

知識共有コミュニティを創り出す人たち(4) 形態素解析による「知識」の分析

People who create knowledge-sharing community(4): What is included in “Chie-Bukuro”?

三浦 麻子*1
Asako Miura

川浦 康至*2
Yasuyuki Kawaura

地福 節子*3
Setsuko Jifuku

大瀧 直子*3
Naoko Otaki

岡本 真*3
Makoto Okamoto

*1神戸学院大学
Kobe Gakuin University

*2東京経済大学
Tokyo Keizai University

*3ヤフー株式会社
Yahoo Japan Corporation

Web-based knowledge sharing communities (KSCs), which are supported by countless and voluntary Internet users, are in widespread use in our life. There are two types of KSCs - (1) Dictionary type and (2) Question & Answer type. We focus on the latter type of KSC in which interpersonal communication of users is salient as a process of information accumulation. A questionnaire survey on participants of Yahoo! Chiebukuro, one of the most popular Q & A type KSCs in Japan, was conducted and logs of their behavior in the community were collected. Based on 7,989 survey samples and logs of their behavior, we found that the Chiebukuro community was activated by both questions by a large amount of newcomers and a large amount of answers by experienced participants. Various kinds of information were exchanged and accumulated actively in the community and interpersonal communication of community participants were developed by aggressive need to information acquisition and subsequent social support. Content analysis of questions and answers posted by participants suggested that there existed some career differences.

1. はじめに

インターネットのさまざまなメディア・技術は、一般の多くの人 (real people) が自由に大勢に向けて情報発信でき (一握りの著名人でなく) 多くの人との共有知識が可視化される社会の実現を目指して開発されている。その前提には、インターネット上では誰もが性善説に基づいた行動をするだろうというポジティブな世界観が存在するように思われる。もちろん、実際のインターネット社会で起きる事象は決して「善」のバラ色に彩られたものばかりではない。しかし、中にはこうしたインターネットの自在性を生かした利用者の情報発信が活発かつ有用なオンライン・コミュニティを創出する例もある。

その代表が本稿で採り上げる知識共有コミュニティである。知識共有コミュニティ (Knowledge-Sharing Community; 以下 KSC と略記) とは、インターネット上の情報データベースのうち、利用者たちによってオンライン上でリアルタイムに情報の登録や更新がおこなわれ、一種のコミュニティとしても機能するもののことである。KSC は、これまでの情報データベースが専門家の「仕事」によって支えられてきたのとは異なり、無数のインターネット利用者による情報提供というボランティア的貢献によって支えられている。

著者は、KSC の中でも特に Q&A タイプのものを採り上げ、そこで共有される知識の姿を明らかにするために、同種のコミュニティ・サービス大手の一つである「Yahoo!知恵袋」の利用者を対象とした意識調査データと参加行動ログデータを紐付けした分析にもとづいて、Q&A タイプの KSC を構成する利用者の特性、そして彼らの実際の参加のようすや参加の理由を検討してきた。本発表では、さらにこうした KSC で共有される知識 (すなわち「知恵袋」の中身) のもつ特性を詳細に考察するために、利用者の参加行動パターン、特に質問・回答数の多さに応じた異同に注目して、投稿内容に関する分析をおこなった。

2. これまでの知見

KSC は、インターネット上で運用される参加型の情報データベースであるという特徴を共有しつつ、(1) 辞書タイプ、(2) Q&A タイプの 2 つに大別される。辞書タイプとは、Wikipedia など、正確かつ詳細な情報の蓄積を目的とした、利用者間の対人コミュニケーションは表出しないものであり、Q&A タイプとは、情報の蓄積過程が利用者間の「質問」と「回答」のやりとりという対人コミュニケーションとして表出するもので、オンライン・コミュニティとしての機能も合わせもつものである。

