
PRONÓSTICO PARA LA INYECCIÓN DE TENSO-ACTIVOS EN POZOS DE PETRÓLEO A PARTIR DE UNA METODOLOGÍA QUE INTEGRA TÉCNICAS DE INTELIGENCIA ARTIFICIAL Y MINERÍA DE DATOS

MARÍA DE LOS A. ALONSO, ARGELIO V. DE LA CRUZ y GRETTEL BARCELÓ

RESUMEN

Se presenta una metodología que integra diversas técnicas de inteligencia artificial para construir un sistema de pronóstico que determina la conveniencia de aplicar una solución con propiedades tenso-activas a un pozo de petróleo, con el objetivo de aumentar la producción de este hidrocarburo. La metodología comienza con el procesamiento de los datos obtenidos de un experimento consistente en la inyección de tenso-activos a un conjunto de pozos en un yacimiento. Se utilizaron diversas técnicas exploratorias de datos, como lo son reconocimien-

to de patrones, selección de variables y métodos para la generación automatizada de hipótesis. La información derivada de este procesamiento fue modelada en una base de conocimiento, que junto con las máquinas de inferencia de un lenguaje, denominado HARIES, permitieron la construcción de un sistema capaz de tomar decisiones en relación a la inyección de tenso-activos y sugerir la tecnología más apropiada a usar en cada contexto. El sistema se aplicó a diversos pozos, obteniendo, en todos los casos, resultados satisfactorios.



a aplicación de agentes tenso-activos en la industria del petróleo ha sido muy diversa y tiene una gran importancia práctica, de ahí la proliferación de su uso en muchas de las etapas del proceso de producción del petróleo (Rivas y Gutiérrez, 1999; Schramm, 2000), y en especial, en la extracción mejorada (Hou *et al.*, 2005; Babadagli *et al.*, 2006; Cuerbelo *et al.*, 2006; Carrero *et al.*, 2007; Rai *et al.*, 2009). Esta situación hace que el tema de los tenso-activos y el análisis de las condiciones que optimizan la producción del crudo, tengan un espacio prioritario para empresas y equipos de investigación (Champion, 2009; Oil Chem, 2009) en respuesta a las necesidades de diversos procesos industriales.

En estudios realizados a un yacimiento de petróleo en Cuba, por parte

del Centro de Investigaciones del Petróleo y el Instituto de Oceanología de la Academia de Ciencias, se inyectaron tenso-activos producidos a partir de algas marinas a un conjunto de 22 pozos de petróleo, con el fin de aumentar las producciones de dicho yacimiento. Si bien la utilización de esta tecnología es mucho más económica que las ofertas disponibles en el mercado mundial, de los experimentos realizados se obtuvo como consecuencia que en todos los casos el efecto no fue positivo y que a pesar de que en más de la mitad de los pozos se presentaron resultados favorables, al contabilizar las producciones totales antes y después del experimento se observó un cierto decremento global de la producción del crudo.

Fue por ello necesario estudiar bajo qué condiciones se podía pro-

nosticar un efecto positivo o negativo *a priori*, para actuar en consecuencia. Este trabajo presenta una metodología basada en técnicas de inteligencia artificial (IA) y minería de datos (MD) para construir un sistema basado en conocimiento, que permite pronosticar cuándo el efecto de la inyección será positivo o negativo y, en caso afirmativo, con cuál tecnología proceder.

Las técnicas empleadas para el análisis exploratorio de los datos obtenidos en los experimentos fueron reconocimiento de patrones (Kanal y Dattatreya, 1992; Berthold y Hand, 1999; De la Cruz y Alonso, 2002a), selección de variables (Witten y Eibe, 2000; Carrasco-Ochoa y Martínez-Trinidad, 2003), y generación automatizada de hipótesis (Louvar y Havranek, 1981; Hájek *et al.*, 1995; Hájek y Holena, 1998; Holena, 1998).

PALABRAS CLAVE / Análisis Exploratorio de Datos / Reconocimiento de Patrones / Selección de Variables / Generación Automatizada de Hipótesis / Sistemas Inteligentes de Pronóstico/

Recibido: 19/05/09. Modificado: 30/09/2009. Aceptado: 02/10/2009.

María de los A. Alonso. Ingeniera Electrónica y Maestra, Instituto Superior Politécnico José Antonio Echeverría, Cuba. Doctora en Ciencias de la Computación, Instituto Politécnico Nacional (IPN), México. Investigadora, Universidad Autónoma del Estado de Hidalgo (UAEH), México. Dirección: Centro de Investigación en Tecnologías de Información y Sistemas, UAEH, Abasolo 600, Pachuca, 42000 Hidalgo, México. e-mail: marial@uaeh.edu.mx

Argelio V. de la Cruz. Licenciado Educación, Instituto Superior Pedagógico Enrique José Varona (ISPEJV), Cuba. Doctor en Ciencias de la Computación, Academia de Ciencias, Cuba. Director General, Innovaciones Tecnológicas Avanzadas S.A. de C.V., México.

Grettel Barceló. Ingeniera en Sistemas Computacionales, IPN, México. Maestra en Ingeniería Eléctrica, Centro de Investigación y Estudios Avanzados, México. Estudiante de Doctorado, IPN, México.

La definición de los pasos de la metodología planteada responde los objetivos trazados para desarrollar el sistema de pronóstico. Brevemente, estos pasos fueron:

1. Definición de objetivos. Establecer la o las necesidades fundamentales del usuario final para determinar los objetivos que debe cubrir el sistema a desarrollar, todo lo cual se tomará en cuenta en las próximas etapas. Para este problema en particular se partió de la interrogante ¿Cómo decidir en qué pozo inyectar para lograr que el efecto sea positivo?

2. Caracterización del objeto de estudio. Es necesario identificar los rasgos distintivos de los pozos empleados en los experimentos iniciales. Este estudio debe definir las variables a utilizar en el análisis exploratorio de los datos y caracterizar los pozos según el tipo de crudo, la geología y la tecnología a emplear.

3. Evaluación de la importancia relativa de las variables para identificar el problema. No todas las variables que se utilizan para describir un problema influyen de igual forma en la caracterización del mismo. Por ello se requiere investigar ¿Cuáles variables son las más informativas respecto al problema planteado? y ¿Qué variables son las más diferenciadoras en el efecto obtenido (positivo o negativo)?

