

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

亀井 靖高

大平 雅雄

伊原 彰紀

小山 貴和子

杉本 真佑

松本 健一

鵜林 尚靖

1

発表者 森岡 勇介

- 短期間かつ限られた開発コストのなかで高品質なソフトウェアを生産することが求められる。
- 行政機関や教育機関においても導入されつつあり，社会的影響は今後も拡大していくものと予想される。
- 世界中に点在する開発者が共同開発を行う分散開発の形態をとる。

- タイムゾーンが違うことによる情報交換のタイムラグによる
対応の遅れが見られる.
- 対応の遅れは開発スピード, 生産性, および, ソフトウェア
品質管理能力の低下など, ソフトウェア開発全体に悪影響を
もたらす可能性がある.

目的

- 情報交換に対する時差の影響を，OSS開発プロジェクトから調査し，定量的な分析に基づいた客観的な知見を得る.
- 得た知見によって，**情報交換の所要時間を短縮化**するための指針を得ることが目的.

仮説

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。

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。

これらの仮説に対して大規模な複数のプロジェクトを分析することで実験的に検証する。

ケーススタディ

分析データの収集と整形 .1

- MLのアーカイブを対象として分析を行う.
- メールヘッダを解析し、送信日時と送信場所のデータを収集する.
- 送信日時は現地時刻とし、送信場所は協定世界時（以降、UTC）と現地時刻の差で表す.

ケーススタディ

分析データの収集と整形 .2

- 例 東部標準時の地点からメッセージを送信した場合、標準時がUTCから5時間遅れているため、UTC-5と表される.
- 以降、地域Aから送信されたメッセージを地域Bから返信したメッセージの対を $A \rightarrow B$ と示す.

