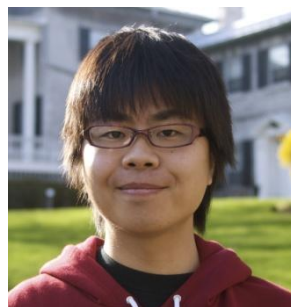


グローバル環境下におけるOSS開発者の 情報交換に対する時差の影響

60151039 東 裕之輔

著者

1. 亀井 靖高
2. 大平 雅雄
3. 伊原 彰紀
4. 小山 貴和子
5. 裕本 真佑
6. 松本 健一
7. 鵜林 尚靖



背景

- ▶ OSS開発プロジェクトの特徴の1つには、世界中に点在する開発者が共同開発を行うという分散開発の形態をとることである。
- ▶ 開発者は北米やヨーロッパだけではなく、アジア、オーストラリア、南アフリカ、南米に点在していることもある。



背景

- ▶ OSS開発プロジェクトでは、このような地域や時差をまたがった環境下で開発が行われているという特徴から、開発者同士の情報交換は主に電子メールなどが用いられている。
- ▶ しかし、グローバル環境下では電子メールにより情報交換を行う場合、時差の影響を受けて情報交換の達成に時間がかかると考えられる。



背景

- ▶ 特に開発者が多いとされるアメリカ、西ヨーロッパ、中国では最大5時間の時差が発生する。
- ▶ リアルタイムな情報交換は困難であり、タイムラグが発生する。
- ▶ タイムラグは、開発スピード、生産性、品質管理能力の低下など悪影響をもたらすといわれている。



本研究の目的

- ▶ OSSプロジェクト内におけるタイムラグによる実態を解明。
- ▶ メーリングリストを用いた情報交換時間を短縮化するための指針を得る。



仮説

- ▶ 2つの仮説
- ▶ 仮説1 : 時差のある地域間の情報交換は返信時間は長くなる。
- ▶ 仮説2 : 返信者の地域によって素早い返信を期待できる時間帯が異なる。
- ▶ この2つを具体的に確かめる。



タイムラグによる予備調査

- ▶ Pythonプロジェクトの情報交換におけるタイムラグの実態を調査した。

From: us@ait.kyushu-u.ac.jp Date: Sun, 30 Jul 2000 18:26:00 -0500 (アメリカ) Subject: [Python-Dev] title messages
From: eu@ait.kyushu-u.ac.jp Date: Mon, 31 Jul 2000 09:43:55 +0200 (ヨーロッパ) Subject: Re: [Python-Dev] title Good morning. messages

