

Q&A サイトに質問と回答が投稿された時刻と間隔についての調査

梅本 顕嗣 石川 尚季 西村 涼 渡辺 靖彦 岡田 至弘

龍谷大学大学院 理工学研究科 情報メディア学専攻

{t11m074,t10m096}@mail.ryukoku.ac.jp, r_nishimura@afc.ryukoku.ac.jp

{watanabe,okada}@rins.ryukoku.ac.jp

1 はじめに

質問を投稿しておく他のユーザが答えてくれる、Yahoo! 知恵袋、教えて! goo、人力検索はてな、OKWave などのコミュニティベース質問応答サービス (以下 Q&A サイト) が、さかんに利用されている。こうした Q&A サイトを利用するユーザの行動を調査することは、豊かなネットワーク社会の基盤を形成するために重要である。そこで本研究では、Yahoo!知恵袋を対象に、ユーザが質問や回答を投稿する時刻とその間隔について調査する。

2 Yahoo!知恵袋の公開データにおける時間の情報

本研究では、Q&A サイトの例として Yahoo!知恵袋^{*1}を取り上げる。Yahoo!知恵袋は、登録されたユーザが質問と回答のメッセージをやりとりすることができる Q&A サイトである。この Yahoo!知恵袋に 2004 年 4 月から 2005 年 10 月までに投稿された質問約 311 万件、回答約 1347 万件が国立情報学研究所から公開されている^{*2}。この Yahoo!知恵袋のデータの「パソコン、周辺機器」「政治、社会問題」カテゴリおよび全カテゴリ (286 カテゴリ) に質問あるいは回答を投稿したユーザの異なりと投稿件数を表 1 に示す。回答が投稿されなかった質問は削除されているので、この公開データに収録されている質問には 1 つ以上の回答が必ずある。また、それぞれの質問に対して投稿された回答の中から、最もよいと判定された回答 1 件がベストアンサーに選ばれる。それぞれの質問について必ず 1 件の回答がベストアンサーに選ばれるので、ベストアンサーの数は質問の数と等しい。以下に示す時間に関する情報も公開データには記録されている。

質問投稿日 質問が最初に投稿された時刻、あるいは最後に更新された時刻

回答投稿日 回答が投稿された時刻

質問最終更新日 回答の受付が締め切られ、ベストアンサーが決定された時刻

「質問投稿日」「回答投稿日」「質問最終更新日」はいずれも日付だけでなく、時刻も秒単位で記録されている。図 1 に質問が投稿されてからの一連のできごととその発生時刻の例を示す。「質問投稿日」には、質問が更新された場合は質

表 1 Yahoo!知恵袋の公開データにおける「パソコン、周辺機器」「政治、社会問題」およびすべてのカテゴリ (286 カテゴリ) における投稿者の異なりと投稿件数

	すべてのカテゴリ	政治、社会問題	パソコン、周辺機器
質問者の異なり	165064	13259	43493
質問の投稿件数	3116009	78777	171848
回答者の異なり	183242	25766	27420
回答の投稿件数	13477785	403306	474687

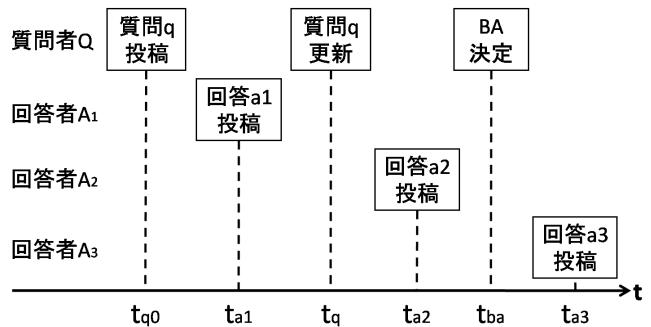


図 1 質問が投稿されてからの一連のできごととその発生時刻の例。質問者 Q が時刻 t_{q0} に質問 q を投稿し、続いて回答者 A_1 が時刻 t_{a1} に回答 a_1 を投稿し、その後質問者 Q が時刻 t_q に質問 q を更新し、それから回答者 A_2 が時刻 t_{a2} に回答 a_2 を投稿した。その後、質問者 Q が時刻 t_{ba} に回答の受付を終了してベストアンサー (BA) を決定し、さらにその後、回答者 A_3 が時刻 t_{a3} に回答 a_3 を投稿した。

