IP レピュテーションの統合方法

ICSS/IA 6月研究会

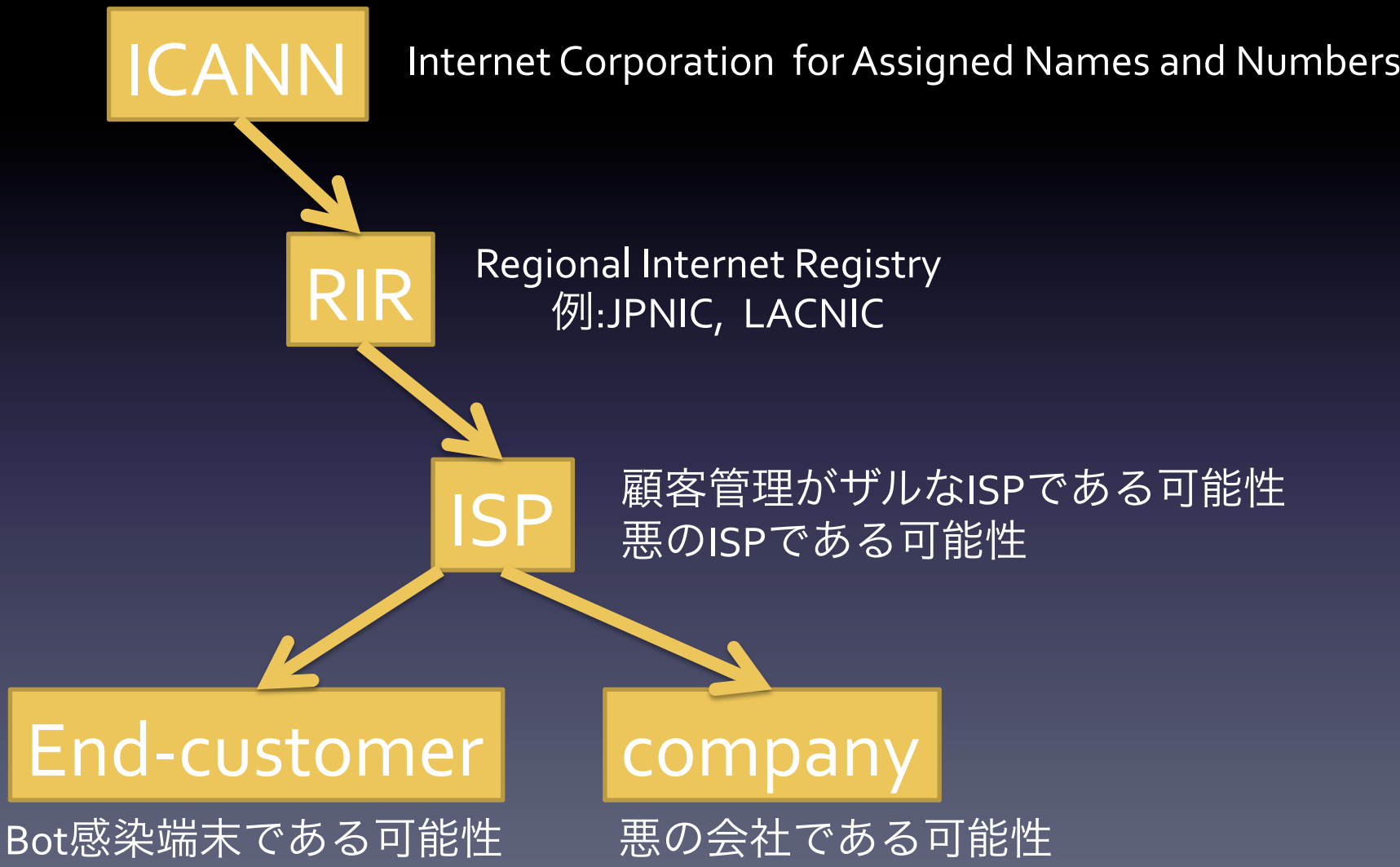
NTTサービスインテグレーション基盤研究所

森　佐藤　高橋　木村　石橋

IP reputation とは？

- IPアドレスの評判 (reputation)データベース
- 端末サーバの評判をIPアドレスによって判別する
- IPアドレスの配分方法により，アドレスブロック毎にそのブロックを使う組織の特性が強く出る傾向があることを利用
  - ある種の「偏見」である可能性もある

