Combining One-Class Classifiers for User Substitution Detection

Oleksiy Mazhelis

mazhelis@jyu.fi
University of Jyväskylä,
Finland

Workshop on Concept Drift
TU/e, Eindhoven, the Netherlands
August 20, 2010

Motivation

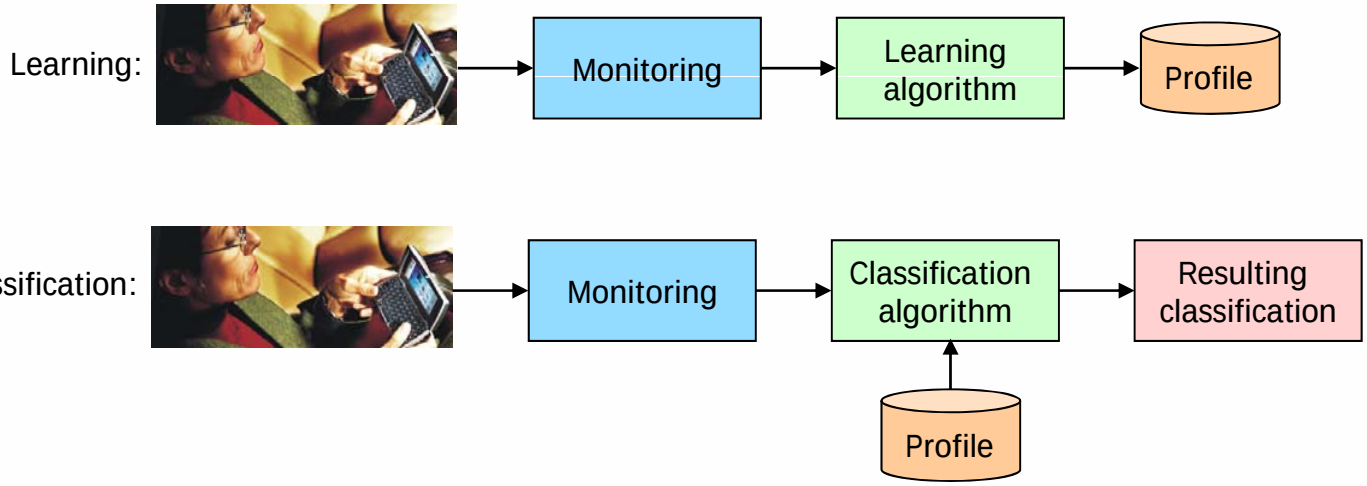
1. Personal names and addresses 86%
2. Business names and addresses 81%
3. Telephone 71%
4. Business diary 59%
5. Personal diary 55%
6. Receive and view emails 45%
7. Entertainment, games, music etc 37%
8. Passwords/PIN numbers 37%
9. Personal images (photographs)
10. Corporate information 27%
11. Bank account details 15%

- Mobile terminals are used to:
  - Store sensitive personal information
  - Store valuable corporate information
  - Receive and view emails…
- Terminals are often lost:
  - “lost mobile devices are just a fact of life”
  - 22% of users experienced a loss or theft of the terminal
  - 85 619 mobile phones lost in taxis in Chicago (the six months period)
  - 54 874 mobile phones lost in taxis in London (the six months period)
- Access to terminals is often poorly controlled
  - “One in four [lost phones or laptops] have absolutely no security on them”
  - 34% of mobile phone users disable PIN authentication
  - 78% of users do not encrypt the information on PDA/Smartphone
- Need for preventive and detective security means

Sources:
http://www.pointsec.com/news/newsreleases/release.cfm?PressId=108
http://www.pointsec.com/news/newsreleases/release.cfm?PressId=386
http://www.pointsec.com/news/newsreleases/release.cfm?PressId=313
http://www.pointsec.com/_file/PointsecNews_3_2006_Global_72dpi.pdf
(Clarke and Furnell 2005)

