

スケジュール情報からのインフォーマルコミュニティ抽出

桑 照宣[†], 鵜飼 孝典[†], 西野 文人[†]
e-mail: {t-kume, ugai, nishino}@jp.fujitsu.com

企業などの組織では、業務上のフォーマルなコミュニティ(プロジェクト) 以外にインフォーマルなコミュニティが存在する。このコミュニティは、プロジェクトメンバが作業を円滑かつ効率的に進めるために重要な役割を果たしている。また、組織運営として、その実態を把握しておくことで、プロジェクトの状況把握、担当者への作業指示や人材・資源の投入などを効率的に行うことができる。しかしながら、このようなインフォーマルなコミュニティは、プロジェクトメンバの作業状況やプロジェクトの進捗状況の変化にあわせて、時々刻々できては消えて行くために、その実態を把握することは困難である。また、特に大きな企業では、その存在を発見することすら困難である。本稿では、様々に変化するコミュニティのある時点の状況を予定表のデータから抽出する方法を提案し、実験結果を報告する。

Finding informal communities from schedule data

Terunobu Kume[†], Takanori Ugai[†], Fumihito Nishino[†]
e-mail: {t-kume, ugai, nishino}@jp.fujitsu.com

In an enterprise organization, a lot of informal communities exist besides formal communities, like as projects, on the business. The communities are playing important roles so that the project's members may proceed work smoothly and efficiently. And the managers can understand the project status, instruct person in charge about the job and manage person and the resource efficiently by understanding the realities. However, it is difficult to understand the real status of informal community, because it may be variable by member's job change and the progress of the project. Moreover, it is difficult in an especially big enterprise even to discover the existence. In this paper, we proposes the method of express the situation of point with the informal communities from the the schedule data, and we describe the experiment result.

[†]株式会社富士通研究所

[†]Fujitsu Laboratories Ltd.

1 はじめに

企業などの組織では、業務上のフォーマルなコミュニティ(プロジェクト)以外にインフォーマルなコミュニティが存在する。インフォーマルなコミュニティとは、個人的な友人関係のことではなく、共通の問題を持った人達が集まり問題解決方法について検討するワークグループや、最新技術などについての意見や情報交換を行う研究会などを指す。このコミュニティは、プロジェクトメンバが作業を円滑かつ効率的に進めるために非常に重要な役割を果たしている。また、作業担当者はプロジェクト状況に応じて、情報収集や相談のために必要なコミュニティを発見し活用する。組織としては人材・資源の投入などを効率的に行うことができる。

しかしながら、このようなインフォーマルなコミュニティは、世の中の流行、プロジェクトメンバの作業状況やプロジェクトの進捗状況の変化にあわせて、時々刻々変化している。そのため、その実態を把握することは困難である。また、特に大きな企業ではプロジェクトだけでも数が多く、その中から発生するインフォーマルなコミュニティに至っては、その存在をコミュニティ外の人が発見することは難しい。

インフォーマルなコミュニティ発見支援の取り組みとして、情報を知っている人を探す KnowWho 技術を応用し、スケジュールの情報や打合せ議事録の参加者の情報から人間関係を抽出し、その人間関係からインフォーマルなコミュニティを発見する片山らの方法 [1] やスケジュールとセンサ情報を組合せることでユーザの活動を推測する岡本らの方法 [2] があり、ある程度の成果が得られている。しかしながら、これらの手法では次の2点で不十分である。

1. 役目を完了したコミュニティが強く表現されてしまい、新たに発生したコミュニティが発見しにくい。
2. コミュニティ発見に影響する情報と影響しない情報を、全ての情報を等価に扱うため、不要な関係を抽出する場合や必要な関係が抽出できない場合がある。

本稿では、時々刻々変化するインフォーマルなコミュニティのある時点の状況を抽出するための方法として、予定表のデータを利用し、情報の経過時間にもなる重み付け手法と、情報の特性による重み付けを行う手法を提案する。

既存の手法と本稿の提案する手法のそれぞれを使い、同じデータを用いて実験を行ったとこ

ろ、時間による重み付けでは、本稿の提案する手法の方が、より現状を反映した結果が得られることがわかった。また、情報の特性による重み付けでは、用いる手法によりいくつかの特性があることがわかった。

本稿は、以下、次の様な構成とする。第2節では、既存の手法でのコミュニティ抽出方法の問題を述べ、第3節で、情報の経過時間にもなる重み付けの手法について述べ、その実験結果を報告する。第4節では、情報の特性を考慮する手法について述べ、いくつか試した結果を報告する。第5節では、まとめと今後の課題を述べる。