4. Identificación de las relaciones entre las variables y el efecto resultante de las inyecciones. Existen relaciones entre las variables que no son explícitas y resultan ser de gran utilidad para alcanzar la solución. De ahí que sea de suma importancia conocer ¿Qué parámetros o combinación de ellos se asocian al efecto positivo o negativo? y ¿Qué hipótesis se pueden generar con relación a esto?

5. Construcción de un sistema automatizado para el pronóstico. Con el conocimiento generado en los pasos anteriores es posible construir un asesor automático que permita ayudar a resolver estos problemas. Sin embargo, para ello se requiere diseñar los componentes y características que debe tener este sistema y utilizar el conocimiento para realizar el pronóstico, alcanzando la máxima objetividad en los resultados.

Para poder ejecutar los pasos anteriores es necesario aplicar un tratamiento híbrido a la información, utilizando diversas técnicas de análisis exploratorio de datos y de ingeniería del conocimiento, que se describen a continuación.

Caracterización del objeto de estudio

Para llevar a cabo un análisis profundo del comportamiento del experimento en cada pozo se consideró necesario comenzar realizando una caracterización de los pozos a través de un conjunto de variables

TABLA I
RELACIÓN DE VARIABLES MEDIDAS A LOS POZOS DE PETRÓLEO

Crudo	Geología	Tenso-activo	Tecnología	Resultado
Densidad	Capa geológica	Tiempo de envejecimiento	Volumen del caldo / metro lineal abierto	Tiempo de duración
Viscosidad	Edad de la roca	Concentración	Volumen del caldo / espesor efectivo	Efecto obtenido
Porcentaje de resinas asfálticas	Contenidos de calizas	pH	Velocidad de inyección	
azufre	arcilla		Tiempo de residencia	
Nº de neutralización	serpentina			
Humectabilidad	areniscas			
Compatibilidad	conglomerados			
Producción	Porosidad de la roca			
antes de la inyección	Saturación			
después de la inyección	Espesor efectivo			

que, en general, están relacionadas con las particularidades del crudo, la geología, la tecnología empleada, las características del tenso-activo (caldo) y el efecto obtenido. Las variables identificadas de tal análisis se muestran en la Tabla I. Estas variables fueron estudiadas para conocer *a priori* la influencia del uso de la nueva tecnología en los pozos de petróleo de acuerdo a sus características, lo cual se traduce en investigar ¿Cómo decidir en qué pozo inyectar tenso-activos para lograr un aumento de la producción en el mismo?

Para poder aplicar las técnicas de MD, los datos fueron organizados en forma de matriz, donde las filas representan los pozos estudiados y las columnas las 30 variables medidas (Tabla I), aunque en cada caso el procesamiento se dirigió a un subconjunto de la información en correspondencia con los objetivos de la metodología empleada. Cabe señalar que en varios pozos algunas variables no se pudieron medir y por tanto, la matriz presentó información incompleta.

En este paso se aplicaron dos procesos, i. preprocesamiento de la información, fase en la cual se realizaron las transformaciones de las variables (estandarización, categorización y dicotomización), el cálculo de semejanzas y los estudios de correlación, y ii. estudio tipológico de los pozos, efectuándose clasificaciones en correspondencia con las características del crudo y la geología.

Para llevar a cabo el análisis de la tipología, se formaron dos submatrices, utilizando en cada caso las variables referidas a las características del crudo y la geología (columnas 1 y 2 de Tabla I). Además, se aplicó un conjunto de siete métodos jerárquicos y otro conjunto de cuatro métodos de partición para cada submatriz de datos (Anderberg, 1973), implementados en el sistema integrado para reconocimiento de patrones SIRP (Alonso *et al.*, 1995). Se tomaron como medidas de semejanza la distancia euclidiana, el coeficiente de divergencia, el coeficiente de Duran y Odell, la distancia de Gower y la distancia para datos heterogéneos. Todo ello, con el objetivo de poder contrastar los resultados y

obtener un esquema clasificatorio más cercano a la realidad.

Los métodos jerárquicos empleados fueron conexión simple, conexión completa, centroide, mediana, promediación entre los grupos formados, promediación dentro del nuevo grupo y Ward, mientras que los métodos de partición fueron McQueen, Forgy, Jancey, K-Means convergente, también propuestos en Anderberg (1973) y Witten y Eibe (2000). Es de señalar que en el caso de los datos relacionados con la geología, solo se pudo utilizar los métodos jerárquicos, con las dos últimas medidas de semejanza mencionadas, dada la existencia de variables no numéricas e información incompleta.

Clasificación usando las características del crudo

La aplicación de los métodos jerárquicos da como resultado una gráfica denominada dendograma. Este diagrama de datos en forma de árbol permite apreciar claramente las relaciones de agrupación entre los datos, e incluso entre grupos de ellos, utilizando criterios de semejanza cuantitativos. Estos métodos han sido ampliamente empleados para el descubrimiento de la estructura de poblaciones desconocidas (González *et al.*, 2004; Martín y García-Moya, 2004; Abbas *et al.*, 2008; Cargnelutti *et al.*, 2008; Sam *et al.*, 2008; Uma *et al.*, 2008; Varin *et al.*, 2008; Wajrock *et al.*, 2008; Yilmaz *et al.*, 2009).

En la Figura 1 se muestra el dendograma de la clasificación ocupando la distancia de De la Cruz para datos heterogéneos (De la Cruz, 1990) y el método de Ward para las variables relacionadas con el crudo. A la izquierda se presentan los identificadores de los pozos y en la parte inferior la escala de semejanza ordenada en 25 intervalos, donde el 1 indica la máxima semejanza y el 25, la mínima. En la figura se aprecia la existencia de tres grupos de pozos en la clasificación hecha sobre las variables relacionadas con el crudo (C), el primero (1C) integrado por siete pozos, el segundo (2C) por cinco y el tercero (3C)

por tres pozos. Cabe indicar que los métodos de reconocimiento no supervisado se aplicaron únicamente sobre quince de los pozos bajo estudio, por ser éstos los que contaban con información suficiente para ello.