Q&A タイプの KSC の特徴は、辞書タイプの KSC に含まれる知識が、「正解」が存在するいわば狭義のそれだけであるのとは異なり、解が一意に定まらない「質問」に対する多様な「解答」も知識の範疇に含まれていることである。こうした知識の獲得には、オンライン・コミュニティが大きな役割を果たすことが期待される。なぜなら、こうした知識は、個人が独力で解を探索しようにも、手がかりが曖昧だったり、利用可能な資源が限定されていたり、あるいはプライバシーに抵触する事柄だったりするために、既存資源を利用することが難しい場合が多い。そのような場合に、Q&A タイプの KSC に質問を投げかけることによって、互いのプライバシーを明かさないう状態で、自らの手のみでは獲得し得ない多様な解を得られる可能性があるからである。

著者ら [Miura et al. 07] は、Q&A タイプの KSC における全般的な知識共有プロセスについて、(1) 正確さの要求される質問より正解の特にならぬ質問の方が答えられやすい傾向、(2) 回答者による情報の正確さの再確認はあまりおこなわれていない傾向、(3) 質問投稿者にとっての重要度や緊急度は回答者によって過少評価されやすい傾向、を明らかにしている。これらのことから、Q&A タイプの KSC では、ある参加者の質問趣旨に沿い、それに回答可能な参加者が、投稿時に既に持っている経験や記憶にもとづく情報を提供するというプロセスで知識が蓄積・共有されていることが推察された。コミュニケーションが重視される Q&A タイプの KSC では、正確さがある

程度担保されている静的データベース、正確さを追求して常に議論・改訂が繰り返される辞書タイプの KSC などとは提供される情報の質そのものが異なる可能性があり、またそのことが「みんなの意見」が「案外たためになる」集団の知恵として多くの利用者に受け入れられて、結果として活発な対人コミュニケーションが展開されていることが考えられる。

さらに、質問や回答の投稿という形で KSC を積極的に利用する人々が、どのような動機で参加しているのかについても検討されている。その結果、物理要因（検索コストの高さ）が質問を、心理要因（困っている人を助けたい）が回答を、それぞれ促している可能性が示された。さらに、コミュニティへの愛着は、質問より回答投稿理由として多く挙げられていたことから、切迫した援助を求める人を自身のコミュニティメンバーとして認め、好意にもとづいて援助を実行している利用者の姿が明らかになった。こうした分析結果は、KSC の参加者、特に積極的な情報提供者の多くが純粋な善意で参加しており、彼らの自由な情報発信が活発なオンライン・コミュニティを創り上げているさまを示すものであった。

では、知識データベースとしての有用さについてはどうだろうか。前述したように、Q&A タイプの KSC では提供される情報の質そのものが従来型データベースや辞書タイプの KSC と異なる可能性が大きく、「掲載されている情報が正確であること」だけが有用さの指標となるわけではない。正解の存在しない問題解決型の質問については、数多くの、そして多様な情報が寄せられることが有用さを高めることにつながることを考えられる。質問紙調査参加者の調査時点までの質問投稿数と投稿した質問の被回答数を比較した結果、平均すると 1 質問に対して 3.5 以上の回答が投稿されている（平均値：質問投稿数 16.0、被回答数 57.1）。この数値が、正解が一意に定まる、すなわち 1 つの（正しい）回答があれば完結しうる質問の被回答数も多分に含んだデータにもとづくものであることを考慮すれば、第二の有用さについても十分なレベルにあると見てよいだろう。

3. 知恵袋の「中身」

このように、量的データを中心としたこれまでの分析の限りにおいて、Yahoo!知恵袋は、利用者相互の善意にもとづく活発な対人コミュニケーションに支えられた有用性の高い KSC であることが示されてきた。では、コミュニティの核となる「知恵袋」の中には、質的に見てどのような情報が詰まっているのだろうか。

著者らは、既に投稿されたログについて、対人コミュニケーションの性差の観点から分析をおこない、ログ内の名詞のみを抽出したデータにもとづく形態素解析の結果から、男性においてはより物質志向、女性においては人間関係志向の名詞が頻出していることを見いだしている [川浦ら 06]。本稿では、交換される知識における利用者の経験による変容に着目して、質問・回答数の多さによって利用者をグルーピングし、まず各群の特徴を明らかにした後に、投稿内容における群間の違いを検討する。