問が最後に更新された時刻 (図 1 では時刻 t_q) が記録されている。一方、質問が更新されなかった場合は質問が最初に投稿された時刻 (図 1 では時刻 t_{q0}) が記録されている。ただし、質問が更新されたかどうかは公開データでは明示されていない。このため、「質問投稿日」に記録されている時刻が、質問が最初に投稿された時刻 t_{q0} なのか、質問が最後に更新された時刻 t_q なのか区別することはできない。質問が最後に更新された時刻よりも前に回答が投稿された場合のみ、「質問投稿日」に記録されている時刻が質問が最後に更新された時刻 t_q であることがわかる。図 1 では、質問が最後に更新された時刻 t_q よりも前の時刻 t_{a1} に回答者 A_1 が回答 a_1 を投稿している。この例のように「回答投稿日」の時刻が「質問投稿日」の時刻よりも前である

^{*1} <http://chiebukuro.yahoo.co.jp>

^{*2} <http://research.nii.ac.jp/tdc/chiebukuro.html>

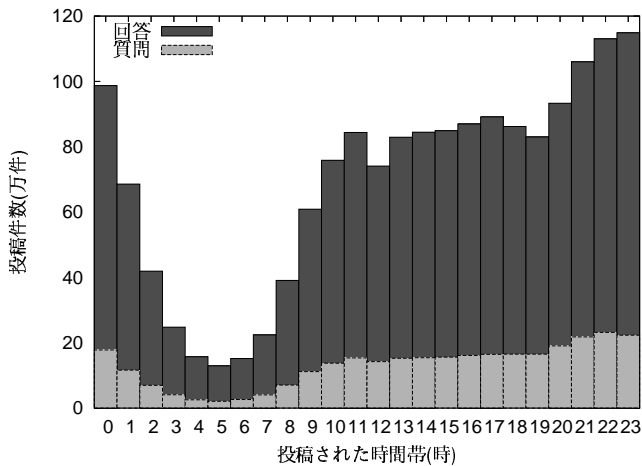


図2 Yahoo!知恵袋に投稿された質問と回答の時刻別の投稿件数 (2004年4月~2005年10月)

場合が、476545件の質問における1706325件の回答で見られた。これに対して、「回答投稿日」の時刻が「質問最終更新日」の時刻よりも後の場合が4095件の質問における4726件の回答で見られた。これは、質問者が回答の受付を終了し、ベストアンサーを決定した時刻よりも後に回答を投稿する回答者がいた場合である。図1では、回答の受付が終了となりベストアンサーが決定された時刻 t_{ba} よりも後の時刻 t_{a3} に回答者 A_3 が回答 a_3 を投稿している。「質問最終更新日」の時刻よりも後に投稿された回答でベストアンサーに選ばれたものは0件である。

3 質問と回答を投稿する時刻の調査

Yahoo!知恵袋に2004年4月から2005年10月までに投稿された質問約311万件と回答約1347万件を対象に、その投稿時刻を調査した。その結果を図2に示す。図2から、深夜から明け方、特に4時~6時は質問件数も回答件数も著しく減少することがわかる。また、12時に回答件数はかなり減少するが、質問件数はあまり減少しない。日本在住の15歳以上の人の約5割が12時から12時30分の間に食事をしている[1]ので、食事の時間には回答の投稿を控えるユーザがいること、一方、問題を抱えているユーザは食事の時間でも質問の投稿を行うことが考えられる。同様に、統計局の調査で15歳以上の約3割が食事をしている19時台でも回答件数が減少している。

4 質問と回答を投稿する間隔の調査

Yahoo!知恵袋の「政治、社会問題」「パソコン、周辺機器」そしてすべてのカテゴリを対象に

1. 質問が投稿されてから回答が投稿されるまでの時間 (図1では $t_{a1} - t_{q0}$ など) (調査結果は図3)
2. 質問が投稿されてからベストアンサーが投稿されるまでの時間 (回答 a_2 がベストアンサーならば、図1では $t_{a2} - t_{q0}$) (調査結果は図4)