IPアドレスの配分


ICANN
Internet Corporation for Assigned Names and Numbers
RIR
Regional Internet Registry
例:JPNIC, LACNIC
ISP
顧客管理がザルなISPである可能性
悪のISPである可能性
End-customer
company
Bot感染端末である可能性
悪の会社である可能性

IP reputation system

- IPアドレスに関する評判の問い合わせに対し，データベースを参照して結果を応答するシステム
- 例)
  - DNSBL (DNS-based block list)
  - DNSWL (DNS-based white list)

Public な BL/WL

- DNSBL
  - Barracudacentral.org
  - Spamhaus.org
  - Uceprotect.net
  - Sorbs.net
  - Abuseat.org
  - Five-ten-sg.com
  - Spamrats.com
  - Anti-spam.org.cn
  - …
- DNSWL
  - Dnswl.org

非常に多数のリストが存在する

商用 IP reputation

- Spamhaus
- Senderbase
- Sophos
- Senderscore
- Commtouch

カスタム BL/WL

- ローカルで観測した情報の履歴を元に，カスタムのIP reputationを構築するアプローチも有効である
  - Esquivel, Mori, Akella, “On the effectiveness of IP reputation for spam filtering”, COMSNETS 2010

疑問

複数の reputation (予想)をどう使えばよいか？

Possible solutions

- 一番精度が良い reputation system を採用
- 多数決（投票）で決める
- その他？

本研究の課題とアプローチ

- 課題：
  - 複数IP reputationの出力をどのように組み合わせるとベストな精度を得ることができるかを明らかにする
    - Reputation of reputation systems
- アプローチ：
  - 投票，パターンマッチング，機械学習による判定方法の提案と実データによる評価

関連研究

- IP reputation の有効性については多数ある
- IP reputationの最適な組み合わせ方法については（著者らの知る限り）存在しない
  - 協調フィルタリング?
  - オンライン学習アルゴリズム

問題の定式化

- IPアドレス一つにつき，m 個の reputation system からの出力 {b1, b2, …, bm} を得る
- bi = {-1, 1}
  - -1 であれば negative (リスト掲載有り),
  - +1 であれば positive (リスト掲載無し)
- 既知のIPアドレスに対し，ラベル Ci={-1,1} を付与
  - -1 であれば white （良性）
  - +1 であれば black（悪性）
- 未知のIPアドレスに対する評判列 B より，ラベルを推定する

提案手法

- 投票 (SV)
- パターンマッチング (PM)
- 教師付き機械学習 (SVM)

単純投票

- 評判列に対し，下記のような出力 y を定義

$$y = \sum_{i=1}^{m} \theta(b_i)$$

$$\theta(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$$

- 出力yが閾値 y* を超えたら C=+1 と判定
  - 閾値以上の投票があった
- 閾値はラベルつきデータを用いて学習

パターンマッチング

- 過去に観測した投票列と同じであれば，その投票列のラベルを採用．異なる場合は距離が一番近い投票列のラベルを採用.

- 距離はコサイン類似度を利用し，距離の値(0~1の値)をその判定に対する確度とする.

機械学習

- SVM (Support Vector Machine)を利用
- 投票列とラベルから成る教師付き信号を用いてSVM を訓練し，ラベルなしデータに対して
- SVMの出力は二値であるが，距離の情報を用いて確率値（スコア）にマッピングする
- SVMパラメタはグリッド探索により最適化

データ (1)

- 企業における電子メール配送ログ
- 各メッセージにスパムであるか通常であるかのラベルがついている
  - ラベルの正当性は高いと仮定
- 1ヶ月のログより，スパムだけを10通以上送ったIPアドレス 3705 と，ハムだけを10通以上送ったIPアドレス 2569 を抽出

データ(2)