- ▶ アメリカの開発者がヨーロッパの開発者にメールを送ってもすぐに返信されることはなかった。
- ▶ ヨーロッパ現地時刻9時頃に返信されるケースが多かった。

分析データの整形方法

- ▶ 本論文ではまず、メールのヘッダ(Date ,From)を解析し、メッセージの送信日時と送信場所のデータを収集する。

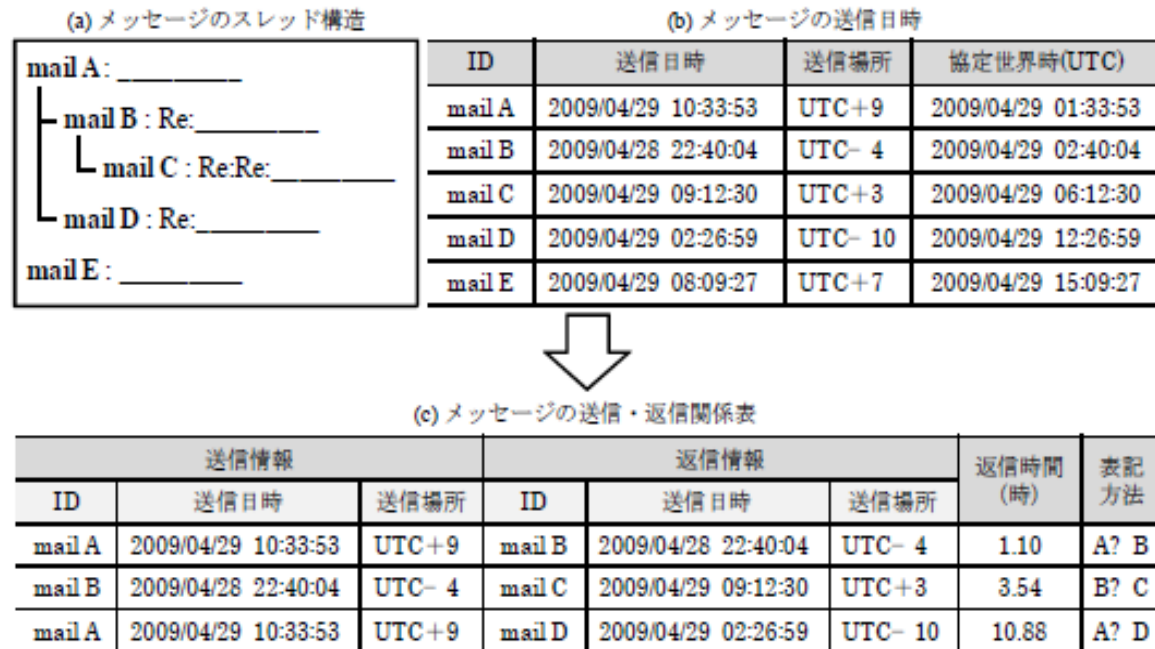


図-2 分析データの整形方法

- ▶ ※メールEには返信がないので図cには現れない。

分析対象プロジェクト

- ▶ Python
- ▶ PostgreSQL
- ▶ Apache HTTP Server
- ▶ Eclipse 各プロジェクト
 - Java Development tools
 - Plugin Development Enviroment
 - Business Inteligence and Reporting
 - Eclipse Modeling Framework



分析対象プロジェクトの分類

- ▶ OSSプロジェクトは種類によって活動時間が異なる。
- ▶ ボランティア型では夕方から夜にかけて情報交換が活発に行われていることが多い。
- ▶ 企業混同型では業務時間中、つまり(朝から夕方にかけて)情報交換が行われることが多い。
- ▶ よって2つに分類することが必要である。



分析対象プロジェクトの分類

- ▶ 本研究では対象となるプロジェクトを2つに分類する。
- ▶ ボランティア型(夜中心)
- ▶ Python, PostgreSQL, Apache HTTP Server(Apache)
- ▶ 企業混同型(朝中心)
- ▶ Eclipse



