

Software Evolution Prediction Using Seasonal Time Analysis : A Comparative Study

Miguel Goulão, Fernando Brito e Abreu

2012 16th European Conference on Software Maintenance and Reengineering p213-p222

蘆田 誠人

背景

- ソフトウェアのプロセス管理者は必要なリソースの割り当てを計画しなければならないため、変更要求の予測を行う必要がある
 - ▶ 変更要求
 - 是正処置
 - 予防処置
- 変更要求の数を容易に予測はできない
 - ▶ 新しいバージョンが公開されたりするため

背景

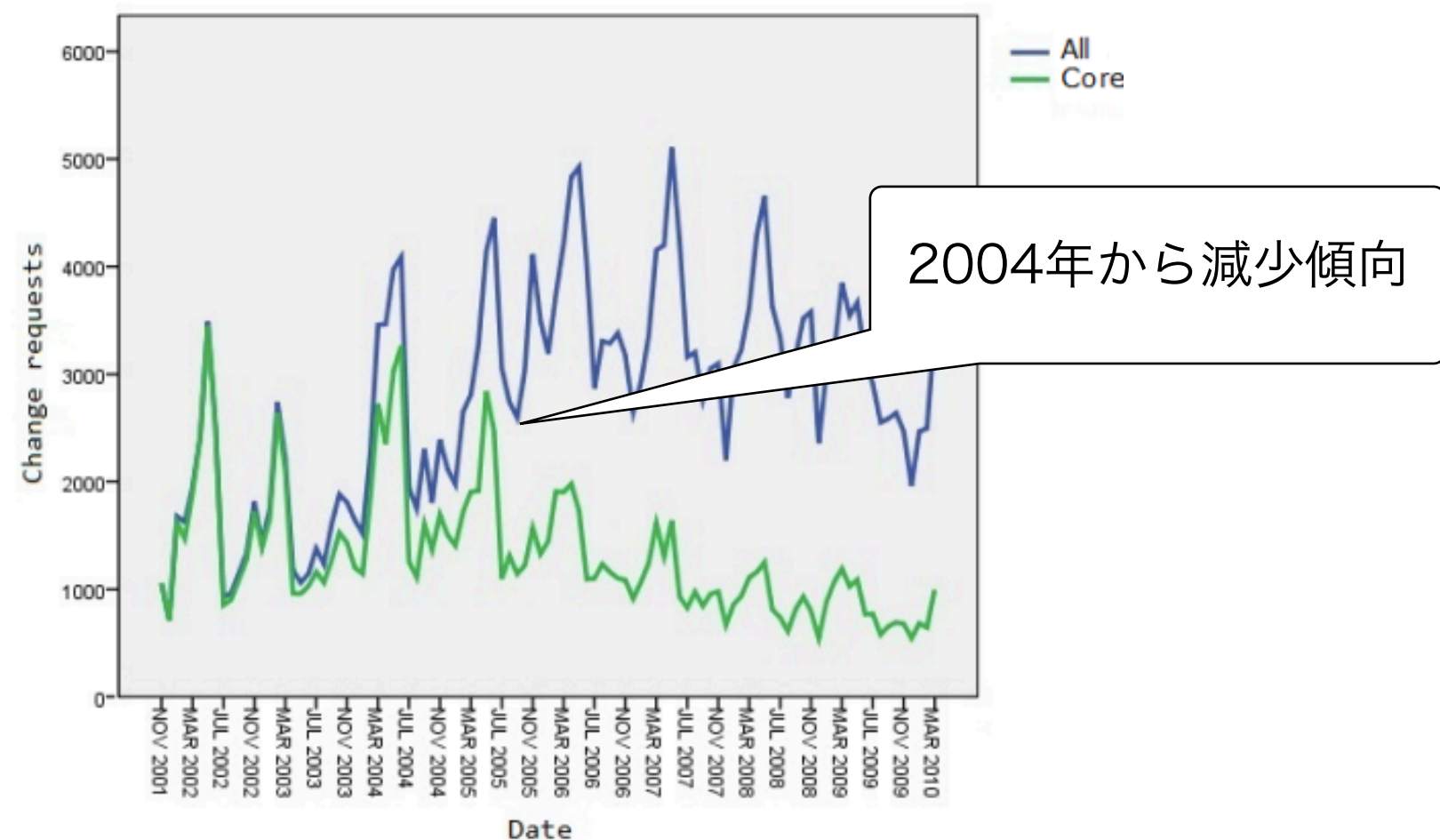
- 新しいバージョンがリリースされる時期が決まっているプロジェクトの変更要求を予測する場合、季節性を考慮したモデルと考慮しないモデルでは予測精度が異なる場合がある
 - ▶ 季節性（トレンド）を無視した予測は一般的に精度が悪くなると言われている

