del sistema con intervenido. Mientras ese accuracy medido sea el mismo que el accuracy que tendría el sistema (intervenido) en los datos con los que se evalúa el modelo, la función sigue calculando el peor caso correctamente.

- Por ley de los grandes números, si la predicción del sistema en una instancia cualquiera es una variable aleatoria iid con media en el accuracy, mientras la cantidad de datos de cada conjunto de evaluación sea lo suficientemente grande, el accuracy de cada conjunto de evaluación va a tender a la média poblacional (accuracy real) y por lo tanto la medición del primer conjunto de datos con intervención va a ser la misma que el accuracy efectivo para el segundo conjunto de datos.

Experimentación

Se evaluó la función de costo para el modelo ejemplo anterior, donde el sistema recibe como entrada imágenes para el modelo 1 y história clínica para el modelo 2. Luego un fusionador recibe los resultados de clasificación de ambos modelos y en base a alguna operación lógica determina el resultado final del sistema (OR o AND).
Este sistema fue simulado con variables aleatorias Bernoulli para cada instancia y cada clase. Se tiene una variable por par de modelo-clase y con correlación entre variables representantes de distintos modelos para la misma clase. Esto es para simular como los modelos pueden tener resultados no independientes dada una instancia.

Experimentación

La experimentación fue separada en dos etapas.

La primera etapa compuesta por una evaluación inicial del sistema sin el modelo de histórias clínicas. Aquí son estimadas las precisiones del sistema al intervenir en los resultados del modelo para cada instancia:

- $m_{0,0}^{S_{*\to p},D}$
- $m_{0,1}^{S_{*\to p},D}$
- $m_{1,0}^{S_{*\to p},D}$
- $m_{1,1}^{S_{*\to p},D}$
- $m_{0,0}^{S_{*\to n},D}$
- $m_{0,1}^{S_{*\to n},D}$
- $m_{1,0}^{S_{*\to n},D}$
- $m_{1,1}^{S_{*\to n},D}$

Experimentación

La segunda etapa corre el modelo en un nuevo conjunto de datos distinto, generado con la misma distribución que el anterior y calcula las confusiónes de cada resultado. Luego evalúa la función de costo de peor caso para ese conjunto de datos.
Por otro lado, simula lo que hubiera pasado realmente con el sistema si esos resultados del modelo fuesen usados.

Experimentación

Al correr el experimento se pudo ver que la función de peor caso, al ser usada en sistemas simulados con una muy mala interacción con el modelo en cuestión, puede llegar a subestimar el costo del sistema. Esto se debe a que cuando el sistema es evaluado sobre el nuevo conjunto de datos, este puede funcionar peor que lo anteriormente medido en la etapa 1 de los experimentos y que luego los cálculos no sean del todo precisos.

Experimentación

Al ver esto se corrieron dos experimentos distintos:

Uno simulando un modelo con valores fijos de precision en ambos modelos y con una relación entre los resultados del modelo y del sistema bastante desfavorable. Su objetivo es analizar la probabilidad de que la estimación del peor caso de la función de costo sea realmente peor que la ejecución del sistema real y como esta probabilidad se comporta al cambiar la cantidad de instancias de testing de ambas fases del experimento.

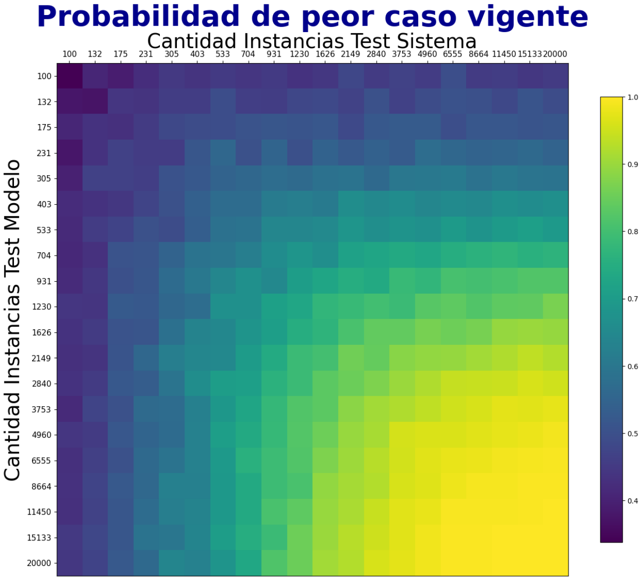
- $\mathrm{P}(EvalPeorCaso^S > m_{0,0}^{S,D} * Costo_{00}^S + m_{0,1}^{S,D} * Costo_{01}^S + m_{1,0}^{S,D} * Costo_{10}^S + m_{1,1}^{S,D} * Costo_{11}^S)$

