



Departament d'Enginyeria Informàtica i Matemàtiques

Construcció d'un Sistema Expert per caracteritzar el risc de contaminació del sòl utilitzant l'eina *FuzzyTech*

TITULACIÓ: Enginyeria Tècnica en Informàtica de Gestió

AUTORS: Alejandra Godia González
DIRECTORS: Dra. Aida Valls, Dra. Marta Schuhmacher

DATA: Juny / 2008.

AGRAÏMENTS

Aquest projecte s'ha pogut fer gràcies al suport del projecte espanyol CTM2007-64490/TECNO (/Desarrollo de un modelo de evaluación de exposición y riesgo por la aplicación de fangos de depuradoras en suelo agrícola, basado en un sistema experto neuro-fuzzy integrado en SIG/).

Personalment vull agrair a la meua família que m'ha estat donant suport durant tot el projecte, a l'Aida Valls per confiar en mi i tirar en davant aquest projecte, i finalment, a la Marta Schumacher per l'interès que ha mostrat a l'hora d'ajudar-me en tot el possible.

1 INTRODUCCIÓ.....	5
1.1 FUZZYTECH.....	7
1.2 OBJECTIUS PFC.....	9
1.3 ESTRUCTURA DEL DOCUMENT	9
PART I. CONCEPTES TEÒRICS	11
2. Lògica Fuzzy	11
2.1 OPERACIONS	12
2.1.1 Complementació.....	12
2.1.2 Unió	13
2.1.3 Intersecció	14
2.2 Fuzzyficació	15
2.3 Defuzzyficació	16
3. Sistemes experts difusos	20
3.1 Introducció als sistemes experts.....	20
3.2 Adquisició del coneixement	23
3.3 Avantatges i inconvenients dels sistemes experts	25
3.4 Sistemes experts difusos	26
3.4.1 Grau de satisfacció de l'antecedent	26
3.4.2 Aplicació d'una regla.....	28
3.4.3 Obtenció de la conclusió	30
3.5 Avantatges i inconvenients dels sistemes experts difusos	32
3.6 Sistema expert en Fuzzytech	32
3.6.1 Definir les variables d'entrada	32
3.6.2 Definir les variables de sortida	35
3.6.3 Definir les variables intermitges	36
3.6.4 Definir el bloc de regles	37
4. Els sistemes neuro-fuzzy	41
4.1 Funcionament d'una neurona.....	41
4.2 La neurona artificial	42
4.2.1 Topologies	43
4.2.2 Funcionament	44
4.2.3 Aprenentatge.....	44
4.2.4 Avantatges i inconvenients de les xarxes neuronals	48
4.3 Neuro-fuzzy en general	49
4.4 Avantatges i inconvenients del sistema neuro-fuzzy	50
4.5 Neuro-fuzzy a Fuzzytech.....	50
4.5.1 Configuració	50
4.5.2 Execució	52
PART II. Aplicació al cas de la categorització del risc de contaminació d'un sòl	55
5. Adquisició del coneixement	55
5.1 Sistema expert previ	55
5.1.1 Aprofitament	56
5.2 Determinació de les regles	58
6. Disseny i implementació en Fuzzytech	60
6.1 Caracterització de la font de contaminació	60
6.1.1 Descripció del projecte	60
6.1.2 Estructura del sistema.....	60
6.1.3 Variables.....	61
6.1.4 Blocs de regles	69

6.2 VECTORS DE TRANSFERÈNCIA	76
6.2.1 Descripció del projecte	76
6.2.2 Estructura del Sistema	76
6.2.3 Variables	77
6.2.4 Blocs de Regles	82
6.3 SUBJECTES A PROTEGIR	88
6.3.1 Descripció del projecte	88
6.3.2 Estructura del Sistema	88
6.3.3 Variables	89
6.3.4 Bloc de Regles	92
6.4 SISTEMA EXPERT JUNT	96
7. JOC DE PROVES DEL SISTEMA EXPERT	98
7.1 ESTUDI D'UN CAS DETALLADAMENT	98
7.2 CARACTERITZACIÓ DE LA FONT DE CONTAMINACIÓ	101
7.3 VECTORS DE TRANSFERÈNCIA	107
7.4 SUBJECTES A PROTEGIR	110
7.5 SISTEMA EXPERT JUNT	112
8. ESTUDI DE L'ÚS DE LES TÈCNIQUES NEURO-FUZZY EN AQUEST PROBLEMA	116
9. CONCLUSIONS	119
10. VALORACIÓ PERSONAL DEL PROJECTE FINAL DE CARRERA	121
11. TREBALL FUTUR	122
12. REFERÈNCIES	123
12.1 Bibliografia	123

