

UNIVERSITATEA „LUCIAN BLAGA” DIN SIBIU
INSTITUTUL DE ORGANIZARE A STUDIILOR
DOCTORALE ȘI POSTDOCTORALE

Domeniul de doctorat Cibernetică și Statistică

TEZA DE ABILITARE

Modelare explicativă și predictivă în tehnici
avansate de analiză și optimizare a sistemelor
socio-economice

Candidat:
Conf. univ. dr. Laura Florentina STOICA

Sibiu 2024

Rezumat

Una dintre distincțiile cheie dintre modelele statistice explicative și modelele predictive este diferența în scopul analizei. Un model explicativ este utilizat pentru testarea ipotezelor cauzale, în timp ce un model predictiv este ideal pentru prezicerea noilor observații și evaluarea nivelurilor de predictibilitate. Noțiunea de cauzalitate între construcțiile teoretice este caracteristică modelului explicativ, în timp ce modelul predictiv depinde de asocierile dintre variabilele măsurabile. Modelarea explicativă se bazează pe o abordare retrospectivă pentru a explica relațiile cauzale de bază dintre constructele teoretice, care este atât descriptivă, cât și explicativă. Pe de altă parte, modelarea predictivă este o abordare prospectivă pentru a prezice rezultatul necunoscut pe baza relațiilor de asociere între variabilele măsurabile.

Evoluția bazelor matematice ale modelelor predictive a fost stimulată de creșterea puterii de calcul a calculatoarelor personale și de disponibilitatea seturilor mari de date în era „big data” în care am intrat.

Sistemele socio-economice sunt sisteme mari care includ domenii sociale, economice, educaționale, științifice, tehnologice și ecologice, implicând diverse aspecte ale activităților umane și mulți factori complecși ai mediului de viață.

Complexitatea sistemelor socio-economice este un subiect efervescent în cercetarea academică actuală. Datorită complexității și naturii lor dinamice, nu există o soluție unică, adecvată tuturor problemelor sistemelor socio-economice. În special, multe dintre aceste sisteme implică schimbări dinamice care vizează oamenii și societatea, iar problemele în sine sunt în continuă schimbare și evoluție, ceea ce duce, de asemenea, la absența unor modele analitice generale precise și complete pentru astfel de sisteme.

Scopul principal al tezei de abilitare este explorarea metodelor și tehnicilor de modelare de analiză a complexității și dinamicii sistemelor socio-economice în condițiile tehnologice actuale și găsirea de soluții eficiente pentru optimizarea acestora cu adaptabilitate dinamică. În complementaritate cu modelarea explicativă și predictivă, o abordare evolutivă este utilizată pentru a optimiza modelele predictive și pentru a genera un plan de producție adaptiv. De asemenea, metode formale bazate pe logici temporale au fost folosite pentru a verifica aplicațiile de afaceri modelate ca sisteme multi-agent.

Teza de abilitare intitulată „Modelare explicativă și predictivă în tehnici avansate de analiză și optimizare a sistemelor socio-economice” prezintă o sinteză a activității academice, a rezultatelor activității de cercetare și a perspectivelor de dezvoltare a carierei profesionale a conf. univ. dr. Laura Florentina Stoica.

Teza de abilitare este împărțită în următoarele secțiuni principale: (1) Realizări academice, științifice și profesionale; (2) Rezultatele cercetării științifice; (3) Perspective de dezvoltare academică, științifică și profesională.

Prima secțiune conține realizările academice, științifice și profesionale ale candidatei.

A doua parte conține principalele rezultate ale cercetării candidatei, conexe temei principale a tezei. Rezultatele cercetării sunt grupate în trei subsecțiuni:

1. Calcul evolutiv în suport decizional și optimizarea modelelor predictive;
2. Verificarea formală cu logici temporale a aplicațiilor de afaceri modelate prin sisteme multi-agent;
3. Modele statistice și instrumente de analiză predictivă ca suport decizional computațional pentru creștere economică și sustenabilitate socială.

Tema „Calcul evolutiv în suport decizional și optimizarea modelelor predictive” include rezultatele cercetării publicate în: Utilizarea algoritmilor genetici și simulării ca suport decizional în strategiile de marketing și planificarea producției pe termen lung, Tuningul mașinilor de învățare extremă cu algoritmi genetici.

Calculul evolutiv oferă o capacitate inerentă de a optimiza mai multe obiective simultan, datorită faptului că se bazează pe o populație în care un individ poate codifica mai multe variabile de optimizare. Prin exploatarea acestor avantaje, am dezvoltat o abordare bazată pe utilizarea modelelor de simulare și a algoritmilor genetici pentru generarea unui plan de producție agregat pentru a maximiza profitul total al unei companii. Metodologia descrisă oferă un instrument care ajută conducerea unei companii să decidă care dintre soluțiile posibile este optimă pentru a deveni planul de producție. Sistemul propus preia date dintr-un sistem de business, include un model de simulare și un algoritm genetic bazat pe un operator de co-mutație pentru a oferi suport decizional în scopul determinării deciziilor optime în ceea ce privește planificarea producției pe termen lung și strategiile de marketing.