3. 質問が投稿されてから最初の回答が投稿されるまでの時間 (図1では $t_{a1} - t_{q0}$) (調査結果は図5)
4. 2件以上の回答が投稿されている質問について、それらの回答が投稿される時間の差 (図1では $t_{a2} - t_{a1}$ など) (調査結果は図6)
5. 質問が投稿されてからベストアンサーが決定されるまでの時間 (回答 a_2 がベストアンサーならば、図1では $t_{ba} - t_{q0}$) (調査結果は図7)
6. ベストアンサーが投稿されてからベストアンサーとして認定されるまでの時間 (回答 a_2 がベストアンサーならば、図1では $t_{ba} - t_{a2}$) (調査結果は図8)

について調査を行った。

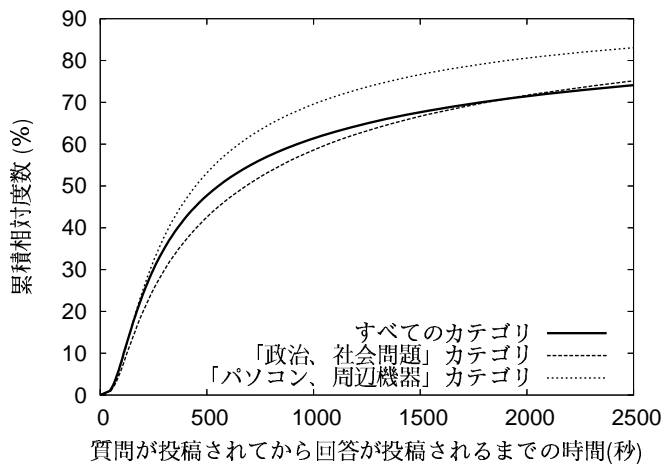
図3、図4、図5、および図7の調査は、どの回答の「回答投稿日」の時刻よりも「質問投稿日」の時刻が前である質問2639464件とその回答10920340件を対象に行った。「政治、社会問題」カテゴリでは質問58025件とその回答267556件、「パソコン、周辺機器」カテゴリでは質問153958件とその回答419431件を対象に調査を行った。「政治、社会問題」カテゴリと「パソコン、周辺機器」カテゴリの調査を行ったのは、投稿される質問や回答に違いがあるからである。「パソコン、周辺機器」カテゴリでは問題解決を目的とする質問が多く投稿される。一方、「政治、社会問題」カテゴリでは、問題解決だけでなく、意見収集や議論を目的とする質問が多く投稿されていた。また、回答間での議論は「パソコン、周辺機器」カテゴリではあまり行われぬが、「政治、社会問題」カテゴリでは長い時間をかけて行われることがある。このため、図3の質問が投稿されてから回答が投稿されるまでの時間でも、図4の質問が投稿されてからベストアンサーが投稿されるまでの時間でも、「政治、社会問題」カテゴリの方が「パソコン、周辺機器」カテゴリの場合よりも時間は長かった。しかし、図5の質問が投稿されてから最初の回答が投稿されるまでの時間では、「政治、社会問題」カテゴリの方が「パソコン、周辺機器」カテゴリの場合よりも時間は短い。この原因として考えられるのは、投稿されたばかりの質問にすぐ回答を投稿することを目的とするユーザである。

2件以上の回答が投稿されている質問は2576718件あり、その回答12938494件を対象に図6の調査を行った。「政治、社会問題」カテゴリでは回答395415件、「パソコン、周辺機器」カテゴリでは回答427049件を対象に図6の調査を行った。図8の調査は、ベストアンサーに認定された回答3116009件すべてを対象に行った。ベストアンサーに認定された回答は「政治、社会問題」カテゴリでは78777件、「パソコン、周辺機器」カテゴリでは171848件あり、それぞれを対象に調査を行った。

図7から、質問の約3割が投稿されてから3時間以内に解決していることがわかる。また、質問を投稿してから1日(86400秒)経過したところに質問を打ち切り、ベストアンサーを決定するユーザが多いこともわかる。