Para tratar de interpretar de forma más objetiva el significado de la clasificación obtenida fueron calculadas algunas medidas estadísticas de tendencia central y de dispersión, tales como mínimo, máximo, media, desviación estándar, asimetría, curtosis y coeficiente de variación. De acuerdo a la interpretación del dendograma y los valores de estas medidas para cada grupo, las variables se pueden caracterizar como sigue:

En el Grupo 1C, las variables bajo estudio se comportan de manera que los valores bajos corresponden al número de neutralización, los valores medios a densidad, viscosidad y resinas, y los valores altos a asfaltenos y azufre. En el Grupo 2C, las variables asfaltenos y azufre corresponden a valores medios, mientras que densidad, viscosidad, resinas y número de neutralización corresponden a valores altos. Por último, el Grupo 3C de pozos se caracteriza, según la tipología del crudo, por valores bajos en los casos de las variables densidad, viscosidad, asfaltenos, azufre y resinas, y valores medios para el número de neutralización.

Con este estudio se logró una caracterización general que ha permitido identificar distintos tipos de petróleo y la calidad del mismo en el conjunto de pozos estudiados.

Clasificación usando las variables geológicas

A diferencia del estudio de las variables del crudo, donde se utilizó el método de Ward, para las variables geológicas (litología) se aplicó el método de promediación dentro del nuevo grupo, manteniendo como medida de semejanza la distancia para datos heterogéneos. En la Figura 2 se presenta el dendograma resultante. En este caso se consideraron 17 pozos. De la gráfica también se evidencia la formación de tres grupos con base en la clasificación

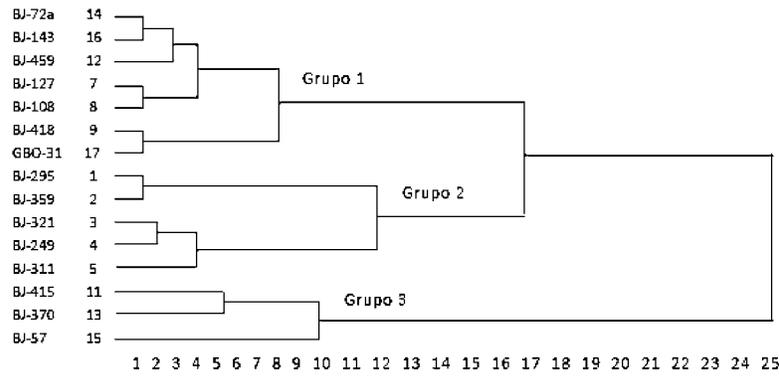


Figura 1. Clasificación de los pozos utilizando las variables relacionadas con el crudo.

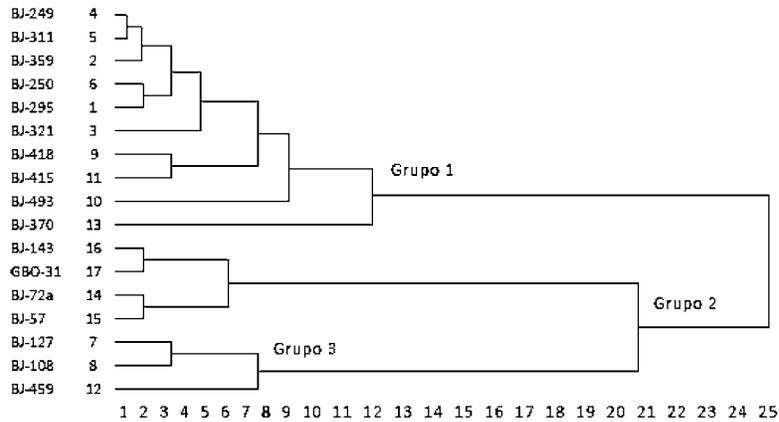


Figura 2. Clasificación de los pozos utilizando las variables relacionadas con la geología.

realizada con las variables relacionadas con la litología (L), 1L con 10 pozos, 2L con cuatro y 3L con tres pozos.

De la misma forma que en el caso anterior, se procedió a estudiar las características de los tres grupos formados por medio de medidas estadísticas. El análisis de estos resultados permitió determinar las características generales de cada grupo, a saber: el grupo 1L representa pozos en rocas calizas fundamentalmente (80%) y una cierta presencia de arcillas, en capas geológicas E, B y C; el grupo 2L se caracteriza por encontrarse en areniscas y conglomerados y pertenecer a las capas A1 y A2; y el grupo 3L está conforma-

- A. *Conjunciones de dos componentes relacionadas con el crudo:*
densidad NO baja & compatibilidad baja ~E+
densidad NO baja & producción original baja ~E+
viscosidad menor de 10,000 & producción original baja ~E+
densidad alta & compatibilidad alta ~E-
contenido de azufre bajo & humectabilidad baja ~E-
- B. *Conjunciones de dos componentes relacionadas con la geología:*
contenido de calizas alto & NO capa geológica E~E-
contenido de calizas alto & saturación de la roca bajo ~E-
contenido de calizas alto & edad de la roca Cretácico Superior ~E- C
- C. *Conjunciones de dos componentes relacionadas con el crudo y la geología:*
contenido de azufre bajo & contenido de calizas alto ~E-
contenido de azufre bajo & espesor efectivo bajo ~E-
compatibilidad baja & contenido de calizas bajo ~E+
producción original alta & contenido de arcillas bajo ~E-
producción original alta & contenido de areniscas bajo ~E-

Figura 3. Ejemplos hipótesis de asociación de dos variables en relación con el crudo y la geología.

do por pozos ubicados en serpentinas (70%) con cierta presencia de arcillas, capas geológicas S, S2 y S3.

Resultados de la Fase de Clasificación

De los resultados en esta primera clasificación se tiene como conclusión:

Resultado 1. Los pozos quedaron clasificados en tres grupos según el tipo de crudo y tres grupos según la geología, los cuales estuvieron integrados por miembros diferentes para cada variante.

Resultado 2. No todos los pozos son del mismo tipo, lo cual no fue tomado en cuenta en los experimentos iniciales, antes de este análisis exploratorio. Por lo tanto, es necesario estudiar la relación existente entre las características del crudo, la geología y el efecto obtenido.