4. 分析データ

Yahoo!知恵袋利用者を対象とした質問紙調査結果と、当該利用者の過去の参加行動、および調査協力時に投稿した質問/回答投稿のログが分析対象となった。2005 年 12 月 6~20 日の 15 日間に Yahoo!知恵袋サイト上で質問/回答を投稿した利用者を対象として、投稿直後に質問紙調査への依頼画面が呈示

された。7989 名の調査協力者（質問投稿者 5515 名、回答投稿者 2474 名）について、知恵袋への参加状況データ（当該時点までの質問投稿数、回答投稿数、投稿した質問の被回答数など）と、調査直前に投稿された質問/回答のログデータを収集した。質問項目などの詳細については [Miura et al. 07] に示したとおりである。

5. 利用者のグルーピング

5.1 質問投稿者

調査に協力した質問投稿者 5515 名を、過去の質問投稿数に基づいて 3 群に分類した。質問投稿数は 1~2878 件で、平均 16.5 (SD 77.40)、中央値 3 であった。質問投稿数が 1 件、すなわち調査協力時が初めての質問投稿だった利用者が 1972 名（全体の 35.8%）に達しており、このうち回答投稿をしたことのないケースが 1800 名であった。すなわち、この 1800 名は知恵袋に初めて参加した際に調査に協力したことになる。これらの利用者は過去に質問/回答投稿をした経験のある利用者とは異なる属性を持っていると考え、1 つの群とした。残りの利用者は、度数分布を参考にして、質問回数が 2~9 回の利用者（2156 名）と 10 回以上の利用者（1387 名）の 2 群に分けた。質問は初投稿だが回答経験のある 172 名は、以降の分析から除外した。

5.2 回答投稿者

調査に協力した回答投稿者 2474 名を、過去の回答投稿数に基づいて 4 群に分類した。回答投稿数は 1~29977 件で、平均 209.3 (SD 1141.57)、中央値 33 であった。回答投稿数が 1 件、すなわち調査協力時が初めての回答投稿だった利用者は 332 名（全体の 13.4%）であった。このうち質問投稿をしたことのないケースは 217 名であった。すなわち、回答投稿者のデータの場合、知恵袋に初めて参加した際に調査に協力した利用者は 217 名だったことになる。これらの利用者は過去に質問/回答投稿をした経験のある利用者とは異なる属性を持っていると考え、1 つの群とした。残りの利用者は、いずれも過去に質問投稿による知恵袋参加経験のある利用者ということになる。残りの利用者は、度数分布を参考にして、回答回数が 2~19 回の利用者（743 名）、20~99 回の利用者（713 名）、そして回答回数が 100 回以上の利用者（686 名）の 3 群に分けた。回答は初投稿だが質問経験のある 115 名は、以降の分析から除外した。

6. 各群利用者の特徴

6.1 質問投稿者

投稿した質問の正解の有無についての群間比較は、表 1 にまとめたとおりである。初参加群以外のデータは、当該質問投稿者が「調査協力時」に投稿した質問に関するデータではあるが、初参加群で他群よりも「正解あり」の質問を投稿する頻度が有意に高いことが示された。

表 1: 投稿された質問の正解の有無：群間比較 (%)

	質問の正答有無	
	正解あり	正解なし
初参加	72.4	27.6
質問 2-9	66.1	33.9
質問 10-	62.9	37.1

また、正解のある質問を投稿したケースに限って、質問前に利用者自らが正解を探索した手段（オンライン/オフラインの9種類）数を比較したところ、質問回数が多くなるほど手段数が減っていた（初参加 1.65 < 質問 2~9 回 1.29 < 質問 10 回以上 1.14）。

質問投稿動機の3因子 [Miura et al. 07] について群別に検討した結果を表2に示す。外発的動機（正解がわからないと困る、わかるといいことがある）は初参加者でもっとも高く、質問回数が多いほど低くなっていったのに対して、内発的動機（ふと気になることを知りたい、日常の疑問の解明）は質問回数が多い利用者ほど高くなっていった。社会的動機（知恵袋を盛り上げたい、注目を集めたい）は、質問回数が10回以上の利用者で他群よりも高かった。

表 2: 質問投稿動機：群間比較

	外発的動機	内発的動機	社会的動機
初参加	3.22a	2.78a	1.32a
質問 2-9	3.14b	2.85b	1.31a
質問 10-	3.00c	3.02c	1.43b