目的

- 季節性を考慮した時系列モデルを用いて、季節性を持つプロジェクトの変更要求の傾向を予測する

分析対象

- Eclipseプロジェクトの変更要求
 - ▶ 2001年から2009年



RQ

- RQ1

- 変更要求の数の季節性を検出できるか

- RQ2

- 時間の経過と共に変更要求の分布に統計的に関連した傾向があるか

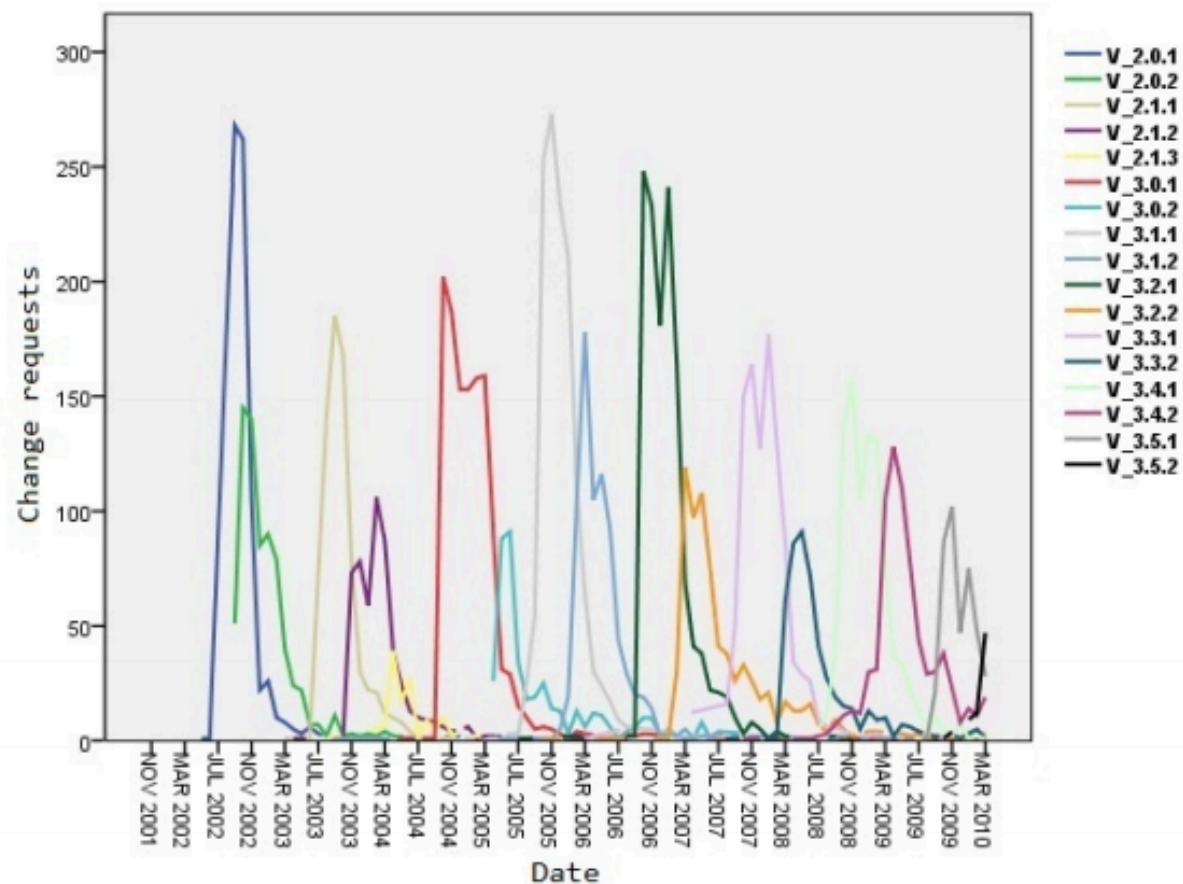
- RQ3

- 季節性を持つARIMAモデルの使用はOSSプロジェクトの変更要求を予測することができるか

結果

- RQ1

- 変更要求の数の季節性を検出できるか



- 毎年4月から6月に多くなる傾向がある

- クラスカル=ウォリス検定
- ヨンクヒール=タプストラ検定

時系列データに増加・減少傾向は存在するかどうかの検定

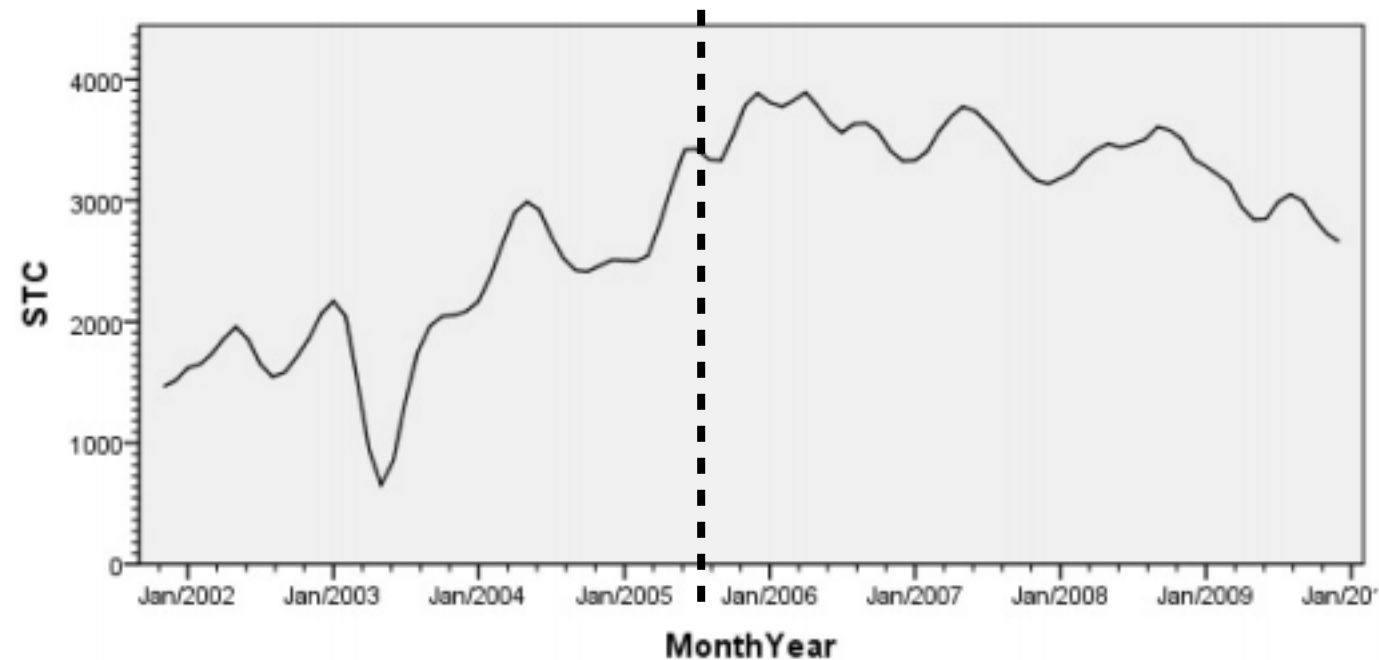
✓ 季節性が存在する結果に

RQ

- RQ2

- 時間の経過と共に変更要求の分布に統計的に関連した傾向があるか

Ver.3.1リリース



- 変更要求の数が年々増加傾向
- Ver.3.0とVer.3.1リリースから要求数は落ち着いている

RQ

- RQ3

- 季節性を持つARIMAモデルの使用はOSSプロジェクトの変更要求を予測することができるか
 - 先行研究[1, 2]で用いられたARIMA(1,1,0)(1,1,0)モデルを使用
 - パラメータ(AR, 差分, MA)(AR季節パラメータ, 季節差分, MA季節パラメータ)

[1]S. Anderson, A. Auquier, W. W. Hauck, D. Oakes, W. Vandaele, and H. Weisberg, Statistical methods for comparative studies: techniques for bias reduction: John Wiley and Sons, ISBN, 1980.