Una tarea inmediata tras obtener los resultados anteriores consistió en la búsqueda de las relaciones potenciales entre las clasificaciones considerando ambos grupos de variables, puesto que se observó que eran diferentes (Figuras 1 y 2). Esta comparación arrojó el resultado siguiente:

Resultado 3- Los integrantes del grupo 1C y 3C del crudo corresponden con el mismo tipo de litología (1L) y el grupo 2C del crudo responde a dos tipos de litologías (2L y 3L).

Evaluación de la importancia relativa de las variables para identificar el problema

El próximo paso después de obtener el esquema clasificatorio de los pozos consistió en el estudio de la importancia relativa de las variables involucradas en la investigación. Para ello se aplicó la teoría de test y testores, con lo cual se obtiene un valor que permite conocer y establecer comparaciones entre las variables implicadas. Esta teoría es aplicada cuando se requiere conocer el peso informacional y/o diferenciante de las variables (Carrasco-Ochoa y Martínez-Trinidad, 2004; Carrasco-

TABLA II
PESOS DIFERENCIANTES DE
LAS VARIABLES RELACIONADAS
CON EL CRUDO Y LA GEOLOGÍA

Identificación	Pesos Diferenciantes
Capa geológica	0,082
Densidad	0,224
Viscosidad	0,215
Porcentaje de resinas	0,274
Porcentaje de asfaltenos	0,576
Porcentaje de azufre	0,207
No. de neutralización	0,22
Humectabilidad	0,394
Compatibilidad	0,579
Producción original	0,337
Contenido de calizas	0,15
Contenido de sílice	0,239
Contenido de arcilla	0,546
Contenido de serpentina	0,11
Contenido de areniscas	0,2
Contenido conglomerados	0,431
Porosidad de la roca	0,292
Saturación de la roca	0,239
Espesor efectivo	0,449
Edad de la roca	0,414

Ochoa y Ruíz-Shulcloper, 2004; Cumplido *et al.*, 2006).

Para ejecutar el cálculo se tomaron todas las variables relacionadas con el crudo y la geología, sumando éstas un total de 20 (Tabla I). Se realizó un análisis previo para determinar los intervalos de valores o umbrales de semejanza a ser utilizados para la comparación entre los valores de cada variable. La variable que se tomó como objetivo fue la relacionada con el efecto, lo cual divide el universo de individuos en dos clases, positivo y negativo.

En la Tabla II se presenta el resultado brindado por esta técnica. Allí se puede apreciar que cada variable se encuentra asociada con un valor o peso diferenciante, siendo éste un número entre 0 y 1. El valor 0 indica que la variable es completamente irrelevante para el proceso y es posible prescindir de ella, mientras que el valor 1 indica que la misma es indispensable. Los demás valores expresan gradaciones de la importancia que poseen las variables para analizar el problema. Este resultado puede ser interpretado como el poder diferenciante que tiene cada variable para establecer si un pozo tendrá un efecto positivo o negativo producto de la inyección.

Como se desprende de la Tabla II, no existe una variable predominante por su alto valor, lo cual corrobora la expectativa inicial, puesto que no se esperaba que una o pocas variables de forma individual permitieran dar solución a un problema de gran complejidad como el estudiado. Esto también

se hace evidente al observar que no hay diferencias notables entre el valor de una variable y su consecutiva, si se ordenan de acuerdo a su peso.

Resultados de la Fase de Selección

Del estudio de las variables, se puede concluir:

Resultado 4. La influencia de las variables sobre el efecto que resulta de la inyección de tenso-activos en un pozo, está dada por todas en conjunto, dado que no se reflejan diferencias contrastantes en su poder informacional.

Un aspecto que resultó de interés fue que al seleccionar los pozos donde se iba a ejecutar el experimento, se tomaba en consideración como variables más importantes a la capa geológica y la humectabilidad, las cuales, por el contrario, no tienen valores relevantes como puede advertirse en la Tabla II.

Identificación de las relaciones entre las variables y el efecto de las inyecciones

El próximo paso en la exploración de los datos consistió en buscar las relaciones existentes entre las distintas variables y el objetivo del estudio que se relaciona con el efecto resultante de la inyección de los pozos con tenso-activos. Para ello se utilizaron diversas variantes de los métodos GUHA (*general unary hypotheses automaton*), los cuales posibilitan la generación automatizada de hipótesis basada en datos empíricos y, por tanto, es una técnica empleada en minería de datos. Producen todas las hipótesis “interesantes” con respecto a alguna pregunta general a investigar y los datos a ser pro-

Si A y B representan sentencias observacionales acerca de un problema real, entonces ejemplos típicos de hipótesis generadas por GUHA son: A está asociado con B (A~B, donde “~” significa “asociado con”) y A es causa de B (A→B, donde “→” significa “implica”)

El objetivo de este paso de la metodología exploratoria fue encontrar asociaciones y causas que permitan pronosticar cuándo el resultado (efecto) será positivo o negativo. Se tomaron los cuantificadores Fisher (asociación) y casi implicación (implicación) con niveles de significación de 0,5, para definir los tipos de hipótesis buscadas.

A continuación se presentan algunos ejemplos de los principales resultados obtenidos para las hipótesis encontradas.

Relaciones entre el crudo y la geología con el efecto

No se encontraron hipótesis que relacionen de forma individual estas variables al nivel de significación definido, lo cual ratifica que no es posible, ni recomendable, tomar en cuenta un solo parámetro como criterio para la elección. Por tal motivo se disminuyó el nivel de significación a 0,2 que todavía sigue siendo alto para explorar relaciones individuales. Para este nivel se obtuvieron tres hipótesis de asociación de una variable, a saber: i. no pertenecer a la capa geológica B~efecto positivo (E+); ii. producción original alta~efecto negativo (E-); y iii. producción original baja~efecto positivo (E+). De esto es posible concluir que:

Resultado 5. Existe una tendencia en los pozos de menor producción a un efecto posi-

Humectabilidad baja & producción original alta →E-
Compatibilidad baja & contenido de calizas bajo →E+
Producción original baja & edad da la roca Jurásico Superior →E+
Viscosidad fuera de [1000 10000] & porcentaje de asfaltenos bajo & contenido de calizas bajo →E+
Porcentaje de asfaltenos bajo & humectabilidad alta & contenido de arcillas alto →E+
Producción alta & contenido de serpentinas bajo & contenido de areniscas bajo →E-
Producción baja & espesor efectivo bajo & 110 edad de la roca Cretácico Superior →E+
NO está en la capa C & porcentaje de azufre bajo & humectabilidad baja →E-

Figura 4. Ejemplos de hipótesis de implicación de dos o tres variables en relación con el crudo y la geología.

cesados. También responden a preguntas tan generales como ¿Existe alguna dependencia, asociación o relación entre los factores o variables estudiadas?, por lo cual han sido utilizados en innumerables aplicaciones (Hájek *et al.*, 1984, 1987; Pecén *et al.*, 1995; Zvarova *et al.*, 1997; Halova *et al.*, 1998, 2002; Harmancová *et al.*, 1998).