注：符号は5%水準の有意差を示す

また、質問投稿者は自らの疑問を解決する手がかりを得る場としてなぜ知恵袋というKSCを選択したのだろうか。群別に検討したところ、初参加群は他群よりも「信頼できる回答をもらえそう」であることを理由として挙げる傾向が強いことがわかった。また「たくさん回答がもらえそう」「すぐ回答がもらえそう」は、質問回数9回以下の利用者で、10回以上の利用者よりも理由として挙げられる傾向が強かった。質問経験の少ない利用者ほど、即時性・正確性をより強く要望していることが示された。「雰囲気の良いコミュニティだから」については、群間差は見られず、全群において他の理由よりもやや低い傾向が見られた。

6.2 回答投稿者

回答を投稿した質問の正解の有無についての群間比較をおこなったところ、有意な差は見られなかった。どの群においても、特に決まった正解のない質問に対する回答が、正解ありの場合をわずかに上回る程度の割合となっていた。

また、正解のある質問に対する回答を投稿したケースに限って、回答前に正解を知っていたかどうか、さらに正解既知の場合は真偽確認/未知の場合は調査をおこなったかどうかを群別に比較したところ、いずれの群でも「真偽を確認せず既知の情報を回答した」利用者が過半数を占め、回答回数群による有意な比率の偏りは見られなかった（表3参照）。

表 3: 回答時の状況と行動：群間比較 (%)

	正解既知		正解未知	
	真偽未確認	真偽確認	調べず回答	調べて回答
初参加	61.4	26.1	8.0	4.5
回答 2-19	59.6	26.9	9.3	4.2
回答 20-99	58.6	29.9	5.9	5.6
回答 100-	58.0	30.6	6.3	5.2

回答投稿動機の4因子 [Miura et al. 07] について群別に検討した結果を表4に示す。援助的動機（質問者の問題を解決

したい、知りたい人を助けるのは当然）についてはどの群でも高く、有意差は見られなかった。互酬的動機（困ったときに教えてもらえる、回答をもらったことへの返報）については、回答回数2-19件（質問経験あり）の利用者が他群よりも高かった。社会的動機（知恵袋を盛り上げたい、行為そのものが楽しい）と報酬的動機（自分の評価を高めたい、感謝されるとうれしい）は、回答回数が多くなるほど高くなる傾向が見られた。

表 4: 回答投稿動機：群間比較

	援助的動機	互酬的動機	社会的動機	報酬的動機
初参加	3.34a	2.13a	2.56a	2.25a
回答 2-19	3.27a	2.36b	2.64a	2.35a
回答 20-99	3.24a	2.28ab	2.67ab	2.46b
回答 100-	3.22a	2.25a	2.74b	2.48b

注：符号は5%水準の有意差を示す

7. 投稿内容の形態素解析

質問/回答投稿ログを対象として群別に形態素解析をおこなった。TinyTextMining^{*1}を用いて得られた解析結果から、名詞のみを抽出した。なお、ログ収集時にいくつかの欠測が生じたため、分析対象となったのは質問投稿ログ5124件、回答投稿ログ2082件である。

