

INFORME LABORATORIO 4 ANÁLISIS DE DATOS
TIC-TAC-TOE ENDGAME DATA SET

CARLOS CÁCERES
DANY RUBIANO

Profesor: Felipe Bello
Ayudantes: Bryan Guzmán
Fernanda Lobos

TABLA DE CONTENIDOS

ÍNDICE DE FIGURAS.....	ii
ÍNDICE DE CUADROS	iii
CAPÍTULO 1. INTRODUCCIÓN.....	1
1.1 MOTIVACIÓN Y ANTECEDENTES	1
1.2 OBJETIVOS	1
1.3 ORGANIZACIÓN DEL DOCUMENTO	1
CAPÍTULO 2. MARCO TEÓRICO.....	3
2.1 REGLAS DE ASOCIACIÓN	3
2.2 MEDIDAS DE CALIDAD Y CONFIANZA	3
2.2.1 Soporte	3
2.2.2 Confianza	4
2.3 MONOTONICIDAD	4
2.4 PROPIEDADES DE LAS MEDIDAS	5
2.4.1 Lift	5
2.4.2 Convicción	5
2.4.3 Laplace	5
2.4.4 Ganancia	6
2.4.5 Métrica de Piatetsky-Shapiro (P-S)	6
CAPÍTULO 3. OBTENCIÓN DE REGLAS.....	7
CAPÍTULO 4. ANÁLISIS DE RESULTADOS	15
CAPÍTULO 5. CONCLUSIONES.....	17
CAPÍTULO 6. REFERENCIAS	19
CAPÍTULO 7. ANEXO: CÓDIGO FUENTE EN R.....	21

ÍNDICE DE FIGURAS

3.1	Información de las reglas Positive.	7
3.2	Información de las reglas Positive.	8
3.3	Distribución de las reglas Positive.	8
3.4	Distribución de las reglas Negative.	9
3.5	Distribución de las reglas Positive escogidas.	11
3.6	Distribución de las reglas Negative escogidas.	11
3.7	Graph de las reglas Positive ordenadas.	12
3.8	Graph de las reglas Negative ordenadas.	13

ÍNDICE DE CUADROS

3.1	Reglas obtenidas - Clase Positive	10
3.2	Reglas obtenidas - Clase Negative	10

CAPÍTULO 1. INTRODUCCIÓN

1.1 MOTIVACIÓN Y ANTECEDENTES

En el desarrollo de experiencias anteriores se ha intentado extraer conocimiento del dataSet tic-tac-toe, mediante diferentes técnicas como el clustering y la regresión logística. Con estas, se han obtenido diversos resultados según la técnica aplicada, pero dentro de lo esperado para el problema de este juego en particular. Ahora es el turno de la técnica de las reglas de asociación, la cual busca descubrir hechos que ocurren en común dentro de un determinado conjunto de datos, es entonces, que con esta técnica se espera descubrir más variables que inciden en el resultado del juego, o confirmar lo ya encontrado.

1.2 OBJETIVOS

El objetivo de este laboratorio es obtener conocimiento del dataSet tic-tac-toe mediante el uso de las reglas de asociación. De esta manera, se busca analizar y comparar los resultados con lo desarrollado en experiencias anteriores usando los métodos de clustering y regresión logística.

1.3 ORGANIZACIÓN DEL DOCUMENTO

El presente documento distribuye su contenido de la siguiente forma, en primer lugar se encuentra un capítulo dedicado a un pequeño marco teórico en el cual se incluyen las definiciones de los conceptos y técnicas a utilizar en el desarrollo de la experiencia.

A continuación, se presenta todo lo referente a los resultados y su proceso de obtención. En lo que sigue, se exponen los análisis de los resultados obtenidos en el desarrollo de los capítulos anteriores, y por último, con lo elaborado, se realiza una síntesis total, la cual es presentada en las conclusiones del presente documento.

CAPÍTULO 2. MARCO TEÓRICO

2.1 REGLAS DE ASOCIACIÓN

En minería de datos y aprendizaje automático, las reglas de asociación se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos. Se han investigado ampliamente diversos métodos para aprendizaje de reglas de asociación que han resultado ser muy interesantes para descubrir relaciones entre variables en grandes conjuntos de datos. Piatetsky-Shapiro describe el análisis y la presentación de reglas 'fuertes' descubiertas en bases de datos utilizando diferentes medidas de interés. Basado en el concepto de regla fuerte, Agrawal presentaron un trabajo en el que indicaban las reglas de asociación que descubrirían las relaciones entre los datos recopilados a gran escala en los sistemas de terminales de punto de venta de unos supermercados. Por ejemplo, la siguiente regla:

$$\{cebollas, vegetales\} \Rightarrow \{carne\}$$

Encontrada en los datos de ventas de un supermercado, indicaría que un consumidor que compra cebollas y verdura a la vez, es probable que compre también carne. Esta información se puede utilizar como base para tomar decisiones sobre marketing como precios promocionales para ciertos productos o dónde ubicar éstos dentro del supermercado. Además del ejemplo anterior aplicado al análisis de la cesta de la compra, hoy en día, las reglas de asociación también son de aplicación en otras muchas áreas como el Web mining, la detección de intrusos o la bioinformática.