2 既存手法の問題点

既存のコミュニティは抽出手法は以下の様な手順で行なっている。

手順1 スケジュールデータからメタ情報の抽出

片山らの方法では、ユーザ自身が打合せのメタ情報として、会議の出席者や議事録などを登録する手法をとっている。また、岡本らの方法では、複数のユーザが入力したスケジュール情報を使い、情報の抜けを補完することでメタ情報の抽出を行う手法をとっている。

手順2 メタ情報から同一の予定を持つ人を抽出

片山らの方法、岡本らの方法とも、メタ情報に入力された出席者を同一の予定を持つ人として抽出する。

手順3 抽出された人の組合せを関係付け

片山らの方法では、抽出された同じ予定を持つ人同士に関係付け、その回数を関係の強さとしている。また、岡本らの方法では、その日時のスケジュールを持つものを表示している。

実際に全59名のユーザで約3年利用した全892件データから、片山らの方法を用いて抽出されたコミュニティの例を図1に示す。各ユーザが線で結ばれ1つの塊となっている部分が、抽出されたインフォーマルなコミュニティである。図1では、2つのインフォーマルなコミュニティが抽出されている。

この結果を実際の状況と照合すると、図1中の実線で囲まれた部分は既に組織変更により他部署へと移動してしまったコミュニティであり、また、破線で囲まれた部分は既に退職した人物であることがわかった。

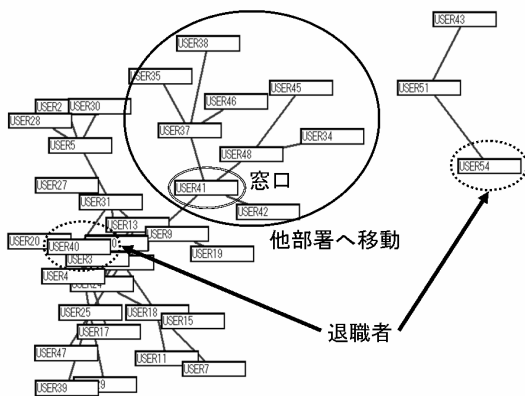


図 1: 既存方法によるコミュニティ抽出

前述の手順により抽出されたコミュニティは、現在の状況を適切には反映しておらず、不要な関係を抽出していることがわかった。この原因は、既存の手法には以下の考慮がないことが考えられる。

- 情報の時間による減衰
- 情報の特性

まず、1つ目の問題は、現状の手法では、過去、現在に関わらず全ての情報を等価に扱い関係を抽出するため、過去に非常に強い関係があった場合に、その関係に引きづられ現在の弱い関係が消えてしまう。つまり、役目を完了したコミュニティが強く表現されてしまうので、新たに発生したコミュニティが見えなくなってしまう。この方法は、形成されたら切れることの少ない人間関係においては、その関係の強弱を抽出するには良い方法であるが、今、活動しているコミュニティを抽出することができない。

また、2つ目の問題は、本来関係として取得すべきでない情報や逆により強く関係を表現しなければならない情報というものを考慮していないため不要な関係が抽出されてしまうということである。例えば、「会議」の特性としては、部会やプロジェクト会などに同時に参加したとしても、それはインフォーマルなコミュニティを抽出するための関係ではない。逆に、ワークグループや研究会であればインフォーマルなコミュニティを抽出するための関係である。

これらの問題を解決する手法として、関係抽出の際に重みを付ける手法を考案し実験を行った。

3 重み付けによる抽出方法

3.1 時間による情報の減衰

第2節で示した問題点の1つである「情報の時間による減衰」を考慮する手法について述べる。

この手法では、スケジュール情報に含まれる日付のデータを活用する。つまり、現在の日付から見てどの程度の時間が経過しているかを計算し、関係抽出の際に、経過時間に応じた重み付けを行う。重み付けの計算方法のモデルとしては、次の様なものが考えられる。(1) ある程度、活動がある最新のコミュニティを抽出したければ、時間経過とともに重みが単調減少するモデルを利用する。また、(2) これから活動が活発になりそうな発足したてのコミュニティを抽出したければ、時間経過が非常に大きければ重みを非常に小さくする様なモデルを利用する。さらに、(3) フォーマルなコミュニティの変化を抽出したければ、組織変更などのタイミングで段階的に重みを下げるモデルが考えられる。

3.2 コミュニティ抽出実験 (時間)