En este estudio, la búsqueda de relaciones entre variables y efecto generó hipótesis de dos tipos, asociación e implicación.

vo y los de mayor producción hacia un efecto negativo.

En el caso de relaciones donde se consideraron dos o tres variables, se obtuvieron alrededor de 125 hipótesis de asociación de dos variables al nivel de significación deseado, algunos ejemplos de las cuales se muestran en la Figura 3, mientras que en la Figura 4 se presentan algunos ejemplos de las hipótesis de implicación de dos o tres variables obtenidas a partir de conjunciones de dos y tres componentes.

Como se puede notar, las hipótesis obtenidas involucran una gran variedad de parámetros y combinaciones de éstos, lo cual brinda una amplia base para el pronóstico en función de las características de cada pozo.

Relaciones entre la tecnología y las características del tenso-activo con el efecto

Aunque al inicio del estudio se planteó un esquema único para la experimentación, relacionada con la producción del tenso-activo y su inyección en los pozos, en la práctica esto no fue posible y casi todos los pozos fueron inyectados con parámetros diferentes. De ahí la necesidad de estudiar las posibles relaciones entre las variables referidas a la tecnología y a las características del tenso-activo con el efecto resultante de la aplicación. En este ejercicio también se emplearon métodos GUHA. Al analizar los resultados, se obtuvo que no existen hipótesis de asociación fuertes. No obstante, se obtuvieron dos hipótesis a un nivel de significación menor, pero que brindan orientaciones interesantes acerca de cómo se asocian las características del tenso-activo con el efecto positivo. En la Figura 5 se presentan las dos hipótesis resultantes de la aplicación de los métodos GUHA:

De las hipótesis obtenidas, se puede concluir:

Resultado 6. Se corrobora la relación entre velocidad de inyección baja y efecto positivo, reportada en la literatura.

Resultado 7. La asociación con el tiempo de envejecimiento bajo apunta a la necesidad de no demorar la inyección una vez concluido el proceso de fermentación.

Resultado 8. Se da respuesta a la falta de necesidad de obtener una alta concentración del caldo, lo cual disminuye el costo del proceso.

Resultado 9. La relación con el pH del caldo alto es esperada por el hecho de que el pH alcalino debe aumentar la humectabilidad, lo cual se asocia con un efecto positivo.

Relaciones entre el crudo y la geología con la tecnología y las características del tenso-activo

Interesados en el estudio de la tecnología y las características del tenso-activo, con vistas a recomendar acerca de sus parámetros, se buscaron las dependencias de éstos con el crudo y la geología. Algunos ejemplos de hipótesis que relacionan las características del pozo (crudo y geología) con la tecnología y las características del tenso-activo, y de lo cual resulta un efecto positivo, se muestran en la Figura 6.

De las hipótesis anteriores se desprende una última conclusión:

Resultado 10. No es posible, ni recomendable plantearse un esquema tecnológico y/o de fermentación del caldo único, ya que su efectividad está en relación con los parámetros del pozo.

Desarrollo del Sistema Automatizado para el Pronóstico

Derivado del estudio realizado se cuenta con gran cantidad de datos acerca del problema; sin embargo, su volumen y variedad hace prácticamente imposible su utilización por parte de los especialistas. De ahí que aún no pueda responderse a la cuestión planteada desde un inicio acerca de ¿Cómo pronosticar, dado los datos de un pozo, si su inyección provocará un efecto positivo o no?

Para dar solución a esta interrogante se procedió a construir una base de

Las proposiciones modeladas fueron de dos tipos: para solicitar información (preguntas) y para brindar resultados (objetivos). Ejemplos de preguntas son “el pozo se encuentra en rocas de edad Jurásica Superior y el contenido de azufre en el crudo es alto.” Esta última se construyó utilizando los grupos definidos como resultado de las medidas estadísticas. Los conceptos “se pronostica efecto positivo en el pozo y se sugiere inyectar con una relación volumen/metro lineal abierto alta” son ejemplos de objetivos.

Por otra parte, las RPG se definieron en base a los resultados obtenidos en los pasos anteriores de la metodología. De la aplicación de los métodos GUHA se consideraron los antecedentes y consecuentes de las hipótesis como los elementos de las reglas. De la técnica de selección de variables se usaron los pesos obtenidos en la Tabla II como las contribuciones al consecuente de las reglas.

Así por ejemplo, si consideramos una de las hipótesis obtenidas: “produc-

Velocidad de inyección baja & tiempo de envejecimiento bajo & concentración del calda bajo ~E+
pH del caldo alto (alcalino) & tiempo de envejecimiento bajo & concentración del caldo bajo ~E+

Figura 5. Hipótesis de asociación de tres variables en relación con la tecnología y el tenso-activo.

conocimiento (BC) utilizando los resultados obtenidos producto del procesamiento de los datos, la cual junto con las máquinas de inferencia del lenguaje HArIES (De la Cruz *et al.*, 1991, 1993; De la Cruz y Alonso, 2002b; Alonso y De la Cruz, 2004; Alonso, 2006), permitieron construir un sistema capaz de aportar soluciones al problema planteado.

Este sistema, además de pronosticar el efecto resultante del procedimiento de inyección a un pozo de petróleo dado, también brinda sugerencias generales sobre la tecnología a utilizar y los parámetros del tenso-activo a producir.

