Internet a klasifikační metody, 1 přednáška

Tři důležité internetové aplikace klasifikačních metod

volitelný předmět pro magisterské studium

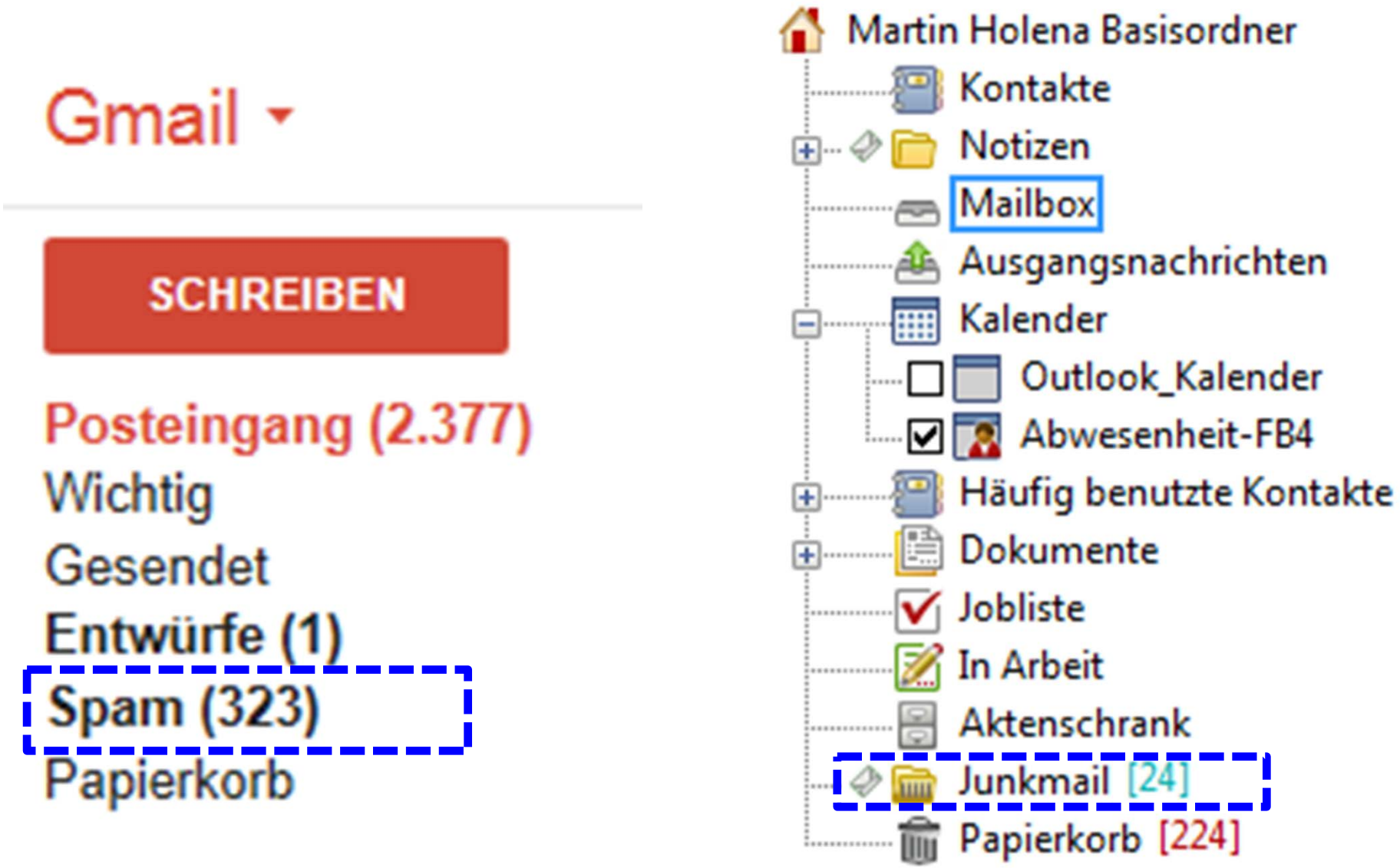
Martin Holeňa

O čem to bude?

- Filtrace spamu
  - klasifikace spamu na základě obsahu + metainformací
- Doporučovací systémy
  - klasifikace při obsahovém filtrování + kolaborativním filtrování
- Systémy pro odhalování hrozeb v síti
  - klasifikace přítomnosti + druhu anomálního chování sítě

Filtrace spamu

- Oddělování spamu od žádoucí pošty (hamu)
- Filtrace se provádí na několika místech:
  - *filtrační servery* předřazené : spam typicky nedoručen
  - servery *příchozí* pošty } spam doručen, ale
  - poštovní *klienti* } explicitně označen | oddělen


Gmail
SCHREIBEN
Posteingang (2.377)
Wichtig
Gesendet
Entwürfe (1)
Spam (323)
Papierkorb
Martin Holena Basisordner
Kontakte
Notizen
Mailbox
Ausgangsnachrichten
Kalender
Outlook_Kalender
Abwesenheit-FB4
Häufig benutzte Kontakte
Dokumente
Jobliste
In Arbeit
Aktenschrank
Junkmail [24]
Papierkorb [224]

