



Sveučilište u Zagrebu

Fakultet elektrotehnike i
računarstva

Zavod za telekomunikacije

Sustav za automatsku detekciju lažnih poruka

Marin Vuković marin.vukovic@fer.hr

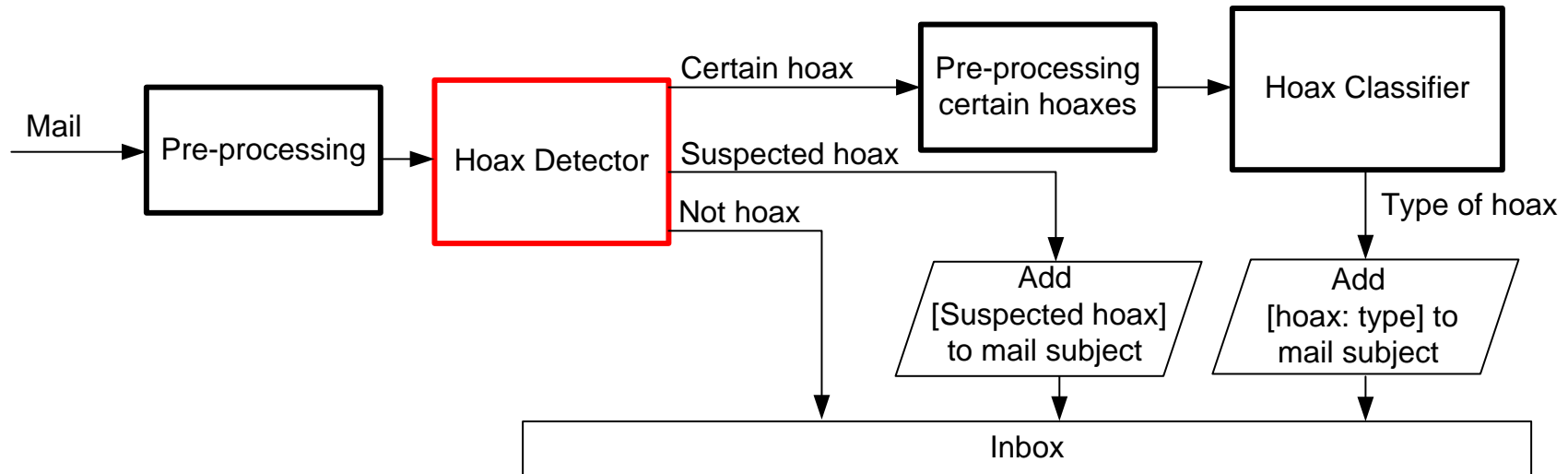
Krešimir Pripužić kresimir.pripuzic@fer.hr

CUC 2010

- ◆ Uvod i motivacija
- ◆ Sustav za uočavanje lažnih poruka
- ◆ Predprocesiranje teksta
- ◆ Evaluacija sustava
- ◆ Zaključak

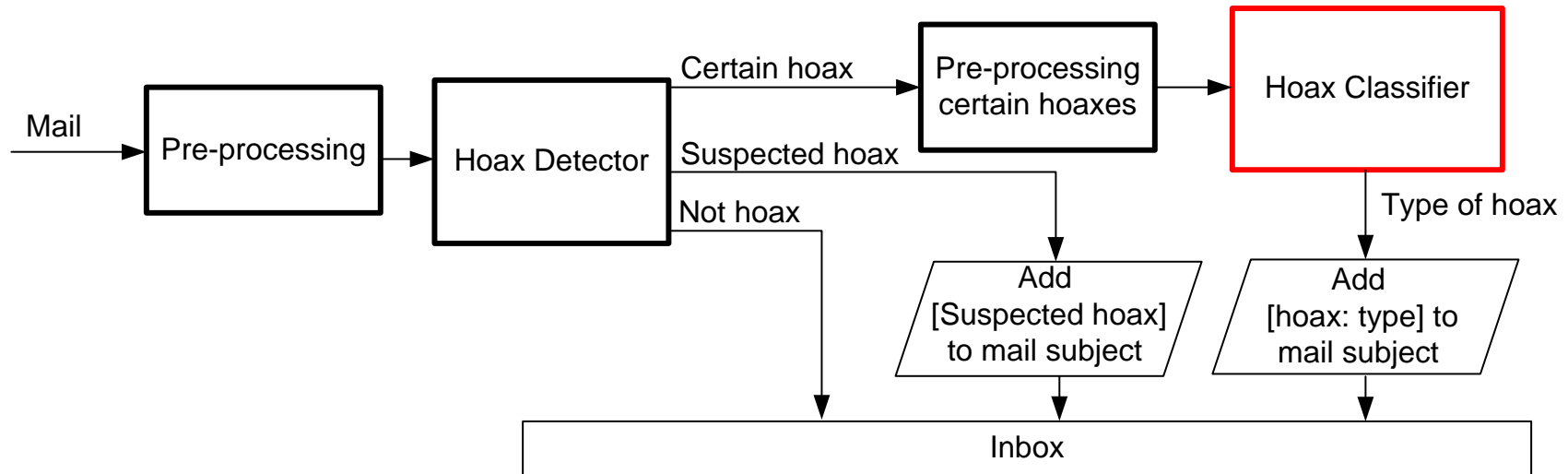
- ◆ Hoax – svaka elektronička poruka koja sadrži lažne informacije s namjerom da korisnika navede na pogrešne zaključke ili radnje