- 公開されている81のDNSBLより，前記のIPアドレスのカバー率が高い15のDNSBLを抽出

b.barracudacentral.org	cbl.abuseat.org
zen.spamhaus.org	blackholes.five-ten-sg.com
pbl.spamhaus.org	dyna.spamrats.com
xbl.spamhaus.org	dnsbl.inps.de
dnsbl-2.uceprotect.net	cdl.anti-spam.org.cn
dnsbl-3.uceprotect.net	residential.block.transip.nl
spam.dnsbl.sorbs.net	noptr.spamrats.com
dnsbl.sorbs.net	

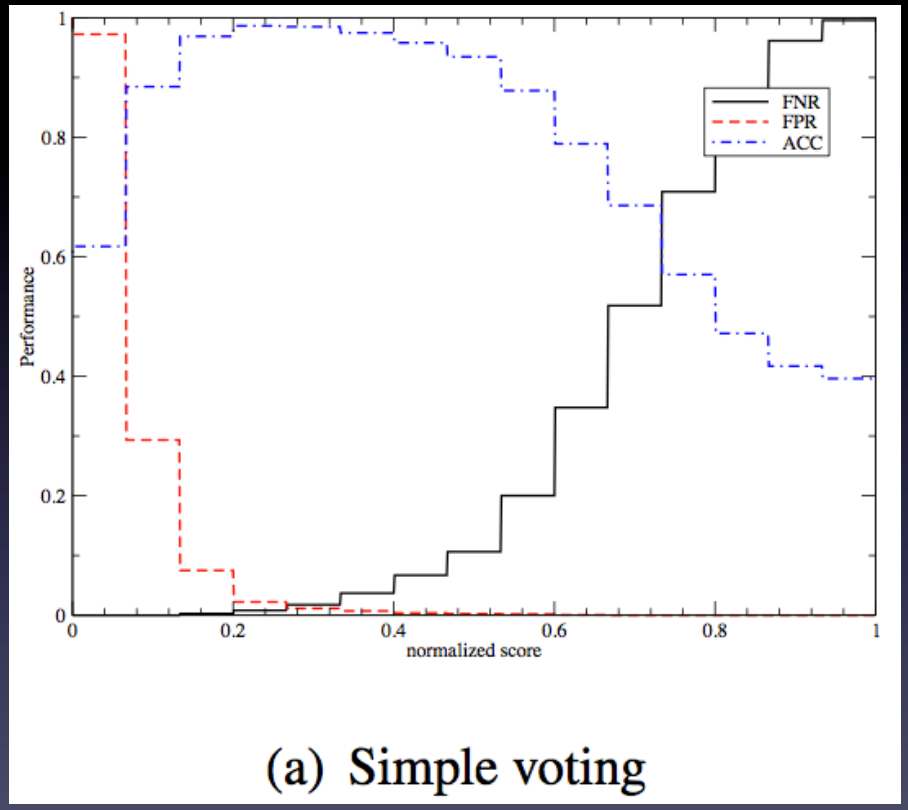
評価方法

- 10-fold cross validation
- ラベルつきデータをランダムに10分割
  - うち9を訓練データ，1をテストデータ
  - この組み合わせを10通り実施し，精度の平均，標準偏差を算出