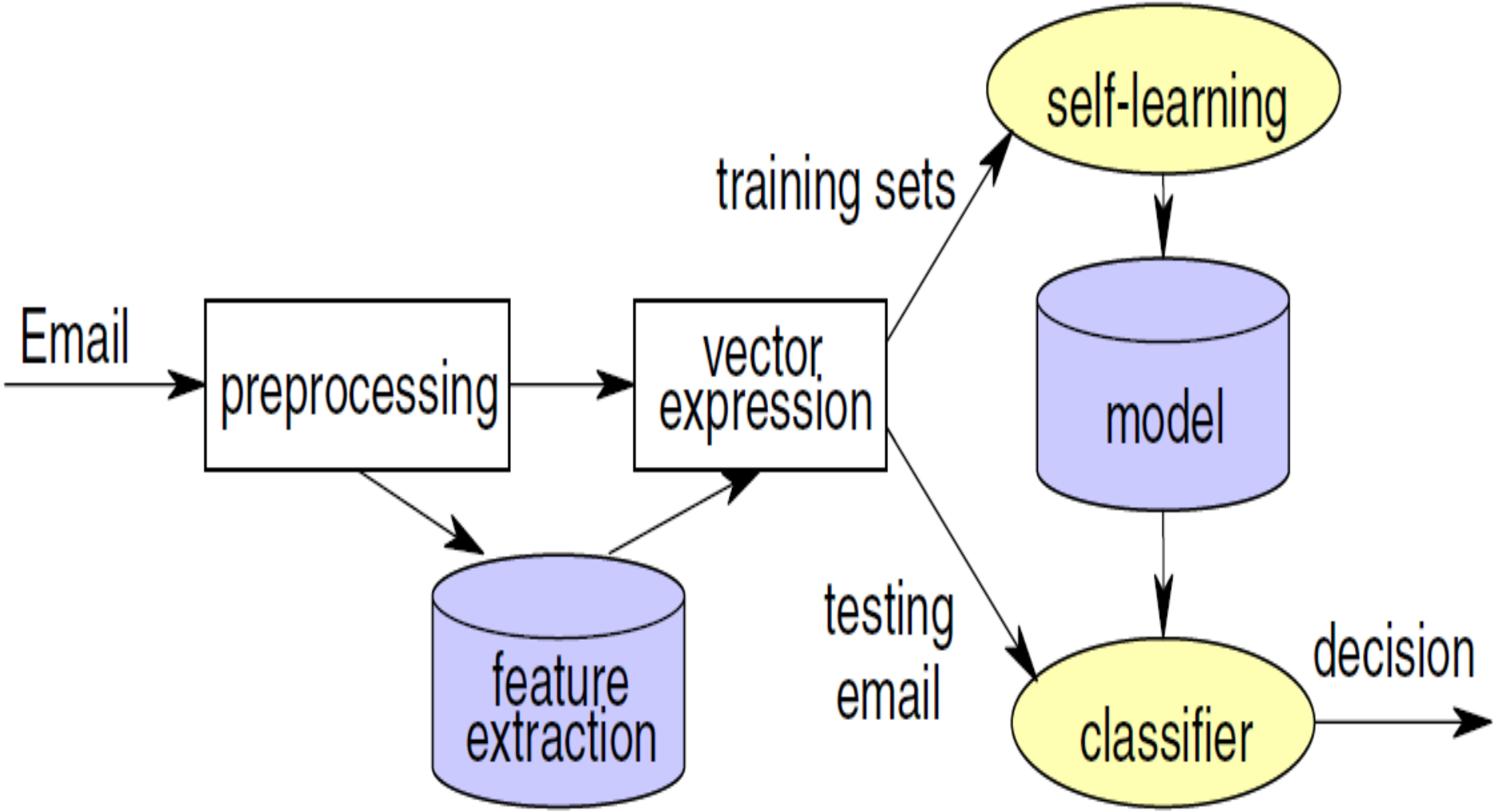
Filtrace spamu

- Oddělování spamu od žádoucí pošty (hamu)
- Filtrace se provádí na několika místech:
  - *filtrační servery* předřazené : spam typicky nedoručen
  - servery *příchozí* pošty } spam doručen, ale
  - poštovní *klienti* } explicitně označen | oddělen
  - servery odchozí pošty: *odchozí (egresní)* filtrace

Proč je filtrace spamu klasifikace?

- Rozpoznání spamu závisí na mnoha příznacích:
  - obsah zprávy: slova, kombinace slov, obrázky
  - postoj adresáta zprávy k jejímu obsahu
  - atributy hlavičky: předmět, odesilatel, podvržené atributy
- *{kombinace hodnot jednotlivých příznaků}→{spam, ham}*
  - každé zobrazení {kombinace}→{třídy} je klasifikátor

Průběh filtrace spamu


self-learning
training sets
Email
preprocessing
vector
expression
model
feature
extraction
testing
email
decision
classifier