Según la definición original de Agrawal el problema de minería de reglas de asociación se define como:

- Sea $I = \{i_1, i_2, \dots, i_n\}$ un conjunto de n atributos binarios llamados items.
- Sea $D = \{t_1, t_2, \dots, t_m\}$ un conjunto de transacciones almacenadas en una base de datos.

Cada transacción en D tiene un ID (identificador) único y contiene un subconjunto de items de I . Una regla se define como una implicación de la forma:

$$X \Rightarrow Y$$

Donde:

$$X, Y \subseteq I$$

$$X \cap Y = \emptyset$$

Los conjuntos de items X y Y se denominan respectivamente "antecedente" (o parte izquierda) y "consecuente" (o parte derecha) de la regla.

(Wikipedia, s.f.).

2.2 MEDIDAS DE CALIDAD Y CONFIANZA

2.2.1 Soporte

El soporte de un conjunto A de transacciones $Sop(A)$, se define como el número de transacciones de los atributos de A que toman el valor verdadero.

El soporte de una regla $A \Rightarrow B$, $Sop(A \Rightarrow B)$, es el número de transacciones en el conjunto S tal que A y B son verdaderos simultáneamente.

Para mantener esta cantidad normalizada se usa, en general, como una proporción de las transacciones conjuntas entre A y B y el número total de transacciones del conjunto S (se indica con n). Al valor normalizado se denominará soporte normalizado

$$Sopn(A \Rightarrow B) = Sop(A \Rightarrow B)/n$$

Nótese que este soporte normalizado es la estimación de la probabilidad de la intersección entre A y B (probabilidad de juntura).

$$s = Sopn(A \Rightarrow B) = \hat{p}(A \cap B) = \frac{Sop(A \Rightarrow B)}{n}$$

2.2.2 Confianza

La confianza de la regla $A \Rightarrow B$, en el conjunto S , es la proporción entre el número de casos de A y B que aparecen conjuntamente en S contenido en el número de casos de A .

Esto es :

$$c = Conf(A \Rightarrow B) = \frac{Sop(A \cap B)}{Sop(A)}$$

Al dividir el numerador y denominador por n .

$$c = \frac{Sop(A \cap B)/n}{Sop(A)/n}$$

Sabiendo que: el soporte de A dividido por n es la estimación de la probabilidad de A , y con la definición de probabilidad condicional se puede obtener:

$$\hat{P}(A) = Sop(A)/n = Sopn(A)$$

$$\frac{Sop(A \Rightarrow B)/n}{Sop(A)/n} = \frac{\hat{p}(A \cap B)}{\hat{P}(A)} = \hat{p}(B/A)$$

Por lo tanto, la confianza de la regla $A \Rightarrow B$ representa la probabilidad que se encuentren los productos B en la transacción dado que ésta también contiene los productos del conjunto A .

(Chacón, s.f.).

2.3 MONOTONICIDAD

Para encontrar todas las posibles reglas se requiere un algoritmo que genere todas las posibles combinaciones y realizar una pre-poda de forma tal que al usar las restricciones se detenga la búsqueda para evitar la explosión combinatoria.

El problema del cumplimiento de las restricciones está asociado con la monotonicidad de la restricción, en

función de la especialización.

Si se tienen dos especializaciones del antecedente, se generan dos reglas tales que $|A_1| < |A_2|$ y dos restricciones o medidas $\mathbf{med}(A_i)$ $i = 1, 2$, asociadas a cada una de las reglas.

- Se dice que la medida es monótona si : $med(A_1) \leq med(A_2)$.
- La medida es anti-monótona si : $med(A_1) \geq med(A_2)$.

Para realizar una pre-poda eficiente se requiere usar restricciones monótonas o anti-monótonas. con lo cual se descartan ramas completas en el proceso de especialización.

(Chacón, s.f.).

2.4 PROPIEDADES DE LAS MEDIDAS

Análisis de monotonicidad de las medidas o métricas de calidad

2.4.1 Lift

Es una medida usada en las herramientas de minería de datos producidas por **IBM**.

$$lift = \frac{nconf(A \Rightarrow B)}{Sop(B)}$$

En términos de probabilidades estimadas se sabe que: $Conf(A \Rightarrow B) = p(A \cap B)/P(A)$ y que $P(B) = Sop(B)/n$. por lo tanto:

$$lift(A \Rightarrow B) = p(B \cap A)/P(A)P(B)$$

Lo cual representa una medida de independencia entre A y B . Esto es, lift tendrá su valor mas bajo (1) cuando A y B sean completamente independientes

2.4.2 Convicción

Una medida similar a la anterior, que mantiene la monotonicidad en confianza:

$$conviccion = \frac{n - Sop(B)}{n[1 - conf(A \Rightarrow B)]}$$

Usando estimación de probabilidades, se tiene:

$$conviccion = \frac{P(A)P(\bar{B})}{P(A \cap \bar{B})}$$

Esto también representa la independencia de A y B además de mostrar su monotonicidad en la confianza