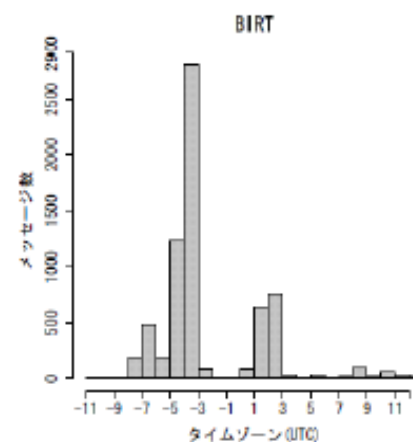
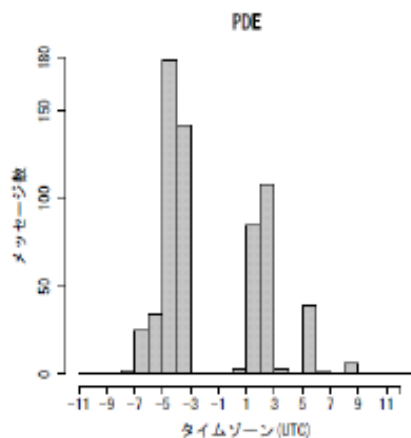
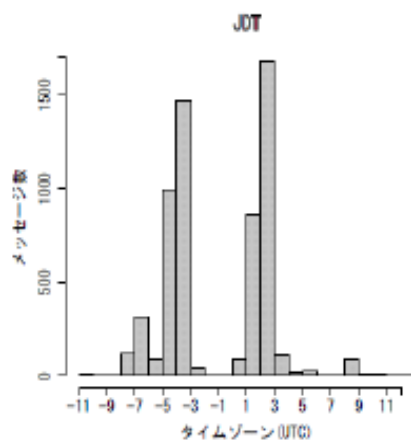
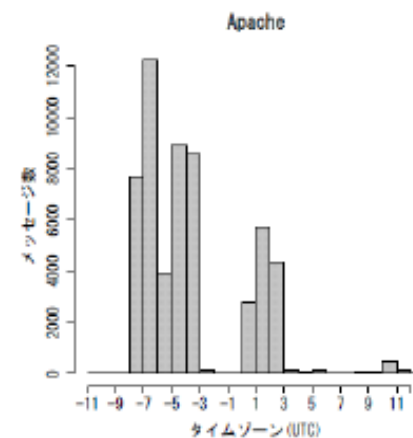
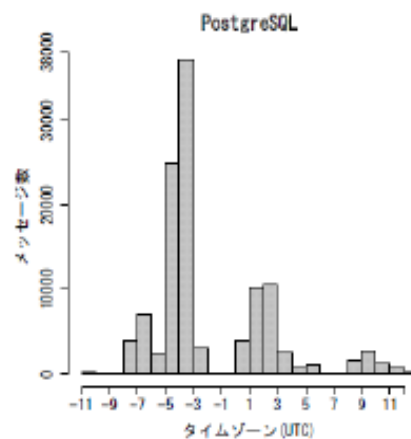
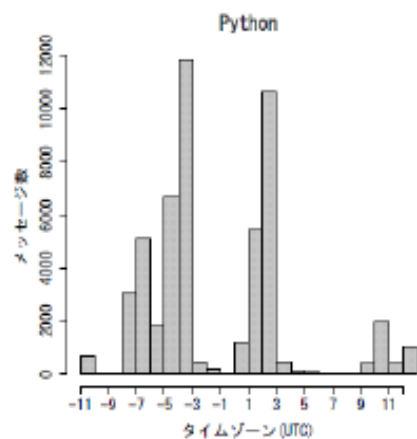
データセット