Průběh filtrace spamu

1. Extrakce *hodnot příznaků* ze zprávy

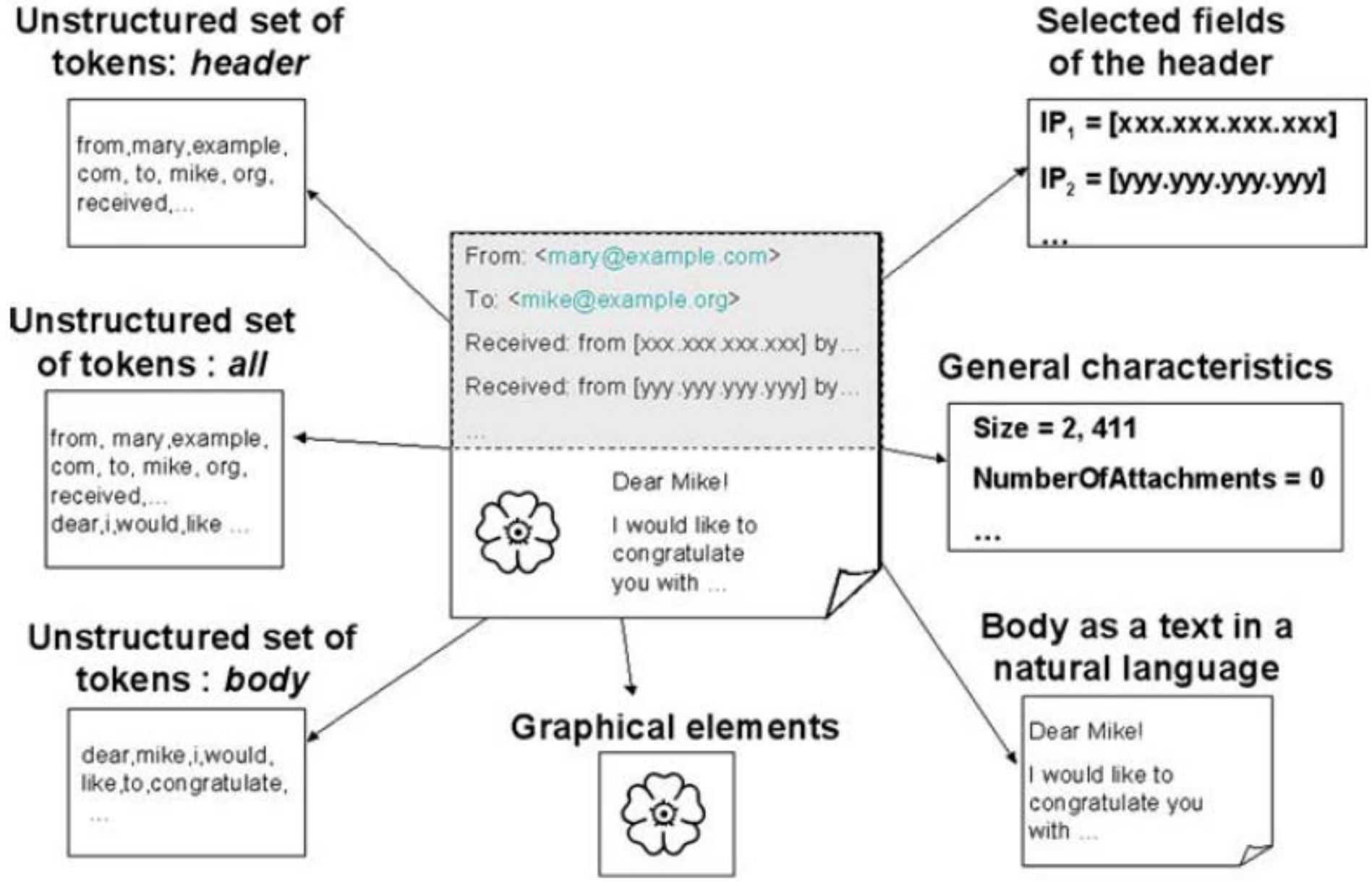
   - použité příznaky určují typ klasifikace:

     podle obsahu zprávy, podle metainformací, hybridní

2. *Učení klasifikátoru* = hledání parametrů jeho modelu

   - každý typ klasifikátoru odpovídá specifickému modelu

3. *Použití* naučeného *klasifikátoru* na nové zprávy


Unstructured set of tokens: *header*
from,mary,example, com, to, mike, org, received,...
Selected fields of the header
IP1 = [xxx.xxx.xxx.xxx]
IP2 = [yyy.yyy.yyy.yyy]
...
From: <mary@example.com>
To: <mike@example.org>
Received: from [xxx.xxx.xxx.xxx] by...
Received: from [yyy.yyy.yyy.yyy] by...
...
Dear Mike!
I would like to congratulate you with ...
Unstructured set of tokens : *all*
from, mary,example, com, to, mike, org, received,... dear,i,would,like ...
General characteristics
Size = 2, 411
NumberOfAttachments = 0
...
Unstructured set of tokens : *body*
dear,mike,i,would, like,to,congratulate, ...
Graphical elements
Body as a text in a natural language
Dear Mike!
I would like to congratulate you with ...