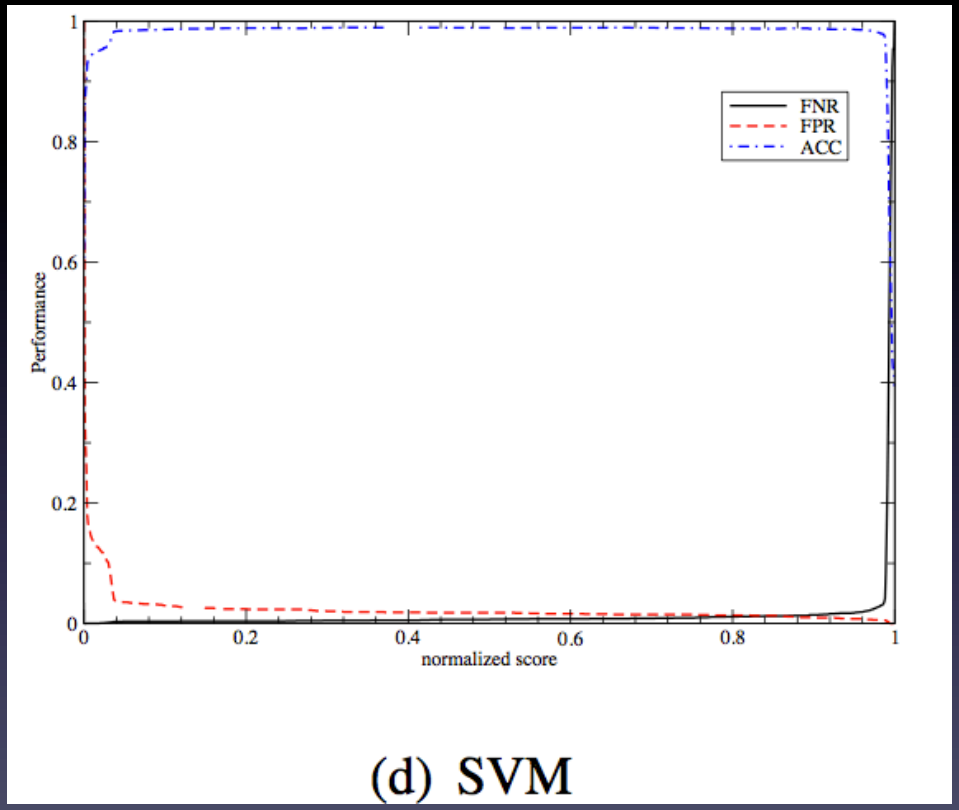
精度の指標

- FPR: False Positive Rate
  - 良性IPアドレスを悪性判定した割合
- FNR: False Negative Rate
  - 悪性IPアドレスを良性判定した割合
- ACC: Accuracy
  - IPアドレスを正しく判定した割合

