

Sisteme informatice bazate pe metodologii inovative de digitizare a patrimoniului cultural; studiu de caz: BYZANTION**

Autori: Anicai Ovidiu**,Horia Moga**,Mihai Andrei**

*Proiectul Byzantion este finantat de ANCS/UEFISCDI si cofinantat de SC Institutul pentru Tehnica de calcul ITC SA in programul Parteneriate în Domenii Prioritare PCCA-2011 Tip 2

**SC Institutul pentru Tehnica de calcul ITC SA

Byzantion (<http://byzantion.itc.ro>) este un portal web pentru gestionarea reprezentărilor iconografice in scopul găzduirii și caracterizării de documente organizate într-o ontologie specifică, definita de text și imagini ale manuscriselor bizantine aflate in proprietatea Bibliotecii Academiei Romane;portalul permite:i) accesul gratuit al publicului numai pentru vizualizare și acces privat securizat pentru utilizatorii autorizați, care editează / actualizeaza conținutul portalului;ii) diferențierea între informațiile autorizate, furnizate din surse științifice sau surse comune; de asemenea permite abordarea problemelor de interoperabilitate în ceea ce privește conținutul de metadate / securitate / comunicare a surselor aflate la dispozitia proiectului.

Pentru a oferi imaginea unei regiuni importante cu valoare istorică, estetică, științifică, etnologică sau antropologică, Byzantion permite construirea unor metadate cu descrieri de manuscrise bizantine și versiuni digitizate ale acestora cu scopul de a reînnoi interesul pentru patrimoniul scris și lingvistic in context european.

BYZANTION isi propune:

1. realizarea unei metodologii de extracție / clasificare / recunoaștere / interpretare a manuscriselor bizantine;
2. realizarea unei accesibilitati facile, prin intermediul unui portal,la cultura bizantină;
3. realizarea unei ontologii integrate și consolidate științific pentru cultura bizantină;
4. creșterea eficienței de informații culturale;
5. oferă un sprijin generalizat accesibilitate la manuscrisele bizantine;

6. creștere semnificativă a calității serviciului cultural pentru școli / universități / muzee / biblioteci;
7. promovarea diversității culturale și a înțelegerii reciproce;
8. explorarea de căi alternative și încurajarea de noi tipuri de colaborare;
9. creșterea masei critice a resurselor și a cunoștințelor despre cultura bizantină;

I. Canonul virtual Byzantion

În urma analizei iconografice **Byzantion** își propune să creeze un canon virtual tematic, bazat în egală măsură pe textele erminiilor și pe trăsăturile comune ale reprezentărilor iconografice analizate

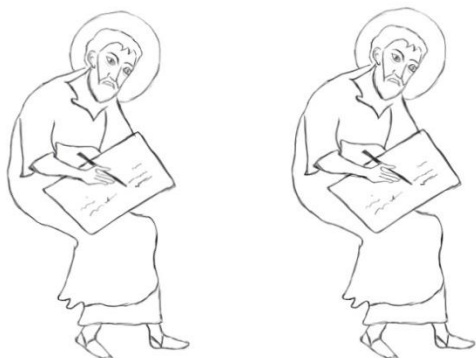


Figura 1 .Exemplificare.Imagini și canon virtual Evangelist Matei

Să admitem că modelul de date constă în interpretarea unei imagini din manuscris și are doar două „rezultate” posibile: fie detectăm, fie nu detectăm o trasatură menționată mai sus.

Suntem obișnuiți să folosim cuvântul „probabilitate” în diverse contexte. De exemplu, ne exprimăm păreri despre probabilitatea ca o anumită trăsătură, și nu alta, să fie cauza trăsăturilor comune pe care le prezintă imaginea care reprezintă evanghelistul. Principala ipoteză a teoriei elementare a probabilităților este următoarea: fiecărui eveniment A – elementar sau nu – i se asociază un număr $P(A)$ cuprins între 0 și 1 (= 100%). Acest număr este numit probabilitatea lui A și nu face altceva decât să exprime „sortii” ca evenimentul A să apară ca rezultat al apariției trăsăturii comune specifice canonului evanghelistului.

Evenimentul (apariția trăsăturii comune) sigur are probabilitatea 1. Evenimentul imposibil are probabilitatea 0. Pentru orice alt eveniment A , oamenii nu-i „cunosc” probabilitatea dar pot s-o **estimeze**.

Cum putem estima probabilitatea unui eveniment A ? Dispunem de trei metode: cea practică, cea logică și cea computațională. Metoda practică constă în repetarea de multe ori a experimentului. Uneori A va apărea ca rezultat, alteori nu. Vom număra de câte ori apare evenimentul A și vom calcula, prin împărțire, frecvența relativă numărul total de încercări

numărul aparițiilor evenimentului astfel:

$$f = \text{numărul aparițiilor evenimentului} / \text{nr total de încercări}$$

Evident, această frecvență relativă depinde puternic de numărul total de încercări. Totuși, oricare frecvență relativă f estimează probabilitatea $P(A)$! Suntem convinși că dacă numărul total de încercări va crește nemărginit, frecvența relativă va tinde spre probabilitatea $P(A)$.