Estructuras de la base de conocimiento

La construcción de la BC se basó en las estructuras de representación de conocimiento disponibles en el lenguaje de representación del conocimiento HArIES. Las proposiciones fueron las estructuras fundamentales utilizadas para representar los conceptos del problema y las relaciones obtenidas entre éstas se modelaron con las reglas de producción generalizadas (RPG; Valdés *et al.*, 1994). Además, se emplearon otras estructuras que implementa el lenguaje para solicitar los valores de las variables concernientes a las características del pozo que se pretende inyectar y para presentar los resultados al usuario que consulta al sistema.

ción original alta & contenido de arcillas bajo ~E-, la regla definida sería:

-Si se satisface con certeza total que la producción original es alta (10) y la roca se caracteriza por bajo contenido de arcilla (22) entonces esto contribuye en 0,337 a la seguridad que se tiene sobre el efecto negativo de la inyección (12).

En esta regla los valores entre paréntesis indican el número de la proposición en la BC. La contribución 0,337 es el mínimo de los pesos diferenciadores de las características implicadas (producción original de 10 y contenido de arcilla de 22) obtenidos en la Tabla II. El valor de la contribución se basa en la teoría de manejo de incertidumbre (Hájek, 1985).

Para representar la regla anterior en el lenguaje HArIES se utiliza la sintaxis

10 & 22 ⇒ 12 (-0,337 0)

De la misma forma se construyen todas las reglas de la BC. Para calcular los pesos con que se cumplen los resultados del pronóstico, se utiliza el mecanismo de encadenamiento hacia atrás.

Los valores de entrada para realizar el pronóstico se pueden obtener de dos fuentes diferentes, por petición directa al usuario o por medio de una base de datos.

Porcentaje de asfáltenos bajo & humectabilidad alta ~ tiempo de residencia bajo
Producción original baja & edad de roca no es Cretácico Superior ~ velocidad de inyección baja
Contenido de areniscas bajo & humectabilidad alta ~ pH del caldo bajo (ácido)

Figura 6. Ejemplos de hipótesis de asociación de dos variables en relación con el crudo, la geología, la tecnología y el tenso-activo.

TABLA III
RESULTADOS OBTENIDOS DURANTE EN EL EXPERIMENTO INICIAL
Y CON EL SISTEMA DE PRONÓSTICO

Producción (en toneladas)	Antes de la inyección	Después de la inyección		Diferencia	
		Experimento	Pronóstico	Experimento	Pronóstico
Diaria	88,14	83,67	118	-4,47	29,86
Tres meses	7932,6	7530,3	10620	-402,3	2687,4
Anual	31730,4	30121,2	42480	-1609,2	10749,6

Estos valores a su vez pueden ser de origen cualitativo o numérico. Los datos cualitativos se solicitan a través de las proposiciones preguntas y los numéricos con estructuras denominadas imagen sensible que permiten además el manejo de multimedia.

Para la construcción dinámica de los resultados que se le presentarán al usuario, se incorporaron a la BC un conjunto de reglas de conclusiones disponibles en el lenguaje. Estos resultados dependen de los valores que posean las diferentes variables asociadas al pozo a inyectar e involucran tres aspectos, a saber: pronóstico, tecnología a emplear y propiedades del caldo. Estos dos últimos solo se presentan en caso de pronosticar un efecto positivo. La Figura 7 muestra un ejemplo del resultado brindado por el sistema para un caso real de inyección en uno de los pozos de la zona estudiada.

La BC del sistema desarrollado consta de 60 proposiciones, 118 reglas de producción generalizadas, 26 estructuras de entrada y 19 reglas de conclusión para salidas.

Conclusiones

La utilidad de la aplicación de diversas técnicas de inteligencia artificial para la solución de problemas ha quedado demostrada porque el ser humano tiene la tendencia de utilizar pocos factores a la vez para analizar los problemas y esto se debe, fundamentalmente, a su limitada capacidad para razonar con múltiples variables simultáneamente. Por otra parte, el humano no considera la importancia de cada variable en particular. De ahí que la elección de los pozos a ser inyectados no siempre fuera exitosa en su etapa experimental.

Los resultados obtenidos utilizando el sistema construido para los 22 pozos coincidieron en todos los casos con el experimento inicial. Posteriormente, se procesaron los datos de otros 12 pozos y el sistema pronosticó efecto positivo en siete y negativo en cinco. Después de inyectar los siete pozos de pronós-

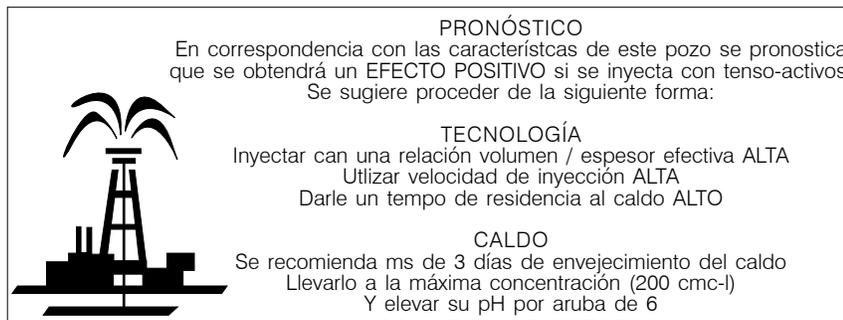


Figura 7. Ejemplo de una pantalla de resultados brindada por el sistema.

tico positivo y dos de los cinco con pronóstico negativo, los resultados del sistema también concordaron incluyendo los de pronóstico negativo, que experimentaron una baja en su producción.

La Tabla III muestra los resultados obtenidos en los 22 pozos durante el experimento inicial y posteriormente utilizando el sistema de pronóstico. En la fase experimental se evidencia la disminución de la producción del crudo tras la inyección, comentada en los antecedentes. Luego de procesar con el sistema la información de cada uno de los pozos, e inyectar con la tecnología y el tenso-activo adecuados a aquellos cuyo diagnóstico resultó positivo, se observa un incremento global en la producción. En ambos casos, se incluyen datos de la proyección del aumento/decremento a tres meses y a un año, tomando como base la producción diaria del crudo.