FPR, FNR のトレードオフ



(a) Simple voting



(d) SVM

閾値によってトレードオフの調節が可能．良い精度を得るためには正しく調節する必要あり

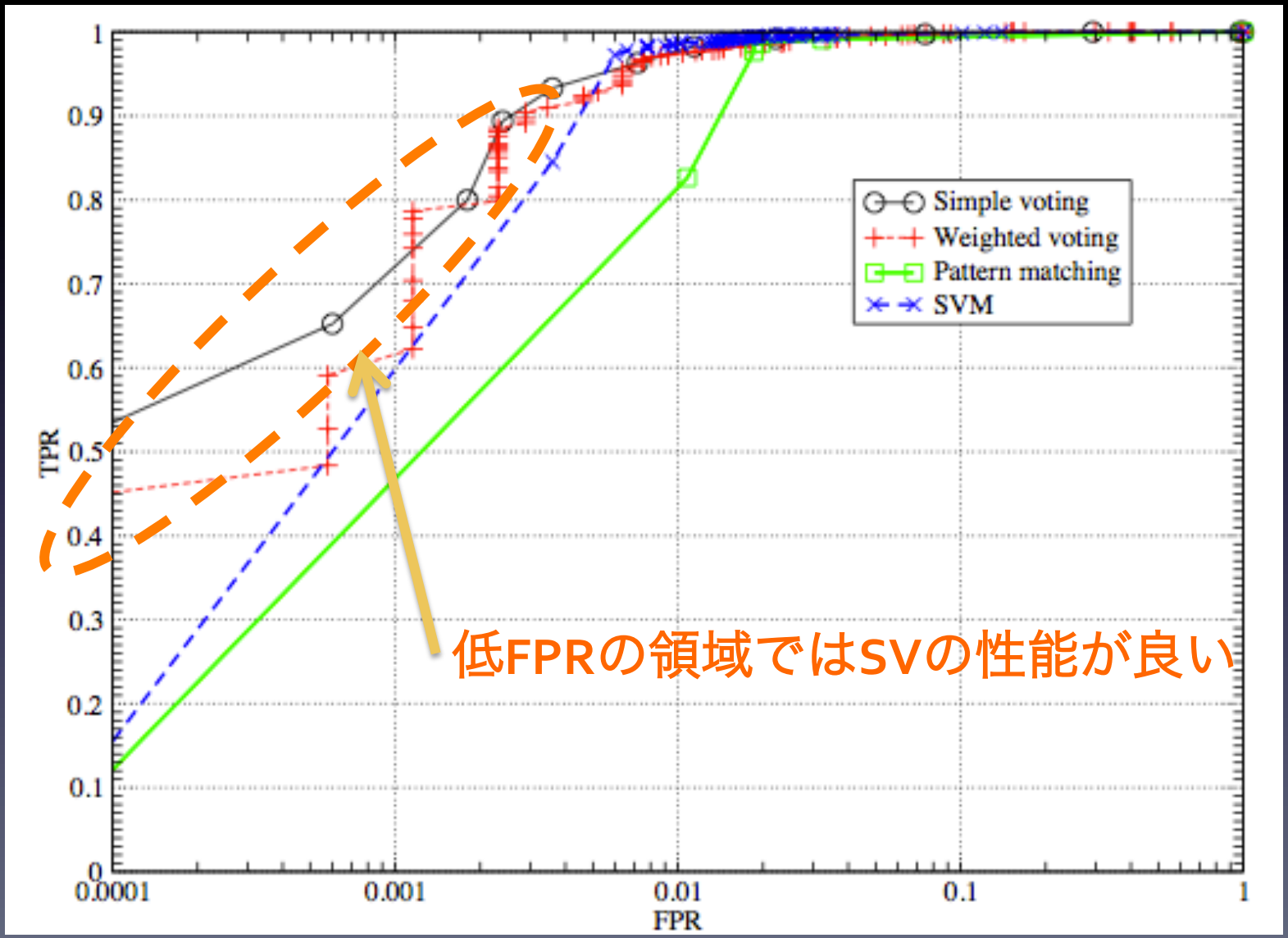
精度に関してパラメタ最適した性能評価結果

	m(FPR)	m(FNR)	m(ACC)	σ(FPR)	σ(FNR)	σ(ACC)
SV	0.0225	0.0108	0.9844	0.0056	0.0041	0.0041
WV	0.0189	0.0182	0.9814	0.0059	0.0067	0.0046
PM	0.0245	0.0239	0.9759	0.0169	0.0167	0.0082
SVM	0.0179	0.0066	0.9889	0.0093	0.0038	0.0035
DNSBL(1)	0.0305	0.0323	0.9683	–	–	–
DNSBL(2)	0.1289	0.0268	0.9329	–	–	–
DNSBL(3)	0.1721	0.0031	0.9303	–	–	–
DNSBL(4)	0.0125	0.1424	0.9086	–	–	–
DNSBL(5)	0.0251	0.1560	0.8954	–	–	–