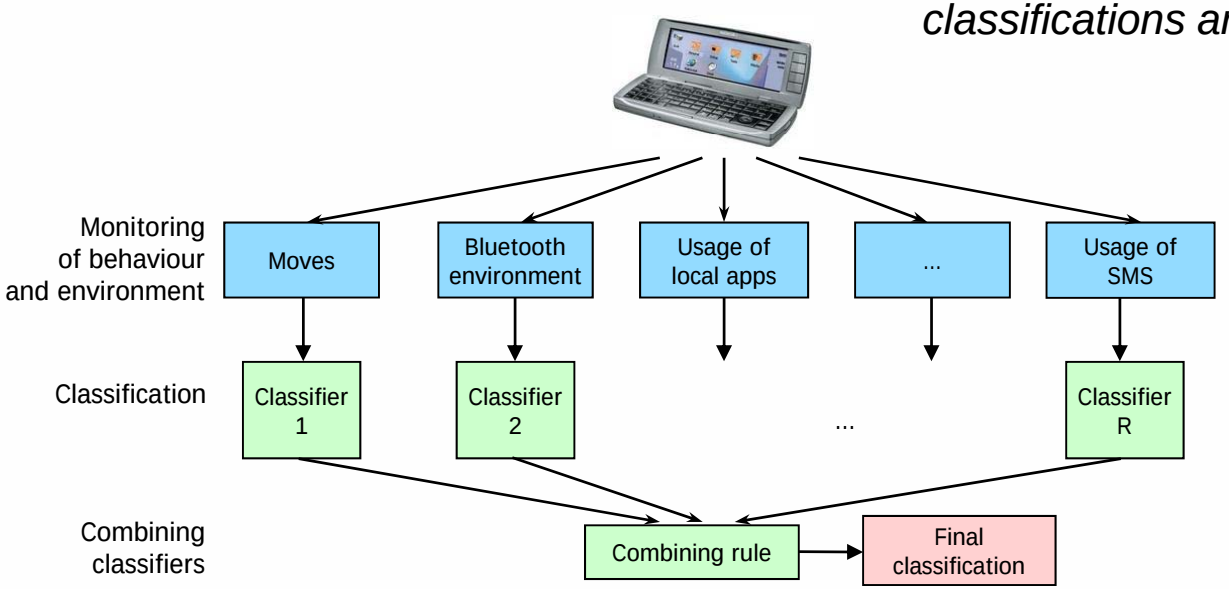
Approach to masquerader detection

- Detection as classification
- A claimant is either user or impostor
- ‘Classification problem: claimant’s behaviour and environment is classified as belonging to the *user class* or to the *impostor class*



Mobile masquerader detection based on combining one-class classifiers

- Data belonging to the user class only may be available for learning
  - Privacy issue
  - Coverage issue
  - *One-class classification*
- Using a single classifier is problematic
  - Different nature or scale of variables
  - Course of dimensionality
  - Partial availability
  - *Use of several classifiers whose classifications are combined*



Questions

- What measures to monitor?
- Which individual classifiers to take?
- How to combine individual classifications?

Measures being monitored

- In total, 9 measures were empirically tested
- Behavioural
  - ARR_CALL Inter-arrival time of calls
  - ARR_SMS Inter-arrival time of SMSs
  - DUR_CALL Duration (length) of calls
  - SPEED 1/timeInCell, where timeInCell < 11 min
  - MOTION Sequences of CellIDs
  - ACT_APP Applications launched at the terminal
- Environmental
  - PLACES CellIDs wherein a terminal is registered > 11 min
  - CONT_NUM ID of terminal being contacted via calls or SMS
  - BT_DEV IDs of Bluetooth devices in the neighbourhood

Individual classifiers

- One-class classifiers - output prior probability estimates $p_i(\mathbf{x}_i|C_U)$
  - Numeric features, temporal relations ignored
  - Symbolic features, temporal relations ignored
  - Symbolic features, temporal relations are important
- Numeric features, temporal relations ignored
  - ARR_CALL, ARR_SMS, DUR_CALL, SPEED
  - K-nearest neighbours classifier
- Symbolic features, temporal relations ignored
  - ACT_APP, PLACES, CONT_NUM, BT_DEV
  - Probability estimators based on histograms
- Symbolic features, temporal relations are important
  - MOTION
  - Classifier based on conventional Markov model

Combining one-class classifiers

- Distance-based schemes
  - Based on Chi-square statistics
  - Based on Hotelling's T^2 statistics
- Rules based on posterior probabilities
  - Various types of vote
  - Product of estimated probabilities (PP rule)
  - Mean of estimated probabilities (MP rule)
  - Modified mean of estimated probabilities rule (modMP rule)