7.1 質問投稿

表5に、質問投稿ログにおける、各群の頻出名詞（3%をしきい値とした）をまとめたものを示す。

表 5: 質問投稿ログの形態素解析：群間比較

名詞	初参加		質問 2-9		質問 10-	
	名詞	%	名詞	%	名詞	%
お願い	17.9	17.9	お願い	16.8	お願い	11.9
出品	9.2	9.2	質問	9.8	質問	9.9
方法	9.1	9.1	場合	8.7	場合	6.5
場合	8.3	8.3	方法	8.6	自分	5.5
質問	8.0	8.0	メール	7.9	購入	5.2
落札	8.0	8.0	使用	6.1	最近	5.0
メール	6.8	6.8	会社	5.5	方法	4.9
購入	6.7	6.7	自分	5.5	パソコン	3.9
自分	6.3	6.3	購入	5.4	会社	3.9
会社	6.1	6.1	子供	4.9	サイト	3.7
パソコン	6.0	6.0	落札	4.7	メール	3.6
連絡	5.9	5.9	最近	4.6	普通	3.6
最近	5.5	5.5	出品	4.6	ご存知	3.3
使用	5.0	5.0	パソコン	4.5	問題	3.3
子供	4.8	4.8	時間	4.4	子供	3.2
ヶ月	4.7	4.7	現在	4.3	ヶ月	3.1
現在	4.5	4.5	ヶ月	3.9	現在	3.1
ご存知	4.4	4.4	サイト	3.7		
時間	4.4	4.4	仕事	3.7		
仕事	3.8	3.8	商品	3.6		
商品	3.8	3.8	感じ	3.3		
電話	3.7	3.7	問題	3.3		
病院	3.7	3.7	予定	3.2		
表示	3.3	3.3	連絡	3.2		
予定	3.2	3.2	ご存知	3.2		
サイト	3.0	3.0	携帯	3.1		
可能	3.0	3.0	表示	3.1		
先日	3.0	3.0				

初参加群や質問回数10回未満群において高頻出の名詞のうち、質問頻度が高い群で出現頻度が低くなっているのは「出品」「落札」「連絡」といった、オークションに関わると思われる単語である。調査実施当時、Yahoo!オークションサービスのトップページ右側「初めの一步」というコーナーから知恵

*1 <http://fieldmining.com/%7Ematumura/ttm/>

袋へのリンクが張られており*2, このリンクによって知恵袋を知り、流入したオークション利用者が多かったことが示唆される。その他の高頻出名詞には顕著な群間差はない。質問頻度が10回以上の群では頻出名詞の数が減少しており、質問内容が多岐にわたるものになっていることが推察される。

7.2 回答投稿

表5に、回答投稿ログにおける、各群の頻出名詞(4%を大きい値とした)をまとめたものを示す。

表6: 回答投稿ログの形態素解析: 群間比較

初参加		回答投稿 2-29		回答投稿 20-99		回答投稿 100-	
名詞	%	名詞	%	名詞	%	名詞	%
場合	13.9	自分	13.5	場合	15.8	場合	15.8
自分	13.0	場合	12.1	自分	12.4	自分	10.4
保険	8.3	必要	6.6	質問	6.9	必要	7.6
必要	7.4	時間	6.4	時間	6.7	時間	6.7
子供	7.4	会社	5.8	必要	5.9	質問	6.5
問題	7.0	可能	5.7	問題	5.7	会社	5.8
相手	7.0	一番	5.6	会社	4.9	日本	5.8
時間	6.5	ヶ月	5.4	保険	4.8	相手	5.2
好き	6.5	好き	5.3	参考	4.6	ヶ月	4.5
効果	6.0	問題	5.0	意味	4.6	関係	4.5
病院	5.1	友達	4.8	落札	4.6	可能	4.4
以外	4.7	以上	4.5	相手	4.4	子供	4.0
連絡	4.2	質問	4.4	メール	4.4	状態	4.0
風邪	4.2	参考	4.4	可能	4.1		
日本	4.2	状態	4.0	子供	4.1		
現在	4.2	気持ち	4.0	感じ	4.1		
感じ	4.2			個人	4.1		
メール	4.2						
ゲーム	4.2						

初参加群において高頻出の名詞のうち、質問頻度が高い群で出現頻度が低くなる傾向が顕著なのは「保険」「効果」「病院」といった単語である。事故や疾病への対応に関わる質問に対する回答投稿であることが推察される。こうしたジャンルは、既存データベースの利用可能性が低く、またプライバシーに抵触する事柄が含まれる場合が多い。すなわち、先述したQ&AタイプKSCのメリットである、互いのプライバシーを明かさないう状態、自らの手のみでは獲得し得ない多様な解を得られる点が生かされる可能性が高い。このことを回答投稿経験のない、あるいは浅い者側の立場から考えてみれば、自らの経験にもとづく既有知識を提供しやす、比較的「敷居の低い」ジャンルなのかもしれない。逆に、質問頻度が高い群で高頻出である一方で、初参加群で出現頻度が低い名詞は「質問」「会社」「可能」であった。