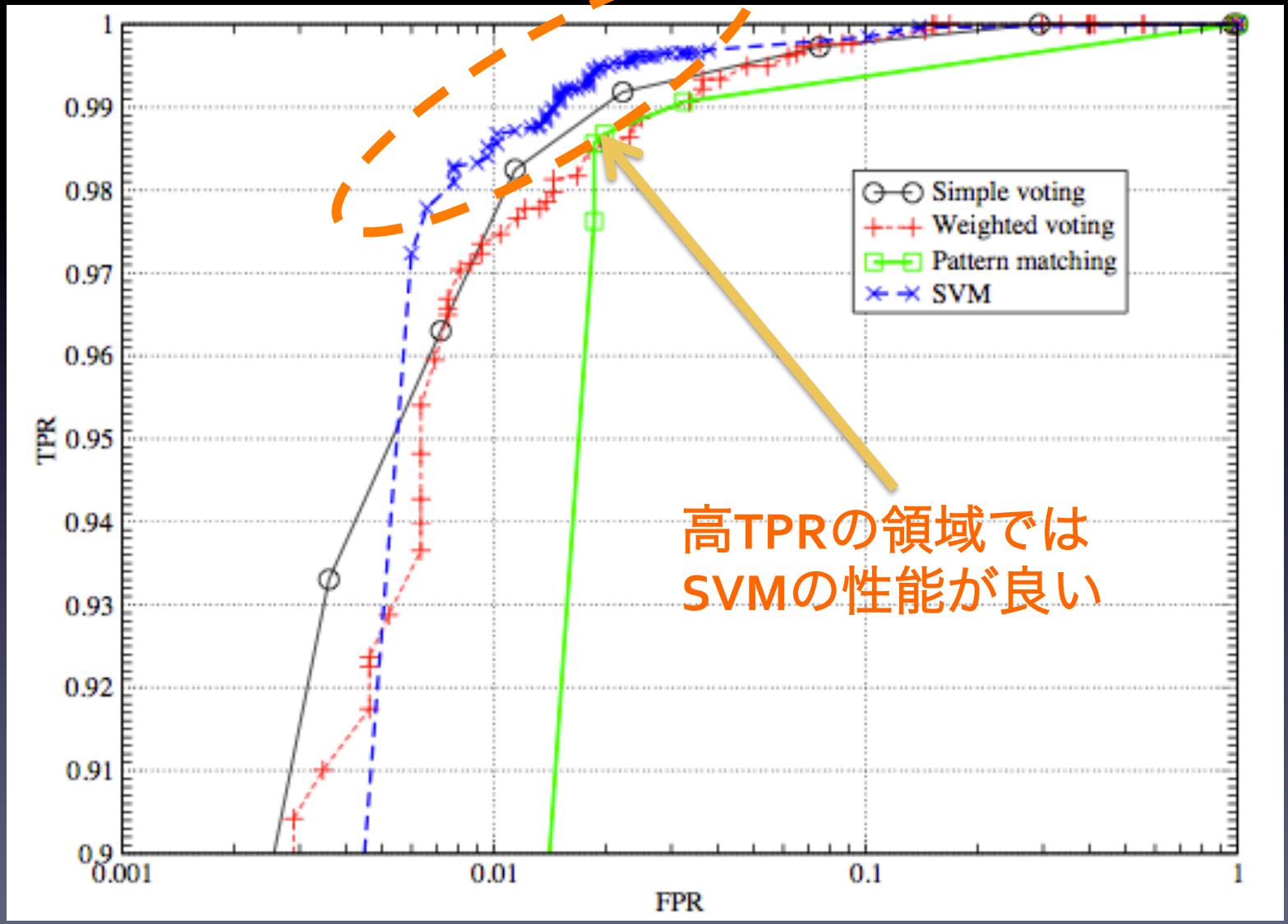
性能評価（メッセージレベル）

	# of detected IPs	# of spam	# of ham
SV	2,543	60,843	761
WV	2,576	61,402	1,622
PM	2,563	60,560	2,941
SVM	2,581	61,848	1,234
DNSBL(1)	2,537	59,652	2,944
DNSBL(2)	2,715	60,503	15,215
DNSBL(3)	2,848	61,641	22,142
DNSBL(4)	2,224	49,519	1,041
DNSBL(5)	2,210	50,594	2,968
Bad IP	2,569	62,003	0
Good IP	1,667	0	167,826
ALL	4,236	62,003	167,826

ROC分析


TPR
FPR
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.0001
0.001
0.01
0.1
1
Simple voting
Weighted voting
Pattern matching
SVM
低FPRの領域ではSVの性能が良い

ROC分析


1
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0.92
0.91
0.9
TPR
0.001
0.01
0.1
1
FPR
Simple voting
Weighted voting
Pattern matching
SVM
高TPRの領域では
SVMの性能が良い

Take away message

- 投票はシンプルであるが，高い精度を与える
  - ただし，適切な閾値学習が必須

- 複数リストをうまく組み合わせることで単一のベストなリストよりも良い結果を得ることが出来る

- 投票はより保守的な利用(低false positive)に，教師付き機械学習(SVM)はアグレッシブな利用(高true positive)に適している

今後の課題

- うまく判定できないケースの説明
- 評価データの拡充
- 方式の改良
- Reputation 統合フレームワークの一般化