Metoda logică de estimare a probabilităților constă în raționamente asupra condițiilor geometrice și fizice. Metoda computațională pentru obținerea probabilității evenimentului A presupune ca „stim” deja (cu alte cuvinte, am estimat anterior) probabilitățile altor evenimente, legate de A .

O prima și imediată relație este așa-numita **relație a complementului**: dacă A este un eveniment căruia îi „cunoaștem” probabilitatea, atunci „cunoaștem” și probabilitatea complementului sau, din:

$$P(\sim A) = 1 - P(A)$$

Relația de adunare: dacă A și B sunt două evenimente arbitrare, atunci

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) .$$

Cu alte cuvinte, „cunoscând” probabilitățile $P(A)$ și $P(B)$, vom „cunoaște” de asemenea pe $P(A \cup B)$, cu condiția să „cunoaștem” și pe $P(A \cap B)$! Ca un caz particular, atunci când A și B sunt exclusive, avem:

$$P(A \cup B) = P(A) + P(B)$$

Probabilitatea $P(A)$ nu este singurul număr care exprimă „sortii de izbândă”. Unii folosesc în mod frecvent **șansa** lui A , notată cu $O(A)$. Acest număr poate fi exprimat ca un raport între probabilitatea evenimentului A și probabilitatea complementarului sau, adică astfel:

$$O(A) = P(A)/(1-P(A))$$

(valabilă, evident, în caz că $P(A) \neq 1$).

Să considerăm o situație în care un eveniment *a priori* C va influența un eveniment ulterior A , iar apariția lui C ne va obliga să re-estimăm sortii apariției lui A . Vom spune că A **este condiționat** de C .

Să admitem că avem de-a face cu imagini care pot avea sau nu trăsătura comună n , și să luăm în considerare un semn-test S , care ar putea să dea rezultat pozitiv sau negativ.

În general, notația $A | C$ se citește „evenimentul A condiționat de către C ”.

$A | C$ este reprezentată de intersecția celor două părți. Prin urmare,

$$P(A | C) = P(A \cap C)/P(C)$$

(atunci când $P(C) \neq 0$).

Doa evenimente A și D sunt numite **independente** (între ele) dacă niciunul nu condiționează pe celălalt, ceea ce înseamnă că realizarea *a priori* a unuia nu modifică probabilitatea celuilalt:

$$P(A | D) = P(A) \text{ și } P(D | A) = P(D).$$

Bazându-ne pe experiența imagistică, vom putea accepta de exemplu ca „apariția trăsăturii n comune în cadranul inferior stâng al imaginii” și „apariția trăsăturii n comune în cadranul inferior drept al imaginii” ar fi evenimente independente.

Relația evenimentelor independente este următoarea:

$$P(A \cap D) = P(A) \times P(D).$$

(Această relație este valabilă doar pentru evenimente independente!)

Acest model probabilistic a fost inițiat pentru ca descrierea canonului virtual thematic (ex. evanghelistul) să poată cuprinde alte trăsături comune tematice specifice.

II. Abordare teoretică a metodologiei SFR (System feature recognition) în determinarea automată a trasaturilor comune canonului virtual Byzantion



Figura 2 Metodologia SFR utilizând sub-griduri

Byzantion a încercat mai multe modele matematice pentru realizarea unui SFR (System Feature Recognition) pentru definirea trasaturilor comune din imagini necesare canonului virtual Byzantion.

În acest scop a fost nevoie de data mining atât în extragerea datelor care permit o anumită clasificare și apoi catalogarea trasaturilor comune imaginilor manuscriselor utilizate cât și aplicarea unei metode primitive CAPTCHA. Forma cea mai răspândită de CAPTCHA se prezintă sub forma unei imagini care conține trasaturi distorsionate pe care programul trebuie să le recunoască. Byzantion a încercat simularea prin segmentarea unei imagini în grid-uri individuale; de fapt pe această dificultate de programare se bazează versiunile moderne de CAPTCHA.

Mecanismul SFR se bazează pe metoda scorurilor (scoring) - tehnică de analiză discriminantă, constituind o metodă de recunoaștere și interpretare a trasaturilor specifice canonului virtual.