- ◆ Hoaxovi
 - Nisu virusi – ne čine štetu računalima aplikacijama
 - Ali – mogu natjerati korisnike da načine štetu osobama, organizacijama, aplikacijama....
 - Čest rezultat – financijski gubitci
 - Utječu na normalan rad i radnu okolinu



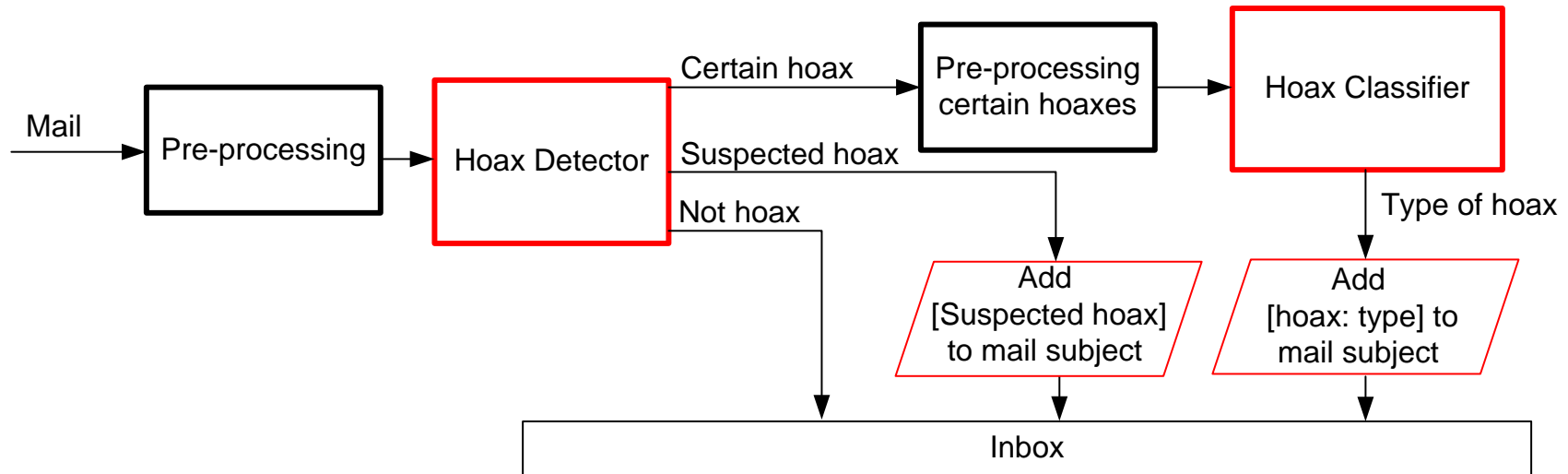
◆ modul Hoax Detector

- FF umjetna neuronska mreža učena s BP algoritmom
- Razlikuje obične poruke od poruka s lažnim sadržajem
- Skup za učenje: 298 lažnih i 1370 običnih poruka
- 1369 ulaznih neurona – odgovara broju učenih n grama iz preprocesiranja
- 2 izlazna neurona – “siguran hoax”, “nije hoax” i “vjerojatan hoax (xy %)”