1 INTRODUCCIÓ

El sòl és un recurs natural, però no sempre renovable. Per aquest motiu i per què en les últimes dècades els sòls han patit una forta degradació per la contaminació humana, la Unió Europea va establir la protecció del sòl i el va integrar com un recurs natural amb el qual s'ha de vigilar [CEC, 2002]. Al 1998 es va fer una llei a Espanya on les comunitats autònomes havien de fer un estudi dels seus sòls, mirar quins d'aquests són contaminats i aplicar la tècnica per remeiar-ho més adient.

Per a aconseguir remeiar un sòl el primer pas consisteix en detectar-ne el nivell de risc de contaminació. La valoració del risc de sòls contaminats implica l'avaluació del perill, de la toxicitat i de l'exposició. El resultat de la valoració del risc d'un sòl és un nivell de risc dins d'un rang predeterminat de valors. Així doncs, el procés de valoració del risc consisteix en classificar un sòl dins d'una categoria a partir dels valors d'un conjunt de variable. Fins ara, els mètodes de classificació que avaluaven els sòls contaminants consistien en comparar els resultats analítics de certes variables amb uns valors límit establerts pels governs i determinar si existia risc o no (dues categories). Si els resultats analítics d'un sòl no superaven els límits establerts, es considerava que no era necessària cap acció de remediació.

Com es pot veure, es tracta d'una aproximació massa simple per a un problema mediambiental complex que depèn de molts factors que estan interrelacionats. Per exemple la percepció del nivell de risc canviarà molt si ens fixem amb condicions hidràuliques, o amb els processos físics, químics i biològics associats. Per tant, és necessari tenir en compte tots els paràmetres que influeixen en la contaminació en el sòl per poder determinar el nivell de risc.

A [Asante-Duah, 1998] s'estableix que les activitats que s'han de fer són: la identificació de les fonts de la contaminació, la definició dels camins dels moviments del contaminant, la identificació de poblacions potencialment en risc, la determinació dels productes químics específics de gran preocupació, la determinació de la freqüència de les exposicions del receptor als agents contaminadors ambientals, l'avaluació dels nivells d'exposició del contaminant, la determinació de la resposta del receptor a les exposicions químiques, i la valoració d'impactes o del dany resultant d'exposicions del receptor al producte químic.

La informació recopilada en totes aquestes activitats té un cert grau d'imprecisió associada. La imprecisió apareix quan els valors amb els que treballem no són magnituds exactes, sinó que treballem amb valors aproximats (més o menys 15 mg/litre) o amb valors qualitatius (poca quantitat). No s'ha de confondre imprecisió amb incertesa. Parlem d'incertesa d'un enunciat quan la informació disponible per avaluar-lo no és complerta. Per exemple, no hem pogut mesurar l'alçada del Joan tot i que s'assembla a la del Pere que és de 1,60, per tant l'enunciat "L'altura del Joan és 1,65" creiem que és correcte però és incert ja que no tenim suficient informació per saber si aquest és el valor exacte o no.

Si es tenen en compte tots aquests factors es necessita manegar una gran quantitat de dades simultàniament, que a més presenten imprecisió, això fa que el procés de classificació dels sòls es compliqui. Per aquest motiu, hi ha una necessitat de desenvolupar noves tecnologies computacionals que permetin el procés manegament de la imprecisió i facilitin la caracterització del risc de llocs contaminats.