În calculul acestui scoring de recunoaștere trasaturi, pentru toate cele x (un nr relativ de subgriduri întretesute) sub-griduri, intervine un **coeficient mediu de probabilitate condiționată** pentru calculul evenimentului complet de recunoaștere. Exemplu:

- pentru exemplul de mai sus avem completate informații doar pentru sub-grid-ul 1, deci vom avea în vedere doar subgrid 1 (notată cu trasatura completă în subgrid 1)

- pentru subgrid-ul 2 avem 2 trasaturi care reprezintă evenimente în Teorema lui Bayes. Toate aceste 2 evenimente formează **evenimentul complet (ex mana stanga in pozitie de scris)** reprezentând fiecare în parte manifestarea unei anumite trasaturi. Luate împreună, ele pot reprezenta manifestarea unui grup mai larg sau mai restrâns de trasaturi, decât luate separat. De aceea, aceste evenimente se numesc evenimente incomplete și sunt interdependente.
- **pentru calcularea evenimentului complet , s-a introdus o variabilă nouă numită coeficient mediu de probabilitate condiționată (CM) ,** reprezentând ponderea respectivului eveniment incomplet în calculul evenimentului complet
- **notăm evenimentul complet cu TSg=trasatura subgrid**
- notăm cele 2 evenimente incomplete (subgrid 1 & subgrid 2) cu S_{kk} $k=1,2$;
- vom avea 2 TSg pentru fiecare eveniment incomplet, CM_j , $j=1,2$
- Recunoasterea prezumtiva este de 2 trasaturi : b_1, \dots, b_n cu $i=1,2$
- la calcularea probabilității evenimentului complet de recunoastere ce condiționează existența trasaturii b_i ,

$$\text{Probabilitatea}(PS/ b_i)= \text{Probabilitatea}(S_{kk}/ b_i) * CM_j ,$$

unde $i=1$ to 2 ; $j=1$ to 2; $k=1$ to 2 .

Acești CM_j (**coeficienți medii de probabilitate condiționată**) sunt predefiniți cu o anumită valoare care poate fi schimbată de utilizatorul SFR după propria decizie accesând *meniul Coeficienți medii pentru scoring recunoastere trasatura* din, intervenind în calculul scoringului după formula din cadranul de mai sus.

Definitie CM_j : Se numesc **coeficienți medii de probabilitate condiționată** acei parametri cu ajutorul cărora se determină o probabilitate condiționată a unui eveniment complet ce condiționează existența unui eveniment probabil (ex: trasatura comuna 1) , în condițiile în care *ev. complet* e format din mai multe evenimente incomplete interdependente și exprimă probabilitatea **medie** (de la 0% la 100%) ca evenimentul incomplet să fie predominant (să influențeze într-un anumit procent) sau nu în evenimentul complet .

O alta abordare de aplicare a metodologiei SFR a fost prin utilizarea modelului Marchuk

Modelul Marchuk s-a implementat pe două cazuri:

- 1) cand α este constant;
- 2) cand α este funcție periodică de timp (eveniment improbabil de modificare trasaturi comune);

Modelul Marchuk

β) coeficientul reactii
 γ) coeficientul vitezei de
 $\alpha(t)$ - coeficientul nașterii
 μ_c - coeficientul al morții
 μ_f - rata de mortalitate
 η - rata de interacțiune
 τ - timpul de viață
 ρ - timpul de viață
 q - viteza de producție
 η - rata de interacțiune

Modelul Marchuk cu $\alpha(t)$ = constant

$$\begin{cases} \dot{V}(t) = (\beta - \gamma F(t))V(t) \\ \dot{C}(t) = \alpha(t)V(t - \tau)F(t - \tau) - \mu_c(C(t) - C^*) \\ \dot{F}(t) = \rho C(t) - (\mu_f + \eta\gamma V(t))F(t) \end{cases}$$

$V(0)$ - concentrația
 $C(0)$ - concentrația
 $F(0)$ - concentrația

V(t), C(t), F(t) - Modelul Marchuk

Soluție staționară

X = (V, C, F) = (, ,)

Semnificații ale parametrilor modelului Marchuk:

1. $V(t)$, este parametrul A la timpul t (in cazul de fata $t=0$) si coordonate x, y
2. $C(t)$, este parametrul B la timpul t (in cazul de fata $t=0$) si coordonate x, y
3. $F(t)$, este parametrul C la timpul t(in cazul de fata $t=0$) si coordonate x, y

Se considera doua tipuri de reactivitate: reactivitate constanta si reactivitate periodica. Se studiaza sistemul de mai jos:

$$\begin{cases} \dot{V}(t) = (\beta - \gamma F(t))V(t) \\ \dot{C}(t) = \alpha(t)V(t - \tau)F(t - \tau) - \mu_c(C(t) - C^*) \\ \dot{F}(t) = \rho C(t) - (\mu_f + \eta\gamma V(t))F(t) \end{cases} \quad (11)$$

unde:

β este coeficientul parametrul A;

γ este un coeficient al expresiei a probabilitatii de intalnire si interactie a A-B;