2.4.3 Laplace

Es una medida monótona en soporte y confianza:

$$Laplace(A \Rightarrow B) = \frac{Sop(A \Rightarrow B) + 1}{Sop(A) + k}$$

Donde la constante k es un entero mayor que 1. Puesto que la confianza $c = p(B/A) = p(A \cap B)/P(A)$ y $Sop(A \Rightarrow B) = p(A \cap B)$, se puede reemplazar en el denominador $P(A) = Sop(A \Rightarrow B)/c$. Quedando :

$$Laplace = \frac{Sop(A \Rightarrow B) + 1}{Sop(A \Rightarrow B)/c + k}$$

2.4.4 Ganancia

$$Gan(A \Rightarrow B) = Sop(A \Rightarrow B) - \theta Sop(A), \text{ con } 0 < \theta < 1$$

Aplicando el mismo concepto anterior $P(A) = Sop(A \Rightarrow B)/c$ y factorizando por $Sop(A \Rightarrow B)$ se tiene:

$$Gan(A \Rightarrow B) = Sop(A \Rightarrow B)(1 - \theta/c)$$

Si el soporte de la regla se mantiene constante, la ganancia disminuirá al disminuir la confianza, puesto que aumenta la relación θ/c , aumentando el sustrayendo, y la diferencia $(1 - \theta/c)$ disminuirá. Lo cual implica que la medida será monótona en confianza para soporte constante y viceversa.

2.4.5 Métrica de Piatetsky-Shapiro (P-S)

$$P - S(A \Rightarrow B) = Sop(A \Rightarrow B) - \frac{Sop(A)Sop(B)}{n}$$

Esta métrica presenta la misma estructura que la anterior al considerar $\theta = Sop(B)/n$, como la probabilidad del consecuente. Las condiciones de monotonidad se pueden estudiar en forma similar a la ganancia.

(Chacón, s.f.).

CAPÍTULO 3. OBTENCIÓN DE REGLAS

Para la obtención de las diferentes reglas de asociación que permitan obtener conocimiento del problema abordado, se utilizan en R los paquetes "arulez" y "arulesViz", de tal forma que se puedan ocupar todas las directivas que estos ofrecen para obtener los resultados pertinentes.

Con la ayuda de la función "apriori", se pudo obtener diferentes reglas para ciertos parámetros de soporte y confianza. Estos parámetros se definieron según el contexto del dataSet, y en el intento y prueba de diferentes configuraciones, de modo que se pudiera obtener un numero de reglas suficientes que permitieran evaluar las dos clases presentes en el problema, "positive" y "negative", que recordando su representación, corresponden a los resultados finales de cada instancia en base al jugador "x".

En definitiva, se tomó como parámetros mínimos un soporte de 0,1 y una confianza de 0,3, obteniendo así un total de 118 reglas, y con el fin de obtener información de las diferentes variables que inciden en cada clase en particular, se dividió el total de reglas según su clase, obteniendo así 104 reglas para la clase "positive" y 14 reglas para "negative". De esta forma, la información para cada tipo de regla se presenta en la figura 3.1 y la figura 3.2.

```
set of 104 rules

rule length distribution (lhs + rhs):sizes
 1  2  3
 1 27 76

      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.000   2.000   3.000   2.721   3.000   3.000

summary of quality measures:
      support      confidence      lift
Min.   :0.1023  Min.   :0.4353  Min.   :0.6662
1st Qu.:0.1075  1st Qu.:0.6398  1st Qu.:0.9790
Median :0.1315  Median :0.7057  Median :1.0800
Mean   :0.1512  Mean   :0.7093  Mean   :1.0855
3rd Qu.:0.1503  3rd Qu.:0.7812  3rd Qu.:1.1956
Max.   :0.6534  Max.   :0.8372  Max.   :1.2812

mining info:
data ntransactions support confidence
DB          958      0.1      0.3
```

Figura 3.1: Información de las reglas Positive.

```

set of 14 rules

rule length distribution (lhs + rhs):sizes
1 2
1 13

    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.000  2.000   2.000   1.929  2.000   2.000

summary of quality measures:
      support      confidence      lift
Min.   :0.1054  Min.   :0.3061  Min.   :0.8832
1st Qu.:0.1172  1st Qu.:0.3162  1st Qu.:0.9124
Median :0.1524  Median :0.4048  Median :1.1680
Mean   :0.1584  Mean   :0.3927  Mean   :1.1332
3rd Qu.:0.1597  3rd Qu.:0.4358  3rd Qu.:1.2576
Max.   :0.3466  Max.   :0.5647  Max.   :1.6295

mining info:
data ntransactions support confidence
DB          958      0.1      0.3

```

Figura 3.2: Información de las reglas Positive.

La distribución para cada tipo de regla según soporte y confianza, se detalla a continuación:

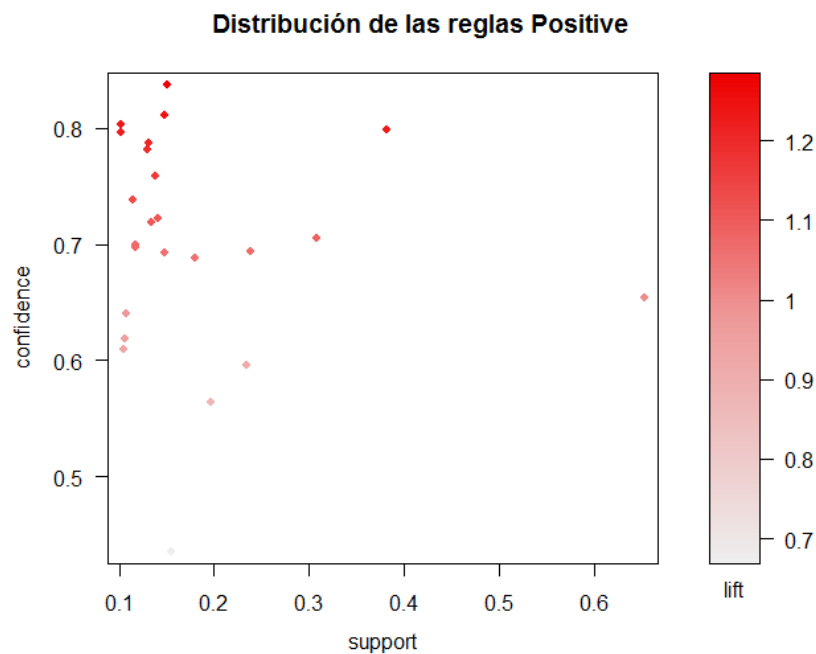


Figura 3.3: Distribución de las reglas Positive.

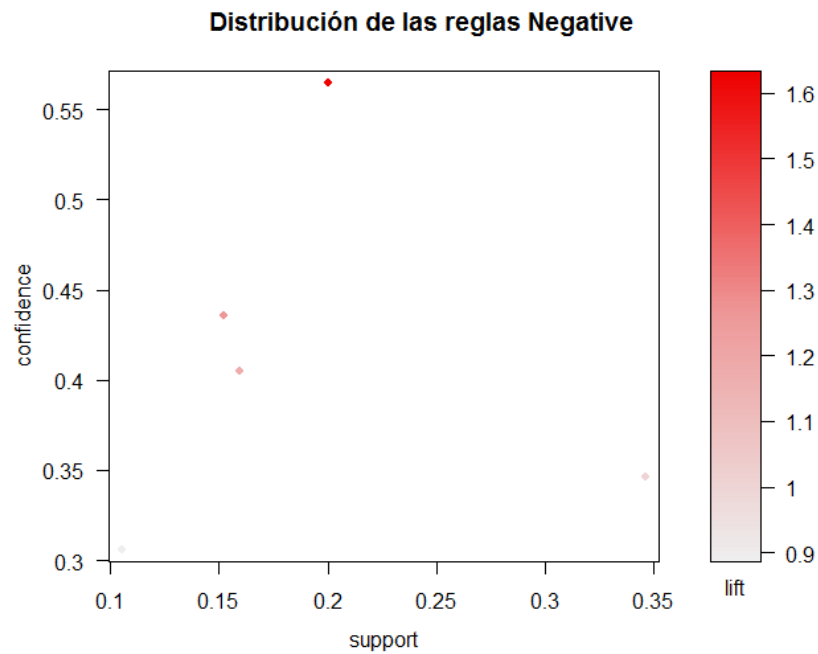


Figura 3.4: Distribución de las reglas Negative.

Si bien un número grande de reglas permiten de alguna forma u otra obtener mayor información acerca de los elementos, en este caso de las jugadas más comunes para cada clase, cabe recordar que como parámetros se tomaron valores de soporte y confianza pequeños, es por ello que en base a los mayores valores soporte se ordenó las diferentes subreglas para cada clase mediante las funciones dispuestas por los paquetes mencionados anteriormente, escogiendo las primeras 9 reglas a modo de analogía a la cantidad total de casillas que tiene el tablero del juego.

En detalle, las reglas obtenidas con una clasificación *positive* son las siguientes:

Cuadro 3.1: Reglas obtenidas - Clase Positive

	lhs	rhs	Soporte	Confianza	Lift
[1]	p2=o	class=positive	0.2390397	0.6939394	1.061971
[2]	p6=o	class=positive	0.2390397	0.6939394	1.061971
[3]	p8=o	class=positive	0.2390397	0.6939394	1.061971
[4]	p4=o	class=positive	0.2390397	0.6939394	1.061971
[5]	p9=x	class=positive	0.3079332	0.7057416	1.080033
[6]	p7=x	class=positive	0.3079332	0.7057416	1.080033
[7]	p3=x	class=positive	0.3079332	0.7057416	1.080033
[8]	p1=x	class=positive	0.3079332	0.7057416	1.080033
[9]	p5=x	class=positive	0.3820459	0.7991266	1.222945

Para la clasificación *negative* son las siguientes:

Cuadro 3.2: Reglas obtenidas - Clase Negative

	lhs	rhs	Soporte	Confianza	Lift
[1]	p7=o	class=negative	0.1524008	0.4358209	1.257580
[2]	p9=o	class=negative	0.1524008	0.4358209	1.257580
[3]	p3=o	class=negative	0.1524008	0.4358209	1.257580
[4]	p1=o	class=negative	0.1524008	0.4358209	1.257580
[5]	p5=o	class=negative	0.2004175	0.5647059	1.629483
[6]	p8=x	class=negative	0.1597077	0.4047619	1.167958
[7]	p6=x	class=negative	0.1597077	0.4047619	1.167958
[8]	p4=x	class=negative	0.1597077	0.4047619	1.167958
[9]	p2=x	class=negative	0.1597077	0.4047619	1.167958

Y su distribución en base al soporte y confianza es:

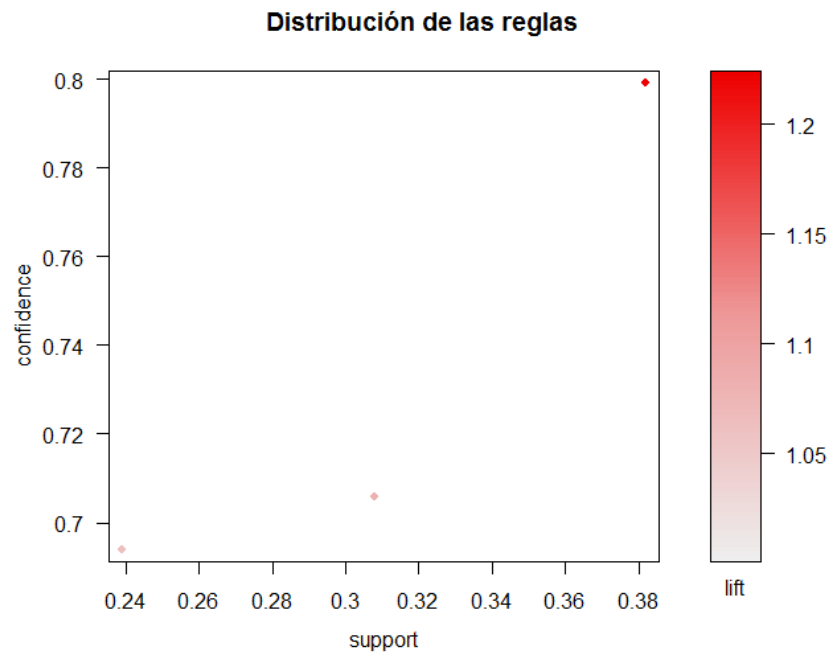


Figura 3.5: Distribución de las reglas Positive escogidas.

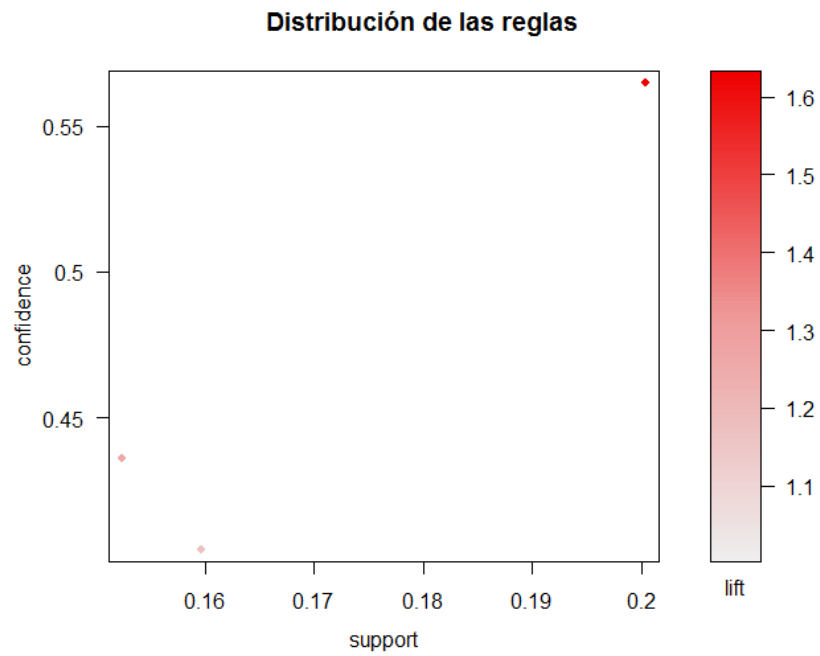


Figura 3.6: Distribución de las reglas Negative escogidas.

En los siguientes gráficos se puede observar de mejor forma la incidencia de las reglas en la clase en base a una representación de circunferencias, en donde su tamaño indica un mayor soporte, y su nivel de oscurecimiento representa el lift correspondiente.