上記手法を用い、時間経過による情報の減衰を考慮することで、抽出されるコミュニティがどのように変化するかを実験する。

入力データとしては比較のために、第2節で用いたデータを利用する。今回は、第3.1節で提案した方法の内、(1)の時間とともに重みが単調減少するモデルに一次関数をあてはめ実験を行う。

まず、既存の手法の手順に従いスケジュール情報からメタデータを抽出し、メタ情報から同一のスケジュールを持つユーザと、日付、予定名を抽出する。抽出されたデータに対し、関係の重み付けを実施する。今回の実験では、スケジュールデータの最新日付と最終日付を先に抽出しておき、データの存在する区間の日数を数え、以下の式で重みを計算する。

$$(\text{重み}) = -\frac{1}{\text{区間の日数}} \times (\text{経過日数}) + 1 \quad (1)$$

式(1)を適用し、得られた関係データを用いて人脈マップを生成すると図2の様になる。

図1と図2を比較すると、図1で問題となっていた、組織変更により他部署へ移動してしまったコミュニティは図1からUSER41が2つのコミュニティをつなぐ役目であったことがわかる。

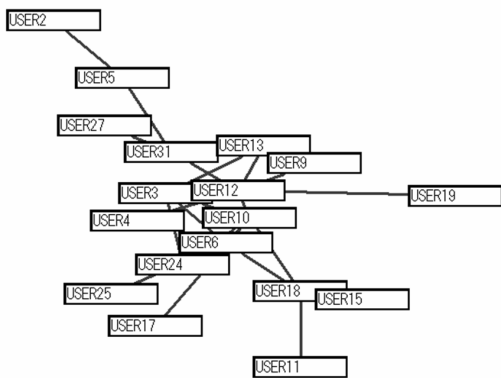


図 2: 時間経過を考慮したコミュニティ抽出

図 2 では窓口であった USER41 が消え、コミュニティが全体から削除された。また、退職した USER40 と USER54 が削除され、USER54 を含んでいた小さなコミュニティが抽出されていないことがわかる。

3.3 実験の評価 (時間)

経過時間にあわせ重みを変化させることと、最新データのみを利用することの違いを確認するために、以下の比較実験を行う。

データに含まれる最新の日付から 3ヶ月分を切り出し、そのデータを使い片山らの手法で抽出したコミュニティが図 3 である。

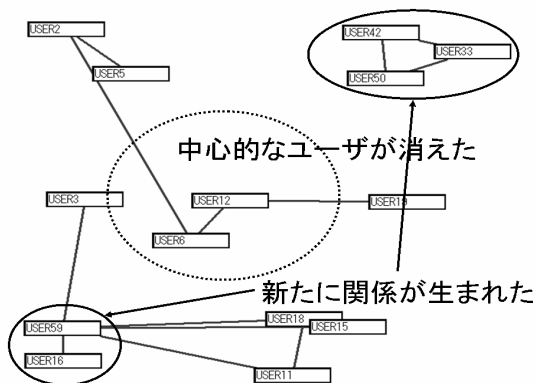


図 3: 最新データによるコミュニティ抽出

図 2 と図 3 を比較すると、図 3 では人間関係の中心 (ハブ) となっている人物が抽出されていない。そのために、本来 1 つであるコミュニティが 3 つに分割されてしまっている。ハブとなる人物は長い時間をかけ様々な人と繰り返し

接点を接点を持っている。そのため、短期間では関係が弱く抽出されないことが原因と考えられる。また、本来抽出するべきでない関係を抽出してしまっている。これは短期間の中でたまたま接点が多い関係が抽出されてしまっていることが原因であると考えられる。つまり、最新データを用いるだけでは現状を反映したコミュニティを抽出することができない。

次に、先の実験とは別のデータとして 15 部署の 1ヶ月分の週報を利用し週報に含まれる次週の予定から岡本らの手法を用いメタ情報の抽出した。週報のデータ数としては 1,337 件で、予定情報としては 136,115 件である。このデータを用い実験を行った結果を示す。図 4 は従来の手法により抽出したもので、図 5 は時間経過を考慮したものになる。図中の点線で囲まれた部分に注目してみると、従来手法では抽出されていた関係が、時間経過を考慮した場合には削除されている。実際に確認をとると、時間経過を考慮した場合のコミュニティが正しいことを確認した。