El grup AGA (Anàlisi i Gestió Ambiental) de la Universitat Rovira i Virgili porta uns quants anys treballant en aquest tema. Un dels projectes de recerca actuals pretén utilitzar eines de suport a la presa de decisions per millorar l'activitat de determinar si un sòl és contaminat.

En aquest sentit, el 2005 es va presentar un altre projecte de final de carrera que utilitzava un sistema expert difús per caracteritzar el risc de contaminació dels sòls [García,2005].

El sistemes difusos permeten definir variables lingüístiques per treballar amb valors imprecisos. Els sistemes experts (SE) imiten el comportament dels experts i es basen en un conjunt de regles que relacionen les variables d'un problema, i permeten classificar-lo.

En aquest sistema expert difús, el coneixement en forma de regles va ser proporcionat per membres del grup AGA. El sistema es va implementar ad-hoc en Java. Les proves inicials van donar molt bones expectatives, doncs resulta ser una aproximació adequada del problema, que permet distingir diversos nivells de risc (es consideraven 4 categories) i a més és una eina fàcil d'utilitzar per usuaris no informàtics [Kumar, 2006].

Ara bé, degut a la complexitat d'aquest domini, els estudis posteriors han donat més informació als experts, que fan que s'hagi de modificar el coneixement introduït en el

sistema. A més, s'ha comprovat que certs casos no es resolen de forma satisfactòria, possiblement pels paràmetres d'implementació del sistema. A l'hora de voler introduir aquests canvis s'han trobat diverses dificultats. En resum, podem dir que aquest software té els següents inconvenients:

- L'execució no és eficient degut a les estructures de dades internes.
- El manteniment del codi és impossible ja que la implementació no segueix els patrons habituals en sistemes de regles, i a més no està documentat.
- No s'utilitza cap motor d'inferència estàndards, i el model implementat no està prou documentat i no es té seguretat de que els paràmetres siguin correctes.

Per aquest motiu, s'ha decidit implementar de nou el sistema expert utilitzant una eina estàndard de construcció de sistemes experts difusos com és l'entorn *FuzzyTech*.

Cal comentar que la dificultat principal en la construcció d'aquest sistema expert és la falta d'una base de casos resolts. Es disposa de molt pocs exemples de casos resolts que indiquin els diferents nivells de risc de contaminant, i tampoc que utilitzin totes les variables. Com s'explicarà més endavant, la manca de casos d'exemple fa que no es pugui utilitzar cap mètode d'Intel·ligència Artificial per aprendre les regles que formen el sistema expert, i que aquestes hagin de ser definides manualment pels experts a partir de la formalització del seu coneixement del domini. Igualment, la manca d'exemples fa que no es puguin usar les eines d'optimització de sistemes experts, com els sistemes *Neuro-Fuzzy*, que necessiten una gran quantitat d'exemples resolts per poder comprovar si les regles són adequades o no.

1.1 FUZZYTECH

FuzzyTech és un entorn de desenvolupament per a sistemes difusos que permet el disseny d'un sistema difús i l'optimització del mateix utilitzant un entrenament basat en xarxes neuronals.

És un programa comercial desenvolupat per INFORM GmbH, que és una companyia fundada pel professor Zimmermann, que va ser un dels més reputables del món dels investigadors en l'àrea de la lògica fuzzy per més de 20 anys. [www.fuzzytech.com].

El *FuzzyTech* facilita la tasca de construir sistemes experts difusos. Aquesta eina incorpora un motor d'inferència i té una interfície gràfica molt potent i fàcil d'entendre. Això permet construir sistemes experts de forma molt ràpida. L'execució del SE és

ràpida i compleix amb els estàndards per al desenvolupament del sistema de la lògica fuzzy.