8. 考察

本稿では、質問投稿者と回答投稿者を、過去の投稿件数に応じてグルーピングした。質問投稿により知恵袋に初めて参加した調査協力者の多さ、回答投稿時に調査に協力した利用者の過去の投稿件数の平均値の高さと分布を見ると、知恵袋というQ&AタイプのKSCが、人数として莫大な質問投稿による新規参加者と、件数として多数の回答投稿をするベテラン利用者によって活性化されていることがわかる。

質問投稿者の群間比較については、利用者たちが、徐々に「正解のない」質問を、自らの直接的な利益に関わる外発的動機からではなく、知的好奇心を満たしたりコミュニティを活性化させることを意図して投稿するようになっている傾向が読み取れる。加えて、正解のある質問であっても、その正解を投稿前に自ら探索しようと試みる手段数も減少し、知恵袋の一次的

情報源としての信頼性が高まっていることを感じさせる。知恵袋を選択した理由の分析や形態素解析の結果から示唆されるように、新規参加者は、今自分が悩んでいたり困っていたりすることについて、より正確な情報をより早く得るための場として知恵袋を利用する傾向が強い。実際、こうした新規参加者のうち、1件の回答も得られなかった割合はわずか8.6%にすぎず、平均2.5件の回答が寄せられている。少なくとも「より早く」情報を得る場として、知恵袋は十分に機能しているようである。

回答投稿者の群間比較については、投稿件数が投稿行動そのものに及ぼす影響はさほど大きくないようである。多くの利用者が、質問者の要望に応じて、自分自身が既にもっている知識や経験を、自らの資源を必要以上に割くことなく投稿している。回答投稿動機の分析結果からは、このデータのみからは因果関係を特定することはできないものの、KSCに投稿するという行為がもつ単純な楽しさや、情報提供という行為がもたらす自己効力感の向上が、回答投稿という形での参加動機を強化している構図が推察されよう。形態素解析の結果、全群で高頻出だった名詞に「自分」があることは、そこで提供されている情報が、回答者自身の経験知である場合が多いことを端的に示している。

こうして量的・質的両面から検討した限り、調査時点での知恵袋というQ&AタイプのKSCは、かなりよいバランスを保ったオンライン・コミュニティであったといえることができるだろう。ネットレイティングス社によって2007年初頭に実施されたインターネット利用動向情報サービス調査は、Q&AタイプのKSCがこの1年で利用者を急増させ、もっとも成長したジャンルカテゴリーのひとつであると報告している*3。知恵袋の利用者は156万人(2006.1) 477万人(2007.1)と3倍以上の増加をみており、これは本調査の実施直後からの急増である。また、知恵袋サービスを展開しているYahoo! JAPANは、従来からインターネット初心者の参加が多く、またサイトでは多種多様なサービス間の連携が積極的に推進されている。本稿の分析結果にもとづいて考えれば、新規参加者が増えることは即時的に正解を求める投稿が増すことを意味しており、このことは活発に回答というサポートをおこなう利用者たちの負担を増やすことになる。同一サイト内からの流入が増えるとするならば、FAQに属するような質問の投稿も多くなることが予想される。今後は、本稿データと比較可能な新たなデータを収集することによって、利用者層の変動によるコミュニティの変容過程についても検討したい。

参考文献

[川浦ら 06] 川浦康至・三浦麻子・地福節子・大瀧直子・岡本真 知識共有コミュニティを創り出す人たち(2):「質問タイプ」から見た参加行動 日本社会心理学会47回大会論文集, 496-497 (2006)

[Miura et al. 07] Miura, A., Kawaura, Y., Jifuku, S., Otaki, N., & Okamoto, M.: People who create knowledge sharing communities. In T. Washio et al. (Eds.) *New Frontiers in Artificial Intelligence (Lecture Notes in Computer Science)*. Springer. Pp.88-98 (2007)

*2 現在はページ下部「Yahoo!サービス」から「オークションについて質問する」際のサービスとして知恵袋がリンクされている。

*3 この調査における「利用者」は、調査期間中に1回以上当該サイトを訪問した個人の推計値で、本稿の「利用者」定義とは異なる。
http://csp.netratings.co.jp/nmr/PDF/Newsrelease0222007_J.pdf