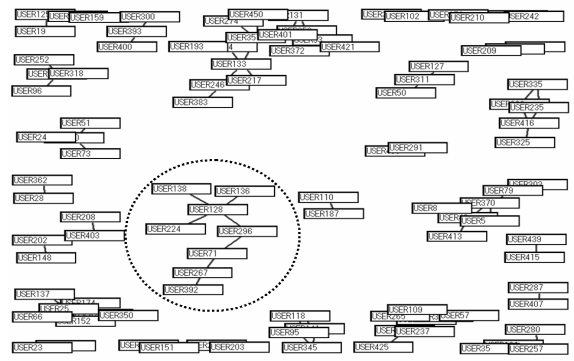


図 4: 時間経過を考慮しない場合

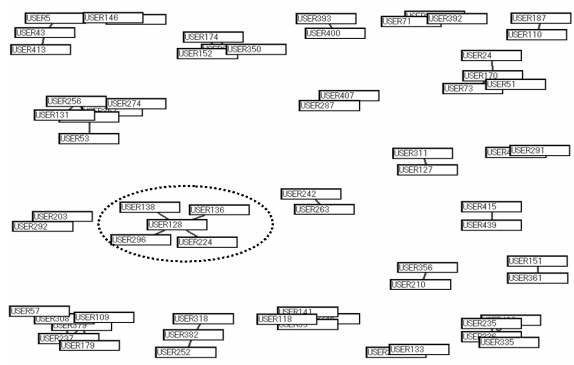


図 5: 時間経過を考慮した場合

このことから、時間経過による情報の減衰を考慮することにより、インフォーマルなコミュニティを抽出するために必要なデータを残しつつ不要なデータを削除することができ、コミュニティ抽出方法として効果的であることがわかる。

3.4 情報の特性

次に「情報の特性」を考慮する手法について述べる。ここでは、予定名などの情報から判断してコミュニティを抽出するのに重要かそうでないかを判断するルールを用いる。今回は、以下の3つの手法を用い、抽出されるコミュニティにどのような違いがあるかを第2節で使用したデータを用いて比較実験を行う。

[ルール1] 特定キーワードによる重み付け
 予定名中に含まれる単語の出現回数が多いものの内、フォーマルなコミュニティを特定する単語とインフォーマルなコミュニティを特定する単語を分類し以下のルールとした。予定名中に「部会」「ブリーフィング」「グループ会議」が含まれている場合は重みを下げ、「研究会」が含まれている場合は重みを上げる。

[実験結果1]
 メタ情報から日時と予定名を抽出し、予定名に含まれるキーワードにより重み付けを行った後、時間経過による重み付けを行った。(図6)

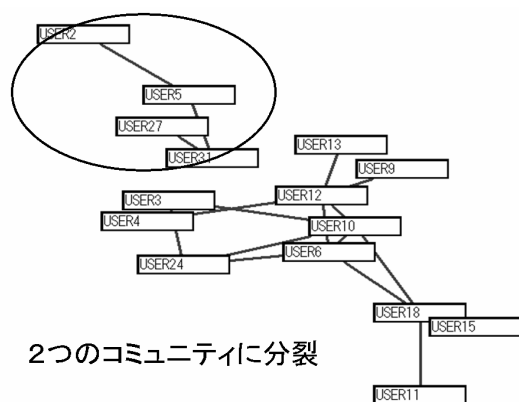


図6: 特定キーワードによる重み付け

[ルール2] 出席者の人数による重み付け
 ある程度の人数を越えた場合は、出席者全員とコミュニケーション発生するとは考えにくい。そこで、全員とコミュニケーションがとれる範囲の出席人数を考え以下のルールとした。出席

者の人数が10人以上の場合は重みを下げ、出席者の人数が5人以下の場合は重みを上げる。

[実験結果2]
 実験結果1の手順とほぼ同様の手順である。予定名を抽出する代りに出席者の人数を抽出し重み付けを行う。

[ルール3] 出席者の所属による重み付け
 出席者の部署が1種類ならば、それはフォーマルなコミュニティでの打合せである可能性が高く、また、部署が2種類ならば、取り引き先と自分の部署である可能性が高いと考え以下のルールとした。出席者の所属の種類が2種類以下なら重みを下げ、3種類以上ならば重みを上げる。

[実験結果3]
 メタ情報から日時と出席者名を抽出し、利用者情報と出席者名を比較し所属を抽出する。日時と所属種類数により重み付けを行う。

3.5 実験の評価 (情報)

実験1~3の結果を図2と比較しながら検証すると、実験1で抽出された図6以外は、ほぼ図2と同じであった。まず変化の無い実験2と実験3の原因について検証する。実験2では出席者の人数を判断基準としていたが、全データ892件中に出席者が10人以上であるものは5件であった。また、逆に5人以下であるものは867件と全体の97.1%であり、このルールにより全体的に重みが上昇しユーザ間の関係が強くなるだけで、コミュニティ抽出としては変化がなかった。今回用いたスケジューラのユーザは10人以上になる様な会議の情報を入力せず、小規模なフォーマルコミュニティの情報のみを入力する傾向にあり、そもそも、データの段階で大人数のフォーマルなコミュニティについての情報が入力されていないことが原因であった。

