

Characterization and prediction of issue-related risks in software projects

Morakot Choetkietikul | Hoa Khanh Dam
Truyen Tran | Aditya Ghose

背景

- ❖ ソフトウェアプロジェクトは費用と納期を超える傾向
 - 66%のプロジェクトが予算超
 - 33%のプロジェクトが納期超
- ❖ 本論では、納期に関するデータを対象にする

2つの寄贈(Contributions)

- ❖ プロジェクトの遅延リスクを構成する不具合の特徴を示す
 - discussion time, waiting time, issues type, number of repetition task, percentage of delayed issues that a developer involved with, developer's workload, issue priority, number of comments, changing of priority, number of fix version, number of affect version, number of issue link, number of blocked issues, number of blocking issues, changing of description, reporter reputation
- ❖ 上記のバグの予測モデルを構築する

データセットとメトリクス

TABLE I: Dataset description

Project	Delayed issues		Non-delayed issues		Total
Apache	111	[2.27%]	4,771	[97.72%]	4,882
Duraspace	314	[9.75%]	2,908	[90.25%]	3,222
JBoss	2,242	[15.08%]	12,627	[84.92%]	14,869
Moodle	214	[2.24%]	9,325	[97.75%]	9,539
Spring	80	[0.96%]	8,238	[99.03%]	8,318
Total	2,961	[7.25%]	37,869	[92.74%]	40,830

- ❖ 3000以上の不具合数を誇るプロジェクトを選定。

データセットとメトリクス

- Discussion time
- Waiting time
- Type
- Reopendされる回数
- Priority
- Priorityの変更
- コメント数
- Fix version数
- Affect version数
- Issue link数
- Block数
- Descroptionの変更
- 報告者の評判*
- 開発者の労働量

$$reputation(D) = \frac{|opened(D) \cap fixed(D)|}{|opened(D)| + 1}$$

リスクの要因

TABLE II: Descriptive ℓ_1 -penalized logistic regression model for risk probability, trained on all issues collected from the five projects

Feature	Apache	Duraspace	JBoss	Moodle	Spring	All
discussion	1.027	-1.368	-3.153	-21.353	-0.371	-7.274
waiting	0	0	0	-15.307	0	-4.495
type:Bug	0	-3.254	-0.959	-2.933	-4.763	-0.859
type:Code Task	N/A	-0.052	N/A	N/A	N/A	0.216
type:Documentation	0	0.359	1.617	N/A	N/A	1.06
type:Epic	0	1.026	0	0	0	0.925
type:Improvement	0.044	-0.745	N/A	-2.833	-4.704	0.271
type:New Feature	0.106	0.097	0	-3.011	-4.495	0.64
type:Story	0	-0.81	0	0	0	-0.342
type:Sub-task	-1.183	0.267	0.94	-2.796	0.001	0.867
type:Task	-0.37	-0.477	0.724	-2.604	-4.534	0.548
type:Technical task	0	0.35	0	0	0	0.567
repetition	-0.391	4.01	3.11	-0.427	6.358	2.878
perofdelay	3.183	4.202	3.088	2.569	22.091	3.361
workload	0.179	-0.911	1.367	-1.104	0.241	1.621

Feature	Apache	Duraspace	JBoss	Moodle	Spring	All
priority:Blocker	0.465	0.657	-0.322	-1.744	4.301	-0.827
priority:Critical	-1.585	0	-0.45	-2.929	3.716	-0.903
priority:Major	0.317	0	-0.453	-2.983	1.571	-0.921
priority:Minor	0	-0.663	-0.438	-2.801	1.778	-0.876
priority:Trivial	-0.229	-1.548	-0.95	-2.639	0	-0.783
no_comment	0	0.001	2.31	-1.43	-2.918	2.826
no_priority_change	-0.001	2.874	-0.234	1.684	1.325	1.166
no_fixversion_change	3.042	0.152	2.081	-1.352	-0.088	2.529
no_affectversion	-0.001	0	-0.903	-11.489	-0.822	-3.413
no_issue link	0.001	0	0	-1.717	0.883	1.268
no_blocking	0	0	-0.175	0.52	0	0.216
no_blockedby	0	-0.236	0.747	2.341	0	1.785
no_des_change	0.001	0.955	0	1.251	0	1.472
reporter_rep	1.355	-0.298	-0.712	2.53	1.623	-0.498

discussion=Discussion time, waiting=Waiting time, type=Issue's type, repetition=Number of repetition tasks, perofdelay=Percentage of delayed issues that a developer involved with, workload=Developer's workload, priority=Issue's priority, no comment=Number of comments, no priority change=Changing of priority, no fixversion=Number of fix version, no affectversion=Number of affect version, no issue link=Number of issue link, no blocking=Number of issues that are blocked by this issue, no blockedby=Number of issues that block this issue, no des change=Changing of description, reporter rep=Reporter reputation, N/A=This feature is not available.