- 主に開発に関する話題(新機能、バグ報告、メンテナンス、リリース)を取り扱うMLまたはニュースグループのアーカイブ。

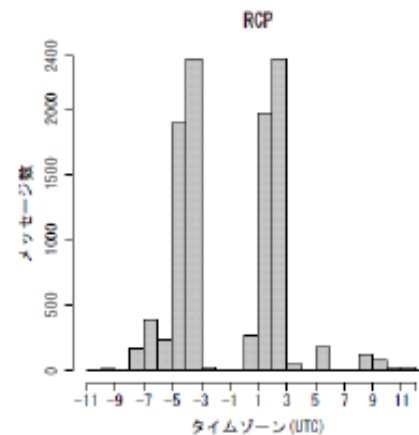
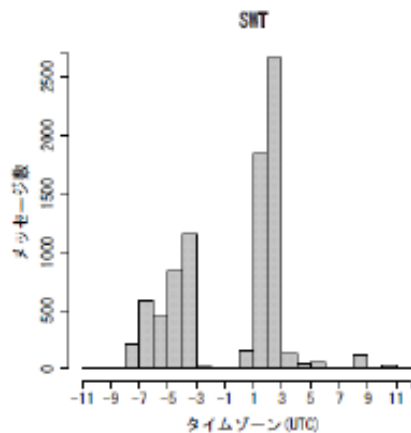
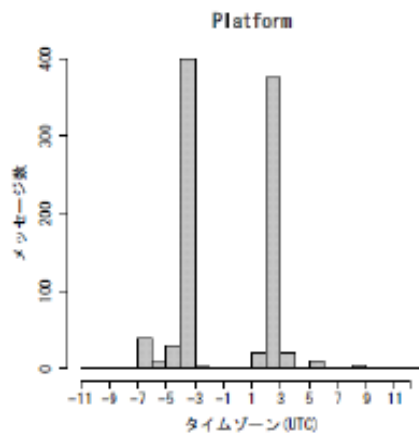
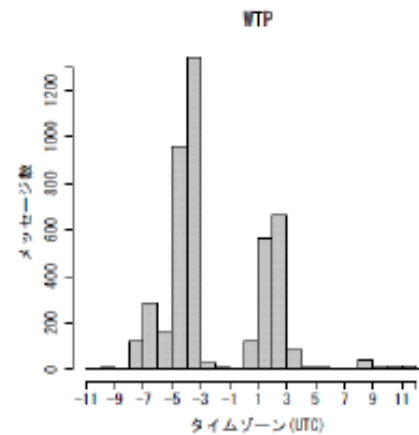
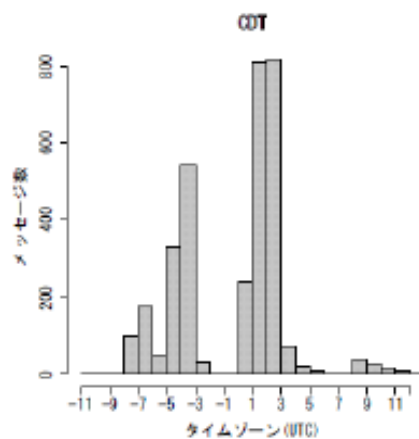
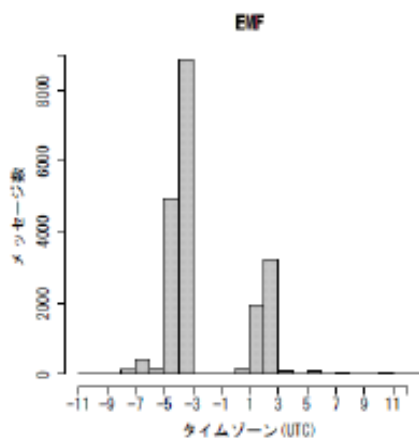
表-2 データセットの概要

分析対象プロジェクト		分析対象 ML/ ニュースグループ	分析期間	全メッセージ (件)	開発者数 (人)	送信・返信 の対(件)	返信 24 時間未 満のデータ数 (件)
Python		python-dev@python.org	1999/04~2009/04	89,301	2,150	56,707	51,830
PostgreSQL		pgsql-hackers@postgresql.org	1997/01~2009/09	184,492	4,731	130,239	114,514
Apache HTTP Server (Apache)		dev@httpd.apache.org	1995/03~2009/08	119,673	2,736	63,293	55,459
Eclipse	Java development tools (JDT)	eclipse.tools.jdt	2003/05~2009/09	24,691	4,717	7,894	5,951
	Plugin Development Environment (PDE)	eclipse.platform.pde	2008/05~2009/09	2,238	578	847	627
	Business Intelligence and Reporting Tools (BIRT)	eclipse.birt	2004/09~2009/09	35,108	3,886	9,058	7,798
	Eclipse Modeling Framework (EMF)	eclipse.tools.emf	2002/10~2009/09	44,216	3,337	22,654	20,259
	C/C++ Development Tooling (CDT)	eclipse.tools.cdt	2002/01~2009/09	19,088	4,290	4,727	3,296
	Eclipse Web Tools Platform Project (WTP)	eclipse.webtools	2003/07~2009/09	19,160	3,658	6,193	4,445
Eclipse Platform	Platform	eclipse.platform	2003/05~2009/09	7,628	1,502	1,152	917
	SWT	eclipse.platform.swt	2003/05~2009/09	45,490	6,963	10,377	8,340
	RCP	eclipse.platform.rcp	2004/09~2009/09	37,583	4,796	13,614	10,243