Filtrace podle obsahu

- Používané příznaky: *slova*, *kombinace* slov, *obrázky*
- Učení klasifikátorů filtrujících na základě obsahu:
  - Možnost *doučování* zpětnou vazbou od adresátů

1. *globální* klasifikátor – zpětnou vazbou všech
2. *lokální* klasifikátory – zpětnou vazbou jednotlivých adresátů

(doučování majitelem sexshopu: „levná viagra“ = ham)

Filtrace podle metainformací

- Metainformace (= informace o zprávě) z *hlavičky*
  - doména, kódování (azbuka), podvržená adresa odesilatele
- Metainformace z *průběhu posílání* zpráv
  - *jedné*: nedodržení SMTP ukončení spojení (QUIT),…
  - *více*: stejná/podobná zpráva mnoha adresátům
- Kombinování s obsahem zprávy → *hybridní* filtry

Příklady spamových filtrů

- Součást běžných *mailových klientů* + *webmailových serverů*
  - Outlook, Groupwise, Gmail, Email.cz (Seznam)
- *SpamBayes* – pythonovský program předřazený klientovi
  - Klasifikuje do 3 tříd: spam, ham, nejistota
- *Další* filtry ve formě samostatných programů:
  - DSPAM, SpamAssassin, Bogofilter, ASSP, Babletext

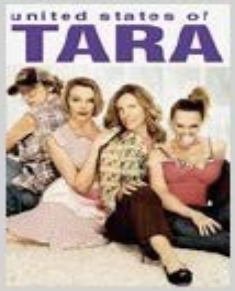
Doporučovací systémy

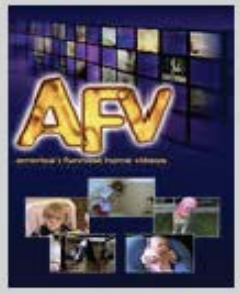
- Doporučují uživateli, co ho pravděpodobně zaujme:

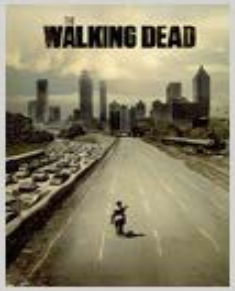
  zboží, literatura, filmy, rekreace, jiní uživatelé

Thousands of movies and TV episodes including these:

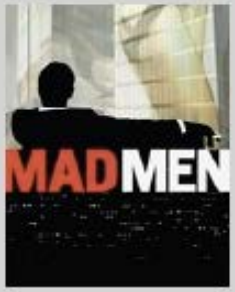
New Arrivals in TV



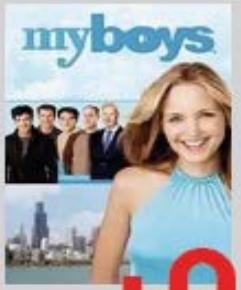




TV Drama



TV Comedy



Children & Family

Netflix

Personen, die Sie vielleicht auch kennen

Sehen Sie Personen aus verschiedenen Bereichen Ihres Berufslebens

Univerzita Karlova v Praze | České vysoké učení | atalysis | FIT CTU in Prague | Vysoká škola ekonomická

Vera Kurkova 2.

scientist at Institute of Computer Science Academy of Sciences of the Czech Republic

Tschechische Republik

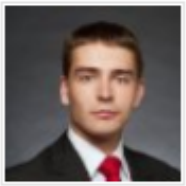
Vernetzen | 4 gemeinsame Kontakte

Ladislav Beneš

chief IT at UI AV

Tschechische Republik

Vernetzen



Martin Štekl 3.

PHP Developer at Designeo Creative s.r.o.

Tschechische Republik

Vernetzen

Petr Hajek 2.

vedecky pracovnik ve společnosti Ustav informatiky AV

Prag, Tschechische Republik

Vernetzen | 1 gemeinsamer Kontakt

Jiri Palek 3.

PhD student at CTU, FNSPE; OSVČ

Bezirk Rakonitz, Tschechische Republik

Vernetzen

Tomáš Kalvoda

teacher, researcher at FIT ČVUT

Prag, Tschechische Republik

Vernetzen

Lukáš Hošek 3.

Programmer, Bohema Interactive Simulations / PhD candidate, MFF UK

Prag, Tschechische Republik

Vernetzen

Jitka Hodalová 2.

secondary school teacher (SOŠ Emila Holuba, s.r.o.)

Bezirk Brünn-Stadt, Tschechische Republik

Vernetzen | 1 gemeinsamer Kontakt

Doporučovací systémy

- Doporučují uživateli, co ho pravděpodobně zaujme:

  zboží, literatura, filmy, rekreace, jiní uživatelé

- Výstup: *1. buď* jen doporučený/é *objekt/y* zájmu

  2. nebo seznam s *odhadem* pravděpodobné *zajímavosti*

- Často *vysvětlují*, proč objekt zájmu doporučují

  → vyšší srozumitelnost doporučení pro uživatele

Frequently Bought Together

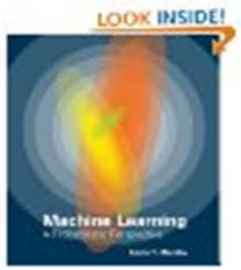
Price For All Three: $181.05

Add all three to Cart | Add all three to Wish List

Show availability and shipping details

- **This item:** Gaussian Processes for Machine Learning (Adaptive Computation and Mac
  Edward Rasmussen Hardcover $35.97
- Pattern Recognition and Machine Learning (Information Science and Statistics) by Chr
  $61.26
- Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Le
  Murphy Hardcover $83.82

Customers Who Bought This Item Also Bought



Machine Learning: A Probabilistic ...
› Kevin P. Murphy
★★★★½ (5)
Hardcover
$83.82

Pattern Recognition and Machine Learning ...
› Christopher M. Bishop
★★★★☆ (67)
Hardcover
$61.26

Probabilistic Graphical Models: Principles and ...
› Daphne Koller
★★★★☆ (16)
Hardcover
$89.89

Proč je doporučování klasifikace?