ケーススタディ

分析データの 収集と整形 .3

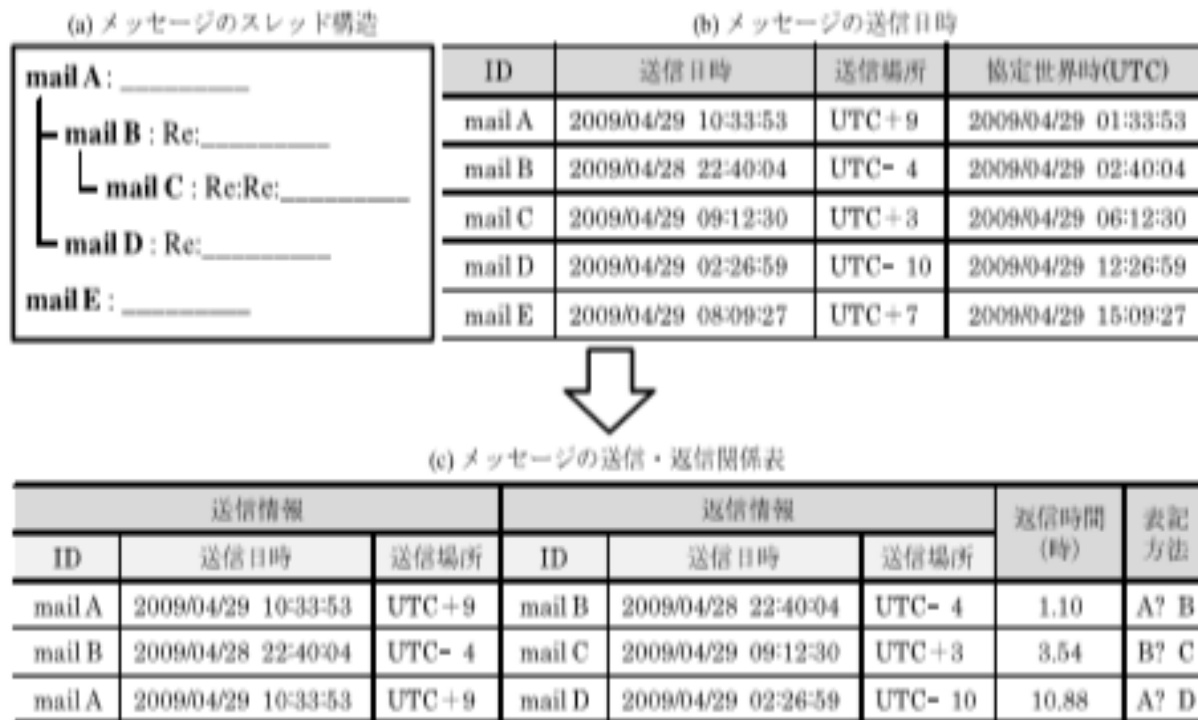


図-2 分析データの整形方法

分析対象プロジェクト

- Python プロジェクト（以降 Python）
- PostgreSQLプロジェクト（以降 PostgreSQL）
- Apache HTTP Server プロジェクト（以降 Apache）

分析対象プロジェクト

Eclipseプロジェクト (以降 Eclipse)

- Java development toolsプロジェクト (以降 JDT)
- Plug-in Development Environment (以降 PDE)
- Business Intelligence and Reporting Toolsプロジェクト (以降 BIRT)
- Eclipse Modeling Frameworkプロジェクト (以降 EMF)
- C/C++ Development Toolingプロジェクト (以降 CDT)
- Eclipse Web Tools Platform プロジェクト (以降 WTP)

Eclipse Platformプロジェクト (以降 Eclipse Platform)

- Platform
- SWT
- RCP

プロジェクトの分類

- ボランティア型OSS開発プロジェクト
- 業務混合型OSS開発プロジェクト

プロジェクトの分類

今回分析を行う対象のプロジェクトの分類

ボランティア型

Python , PostgreSQL , Apache

業務混合型

Eclipse(JDT ,PDE等)

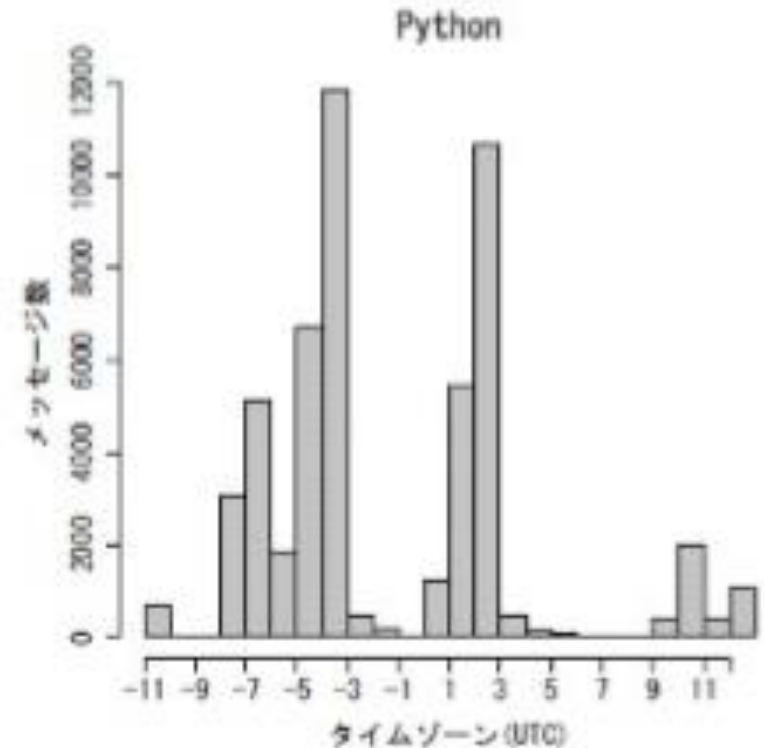
基本統計量

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	1,152	917
		SWT	eclipse.platform.swt	2003/05～2009/09	45,490	6,963	10,377
		RCP	eclipse.platform.rcp	2004/09～2009/09	37,583	4,796	13,614

基本統計量

- 縦軸がメッセージ数, 横軸は送信場所
- UTC-3とUTC+2のメッセージ数が明らかに多い.

他のプロジェクトでは
どうか？



基本統計量

全プロジェクトの半数以上が2つの地域から送返信されている.