各時差帯におけるメッセージ数



各時差帯におけるメッセージ数



各時差帯におけるメッセージ数

- ▶ とくにメッセージ数が多い時差帯はUTC-8~-4とUTC0~3付近である。
 1. UTC-8~-4(北アメリカ・南アメリカ=米大陸地域)
 2. UTC+0~+3(ヨーロッパ・アフリカ=欧・アフリカ地域)
- ▶ 全プロジェクトの半数以上のメールがこれらの地域から送られており、この2地域を分析することがプロジェクトの理解につながるとする。



分析対象地域と統計量

表-3 分析対象地域と統計量（米：米大陸地域、欧：欧・アフリカ地域）

分析対象プロジェクト		分析対象地域		メッセージ数（件）		開発者数（人）		
		米	欧	米	欧	米	欧	
Python		UTC-8～UTC-4	UTC+0～UTC+3	28,574	17,748	679	575	
PostgreSQL		UTC-5～UTC-4	UTC+0～UTC+3	61,875	27,318	537	795	
Apache HTTP Server (Apache)		UTC-8～UTC-4	UTC+0～UTC+2	41,302	12,840	598	451	
Eclipse	Java development tools (JDT)	UTC-5～UTC-4	UTC+1～UTC+2	2,468	2,547	226	320	
	Plugin Development Environment (PDE)	UTC-5～UTC-4	UTC+1～UTC+2	321	193	28	68	
	Business Intelligence and Reporting Tools (BIRT)	UTC-5～UTC-4	UTC+1～UTC+2	4,042	1,401	152	287	
	Eclipse Modeling Framework (EMF)	UTC-5～UTC-4	UTC+1～UTC+2	13,807	5,219	165	510	
	C/C++ Development Tooling (CDT)	UTC-5～UTC-4	UTC+1～UTC+2	872	1630	119	258	
	Eclipse Web Tools Platform Project (WTP)	UTC-5～UTC-4	UTC+1～UTC+2	2,302	1,225	171	285	
	Eclipse Platform	Platform	UTC-4	UTC+2	431	398	64	100
		SWT	UTC-8～UTC-4	UTC+1～UTC+2	3,248	4,517	410	613
		RCP	UTC-5～UTC-4	UTC+1～UTC+2	4,269	4,354	202	570

仮説 1

- ▶ (仮説1): 時差のある地域間の情報交換は返信時間が長くなる。
- ▶ (分析): 同一タイムゾーン内(時差帯)の返信時間の差が統計的に有意であるかどうかを検定する。
- ▶ 検定にはウィルコクソンの順位和検定を有意水準5%で用いる。



(仮説 1) 検定を用いた分析結果

表-4 地域別の返信時間帯（米：米大陸地域、欧：欧・アフリカ地域）

分析対象プロジェクト		米→米 (時)	米→欧 (時)	p値*	欧→欧 (時)	欧→米 (時)	p値*
Python		1.24	2.07	0.00	1.59	1.80	0.57
PostgreSQL		0.76	1.88	0.00	1.19	1.42	0.95
Apache HTTP Server (Apache)		0.87	1.75	0.00	1.30	1.70	0.00
Eclipse	Java development tools (JDT)	1.91	6.88	0.00	2.75	3.39	0.91
	Plugin Development Environment (PDE)	2.10	3.86	0.17	1.69	5.12	0.00
	Business Intelligence and Reporting Tools (BIRT)	2.33	10.41	0.00	1.48	4.72	0.00
	Eclipse Modeling Framework (EMF)	0.93	1.72	0.00	0.97	1.00	0.69
	C/C++ Development Tooling (CDT)	2.16	8.56	0.00	2.42	4.70	0.02
	Eclipse Web Tools Platform Project (WTP)	2.06	6.13	0.00	3.04	4.27	0.15
	Eclipse Platform	Platform	1.88	0.08	2.12	3.56	0.00
		SWT	2.42	0.00	0.96	3.42	0.00
		RCP	2.90	0.01	1.52	4.38	0.00