- Rozhodnutí zda + co doporučit závisí na:
  - *příznacích* popisujících *obsah* objektu zájmu
  - příznacích *chování uživatele*, kterému je doporučováno
  - příznacích *chování ostatních* uživatelů
- → 2 typy rozhodování zda + co doporučit, podle

příznaků: 1. obsahové filtrování, 2. kolaborativní filtrování

Obsahové filtrování

- Doporučovaný objekt vybrán podle *podobnosti* s
  - objekty vyhodnoceými jako *zajímající* uživatele *dosud*:

    explicitně (hodnocení) | *implicitně* (koupě, stažení, streamování)
  - někdy: explicitním stanovením určitych příznaků uživatelem
- *Příklady*: Internet Movie Database – doporučování videí,

  Pandora Radio – skladby „podobné“ původně zvoleným

Kolaborativní filtrování

- Princip: uživatelé s *podobnou historií chování*

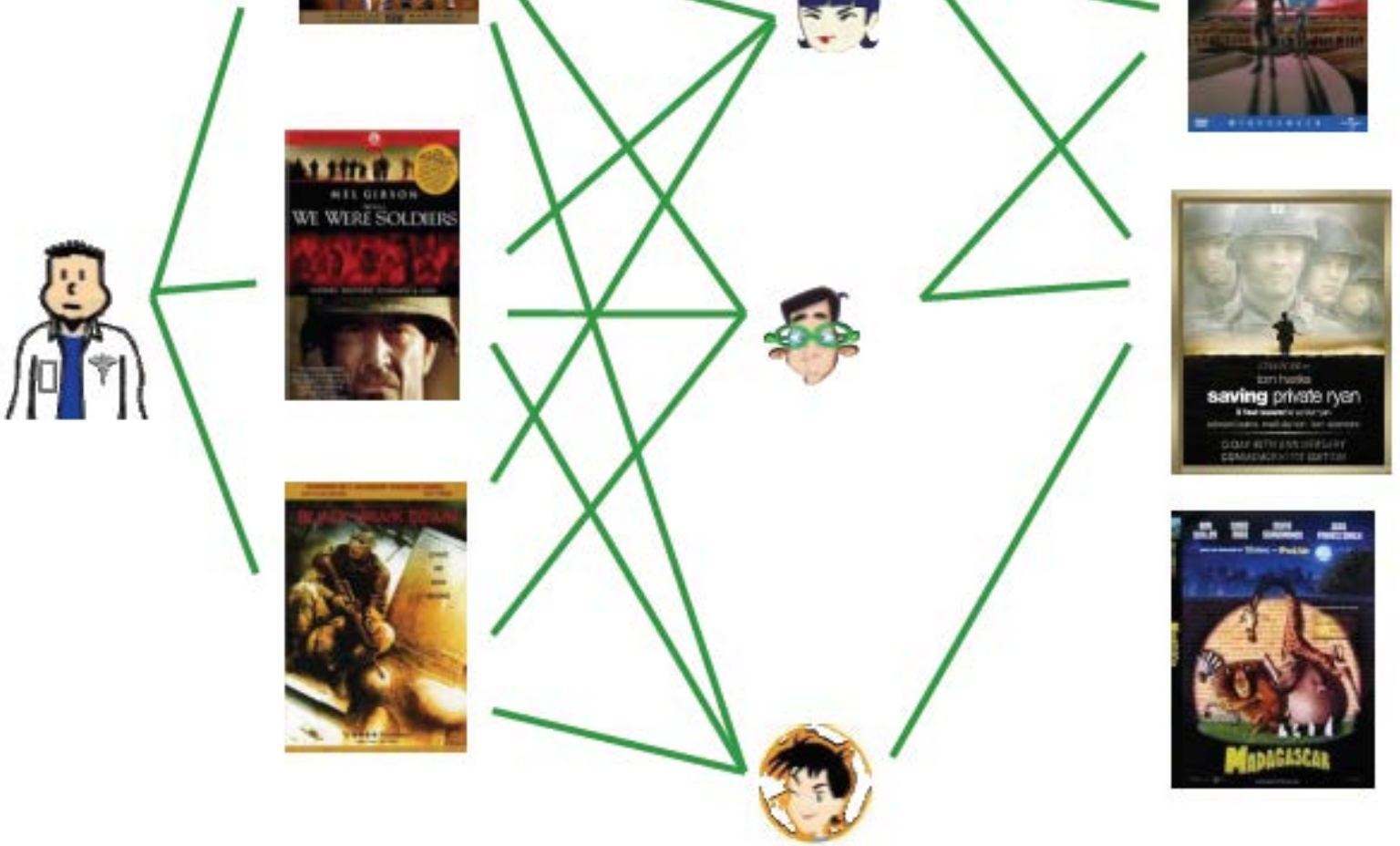
  se často *zajímají* o *podobné* objekty