$\alpha(t)$ este o constanta sau o functie periodica neteda;

τ este intarzierea constanta a reactiei A;

Solutia este MM poate avea doua solutii stationare:

- una $(0, C^*, F^*)$ ce descrie starea de sanatate a organismului, in care pentru acest punct nu poate avea loc o bifurcare, deoarece valorile caracteristicilor sunt reale.
- O altfel de comportare este posibila pentru a doua solutie stationara a MM., care este:

$$\bar{X} = (\bar{V}, \bar{C}, \bar{F}) = \left(\frac{\mu_c \mu_f (\beta - \gamma F^*)}{\beta (\alpha \varrho - \eta \gamma \mu_c)}, \frac{\alpha \beta \mu_f - \eta \gamma^2 \mu_c C^*}{\gamma (\alpha \varrho - \eta \gamma \mu_c)}, \frac{\beta}{\gamma} \right).$$

Aceasta stare stationara descrie forma cronica a maladii si exista numai daca $\alpha \varrho > \eta \gamma \mu_c$ and $\beta > \gamma F^*$ or $\alpha \varrho < \eta \gamma \mu_c$ si $\beta < \gamma F^*$.

Metoda scoringului poate fi completata cu metoda substituirilor în lanț care permite măsurarea contribuției diferiților factori la formarea sau modificarea rezultatului față de nivelul de bază.

Se aplică în cazul în care relațiile de tip determinist dintre factorii de influență au forma de produs sau raport.

Aplicarea corectă a metodei necesită respectarea următoarelor criterii de ordin procedural:

- substituirea factorilor se face, obligatoriu, în ordinea condiționării lor: întâi factorii cantitativi, apoi factorii cu rol structural iar în final factorii cu rol calitativ;
- operațiunile de substituire se fac succesiv, nu simultan;
- un factor substituit rămâne în această poziție și în operațiile ulterioare.

Unul din marile neajunsuri ale teoremei lui Bayes in utilizarea modelului matematic este acela că trasaturile comune trebuie să fie mutual exclusive, condiție care micșoreaza drastic aria de aplicabilitate a teoremei.

O soluție în acest sens o oferă modelul logistic: se introduce noțiunea de antiprobabilitate și antiprobabilitate condiționată. O metodă este și discriminarea logistică (Anderson, 1982)

3. Concluzii

1. Platforma software BYZANTION este de tip multi-modul pentru a asigura maximum de flexibilitate și funcționalitate: o parte dintre module ruleaza sub portalul BYZANTION, altele ruleaza ca aplicație independentă (de analiza imagistica); Dezvoltarea modulelor de analiza a fost realizata pe baza rezultatelor studiilor desfășurate timp de mai mulți ani de echipele de proiect ale partenerilor (Biblioteca Academiei Române, ITC**).
2. Metodologia SFR a avut o abordare teoretica pentru a reusi realizarea unui mecanism de recunoastere a trasaturilor imaginilor stilizate. Modelul curent de prezentare a descrierii și analizei ilustrației în vederea comparării ei cu ilustrații tematice similare se face printr-o catalogare ierarhizată într-un sistem de catalogare universal acceptat (ex. UNIMARC), în care primul nivel de catalogare este reprezentat de descrierea bibliografică a manuscrisului, iar cel de al doilea nivel de catalogare corespunde descrierii și analizei iconografice a ilustrațiilor manuscrisului, ca părți componente ale acestuia.
3. Din compararea descrierilor și analizelor iconografice ale unor ilustrații reprezentând aceeași temă iconografică s-a construit o **image-numitor comun**, care ”suprapunându-se” ilustrațiilor reale permite utilizatorilor proiectului studii comparative ale similarităților și disimilarităților ilustrațiilor reale, ceea ce creaza premisele unei mai bune cunoașteri a stilurilor diferitelor școli, epoci și spații subsumate arealului bizantin și post-bizantin.
4. Corpusul **imaginilor numitor-comun**, care sunt practic imagini virtuale ale unor teme iconografice constituie **canonul virtual** .

4. Bibliografie

- J. Hale, Theory of Functional Differential Equations Springer, New York (1977)
- U. Forsys Mathematical Model of an Immune System with Random Time of Reaction Applicationes Mathematicae, Wroclaw (1993), pp. 521–536

- Stone, JV (2013). Chapter 1 of book "[Bayes' Rule: A Tutorial Introduction](#)", University of Sheffield, Psychology
- Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin (2003), "Bayesian Data Analysis", Second Edition, CRC Press.
- Peter M. Lee (2012), "Bayesian Statistics: An Introduction", Wiley.
- Hazewinkel, Michiel, ed. (2001), „[Bayes formula](#)“, *Encyclopaedia of Mathematics*, Kluwer Academic Publishers, [ISBN 978-1556080104](#)