[2]R. McCleary and R. Hay, Applied time series analysis for the social sciences: SAGE Publications, ISBN, 1980.

RQ

- RQ3

- 季節性を持つARIMAモデルの使用はOSSプロジェクトの変更要求を予測することができるか

✓ 決定係数は他モデルより高め

Model	Model Fit statistics							Ljung-Box Q(18)			Outliers
	Stationary R-squared	R-squared	RMSE	MAPE	MAE	MaxAPE	MaxAE	Statistics	DF	Sig.	
<i>ARIMA(110)(110)</i>	<i>0.743</i>	<i>0.872</i>	<i>384.806</i>	<i>11.464</i>	<i>293.043</i>	<i>47.106</i>	<i>848.347</i>	<i>15.850</i>	<i>16</i>	<i>0.463</i>	<i>6</i>
ARIMA(110)(000)	0.149	0.681	611.421	19.579	469.599	170.267	1578.374	28.649	17	0.038	1
ARIMA(010)(010)	0.519	0.760	510.102	13.729	375.977	56.549	1282.146	29.103	18	0.047	4
ARIMA(011)(000)	0.155	0.683	609.299	19.521	468.542	160.486	1487.703	27.976	17	0.045	1
ARIMA(505)(000)	0.793	0.793	526.691	16.844	389.908	100.130	1288.319	11.002	8	0.202	1

RQ

- RQ3