次に、実験3では出席者の所属部署の種類を判断基準としていたが、全59名のユーザの所属を調査すると、組織変更などもあり多少の増減はあるものの最大でも3部署であった。このルールを適用するにはユーザの所属部署の範囲が狭く、このルールによりユーザ間の重みが下降するもののコミュニティ抽出としては変化がなかった。また、先に述べたように、ユーザが情報入力の際に小規模なフォーマルなコミュニティの情報を入力する傾向にあることも原因であった。

最後に、図2と図6を比較すると、図2では1つのコミュニティが図6では分割され2つに

なっている。実際と比較するとこれは間違いであり図2が正しい。繰返しになるが、この原因はユーザが入力する情報が主に小規模のフォーマルなコミュニティについての情報であるために、データに偏りがあると考えている。このルールにおいては、予定名に含まれるキーワードの内、出現回数の多いものを分類し利用したが、それ以外のキーワードも顧客名やプロジェクト名である場合が多い。全データ数 892 件中のキーワードの分類結果は表1の通りである。

表 1: キーワード数

キーワード	出現回数	%
プロジェクト名系	503	56.3
顧客名系	207	23.2
その他	33	3.69
部会	18	2.01
ブリーフィング	27	3.02
グループ会議	99	11.0
研究会	5	0.56

ユーザが入力した情報から判断しているために正確に分類されていない部分もあるが、その他に含まれているデータがインフォーマルなコミュニティを抽出する上で重要なデータであると考えている。今回のデータではその他に含まれるデータ数が少なくデータとして不充分であった。

4 考察

時間による情報の減衰の考慮に関しては、従来手法よりもより精度の高いコミュニティを抽出することができ、その有効性を確認することができた。

しかしながら、情報の特性の考慮に関しては、今回のデータでは片山らの方法を用いた場合に、その有効性を確認することができなかった。今回実験に用いたデータにはフォーマルなコミュニティ関連のデータが入力され、インフォーマルなコミュニティ関連のデータは入力されていなかった。また、フォーマルなコミュニティも人数の多い打合せなどは入力されず、少人数のフォーマルなコミュニティの情報が主なデータになる。これはデータを取得したスケジュール管理システムの情報入力のためのインターフェースにも問題があると思われるが、参加人数が多い場合には出席者の情報などを入力するのが面倒であり、また、打合せ自体に受け身で参加しているために自らデータを入力しないと考

れる。そのため、少人数のメタ情報が主になっていると考えている。この傾向はスケジュールシステムの利用者の意識、組織の雰囲気などにより全く異なると思うが、今回の実験では、片山らの人手による入力情報のみを利用する手法はフォーマルなコミュニティを抽出するには有効であるが、インフォーマルなコミュニティの抽出には向かなかった。

今回の実験方法の内、データ収集方法を岡本らの方法に変更することで、系統的にメタ情報を取得すれば、片山らの手法と比べ種々雑多なデータが収集されると考えられる。この場合には、情報の特性を考慮しインフォーマルなコミュニティ抽出に不要なデータを削除することが有効であると考えられる。

5 まとめと今後の課題

今回の実験結果をまとめると表2の様になる。

表 2: 実験結果まとめ

	片山らの手法	岡本らの手法
時間経過		
情報特性	x	

時間経過による情報の減衰に関しては良好な結果を得ることができたが、情報の特性を考慮する実験に関しては、あまり良好な結果ではなかった。

今後は、岡本らの手法によるデータ抽出方法を実験し、その精度について調査を行うことで、片山らの手法と岡本らの手法の特徴について検証が必要であると考えている。また、情報の特性として、「フォーマル」と「インフォーマル」を分割するために必要なものについても、より精度の高いルールを検討する必要があると考えている。

参考文献

- [1] 片山佳則, 小櫻文彦, 井形伸之, 渡部勇, 津田宏. セマンティックグループウェア workware++ と knowwho 検索への応用. 情報処理学会研究報告, 2003-FI-71(2), 2003-DBS-130(2), pp. 9-16, 5 2003.
- [2] 岡本昌之, 梅木秀雄. スケジュールとセンサ情報を活用した状況認識手法の提案. グループウェアとネットワークサービスワークショップ 2005, Vol.2005, No.14, pp. 13-18, 10 2005.