1. UTC-8からUTC-4
(北アメリカ・南アメリカ地域. 以降 米大陸地域)
2. UTC+0からUTC+3
(ヨーロッパ地域・アフリカ地域. 以降 欧・アフリカ地域)

これら 2 地域を分析対象地域として
分析を進める

実験結果 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

返信時間が24時間以上のものは地理的要因ではない要因があるとみなし、検証の対象外としている。

実験結果 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

仮説 1 を検証するため、タイムゾーンが異なる場合それぞれ返信時間の分布を確認し、**時差の有無による返信時間の差**が統計的に有意かどうか検定する。

結果の前に . . .

ウィルコクソンの順位和検定

- ノンパラメトリック検定の一つ
- 全データまたは各水準における各データの**大小の順位、順序尺度**を利用する。
- 主にサンプル数が少なく得られた**データに正規性を仮定**できないときに用いられる。

<https://kusuri-jouhou.com/statistics/wilcoxon.html>

結果

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	4.28	0.03	2.12	3.56	0.00
		SWT	2.42	3.25	0.00	0.96	3.42	0.00
		RCP	2.90	3.54	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

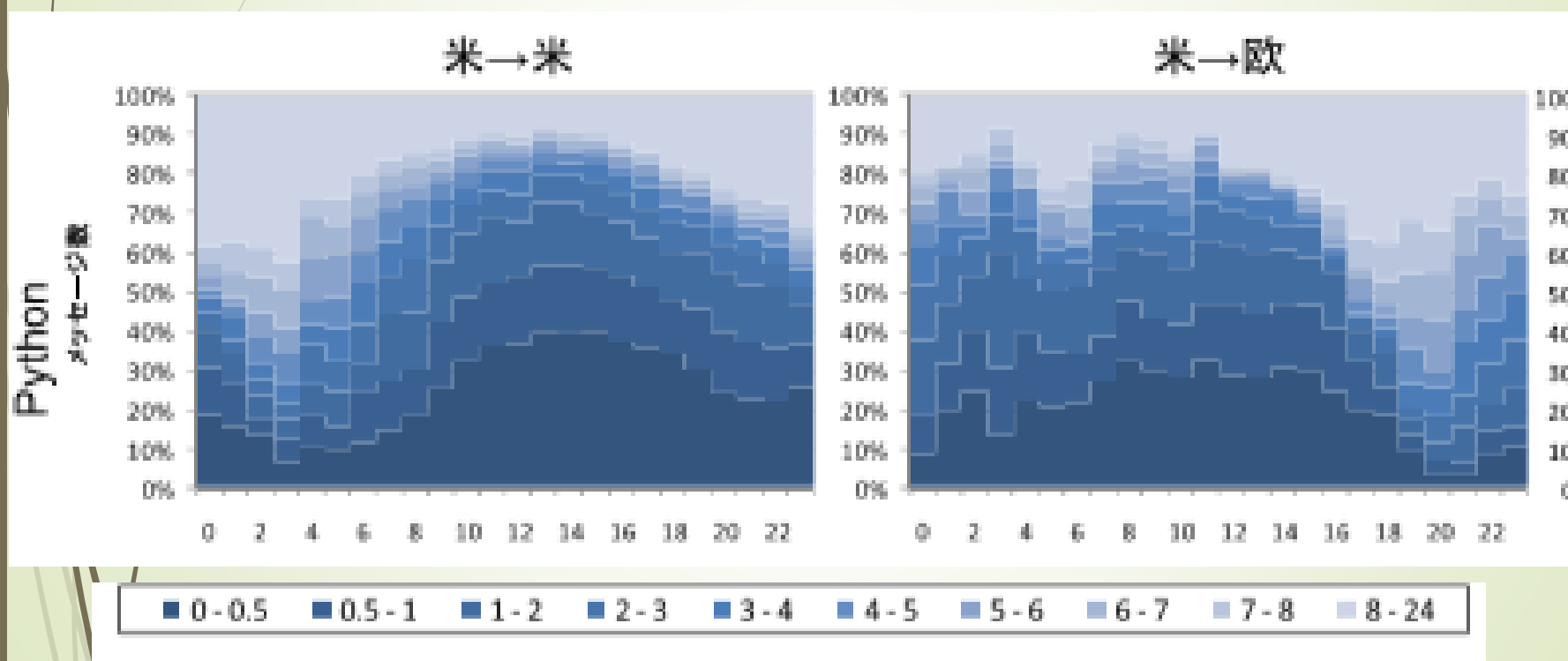
実験結果 仮説 2

仮説 2 返信者の地域によって素早い返信が期待できる時間帯が異なる

分析 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.