- 季節性を持つARIMAモデルの使用はOSSプロジェクトの変更要求を予測することができるか

✓ 相対誤差も他モデルよりも小さい

Model	Model Fit statistics							Ljung-Box Q(18)			Outliers
	Stationary R-squared	R-squared	RMSE	MAPE	MAE	MaxAPE	MaxAE	Statistics	DF	Sig.	
<i>ARIMA(110)(110)</i>	<i>0.743</i>	<i>0.872</i>	<i>384.806</i>	<i>11.464</i>	<i>293.043</i>	<i>47.106</i>	<i>848.347</i>	<i>15.850</i>	<i>16</i>	<i>0.463</i>	<i>6</i>
ARIMA(110)(000)	0.149	0.681	611.421	19.579	469.599	170.267	1578.374	28.649	17	0.038	1
ARIMA(010)(010)	0.519	0.760	510.102	13.729	375.977	56.549	1282.146	29.103	18	0.047	4
ARIMA(011)(000)	0.155	0.683	609.299	19.521	468.542	160.486	1487.703	27.976	17	0.045	1
ARIMA(505)(000)	0.793	0.793	526.691	16.844	389.908	100.130	1288.319	11.002	8	0.202	1

まとめ・所感

- OSSプロジェクトにおける変更要求の数の予測を時系列モデルの一つであるARIMAモデルで行なった
- 先行研究で提案されていたモデルは他モデルに比べ、決定係数が高く、誤差も小さかった
 - ▶ 予測に有用だと言える
- ARIMAモデルのパラメータが気になった
 - 1980年の本を引用している、大丈夫？