  - historie chování: zadávané dotazy, prolézané stránky,

    projevený zájem o stejné objekty: explicitně, implicitně

- Problém náběhu systému: o  málo dat

- *Hybridní* systémy: obsahové i kolaborativní filtrování


STAR WARS
MEL GIBSON
WE WERE SOLDIERS
SPIDER-MAN
DUNE
tom hanks
saving private ryen
MADAGASCAR

Kolaborativní a hybridní příklady

- *Amazon*: 2 způsoby využití informace kupující → zboží
- *Last.fm*: skladby dle podobné historie poslouchání
- *LinkedIn*, *Facebook*: doporučují kontakty + členské skupiny

Personen, die Sie vielleicht auch kennen

Sehen Sie Personen aus verschiedenen Bereichen Ihres Berufslebens

Univerzita Karlova v Praze | České vysoké učení | atalysis | FIT CTU in Prague | Vysoká škola ekonomická

Vera Kurkova 2.

scientist at Institute of Computer Science Academy of Sciences of the Czech Republic

Tschechische Republik

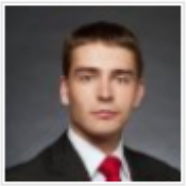
Vernetzen | 4 gemeinsame Kontakte

Ladislav Beneš

chief IT at UI AV

Tschechische Republik

Vernetzen



Martin Štekl 3.

PHP Developer at Designeo Creative s.r.o.

Tschechische Republik

Vernetzen

Petr Hajek 2.

vedecky pracovnik ve společnosti Ustav informatiky AV

Prag, Tschechische Republik

Vernetzen | 1 gemeinsamer Kontakt

Jiri Palek 3.

PhD student at CTU, FNSPE; OSVČ

Bezirk Rakonitz, Tschechische Republik

Vernetzen

Tomáš Kalvoda

teacher, researcher at FIT ČVUT

Prag, Tschechische Republik

Vernetzen

Lukáš Hošek 3.

Programmer, Bohema Interactive Simulations / PhD candidate, MFF UK

Prag, Tschechische Republik

Vernetzen

Jitka Hodalová 2.

secondary school teacher (SOŠ Emila Holuba, s.r.o.)

Bezirk Brünn-Stadt, Tschechische Republik

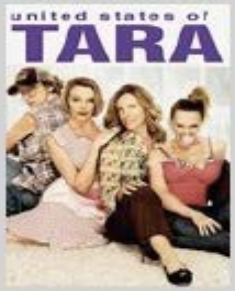
Vernetzen | 1 gemeinsamer Kontakt

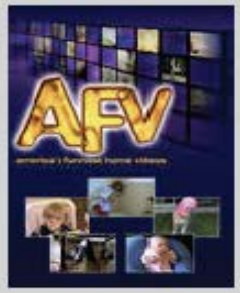
Kolaborativní a hybridní příklady

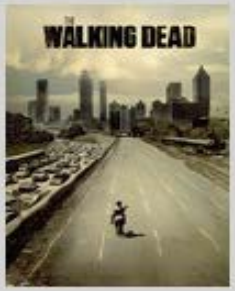
- *Amazon*: 2 způsoby využití informace kupující → zboží
- *Last.fm*: skladby dle podobné historie poslouchání
- *LinkedIn*, *Facebook*: doporučují kontakty + členské skupiny
- *hybridní systémy* Netflix, See This Next:
  - videa podobná těm, co uživatel chválil
  - + dle uživatelské podobnosti historie půjčování + hledání

Thousands of movies and TV episodes including these:

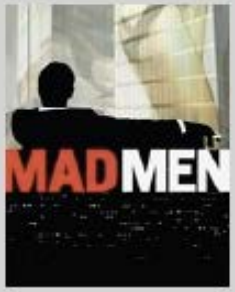
New Arrivals in TV



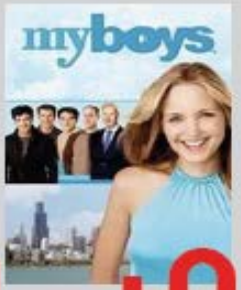




TV Drama



TV Comedy

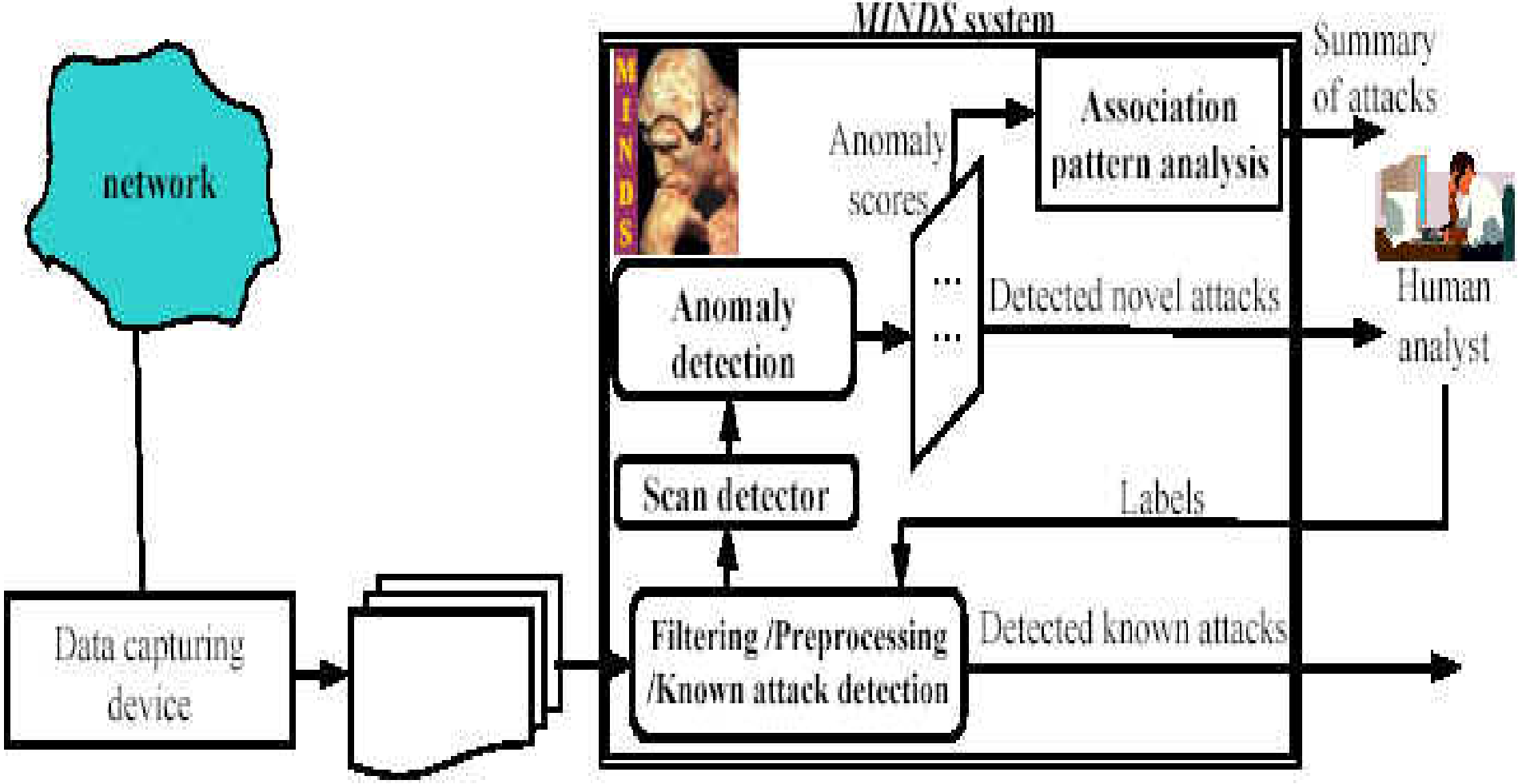


Children & Family

Netflix

Systémy pro odhalení hrozeb v síti

- Systémy odhalující v síti pokusy o:
  - neoprávněný síťový *přístup* (remote to local)
  - neoprávněná *administrátorská práva* (local to root)
  - *zabránění využívání zdrojů* (denial of service)
  - shromažďování informací o slabinách zabezpečení sítě
- Příklad: Minnesota Intrusion Detection System (MINDS)


MINDS system
network
Data capturing device
Filtering /Preprocessing /Known attack detection
Scan detector
Anomaly detection
Anomaly scores
Association pattern analysis
Summary of attacks
Detected novel attacks
Human analyst
Labels
Detected known attacks

Proč je odhalování hrozeb klasifikace?

- K odhalení přítomnosti hrozby v síti

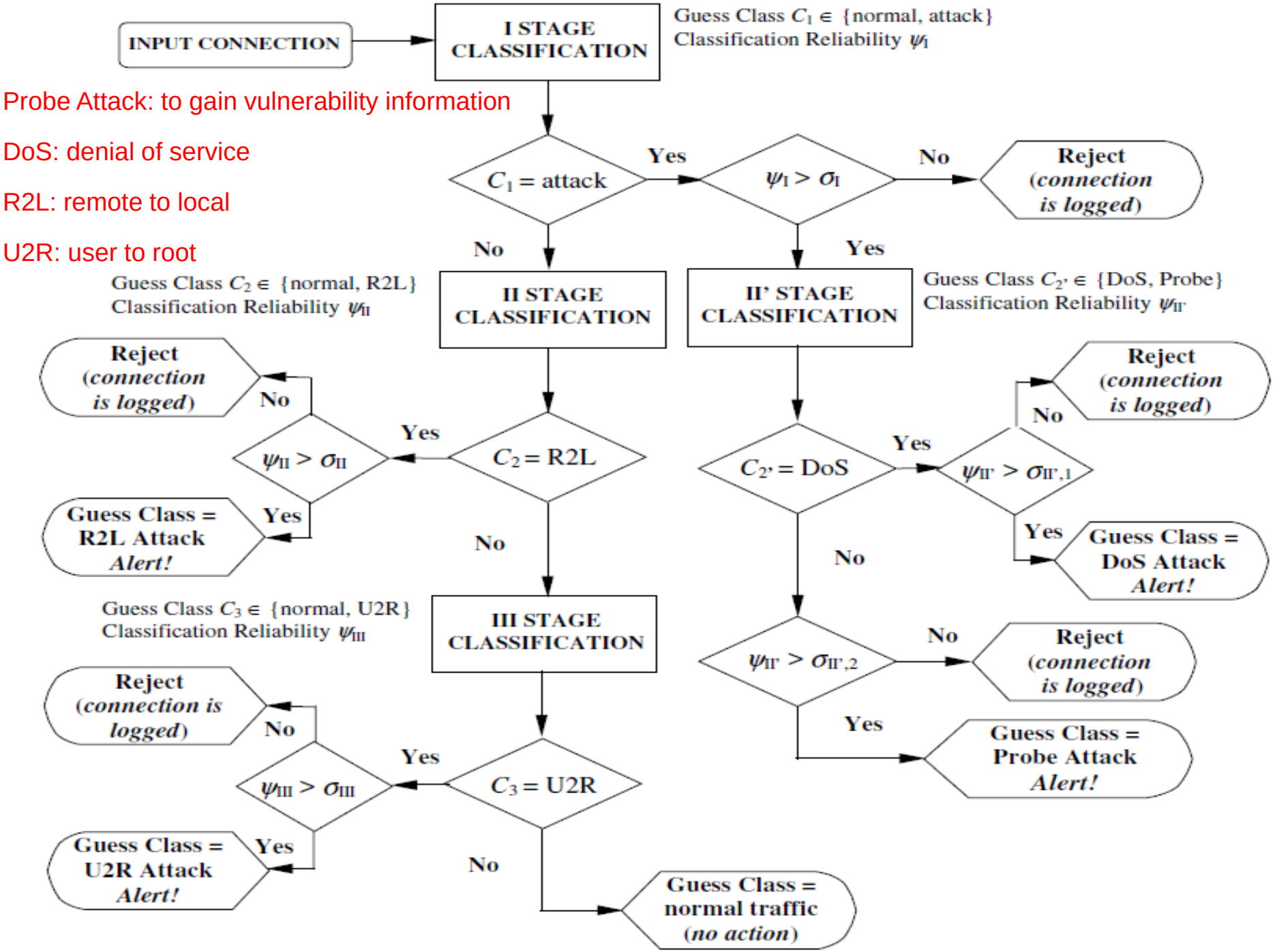
  je třeba znát hodnoty řady příznaků:

  - zdrojové + cílové adresy, protokoly, počty paketů, …

- {kombinace hodnot příznaků} → {normál, anomálie}

  {normál, R2L, L2R, DoS,…}

- *Signatury: známé* , s pravděpodobností ≐100%


INPUT CONNECTION
I STAGE CLASSIFICATION
Guess Class $C_1 \in$ {normal, attack}
Classification Reliability ψ_I
Probe Attack: to gain vulnerability information
DoS: denial of service
R2L: remote to local
U2R: user to root
C_1 = attack
Yes
$\psi_I > \sigma_I$
No
Reject (connection is logged)
No
Yes
Guess Class $C_2 \in$ {normal, R2L}
Classification Reliability ψ_{II}
II STAGE CLASSIFICATION
II' STAGE CLASSIFICATION
Guess Class $C_{2'} \in$ {DoS, Probe}
Classification Reliability $\psi_{II'}$
Reject (connection is logged)
No
$\psi_{II} > \sigma_{II}$
Yes
C_2 = R2L
Guess Class = R2L Attack Alert!
Yes
No
$C_{2'}$ = DoS
Yes
$\psi_{II'} > \sigma_{II',1}$
Reject (connection is logged)
No
Yes
Guess Class = DoS Attack Alert!
No
Guess Class $C_3 \in$ {normal, U2R}
Classification Reliability ψ_{III}
III STAGE CLASSIFICATION
$\psi_{II'} > \sigma_{II',2}$
No
Reject (connection is logged)
Yes
Guess Class = Probe Attack Alert!
Reject (connection is logged)
No
$\psi_{III} > \sigma_{III}$
Yes
C_3 = U2R
Guess Class = U2R Attack Alert!
Yes
No
Guess Class = normal traffic (no action)

Souvislost s data-miningem

- Příznaky používané systémem pro odhalení hrozeb

  ke klasifikaci vytváří tzv. auditová data

- Objevování nových znalostí v = data mining
  - česky: dobývání znalostí, vytěžování dat, datokopectví
  - popisné statistiky, modely časových řad, shlukování, …
- Systém ADAM (Audit Data Analysis & Mining)

Co vám tento předmět chce dát?

- Povědomí o velkém spektru klasifikačních metod
  - o jejich *zajímavosti, důmyslnosti, užitku* pro internet
- Základní vybavení pro jejich tvůrčí používání
  - s *porozuměním*, vhledem pod povrch, samostatně
  - nejenom na problémy řešené jimi *dnes*
  - i problémy, které potkáte za *10−20 roků*

O čem si proto budeme povídat?

1. Základní *koncepty* týkající se klasifikace
2. Hlavní typy klasifikačních *metod*
3. *Přesnost* klasifikace nově příchozích dat
4. *Srozumitelnost* výsledku klasifikace pro uživatele
5. Spojování klasifikátorů do *týmů*

♦ Dnešní *3 aplikace* všechno ilustrují, nikoliv vymezují

Cvičení: peer přednášky

- 6 cvičení střídavě s přednáškami
  - ve výjovém prostředí Matlab (seznámení: 1. cvičení)
- 2 stupně obtížnosti procvičované látky:
  - jednoduché příklady procvičující probírané téma
  - 1 náročnější komplexní úkol jako semestrální práce
  - při cvičeních i konzultace vybraných semestrálek