Am descris o metodă de optimizare a predicțiilor realizate cu mașini de învățare extremă (ELM) prin optimizarea structurii acestora. Metoda se bazează pe algoritmi genetici pentru a determina numărul optim de noduri ascunse și, de asemenea, pentru a determina funcția de activare adecvată. În studiul nostru, ELM-urile au fost optimizate prin algoritmul genetic Breeder. Pentru evaluarea rezultatelor s-a realizat o comparație cu predicțiile obținute prin ecuații de regresie liniară multiplă (MLRE).

Subiectul „Verificarea formală cu logici temporale a aplicațiilor de afaceri modelate prin sisteme multi-agent” include următoarele rezultate ale cercetării: Implementarea unui verificador de modele ATL utilizând concepte de algebră relațională, Verificarea modelelor ATL în cloud, Verificarea agenților JADE utilizând modelele ATL, Verificador de modele CTL distribuit, Modelarea comportamentului brokerului utilizând un agent BDI.

Verificarea unui sistem software implică confirmarea faptului că sistemul în cauză se comportă așa cum a fost proiectat să se comporte. Validarea proiectării implică verificarea dacă designul unui sistem satisface specificarea cerințelor acestuia. Ambele sarcini, verificarea

sistemului și validarea proiectării pot fi realizate în mod temeinic și fiabil folosind metode formale bazate pe modele.

Principala preocupare a metodelor formale în general, și a verificării modelelor în special, este de a contribui la proiectarea sistemelor corecte, fără erori. Detectarea și eliminarea erorilor cât mai devreme posibil în ciclul de proiectare este în mod clar un imperativ economic.

Folosind verificatorul nostru de modele ATL (Alternating-Time Temporal Logic), bazat pe ANTLR (Another Tool for Language Recognition) și expresii algebrice relaționale traduse în interogări SQL, am dezvoltat o metodologie pentru verificarea automată a sistemelor multi-agent, dezvoltate cu framework-ul JADE.

O altă direcție de cercetare prezentată în această subsecțiune se referă la implementarea unui verificator de modele CTL (Computation Tree Logic) original, care permite o proiectare interactivă a modelelor CTL și se bazează pe arhitectura client/server. Sunt disponibile API-uri Java și C# pentru construcția programatică a modelelor de mari dimensiuni.

Pe baza formalismului logicii BDI_CTL - o logică multi-modală care combină logica CTL cu operatori modali pentru reprezentarea stărilor mentale de convingere (B), dorință (D) și intenție (I) - am folosit propriul verificator de modele CTL într-un studiu de caz pentru determinarea comportamentului unui broker folosind un agent BDI.

Subiectul „Modele statistice și instrumente de analiză predictivă ca suport decizional computațional pentru creștere economică și sustenabilitate socială” include rezultatele cercetării publicate în: Orașe inteligente și conștientizarea comunităților durabile în legătură cu programele de răspuns la cerere: procesarea datelor cu analiza factorului de confirmare de ordinul întâi și ierarhic, Descoperirea sensului datelor prin vizualizare interactivă, Analizarea pieței de valori ca pârgie economică, folosind un model calitativ și respectiv cantitativ, Instrument integrat pentru analiză predictivă asistată, Învățare automată automatizată pentru dezvoltarea modelelor predictive ale sindromului metabolic la pacienții cu boală parodontală.

Am propus un model de măsurare statistică folosind o structură de covarianță, în mod specific analiza factorului de confirmare de ordinul întâi (CFA) folosind procedura SAS CALIS pentru a identifica factorii care ar putea contribui la schimbarea atitudinii în cadrul comunităților energetice.

Sustenabilitatea socială înseamnă identificarea și gestionarea impactului afacerilor, atât pozitiv, cât și negativ, asupra oamenilor. Acțiunile pentru realizarea sustenabilității sociale pot debloca noi piețe, pot ajuta la reținerea și atragerea partenerilor de afaceri, pot fi sursa de inovare pentru noi linii de produse sau servicii, pot schimba mentalitatea și conștientizarea consumatorilor de energie electrică.

În simularea întreprinsă, au fost analizate două seturi complexe de date ale unor chestionare create de Comisia Irlandeză pentru Reglementarea Energiei (CER), care demonstrează influența unor variabile exogene asupra itemilor chestionarelor. Rezultatele au demonstrat că există o relație relevantă între factorii social-economici și comportamentali și variabilele observate. În plus,

modelele s-au potrivit bine pe datele respective, conform măsurătorilor realizate prin indicatorii de performanță.

Sustenabilitatea socială este un concept centrat pe om, care vizează asigurarea celei mai înalte calități a vieții în diferite aspecte care influențează viața cotidiană a oamenilor. Mai precis, sustenabilitatea socială poate fi definită ca o măsură a bunăstării umane. Câteva exemple generale se referă la probleme precum egalitatea socială, îngrijirea sănătății, incluziunea, diversitatea în locurile de muncă și educația.