*p値：有意水準5%によるウィルコクソンの順位和検定

P値が0.05以下のところは太字、有意差あり

仮説 1 の検証

- ▶ 米⇒欧には、1つのプロジェクトを除いて時差が影響するという結果が得られた。
- ▶ しかし欧⇒米には、有意差が見られるものが半分程度であった。
- ▶ 仮説1は弱くではあるが支持されたといえる。

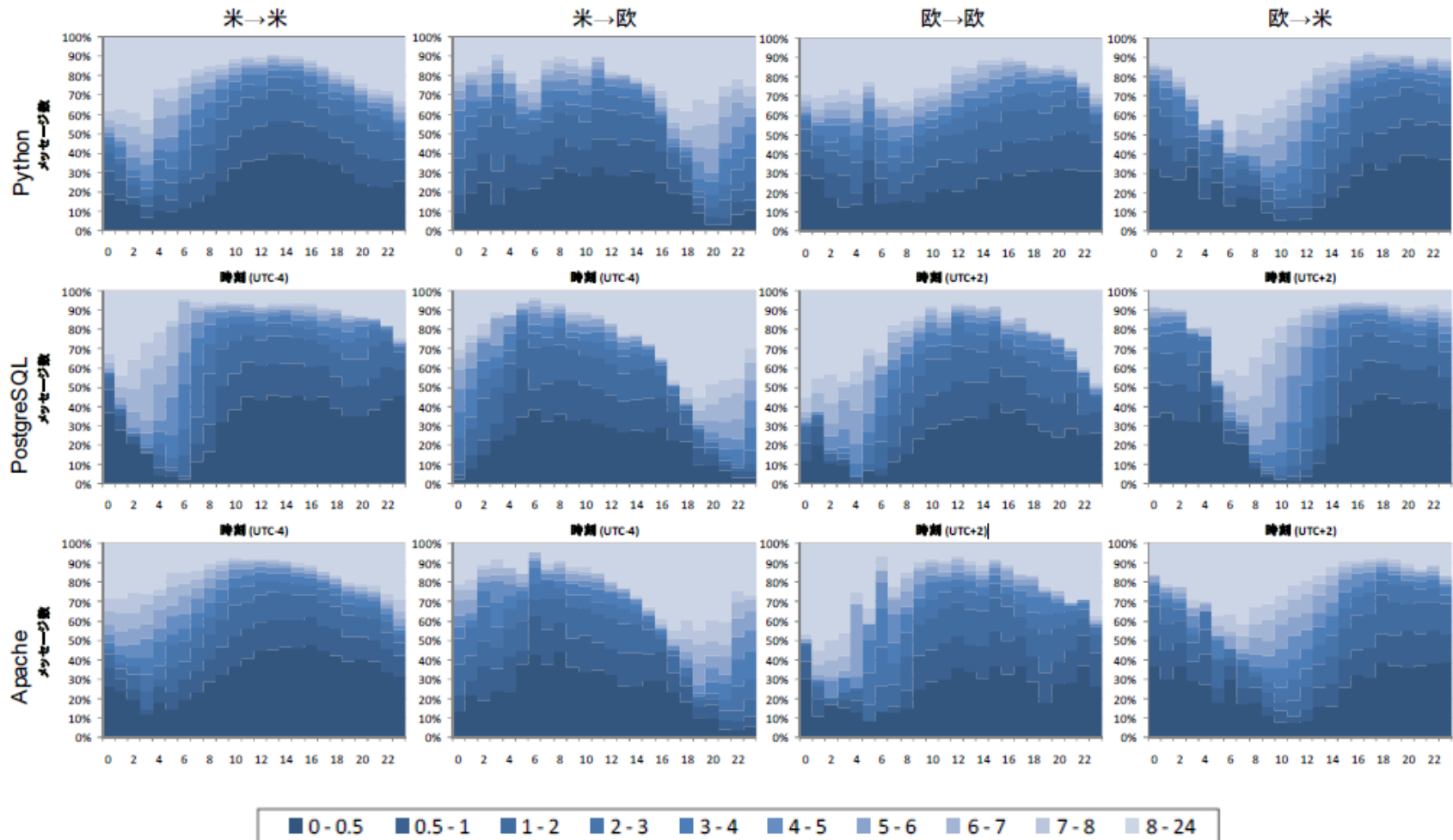


仮説 2

- ▶ (仮説2): 返信者の地域によって素早い返信を期待できる時間帯が異なる。
- ▶ (分析): 仮説2を証明するために、各送信時刻における返信時間の分布を示し、返信の早い時間帯を確認する。
- ▶ 返信者の地域によって素早い返信を期待できる時間帯が異なれば、仮説2が支持されたとみなす。