Experimentación

El otro simulando un sistema con precisiones de ambos modelos y con una interacción entre el sistema y el modelo aleatorias, cuyo objetivo es evaluar que tanto se llega a acercar la evaluación real del sistema a la estimación de peor caso.

- $$\frac{Costo^S_{00} * m^{S,D}_{0,0} + Costo^S_{01} * m^{S,D}_{0,1} + Costo^S_{11} * m^{S,D}_{1,1} + Costo^S_{10} * m^{S,D}_{1,0}}{EvalPeorCaso^S}$$

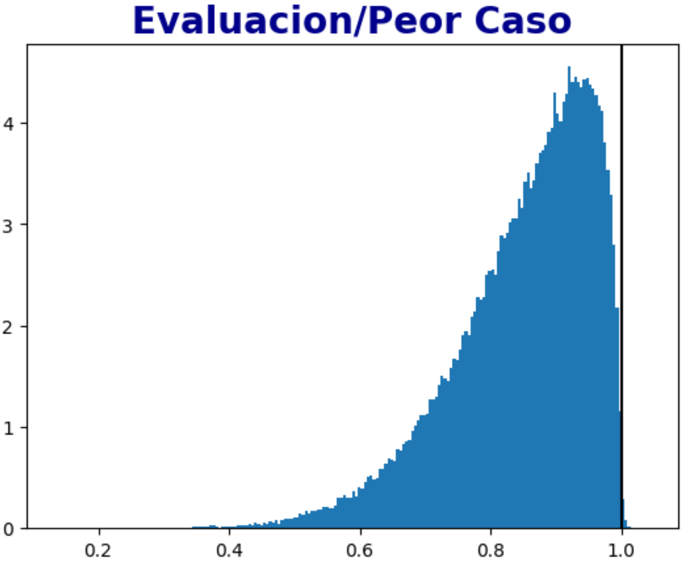
Resultados


Probabilidad de peor caso vigente
Cantidad Instancias Test Sistema
100 132 175 231 305 403 533 704 931 1230 1626 2149 2840 3753 4960 6555 8664 11450 15133 20000
Cantidad Instancias Test Modelo
100 132 175 231 305 403 533 704 931 1230 1626 2149 2840 3753 4960 6555 8664 11450 15133 20000
1.0
0.9
0.8
0.7
0.6
0.5
0.4

Resultados

En el gráfico anterior se ven los resultados del primer experimento, donde se grafica la probabilidad de que el peor caso sea efectivamente mayor a la evaluación del sistema. El sistema es uno compuesto por ambos modelos con accuracies de 50% para ambas clases, con correlación de 0.95 para instancias de clase Positive y -0.95 de clase Negative y un fusionador de tipo OR. Se puede apreciar como la probabilidad se acerca rápidamente a 1 a medida que ambas cantidades de instancias aumentan. Se ejecutó 1000 veces para cada par de cantidades de instancias.

Resultados



Resultados

En el gráfico anterior se ven los resultados del segundo experimento, corrido con una cantidad de instancias fija en 20000 para cada experimento. Se corrió el experimento 100000 veces para graficar la distribución de la relación entre ambos valores. Aquí se puede apreciar como la relación entre el costo del sistema y la estimación de peor caso tiene como cota superior el valor 1, salvo algunos valores atípicos que levenemte lo superan.

Preguntas?