Cercetarea noastră s-a concentrat pe aspectele legate de sănătatea umană, care au un impact direct general asupra bunăstării umane și nu ar trebui ignorate în favoarea prosperității economice pe termen scurt. În domeniul sănătății, asigurarea bunăstării individuale și a comunităților este un component fundamental al sustenabilității sociale. Acest lucru implică promovarea accesului la servicii de sănătate de calitate, încurajarea unor stiluri de viață sănătoase, abordarea determinantelor sociale ale sănătății și promovarea echității în sănătate.

Ne-am propus să dezvoltăm modele avansate de învățare automată care identifică eficient predictorii cheie ai sindromului metabolic, incorporând date despre starea generală de sănătate și bunăstarea generală, punând un accent semnificativ pe explicarea detaliată a predicțiilor generate de modele. Am studiat un grup de 296 de pacienți, internați în SCJU Sibiu, cu vârste cuprinse între 45 și 79 de ani, dintre care 57% aveau sindrom metabolic. Pacienții au răspuns la un chestionar dedicat, împreună cu un chestionar standard EuroQol 5-Dimensiuni 5-Nivele (EQ-5D-5L). EQ-5D este o măsură standardizată a stării de sănătate dezvoltată de Grupul EuroQol în scopul de a oferi o măsură simplă și generică a sănătății pentru evaluarea clinică și economică. Am folosit framework-uri de Învățare Automată Automatizată (AutoML) pentru a construi modele predictive în scopul de a determina care dintre factorii de risc prezintă cea mai robustă asociere cu sindromul metabolic. Pentru a avea încredere în rezultatele furnizate de modelele de învățare automată oferite de conductele AutoML, am folosit valorile SHapley Additive exPlanations (SHAP) pentru interpretarea acestor modele, dintr-o perspectivă globală și locală. Rezultatele obținute confirmă că severitatea bolii parodontale, riscul cardiovascular ridicat și scorul scăzut EQ-5D-5L au cea mai mare impact în apariția sindromului metabolic.

Pe măsură ce tot mai multe companii încep să înțeleagă beneficiile și valorile pe care analiza predictivă le poate aduce, devine evidentă creșterea tendinței de adoptare a acesteia.

În contextul data mining, analiza predictivă se asociază cu analiza statistică pentru a oferi o combinație de tehnici pentru descoperirea cunoștințelor. Analiza predictivă îi ajută pe factorii de decizie să facă alegeri și să rezolve probleme care au un impact de lungă durată.

Analiza predictivă este realizată în esență prin învățare automată și fundamentată prin tehnici statistice pentru modelarea, analiza și deducerea cunoștințelor.

Accentul puternic pus pe metodele de vizualizare a datelor în predicțiile analitice se reflectă în mod evident asupra popularității diagramelor de dispersie în multe industrii. Vizualizările facilitează descoperirea tiparelor și tendințelor și identificarea oportunităților de analiză ulterioară aprofundată.

Interacțiunea în vizualizarea datelor permite explorarea și descoperirea rapidă a tiparelor de date pe care utilizatorul nu le-ar fi intuit fără acest suport. Funcționalitățile analizate în instrumentele și aplicațiile de vizualizare a datelor sunt: *tooltips*, filtrarea datelor, *drill down / drill up* etc.

Industria în ansamblu va continua să evolueze în jurul conceptului *Software Oriented Architecture* (SOA) oferind organizațiilor serviciile de *Predictive-Analytics-as-a-Service* (PAaaS). Comunitatea de cercetare trebuie să se concentreze mai mult pe serviciile SOA decât pe metodele de modelare și perfecționarea algoritmilor, deoarece a fost atins un platou în aceste dimensiuni. Prin urmare, eforturile noastre se vor concentra pe creșterea gradului de disponibilitate și accesibilitate a acestor tehnologii pentru organizații, proiectând și implementând un instrument integrat pentru a furniza predicții analitice în mod asistat (ITAPA) care va funcționa în domeniul previziunii financiare, capabilitatea de modelare predictivă asistată fiind atinsă prin ghidarea utilizatorilor de afaceri în cadrul unui workflow de analiză predictivă.

Am furnizat o perspectivă asupra creșterii economice, bazându-ne pe influența și utilizarea pieței de valori ca pârgie economică. Au fost utilizate două metode: una cantitativă, determinată de un model de regresie liniară multiplă, și una calitativă care desemnează un model vectorial sustenabil pentru generarea creșterii economice. Rezultatele arată o caracteristică probabilistică predominantă a metodelor cantitative, în contrast cu flexibilitatea și complexitatea metodei calitative de obținere a creșterii economice. Metoda calitativă reușește să surprindă o pluralitate de perspective, precum și interrelațiile dintre parametrii determinanți, fiind astfel un sistem complex de ecuații simple, spre deosebire de metoda cantitativă, care se definește ca un sistem simplu de ecuații complexe.

Teza de abilitare se încheie cu secțiunea 3. Aceasta conține perspectivele de dezvoltare a carierei academice, științifice și profesionale ale candidatei.