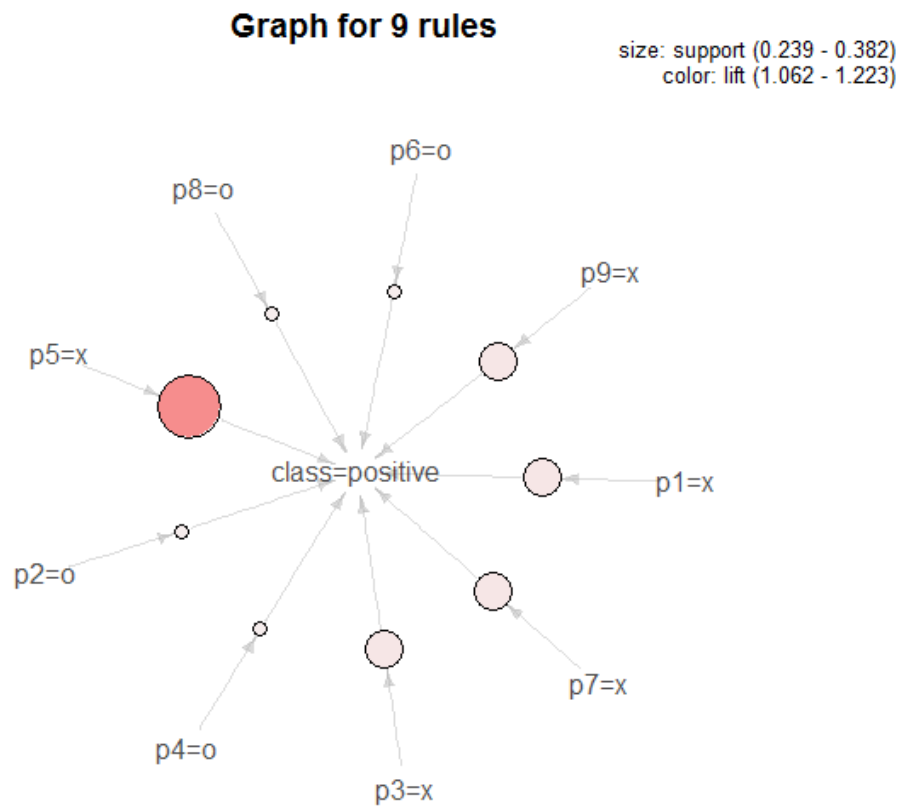


Figura 3.7: Graph de las reglas Positive ordenadas.

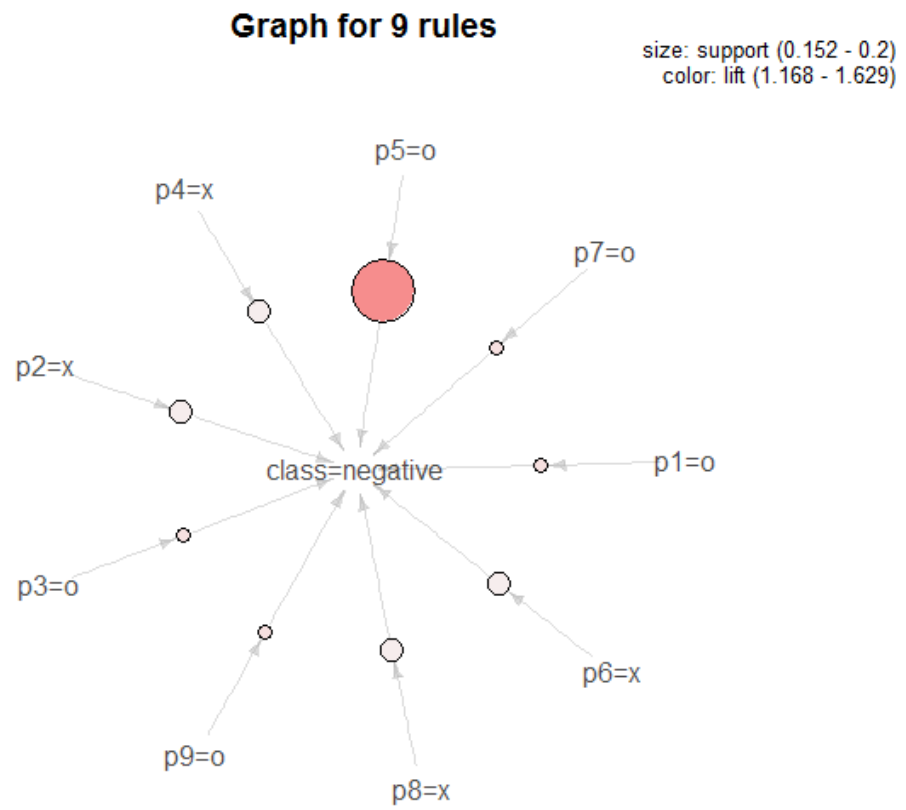


Figura 3.8: Graph de las reglas Negative ordenadas.

CAPÍTULO 4. ANÁLISIS DE RESULTADOS

Considerando que se tienen 9 variables que pueden tomar 3 valores cada una y 8 posibles combinaciones de casillas que finalizan con la victoria de x , se obtiene una cantidad reducida de reglas para ambas clasificaciones. Esto se refleja mucho en la distribución de reglas con la clasificación *negative* y un poco menos en la distribución de la otra clasificación en donde es importante destacar que tienen un alto valor de *lift* en general.

Como se mencionó, por la cantidad de combinaciones ganadoras limitadas en tic-tac-toe, las reglas que se obtuvieron fueron pocas y no tan generales; por ejemplo, en la figura 3.2 el antecedente es conformado solo por una variable y no combinaciones de estas pues para que se diera esta situación el soporte y la confianza deberían mínimos ser aún menores o 0. Lo mismo ocurre para la figura 3.1 en donde se muestran las reglas obtenidas con la clasificación *positive*.

La distribución de las reglas Positive, indicadas en la figura 3.3, muestra que existen muchas reglas que tienen un soporte bajo a pesar de que su valor de confianza es mayor a 0.6, pero es necesario encontrar aquellas reglas que se maximizan dentro de los parámetros posibles en torno a su confianza y soporte, de esta forma poder obtener aquellas reglas que tienen una mayor significancia. Así mismo, se da el caso para las reglas Negative.