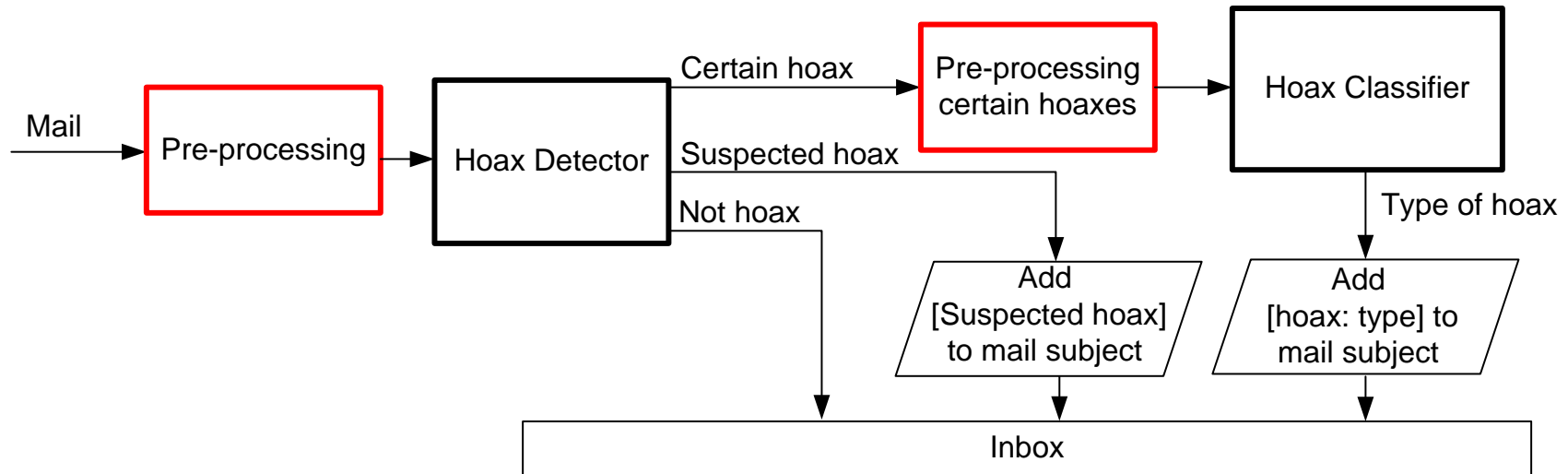
◆ Modul Hoax Classifier

- Arhitektura SOM
- Klasifikacija hoaxeva u poznate grupe
- Skup za učenje: 298 lažnih poruka
- 1880 ulaznih neurona – odgovara broju n-grama iz predprocesiranja sigurnih lažnih poruka
- 20 izlaznih neurona – maksimalno 20 grupa



◆ Prednosti:

- 1. sustav se tijekom rada nadograđuje novim porukama
- 2. korisnicima je dostupna informacija o vrstama lažnih poruka te se na taj način podiže svijest o postojanju i prevenciji takvih poruka



- ◆ 1. Normalizacija teksta
 - Uklanjanje velikih slova
 - Uklanjanje posebnih znakova
 - Uklanjanje dijakritika
- ◆ 2. Tokenizacija teksta
 - ◆ N-gram tokenizacija

- ◆ N-grami ograničeni na $n=\{3,\dots,8\}$
- ◆ Kraći n-grami su preopćeniti i imaju prenisku entropiju
- ◆ Dulji n-grami su prespecifični i nisu pogodni zbog generalizacije kod velikog broja različitih dokumenata
- ◆ Ukupan broj n-grama u obje kolekcije: 79448

Detekcija lažnih poruka:

Pogr. prep.	Vjerojatni hoax	Pogr. ne prep.	Ispravno prep.
4,90%	19,70%	1,54%	73,86%

Klasifikacija lažnih poruka:

ID grupe	Tema poruke	Broj poruka	Udio u skupu lažnih poruka
Group 6	Ulančana pisma – molitve (na hrvatskom)	26	8,70%
Group 9	Ulančana pisma – molitve (na engleskom)	21	7%
Group 11	Traženje pomoći za operacije (na hrvatskom)	20	6,70%
Group 17	Upozorenja o proizvodima (na hrvatskom)	20	6,70%

- ◆ Detekcija i klasifikacija lažnih poruka na temelju sličnosti s poznatim lažnim porukama
- ◆ Podaci se predprocesiraju tehnikama rudarenja podataka
- ◆ Ograničenja
 - Lažna poruka koja nema nikakve sličnosti s poznatim porukama neće biti prepoznata
- ◆ Daljnja istraživanja
 - Razvoj modela kojim bi se smanjio udio vjerojatnih lažnih poruka
 - Implementacija sličnog sustava za druge oblike tekstualne komunikacije

Hvala, pitanja?