(仮説 2) 分析結果 1

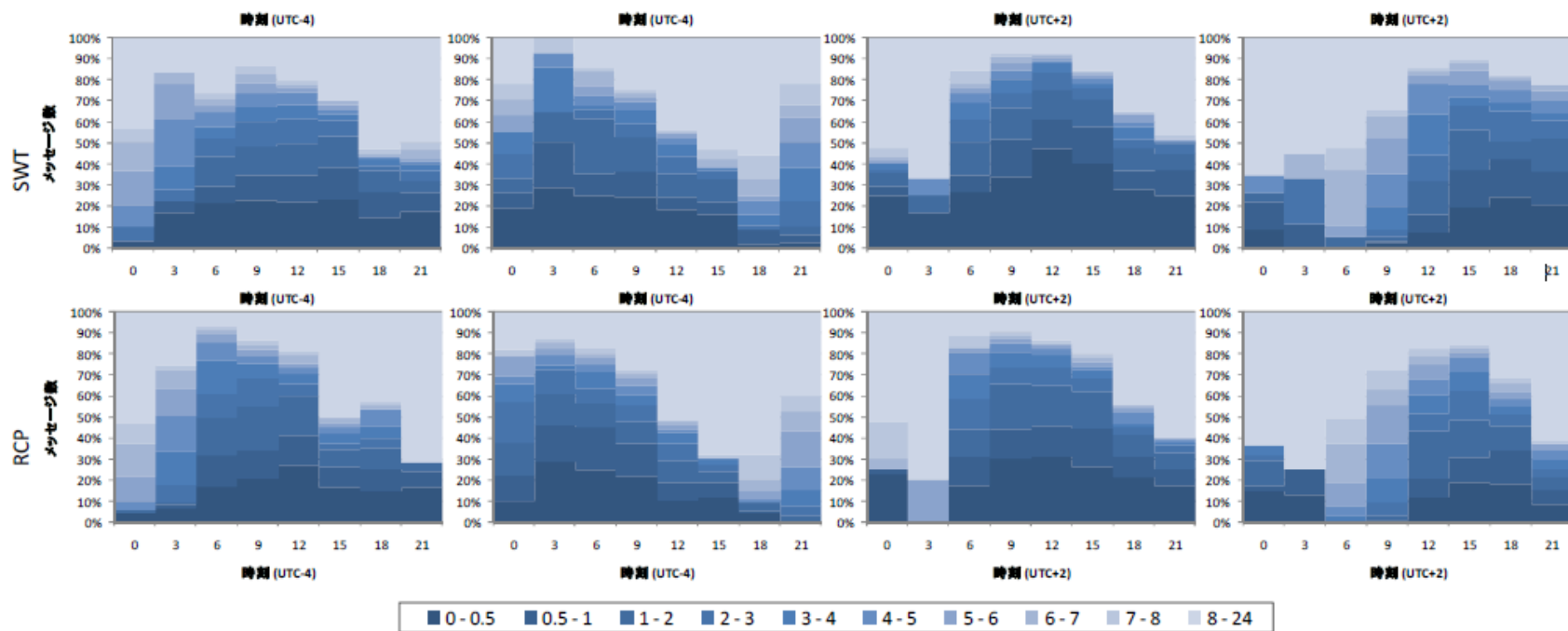


分析結果 1 より

- ▶ PostgreSQLの米大陸地域同士で情報交換を行う場合（米→米）、10時から19時頃にメッセージを送信すると約60%の確率で送信してから1時間未満に返信される。
- ▶ 一方、米大陸地域から欧・アフリカ地域に対してメッセージを送信する場合（米→欧）、米大陸地域の5時から12時頃にメッセージを送信すると約50%の確率で送信してから1時間未満に返信される。



(仮説 2) 分析結果 2



分析結果 2 より

- ▶ 図-4-1のPython, Apacheプロジェクトの米→欧に着目すると, 米大陸地域の現地時刻6時から15時(欧・アフリカ地域の12時から21時)にメッセージを送信すると早く返信がもらえる.
- ▶ Eclipse Platform, SWT, RCPプロジェクトの米→欧に着目すると, 米大陸地域の現地時刻3時から12時(欧・アフリカ地域の9時から18時), つまり, 現地の業務時間内にメッセージを送ることがよいことがわかる

仮説 2 の検証

- ▶ 送信者の地域によって素早い返信を期待できる時間帯が違っていたことが分かった。
- ▶ これにより仮説2は正しかったとする。



素早い返信が期待できる時間帯

表-5 素早い返信が期待できる時間帯（米：米大陸地域、欧：欧・アフリカ地域）

分析対象プロジェクト			返信の早い時間帯（現地時刻）			
			米→米	米→欧	欧→欧	欧→米
Python			9～20	7～16	14～23	16～2
PostgreSQL			10～19	5～12	10～18	16～0
Apache HTTP Server (Apache)			10～19	5～13	9～19	18～1
Eclipse	Java development tools (JDT)		9～21	3～15	9～18	15～21