Hecho el proceso anterior, si se hace un enfoque solo en las reglas en el **cuadro 3.1** se puede observar que:

- Las primeras 4 reglas con antecedente $p2$, $p6$, $p8$ y $p4$ que tienen el valor o y son clasificadas como *positive* cada una con un soporte de 0,23 aproximadamente además de una confianza de 0,69; esto ayuda a reflejar cuáles posiciones se ocupan con valor o cuando x es el ganador del juego.
- De la misma forma, las siguientes 4 reglas que tienen de antecedente $p9$, $p7$, $p3$ y $p1$ con el valor x , en donde cada una tiene un soporte de 0,30 y confianza de 0,70 aproximadamente se tiene que la clasificación es *positive* (x gana); que según las estadísticas ya mencionadas, son las casillas con una probabilidad intermedia de formar combinaciones ganadoras (cada una con 3 posibles estados de victoria). Se observa en estos últimos casos que el valor de *lift* aumenta respecto al *lift* de las 4 iniciales por lo que la regla es de *mayor calidad*.
- Por último está la regla [9] con antecedente $p5 = x$ (casilla central del tablero). Esta regla tiene un soporte de 0,38 y confianza de 0,79 aproximadamente lo que la destaca por sobre las otras por tener los valores más altos. Esta casilla es la única de las 9 que tiene 4 posibles combinaciones ganadoras por lo que estadísticamente es más probable que el jugador que la tenga marcada sea ganador y así se refleja en esta regla lo que la lleva a tener el mayor *lift* encontrado con un valor de 1,22.

Entonces en orden descendente las casillas que en general se asocian a jugadas ganadoras son ($p5$), ($p1$, $p3$, $p7$ y $p9$), siendo ($p2$, $p4$, $p6$ y $p8$) las que son generalmente ocupadas por el jugador que va a perder, siendo en este caso según la clase, el valor o .

Esta idea se puede reafirmar con lo que se puede observar en la figura 3.7, en la que se puede observar que la circunferencia perteneciente a la variable $p5=x$ es la que presenta un mayor soporte y lift, indicando una mayor significancia y hace intuir que existe una relación entre la clase y la variable. Esto mismo pero en menor medida se percibe para las variables $p1$, $p3$, $p7$ y $p9$ con valor de x .

En el caso del **cuadro 3.2** el panorama es similar.

- Para las reglas [1], [2], [3] y [4] se tienen las variables $p7$, $p9$, $p3$ y $p1$ con el valor o y clasificadas como *negative* (x no gana). Esto se explica de igual forma que el caso anterior, estas casillas son las que tienen una probabilidad intermedia de ganar por la cantidad de combinaciones que llevan a la victoria y las incluyen. En este caso como tienen el valor o , se puede haber dado que este jugador gane terminando en una derrota para x y por esta razón es clasificada como *negative*.
- Para la regla [5] ocurre exactamente lo mismo, solo que los valores de soporte, confianza y confidencia son mayores que los anteriores ya que se trata de la casilla central.
- Para las reglas [6], [7], [8] y [9] están en el antecedente las variables $p8$, $p6$, $p4$ y $p2$ respectivamente, con un valor x y clasificadas como *negative*; como ya se dijo, estas casillas con tan solo 2 combinaciones ganadores cada una son las menos probables de ganar, estadísticamente demostrado en el Informe 1. Esto explica la clasificación pues x al tener marcada alguna de estas casillas reduce sus posibilidades de ganar pudiendo producirse un empate o la victoria del contrincante.

La figura 3.8, al igual que para las reglas asociadas a la clase Positive, refuerza las observaciones establecidas para las reglas Negative, y que la regla que presenta la variable $p5=0$, es la que otorga una mayor significancia dentro de la clase, presentando los valores más altos para el soporte, la confianza y el lift.

CAPÍTULO 5. CONCLUSIONES

Luego de analizar las reglas obtenidas, se puede decir que las reglas de asociación ayudan a confirmar lo que se puede observar en la realidad al jugar *tic-tac-toe*. Se confirma entonces que la posición central es la que entrega una mayor incidencia en el resultado final representado por la clase. Así mismo, se pueden observar las distintas jugadas reflejadas por las reglas obtenidas para cada clase, con las jugadas diagonales como principal asociación para la definición de la clase según sea el valor que tome el jugador (x u o). Es entonces que cuando el juego resulta en que x ocupa la casilla central y que existe una jugada diagonal para este jugador, lo que resta es que las casillas p2, p4, p6 y p8 son ocupadas por el valor de o , y resulta obvio que según las normas del juego, es el jugador x quien gana la partida y por lo tanto, la clase resulta en "positive". Esto mismo se da en el caso contrario, resultando "negative" la clase cuando el jugador o ocupa la casilla central y juega en diagonal.

Si bien las reglas obtenidas no fueron muchas, las con mayor soporte y confianza permiten establecer jugadas que lleven a la victoria y también las que probablemente terminen en una derrota, así mismo, la propiedad del lift para todas las reglas restantes resulta en valores mayores a 1, con lo que se puede inferir que todas estas aparecen una cantidad de veces superior a lo esperado bajo condiciones de independencia (por lo que se puede intuir que existe una relación que hace que las casillas correspondientes se encuentren en el conjunto más veces de lo normal para cada clase).