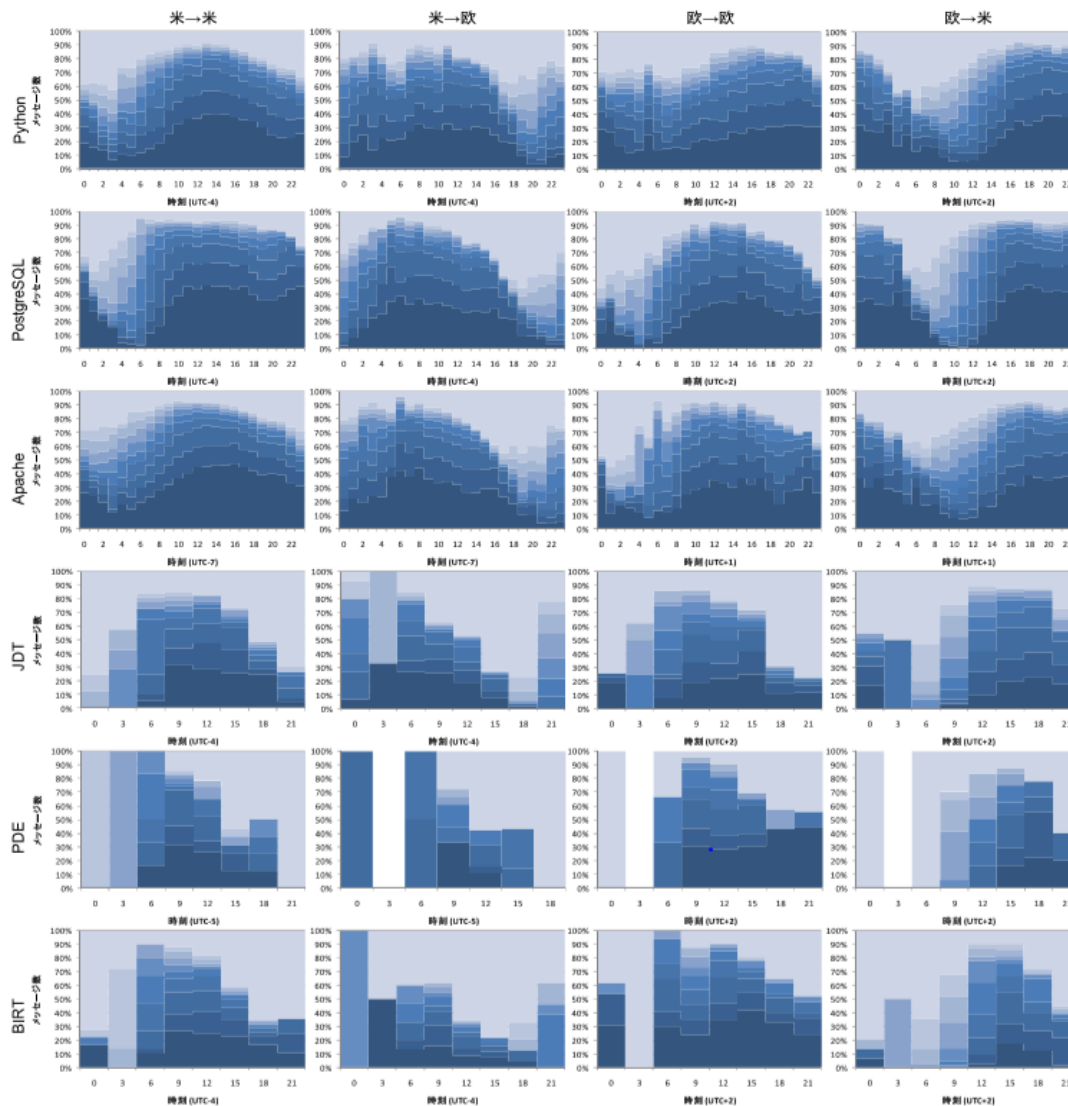
実験結果

仮説 2



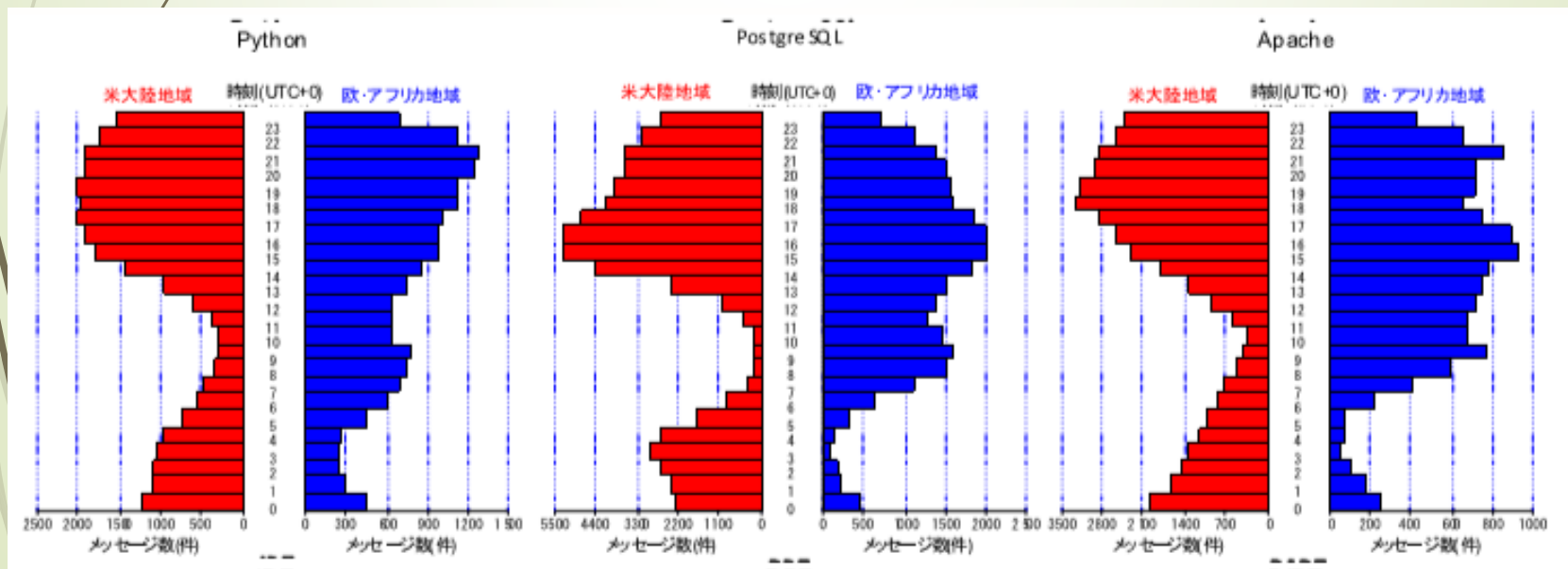
実験結果

仮説 2



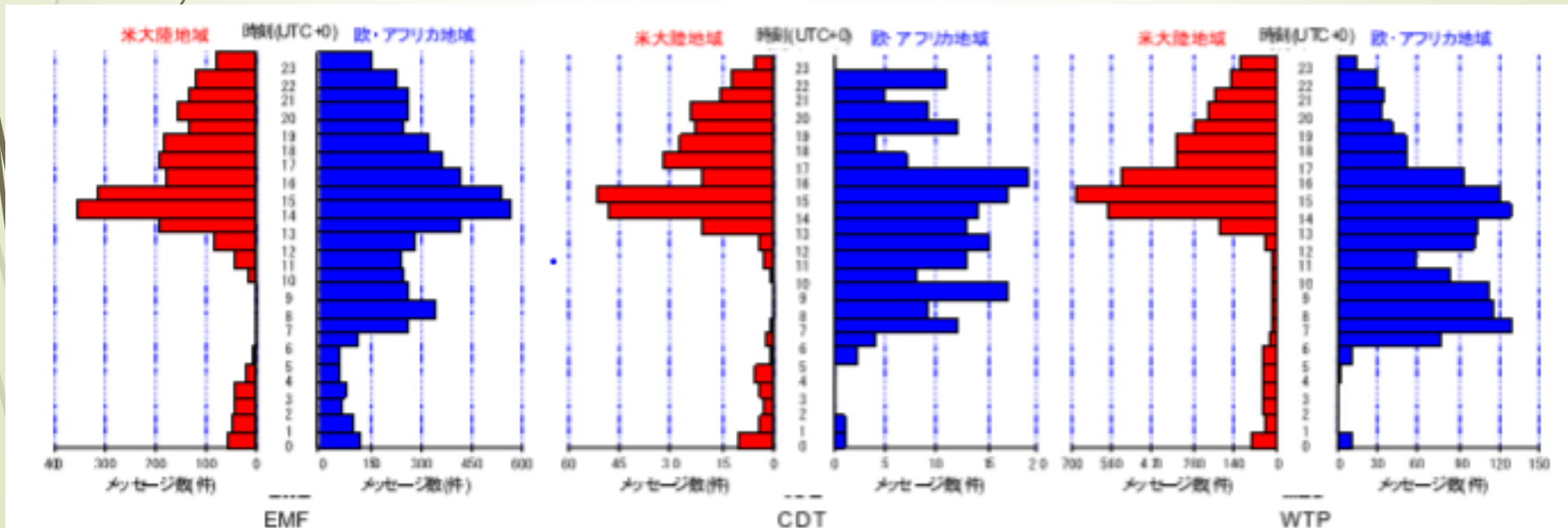
ボランティア型OSS開発プロジェクト

仮説 1 より，時差のある地域間と時差のない地域間の返信時間の差に有意差はあったが，**1 時間程度**であった．



業務混合型OSS開発プロジェクト

下の図より、ボランティア型に比べて情報交換が
活発な時間帯が短いことが分かった。



業務混合型OSS開発プロジェクト

表-6 各地域のトップドメイン（企業と混合のOSS開発プロジェクト）

分析対象プロジェクト	米大陸地域	欧・アフリカ地域