Combining one-class classifiers

- Product of estimated probabilities rule (PP)
- Mean of estimated probabilities rule (MP)
- Modified MP rule (modMP)

$$u_{\mathrm{pp}} = \frac{\prod_{i=1}^{R} p(\mathbf{x}_i|C_U)}{\prod_{i=1}^{R} p(\mathbf{x}_i|C_U) + \prod_{i=1}^{R} p(\mathbf{x}_i|C_I)},$$

$$u_{\mathrm{mp}} = R^{-1} \sum_{i=1}^{R} p(\mathbf{x}_i|C_U),$$

$$u_{\mathrm{mc}} = R^{-1} \sum_{i=1}^{R} \frac{p(\mathbf{x}_i|C_U)}{p(\mathbf{x}_i|C_U) + \overline{p}(\mathbf{x}_i|C_U)},$$

Combining rules: Modified MP rule

- Mean Probabilities (modified) rule
  - Introduced in (Mazhelis & Puuronen, 2004)
  - Takes the prior probability estimates $p_i(\mathbf{x}_i|C_U)$ as input
  - Modifies them (normalises)

$$u_i(p(\mathbf{x}_i|C_U)) = \frac{1}{1+\exp\left(-\ln\frac{p(\mathbf{x}_i|C_U)}{\overline{p}(\mathbf{x}_i|C_U)}\right)} = \frac{p(\mathbf{x}_i|C_U)}{p(\mathbf{x}_i|C_U)+\overline{p}(\mathbf{x}_i|C_U)}$$

  - Averages the modified estimates $u_i(p(\mathbf{x}_i|C_U))$
  - Compares the obtained average with a threshold

Data used

- Dataset is collected during two field studies
  - ContextPhone software used for data collection
  - 3 groups of users (12 users in total) are monitored for some months
  - http://www.cs.helsinki.fi/group/context/data/
  - Behavior and environment of two user groups (nine users) is used
- Data are processed in sliding windows
  - length 1800s, incremented by 900s
- For each user, file is split into training and classification parts
  - *Training phase:* the records are used to build the classifiers' models (for probabilities/likelihoods)
  - *Classification phase:* each classifier outputs the value of probability or likelihood, or non-classification

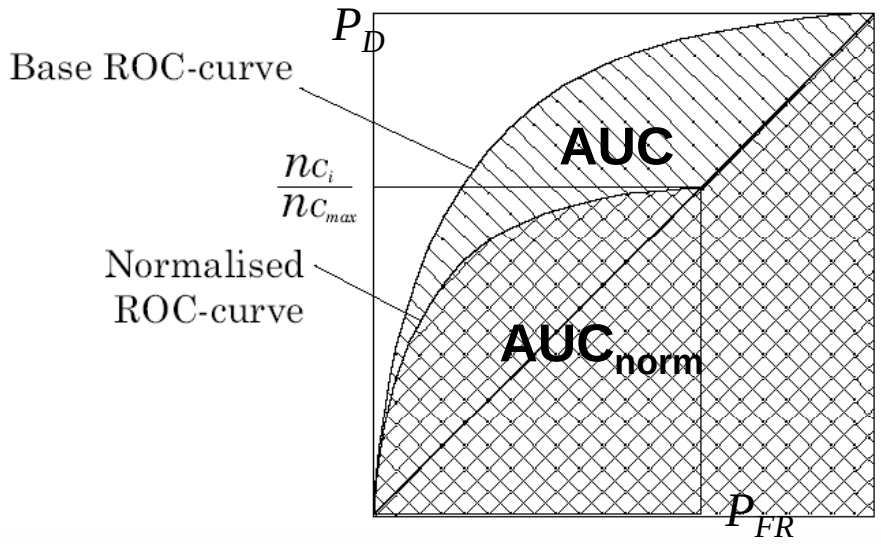
Experimental settings

- Goal of experiments
  I. Assessing accuracy of individual classifiers
  II. Ensembles: selecting classifiers based on ranking
  III. Ensembles: comparing combining rules