❖ 重み付けを行って、予測する

$$f(x) = w_0 + \sum_{i=1}^p w_i x_i$$

$$L(w) = \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y^i f(x^i))) + \lambda \sum_j |w_j|$$

予測モデル構築

- ❖ L1-panalized logistic regression modelを用いる
- ❖ それぞれのアルゴリズムで予測モデルを構築する。
 - ❖ ランダムフォレスト
 - ❖ ニューラルネットワーク
 - ❖ 決定木(C4.5)
 - ❖ ナイーブベイズ
 - ❖ NB木

MMCEとMMAE

- ❖ Macro-averaged Mean Cost-Error

$$\text{MMCE} = \frac{1}{3} \sum_{k=1}^3 \frac{1}{n_k} \sum_{i=1}^n |R\bar{E}_i - C| \delta [y^i = k]$$

- ❖ Macro-averaged Mean Absolute Error

$$\text{MMAE} = \frac{1}{3} \sum_{k=1}^3 \frac{1}{n_k} \sum_{i=1}^n |\hat{y}^i - k| \delta [y^i = k]$$

結果

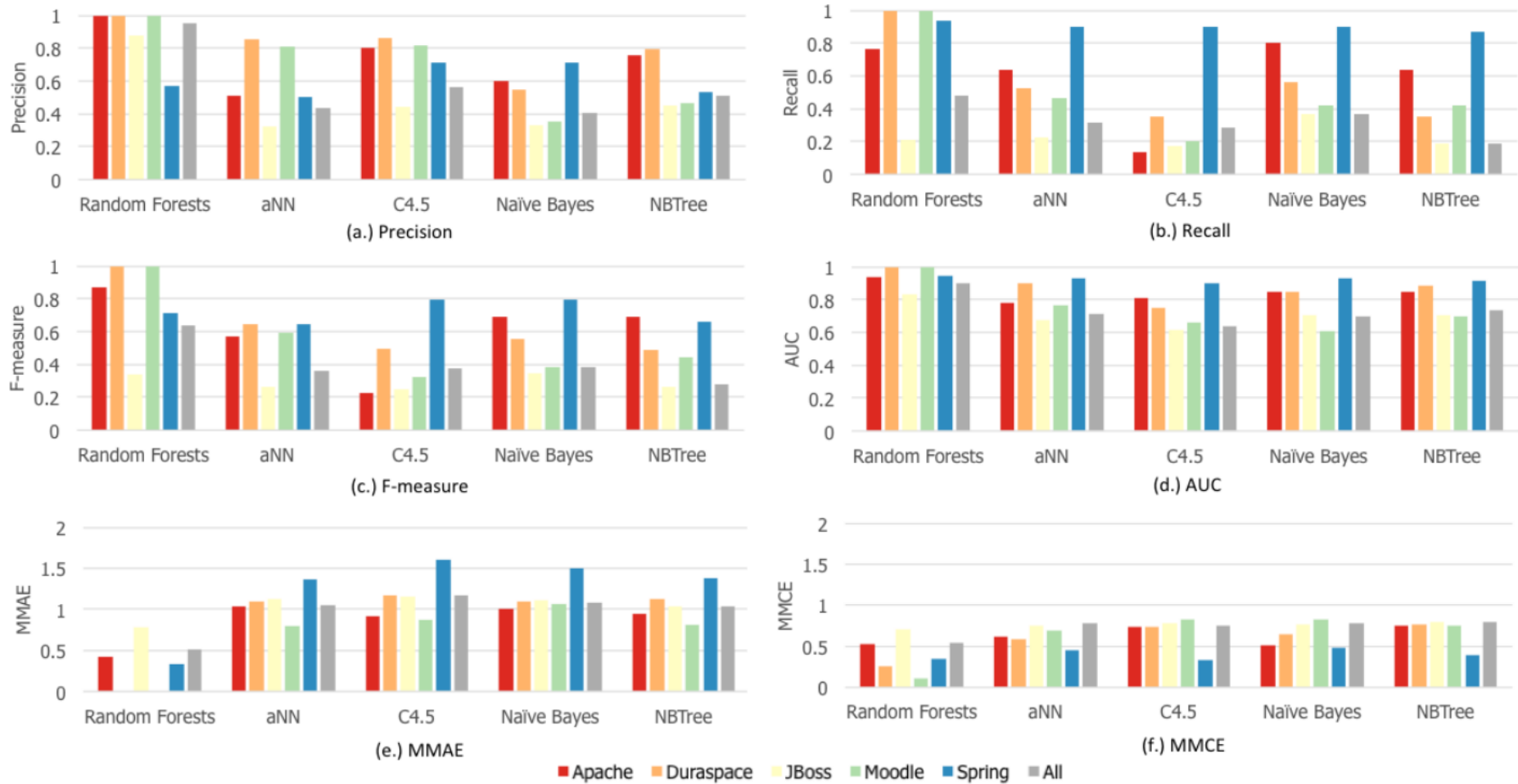


Fig. 1: Evaluation results
(Precision/Recall/F-measure/AUC, the higher the better. The MMAE/MMCE, the lower the better.)

まとめ

- ❖ 重みを使った予測モデル構築
 - プロジェクトによってメトリクスの影響度が違っている
- ❖ Top 3 factor
 - 開発者が遅延不具合に巻き込まれている比率
the percentage of delayed issues that a developer involved with
 - discussion time
 - Reopenされる回数
the number of times that issue is reopen
- ❖ ランダムフォレストが優秀
 - Precision=0.9, F-measure=0.8

Macro-averaged Mean Absolute Error:

$$\text{MMAE} = \frac{1}{3} \sum_{k=1}^3 \frac{1}{n_k} \sum_{i=1}^n |\hat{y}^i - k| \delta [y^i = k]$$

For example, an issue is predicted to be 10% in major delayed, 60% in minor delayed, and 30% in non-delayed. Thus, the predicted class of this issue is minor delayed ($k = 2$). Assume that the actual class of the issue is non-delayed ($k = 1$), then the distance between actual and predicted classes of this issue is 1.

For example, an issue is predicted to be 50% in major delayed, 30% in minor delayed, and 20% in non-delayed. Assume that the issue actually caused a minor delay (i.e. the true class is minor delayed) and the costs C_1 (no delay), C_2 (minor delay), and C_3 (major delay) are respectively 0, 1, and 2. The predicted risk exposure $\bar{RE}$ is 1.3 and the actual risk exposure is 1 (see Section IV-B for how a risk exposure is calculated). Hence, the MMCE error between actual and predicted risk exposure for this issue is 0.3.