Aquest entorn de desenvolupament permet:

- La definició del sistema expert: introduint les variables, les regles i els operadors que defineixen el comportament del sistema.
- Debugar un sistema expert: hi ha diversos mètodes per poder comprovar si el sistema expert funciona com tu vols, és a dir pots fer un joc de proves i utilitzar-lo de diferents maneres.
- La documentació d'un sistema expert: el sistema permet generar tota la documentació de forma automàtica i ho guarda en un document on hi ha totes les variables amb les seves gràfiques, totes les regles que hi ha, etc.
- Millora del sistema expert amb tècniques *Neuro-Fuzzy*: és una eina on s'utilitza xarxes neuronals per millorar el sistema expert creat.
- La generació de codi: *fuzzyTech* genera una solució completa de lògica fuzzy en codi font. Es pot utilitzar aquest codi font per compilar i enllaçar amb el codi principal que es tingui, cridant el sistema de lògica fuzzy com una funció simple. Depenent de la edició, es pot generar codi en C, codi font en assemblador per a processadors, o blocs de funció per a controladors de lògica programable.
- Programabilitat i adaptació: *fuzzyTech* ve amb una interfície de programació d'aplicacions que et permet controlar la majoria de les funcions en *fuzzyTech* del teu software. Aquest API és subministrat com un DLL anomenat FTRCU (Unitat de Control Remot de *fuzzyTech*) el qual conté totes les funcions que necessites per integrar completament el *fuzzyTech* amb el teu programa.

Cal destacar que *FuzzyTech* és el líder del mercat en lògica fuzzy ja que ha permès exitosament més aplicacions de lògica fuzzy que qualsevol altra eina. També perquè és autoritzat per molts fabricants de sistema de control. Una altra raó per ser líder és que reparteix més ràpid i més sistemes compactes que qualsevol altra eina de lògica fuzzy.

1.2 OBJECTIUS PFC

L'objectiu d'aquest PFC és construir un sistema expert difús que ens ajudi a decidir quin és el nivell de risc ambiental d'un sòl contaminat utilitzant l'entorn de desenvolupament *FuzzyTech*.

Aquest objectiu es pot descompondre en els següents subobjectius:

- Adquirir els coneixements del tema, és a dir, entendre i aprendre la matèria relacionada amb sistemes experts i amb xarxes neuronals.
- Obtenir i modelar el coneixement del domini, utilitzant com a base del projecte final de carrera “*Sistema experto para la priorización de suelos contaminados*” [García, 2005], i el suport dels experts del grup AGA..
- Aprendre a utilitzar l'entorn de desenvolupament de *FuzzyTech*, tant en la part de sistemes experts com en la de *Neuro-Fuzzy*.
- Dissenyar i construir el Sistema Expert usant totes les eines de *FuzzyTech*.

1.3 ESTRUCTURA DEL DOCUMENT

Aquest document s'estructura en dues parts. La primera presenta els conceptes teòrics que s'han estudiat en el projecte final de carrera, i també presenta l'eina *FuzzyTech* en general. La segona part està centrada en l'aplicació pràctica d'aquests coneixements i d'aquesta eina al cas de la categorització del risc de contaminació del sòl.

En el capítol 2 s'expliquen els conceptes bàsics de la lògica difusa (Fuzzy), que seran utilitzats posteriorment. El capítol 3 està dedicat a l'explicació dels sistemes experts difusos que utilitzarem en aquest projecte. En aquest capítol s'explica com crear el sistema expert utilitzant l'eina *FuzzyTech*. En el capítol 4 es fa una introducció al sistema *Neuro-Fuzzy* on s'especifica el funcionament de les neurones, i on s'explica com crear un sistema neuronal fuzzy amb l'eina *FuzzyTech*. Dintre del capítol 5 podem trobar l'explicació de la estructura del Sistema Expert que hem creat i les diferents formes de crear les regles. En el capítol 6 estan definides i explicades totes les variables i regles utilitzades en el Sistema Expert creat. El capítol 7 està dedicat a comprovar el seu bon funcionament amb els jocs de proves pertinents. En el capítol 8 es fa un estudi d'un petit Sistema Expert dels que hem creat aplicant-li les tècniques del *Neuro-Fuzzy*

per comprovar si millora. Al capítol 9 trobem les conclusions d'aquest PFC. En el capítol 10 la valoració personal del projecte final de carrera realitzat. El capítol 11 està dedicat a possibles millores d'aquest PFC per un futur. I finalment, en el capítol 12 trobem totes les referències utilitzades per elaborar aquest PFC.