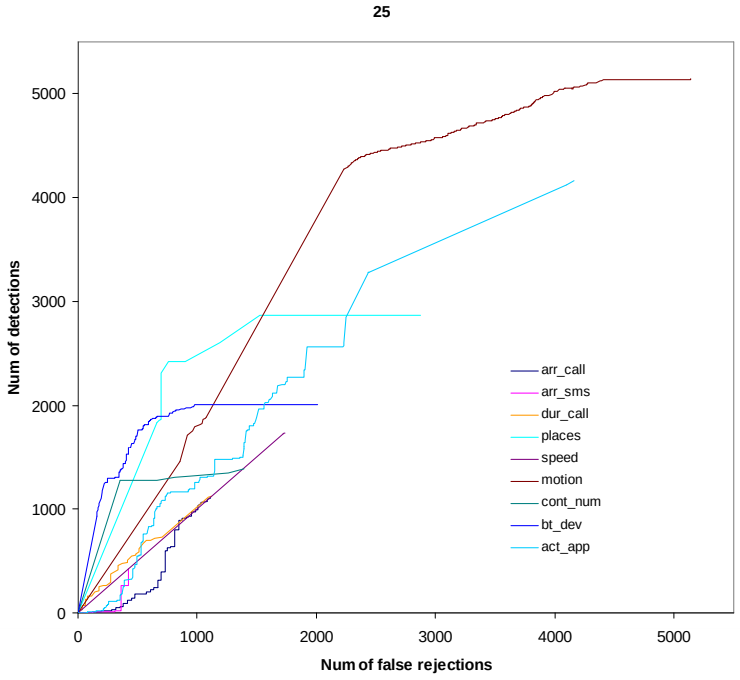
- Evaluation criteria
  – ROC, AUC, partial AUC (p-AUC)
  – Normalized ROC and AUC (averaged)

Normalized ROC curve

- Normalized ROC and normalized AUC (averaged)

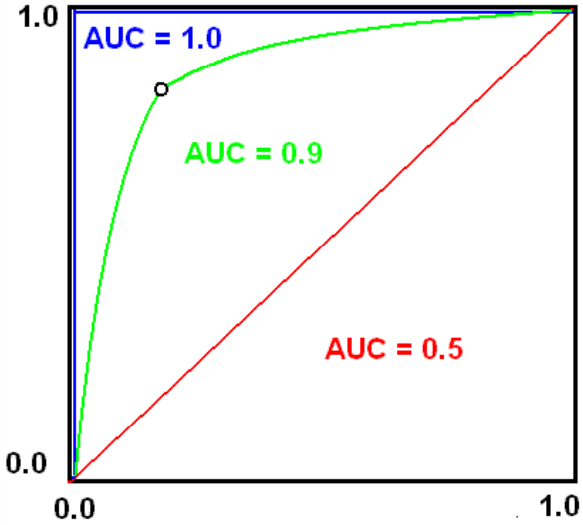


- Example: Normalized ROC-curves for individual classifiers (user 25)



Evaluation criteria

- Accuracy of detection
  - ROC, AUC
  - Partial AUC (p-AUC)
- Comparison
  - modMP vs. MP; modMP vs. PP
  - t-test for means of two paired samples)
  - the Wilcoxon Signed Ranks Test
  - the Sign Test



I. Accuracy of individual classifiers

- Accuracy (averaged AUC and AUC_{norm}) of individual classifiers

Classifier	ARR_CALL	ARR_SMS	DUR_CALL	PLACES	SPEED	MOVE	CONT_NUM	BT_DEV	ACT_APP
AUC	0.501	0.501	0.504	0.743	0.537	0.578	0.806	0.709	0.544
AUC^{norm}	-	-	0.500	0.543	0.506	0.549	0.518	0.536	0.513

- Two classifiers (ARR_CALL and ARR_SMS) are excluded
- The remaining are used to build ensembles of classifiers

- The accuracy ranking differs depending on whether AUC or AUC_{norm} is used

II. Selecting classifiers for an ensemble

- Classifiers are ranked based on
  - AUC
  - $AUC_{Normalized}$
  - $(AUC - 0{,}5) \times$ #classifications

Classifier	DUR_CALL	PLACES	SPEED	MOVE	CONT_NUM	BT_DEV	ACT_APP
Rank acc. AUC	7	2	6	4	1	3	5
Rank acc. AUC_{norm}	7	2	6	1	4	3	5
Rank acc. AUC_{norm} x n_C	7	1	6	4	3	2	5

II. Selecting classifiers for an ensemble (cont)

- Accuracy (averaged AUC and AUC_{norm}) of classifier ensembles
- Ensemble gives output, if one or more individual classification is available

Classifiers combined	AUC	AUC_{norm}
PLACES+CONT_NUM	0.7490	0.5742
PLACES+BT_DEV	0.7521	0.6080
CONT_NUM+BT_DEV	0.7316	0.5719
PLACES+MOVE	0.6119	0.5738
PLACES+MOVE+BT_DEV	0.6788	0.6397
PLACES+CONT_NUM+BT_DEV	**0.7637**	**0.6452**
PLACES+CONT_NUM+BT_DEV+ACT_APP	0.7139	0.6673
PLACES+MOVE+CONT_NUM+BT_DEV	**0.6904**	**0.6687**
PLACES+MOVE+CONT_NUM+BT_DEV+ACT_APP	**0.6854**	**0.6849**
DUR_CALL+PLACES+SPEED+MOVE+CONT_NUM+BT_DEV+ACT_APP	0.6658	0.6658