Java development tools (JDT)	企業(22.4%)	企業(23.6%)
Plugin Development Environment (PDE)	企業(28.0%)	その他(15.5%)
Business Intelligence and Reporting Tools (BIRT)	企業(67.8%)	その他(13.2%)
Eclipse Modeling Framework (EMF)	企業(73.1%)	その他(13.8%)
C/C++ Development Tooling (CDT)	企業(35.8%)	企業(15.3%)
Eclipse Web Tools Platform Project (WTP)	企業(31.2%)	その他(11.4%)

各地域のドメインから企業の開発者の数を調査した結果

考察

米大陸地域の開発者は10時から12頃,
欧・アフリカ地域の開発者は18頃頃,

この時間帯にメッセージを送信することで素早い返信が期待できる.

開発プロジェクトの改善案

米大陸地域

12時頃が時差の影響を受けづらいので、昼食の時間をずらす、もしくは活用する。

欧・アフリカ地域

18時頃が時差の影響を受けづらいので、業務開発者は業務終了時間を1時間ほど遅らせ、ボランティア開発者は業務終了後に情報交換を行う。

まとめ

- 時差のある地域間の情報交換は返信時間が長くなる.
- 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る.
- 少し工夫をすることで、情報交換の所要時間を短縮化することができる.
- メーリングリスト以外の情報交換手段について分析することが今後の課題の一つに挙げられる.