Una conclusión importante derivada del análisis exploratorio fue el conocer que existe una tendencia en los pozos de menor producción a un efecto positivo y en los de mayor producción hacia un efecto negativo.

Basado en los resultados obtenidos y la generalidad de los métodos empleados, se puede decir que el procedimiento utilizado constituye una metodología de integración híbrida de diversas técnicas, que puede ser utilizada en otros problemas de diversos ámbitos.

REFERENCIAS

Abbas SJ, Farhatullah, Marwat KB, Khan IA, Munir I (2008) Molecular Analysis of Genetic Diversity in Brassica Species. *Pak. J. Bot.* 41: 167-176.

- Alonso MA, De la Cruz AV, Ponce E (1995) SIRP: Sistema Integrado para Reconocimiento de Patrones con Inteligencia Artificial. En *Mem. Cong. Int. Reconocimiento de Patrones*. Cuba. pp. 123-129.
- Alonso MA, de la Cruz AV, Gutiérrez A (2004) Knowledge Representation Language: HArIES. En *Proc. 8th World Multiconf. Systemics, Cybernetics and Informatics*. EEUU. pp. 358-361
- Alonso MA (2006) *Representación y Manejo de Información Semántica en Interacción Hombre-Máquina*. Tesis. Instituto Politécnico Nacional. México: 314 pp.
- Anderberg MR (1973) *Cluster Analysis for Applications*. Academic Press. Nueva York, EEUU. 359 pp.
- Babadagli T (2006) Evaluation of the critical parameters in oil recovery from fractured chalks by surfactant injection. *J. Petrol. Sci. Eng.* 54: 43-54.
- Berthold M, Hand DJ (1999) *Intelligent Data Analysis*. Springer. Berlín, Alemania. 400 pp.
- Cargnelutti A, Ribeiro ND, Padilha dos Reis RC, de Souza JR, Jost E (2008) Comparison of cluster methods for the study of genetic diversity in common bean cultivars. *Cien. Rural* 38: 2138-2145.
- Carrasco-Ochoa JA, Martínez-Trinidad JF (2003) Editing and training for ALVOT, an evolutionary approach. En *Intelligent Data Engineering and Automated Learning Book Series: Lecture Notes in Computer Science*. Vol. 2690, pp. 452-456.
- Carrasco-Ochoa JA, Martínez-Trinidad JF (2004) Feature selection for natural disaster texts classification using testers. En *Intelligent Data Engineering and Automated Learning Ideal 2004 Book Series: Lecture Notes in Computer Science*. Vol. 3177, pp. 424-429.
- Carrasco-Ochoa JA, Ruiz-Shulcloper JR (2004) Sensitivity analysis of fuzzy Goldman typical testers. *Fuzzy Sets Syst.* 141: 241-257.
- Carrero E, Queipo NV, Pintos N, Zerpa LE (2007) Global sensitivity analysis of Alkali-Surfactant-Polymer enhanced oil recovery processes. *J. Petrol. Sci. Eng.* 58: 30-42.
- Champion (2009) *Líneas de Producto: Surfactantes*. www.champ-tech.com/onec_prod_surf_es.asp (Cons. 27/08/2009).
- Cumplido R, Carrasco-Ochoa JA, Feregrino C (2006) On the design and implementation of a high performance configurable architecture for testor identification. En *Progress in Pattern Recognition, Image Analysis and Applications Book Series: Lecture Notes in Computer Science*. Vol. 4225, pp. 665-673.
- Curbelo FDS, Barrós EL, Dutra TV, Castro TN, Garnica AIC (2006) Oil recovery by bionic and non-ionic surfactants and adsorption in sandstones. *Afinidad: Rev. Quím. Teór. Aplic.* 63: 291-295.
- De la Cruz AV (1990) SRPS: Un sistema para reconocimiento de patrones supervisado. En *Mem. Cong. Informática '90*. Cuba. pp. 227-240.
- De la Cruz AV, Valdés JJ, Pérez A, Jócik E, Balsa J, Rodríguez A (1991) The ARIES environment for the development of knowledge based expert systems. *Industrial Applications of Artificial Intelligence*. pp. 203-209.
- De la Cruz AV, Valdés JJ, Pérez A, Jócik E, Balsa J, Rodríguez A (1993) *Fundamentos y Práctica de la Construcción de Sistemas Expertos*. Academia. Cuba: 300 pp.
- De la Cruz AV, Alonso MA (2002a) Reconocimiento de Patrones Auto-Supervisado. Método de las Clases Cerradas. En *Reconocimiento de Patrones. Avances y Perspectivas*.