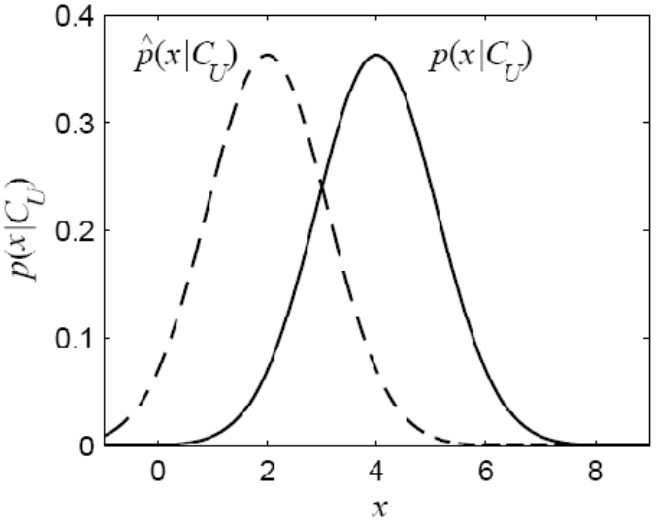
- Accuracy (AUC_{norm}) grows according to the third rank: (AUC – 0,5) × #classifications
- Best accuracy is achieved with five classifiers, deteriorates after that
- Best ROC for 5 classifiers overpass best ROC for 3 classifiers only for large FR

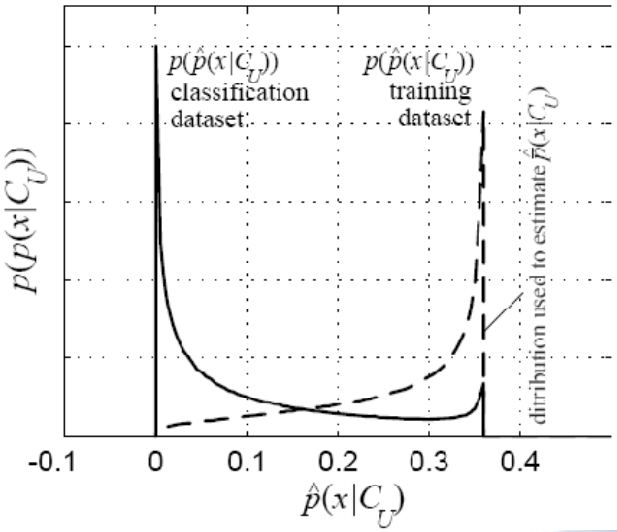
III. Comparing combining rules

- Comparing rules:
  - i) modMP vs. MP; ii) modMP vs PP
  - 2-5 classifiers/ensemble: PLACES, MOVE, CONT_NUM, BT_DEV, ACT_APP
  - Full and partial AUC
- Full AUC:
  - 2 classifiers:
    - ModMP outperforms PP
    - ModMP and MP give similar accuracy
  - 3-5 classifiers:
    - ModMP outperforms MP
    - ModMP and PP give similar accuracy
- Partial AUC (p=0.3)
  - 2-3 classifiers:
    - ModMP outperforms PP
    - ModMP and MP give similar accuracy
  - 4-5 classifiers
    - ModMP outperforms MP
    - ModMP and PP give similar accuracy

III. Comparing combining rules (cont)

- Comparing modMP vs. MP & PP
  - Modified MP rule provides more accurate than or as accurate results as the other rules
- Sensitivity to estimation errors
  - Modified MP rule has one extra parameter to estimate
  - Accuracy of modified MP rule depends on how well its parameter is estimated





Conclusions and further work

- ☑ Differentiating users by behaviour and environment is possible
- ☑ Ensemble to include accurate and “productive” classifiers
- ☑ Three combining rules (MP, PP, modMP) compared
- ☑ Mobile MP rule appears the most reasonable one
- ☑ Accurate estimation of the parameter of modMP rule is important (e.g. by using incremental estimation)

- ☐ Comparing with other combining schemes
- ☐ Validating results using other datasets
- ☐ Improving individual classifiers
- ☐ Response to detected attacks
- ☐ User acceptability tests

Thank you!