En comparación con los métodos utilizados en experiencias pasadas: clustering y regresión logística. Con el primero no se puede realizar una comparación relevante pues la naturaleza del data-set no permitió desarrollarlo correctamente. Sin embargo, para la regresión logística se reafirma el conocimiento obtenido, esta vez también argumentado en las posiciones ocupadas por el otro jugador que es el que pierde el juego al final. Además por la cantidad de combinaciones posibles de estados finales de los tableros, al igual que en la experiencia anterior donde esto se vio reflejado en el área bajo la curva *ROC* (poco más de 0,6; bajo, considerando el óptimo (1) y el mínimo (0,5)), en esta oportunidad se refleja en los valores mínimos de soporte y confianza que son bastante pequeños dada la gran cantidad de instancias que puede tener un tablero. No obstante, dados estos valores tan bajos que desde un punto de vista ingenuo podrían quitarle importancia al análisis, la información obtenida dadas las características del estudio nos parece relevante y realista pues se adecuaba al contexto del juego.

En un punto de vista más crítico acerca del problema abordado, las características del dataSet en general son muy particulares pues cada elemento representa el estado final de un tablero de tic-tac-toe y solo eso. No se cuenta con información adicional como podría ser el orden en el que se jugaron las casillas, dato que podría ser de mucha importancia para un análisis más detallado. Esto deriva en que aunque se reafirma tal vez con mayores razones el conocimiento obtenido en experiencias anteriores, lo cual no deja de ser interesante, no se puede ahondar más allá en cuanto a la búsqueda de más información que de pie a extraer un mayor conocimiento, y no solamente ver aquellas casillas que dada su ocupación conlleven a ganar el juego.

CAPÍTULO 6. REFERENCIAS

Aha, D. W. (1991, Agosto). Index of /ml/machine-learning-databases/tic-tac-toe. <https://archive.ics.uci.edu/ml/machine-learning-databases/tic-tac-toe/>.

Chacón, M. (s.f.). Reglas de asociación. http://www.udesantiagovirtual.cl/moodle2/pluginfile.php?file=%2F115778%2Fmod_resource%2Fcontent%2F0%2FCapitulo%20VI%20Reglas%20de%20Asociacin.pdf.

UCI, M. L. R. (s.f.). Tic-tac-toe endgame data set. <https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame>.

Wikipedia. (s.f.). Reglas de asociación. https://es.wikipedia.org/wiki/Reglas_de_asociacin.

wonderopolis.org. (Febrero 2016). How old is tic-tac-toe? Recuperado desde <http://wonderopolis.org/wonder/how-old-is-tic-tac-toe>

CAPÍTULO 7. ANEXO: CÓDIGO FUENTE EN R

```
DB <- read.table("tic-tac-toe.data", header=FALSE, sep=" ",
                 col.names=c("p1", "p2", "p3", "p4", "p5", "p6", "p7", "p8", "p9", "class"))

library(arulesViz)
library(arules)

reglas = apriori(DB, parameter=list(support=0.1, confidence=0.3))

reglasPositive<-subset(reglas, subset= rhs %in% "class=positive")
summary(reglasPositive)
inspect(reglasPositive)
plot(reglasPositive, main= "Distribución de las reglas Positive")

reglasNegative<-subset(reglas, subset= rhs %in% "class=negative")
summary(reglasNegative)
inspect(reglasNegative)
plot(reglasNegative, main= "Distribución de las reglas Negative")

#####
#Reglas Positive

reglaPos1 = head(sort(reglasPositive, by="lift",decreasing=TRUE),10)
inforegla1Pos<-head(quality(reglaPos1),10)

reglaPos2 = head(sort(reglasPositive, by="support",decreasing=TRUE),10)
inforegla2Pos<-head(quality(reglaPos2),10)

reglaPos3 = head(sort(reglasPositive, by="confidence",decreasing=TRUE),10)
inforegla3Pos<-head(quality(reglaPos3),10)

summary(reglaPos2) #Positives

subreglasPos = reglasPositive[quality(reglasPositive)$confidence>0.65 &
                             quality(reglasPositive)$support >0.23 &
```

```

quality(reglasPositive)$lift >1]

summary(subreglasPos)
inspect(subreglasPos)
plot(subreglasPos, main= "Distribución de las reglas")
subreglas_ordenadasPos <- head(sort(subreglasPos, by="support"), 10)
plot(subreglas_ordenadasPos, method="graph")

#####
#Reglas Negative

reglaNeg1 = head(sort(reglasNegative, by="lift", decreasing=TRUE), 10)
inforegla1Neg<-head(quality(reglaNeg1), 10)

reglaNeg2 = sort(reglasNegative, by="support", decreasing=TRUE)
inforegla2Neg<-head(quality(reglaNeg2), 10)

reglaNeg3 = head(sort(reglasNegative, by="confidence", decreasing=TRUE), 10)
inforegla3Neg<-head(quality(reglaNeg3), 10)

summary(reglaNeg2) #Negatives

subreglasNeg = reglasNegative[quality(reglasNegative)$confidence>0.34 &
                                quality(reglasNegative)$support >0.15 &
                                quality(reglasNegative)$lift >1]

summary(subreglasNeg)
inspect(subreglasNeg)
plot(subreglasNeg, main= "Distribución de las reglas")
subreglas_ordenadasNeg <- head(sort(subreglasNeg, by="support"), 10)
plot(subreglas_ordenadasNeg, method="graph")

# Reglas Juntas, a modo de comparación
reglasJuntas <- c(subreglas_ordenadasNeg, subreglas_ordenadasPos)
plot(reglasJuntas, method="graph")

```