- Instituto Politécnico Superior. México. Vol. 1, pp. 325-336.
- De la Cruz AV, Alonso MA (2002b) The HARies environment (v6.00) for the development of intelligence systems. *Hifen* 26: 184-186.
- González LA, Prieto A, Molina C, Velásquez J (2004) Los reptiles de la Península de Araya, Estado Sucre, Venezuela. *Interciencia* 29: 428-434.
- Hájek P (1985) Combining functions for certainty factors in consulting systems. *Int. J. Man-Mach. Stud.* 22: 59-76.
- Hájek P, Holena M (1998) Formal logics of discovery and hypothesis formation by machine. En *Proc. 1st Int. Conf. on Discovery Science*. Japan. pp. 291-302.
- Hájek P, Valdés JJ, De la Cruz AV (1984) *El método GUHA de formación automatizada de hipótesis y sus aplicaciones en geofísica y geoquímica*. Informe Técnico del Instituto de Matemática de la Academia de Ciencias de Checoslovaquia. 25 pp.
- Hájek P, Sochorová A, Zvárová J (1995) GUHA for personal computers. *Comput. Stat. Data Anal.* 19: 149-153.
- Halova J, Strouf O, Zak P, Sochozova A, Uchida N, Yuzuri T, Sakakibara K, Hirota M (1998) QSAR of catechol analogs against malignant melanoma using fingerprint descriptors. *Quant. Struct.-Activ. Relat.* 17: 37-39.
- Halova J, Zak P, Stopka P, Yuzuri T, Abe Y, Sakakibara K, Suezawa H, Hirota M (2002) Structure-Sweetness Relationships of Aspartame derivatives by GUHA. *Discovery Science, Book Series: Lecture Notes in Computer Science*. Vol. 2534, pp. 291-296.
- Harmancová D, Holena M, Sochorová A (1998) Overview of the GUHA method for automating knowledge discovery in statistical data sets. En *Proc. Int. Conf. on Knowledge Extraction from Statistical Data*. Luxembourg. pp. 39-52.
- Holena M (1998) Fuzzy hypotheses for GUHA implications. *Fuzzy Sets Syst.* 98: 101-125.
- Hou J, Liu Z, Zhang S, Yue X, Yang J (2005) The role of viscoelasticity of alkali/surfactant/polymer solutions in enhanced oil recovery. *J. Petrol. Sci. Eng.* 47: 219-235.
- Kanal LN, Dattatreya GR (1992) Pattern recognition. En *Encyclopedia of Artificial Intelligence*. 2^{da} ed. Wiley. Nueva York, EEUU. Vol. 2, pp. 1116-1129.
- Louvar B, Havranek T (1981) GUHA Methods for Automatic Generation of Hypotheses for Exploratory Statistical Studies. *Biometrics* 37: 600-606.
- Martín F, García-Moya E (2004) Diversidad de especies perennes y su relación con el ambiente en un área semiárida del centro de México: Implicaciones para la conservación. *Interciencia* 29: 435-441.
- Oil Chem (2009) Innovative Surfactants for Enhanced Oil Recovery (EOR). Oil Chem Technologies. www.oil-chem.com/eor.htm (Cons. 27/08/2009).
- Pecen L, Ramesová N, Pelikán E, Beran H (1995) Application of the GUHA method on financial data. *Neur. Netw. World* 54: 565-571.
- Rai K, Johns RT, Lake LW, Delshad M (2009) Oil-Recovery Predictions for Surfactant Polymer Flooding. En *Proc. SPE Annual Technical Conference and Exhibition*. EEUU. pp. 4-7.
- Rivas H, Gutiérrez X (1999) Los surfactantes: Comportamiento y algunas de sus aplicaciones en la industria petrolera. *Acta Cient. Venez.* 50: 54-65.
- Sam V, Tai CH, Garnier J, Gibrat JF, Lee B, Munson PJ (2008) Towards an automatic classification of protein structural domains based on structural similarity. *BMC Bioinform.* 9. Article N° 74.
- Schramm LL (2000) *Surfactants: Fundamentals and Applications in the Petroleum Industry*. Cambridge University Press. Cambridge, RU. 621 pp.
- Uma S, Saraswathi MS, Siva SA, Vadhana MSD, Manickavasagam M, Durai P, Lourdasamy A (2008) Diversity and phylogenetic relationships among wild and cultivated bananas (Silk-AAB) revealed by SSR markers. *J. Hort. Sci. Biotechnol* 83: 239-245.
- Valdés JJ, De la Cruz AV, Jocik E, Balsa J, Rodríguez A (1994) *Ingeniería del Conocimiento en el Medio Ambiente ARIES*. Academia. Cuba: 230 pp.
- Varin T, Saettel N, Villain J, Lesnard A, Dauphin F, Bureau R, Rault S (2008) 3D pharmacophore, hierarchical methods, and 5-HT4 receptor binding data. *J. Enz. Inhib. Med. Chem.* 23: 593-603.
- Wajrock S, Antille N, Rytz A, Pineau N, Hager C (2008) Partitioning methods outperform hierarchical methods for clustering consumers in preference mapping. *Food Qual. Pref.* 19: 662-669.
- Witten IH, Eibe F (2000) *Data Mining*. Morgan Kaufmann. California, EEUU. 371 pp.
- Yilmaz KU, Ercisli S, Asma BM, Dogan Y, Kafkas S (2009) Genetic relatedness in *Prunus* Genus revealed by inter-simple sequence repeat markers. *Hortscience* 44: 293-297.
- Zvarova J, Preiss J, Sochorova A (1997) Analysis of data about epileptic patients using the GUHA method. *Int. J. Med. Informat.* 45: 59-64.

FORECAST FOR SURFACTANT INJECTION IN OIL WELLS FROM A METHODOLOGY CONSISTING OF ARTIFICIAL INTELLIGENCE AND DATA MINING TECHNIQUES

María de los A. Alonso, Argelio V. de la Cruz and Grettel Barceló

SUMMARY

A methodology is presented that integrates diverse artificial intelligence techniques in order to build a forecast system that determines the convenience of applying a solution with tensoactive properties to an oil well, so as to increase oil production. The methodology begins by processing the data obtained from an experiment consisting in the injection of tensoactive products into a group of wells in an oil field. Different exploratory techniques were used, such as pattern recognition, selection of variables, and methods for the automatic generation

of hypotheses. The information obtained through such processing was modeled in a knowledgebase which, together with the inference machinery of a language named HARies, permitted the construction of a system capable of decision-making in relation to the injection of tensoactive substances and of suggesting the most appropriate technology to be used in each instance. The system was applied to different wells, obtaining in satisfactory results in every case.

PROGNÓSTICO PARA A INJEÇÃO DE TENSOATIVOS EM POÇOS DE PETRÓLEO A PARTIR DE UMA METODOLOGIA QUE INTEGRA TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL E MINERIA DE DADOS

María de los A. Alonso, Argelio V. de la Cruz e Grettel Barceló

RESUMO

Apresenta-se uma metodologia que integra diversas técnicas de inteligência artificial para construir um sistema de prognóstico que determine a conveniência de aplicar uma solução com propriedades tensoativas em um poço de petróleo, com o objetivo de aumentar a produção deste hidrocarboneto. A metodologia começa com o processamento dos dados obtidos de um experimento consistente na injeção de tensoativos a um conjunto de poços em uma jazida. Utilizaram-se diversas técnicas exploratórias de dados, como o reconhecimento de

padrões, a seleção de variáveis e os métodos para geração automatizada de hipóteses. A informação derivada deste processamento foi modelada em uma base de conhecimento que, junto com as máquinas de inferência de uma linguagem denominado HARies, permitiram a construção de um sistema capaz de tomar decisões em relação à injeção de tensoativos e sugerir a tecnologia mais apropriada a usar em cada contexto. O sistema foi aplicado em diversos poços, obtendo, em todos os casos, resultados satisfatórios.