	Plugin Development Environment (PDE)		6～12	6～12	9～18	15～21
	Business Intelligence and Reporting Tools (BIRT)		6～15	3～12	6～21	12～21
	Eclipse Modeling Framework (EMF)		6～18	3～12	6～18	12～0
	C/C++ Development Tooling (CDT)		6～18	6～12	9～18	15～21
	Eclipse Web Tools Platform Project (WTP)		6～18	6～12	9～18	12～21
	Eclipse Platform	Platform	9～18	3～12	6～18	15～0
		SWT	9～18	3～12	9～18	15～0
		RCP	6～15	3～12	6～18	12～21

まとめ

- ▶ ボランティア型OSS開発プロジェクトでは
- ▶ 返信時間が深夜から早朝となる時間帯にメッセージを送信すると、日中に比べて開発者間の情報交換にタイムラグが発生する可能性が高いことが分かった。
- ▶ 米大陸地域では10時~12時頃、欧・アフリカ地域では18時頃であることが分かった。



まとめ

- ▶ 企業混合型OSS開発プロジェクトでは
- ▶ 返信地域が夕方から早朝である時間帯にメッセージを送信すると、情報交換にタイムラグが発生する可能性が高いことが分かった。
- ▶ 時差の影響を受けにくい時間帯は、米大陸地域では9時から12時頃、欧・アフリカ地域では15時から18時頃であることが分かった。



問題点

- ▶ あらかじめ時差を考えてメールを送信している場合は考慮されていない。(自分の考察)
- ▶ メーリングリスト以外の情報交換手段(チャット等のインスタントメッセージ)について分析することも今後の課題である。