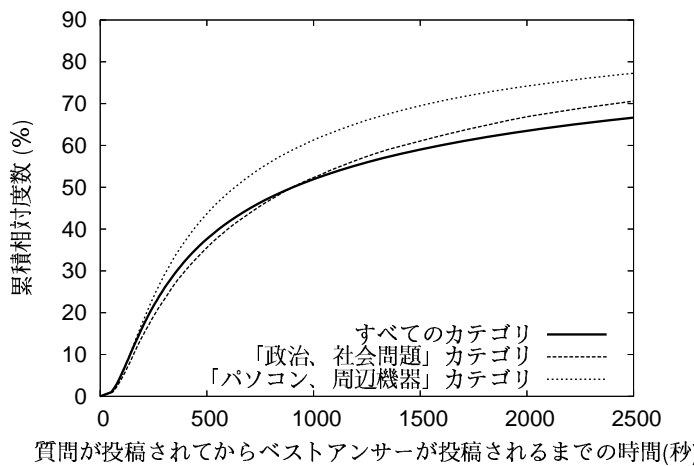
参考文献

- [1] 総務省統計局: 社会生活基本調査 平成13年, (2003).



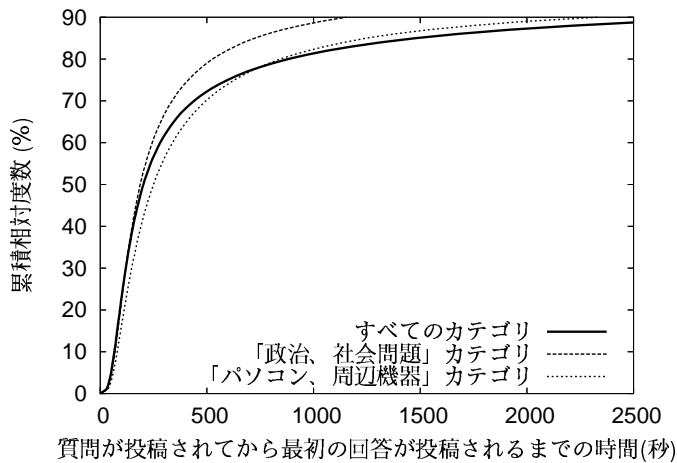
累積相対度数	すべてのカテゴリ	政治、社会問題	パソコン、周辺機器
10	110	125	112
20	171	201	168
30	249	302	234
40	363	451	321
50	554	683	447
60	923	1067	650
70	1784	1810	1021
80	4386	3562	1905
90	15488	10153	5873
100	2921717	2393112	2210490

図3 質問が投稿されてから回答が投稿されるまでの時間(秒)



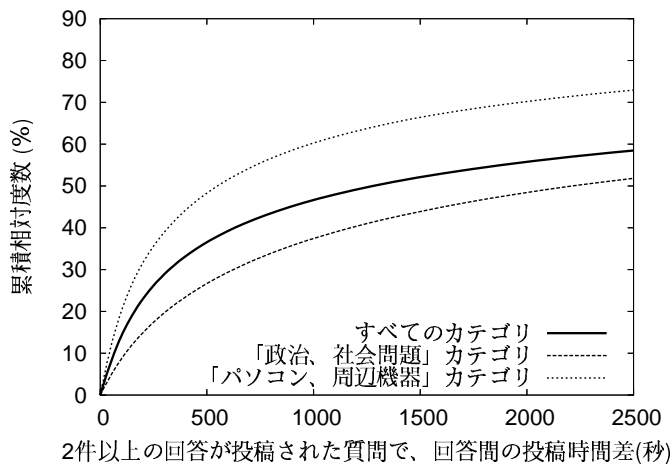
累積相対度数	すべてのカテゴリ	政治、社会問題	パソコン、周辺機器
10	140	156	139
20	232	260	217
30	359	398	311
40	556	598	439
50	903	902	631
60	1594	1421	946
70	3227	2413	1542
80	7787	4821	3151
90	26542	14598	9991
100	2407916	2384831	2210490

図4 質問が投稿されてからベストアンサーが投稿されるまでの時間(秒)



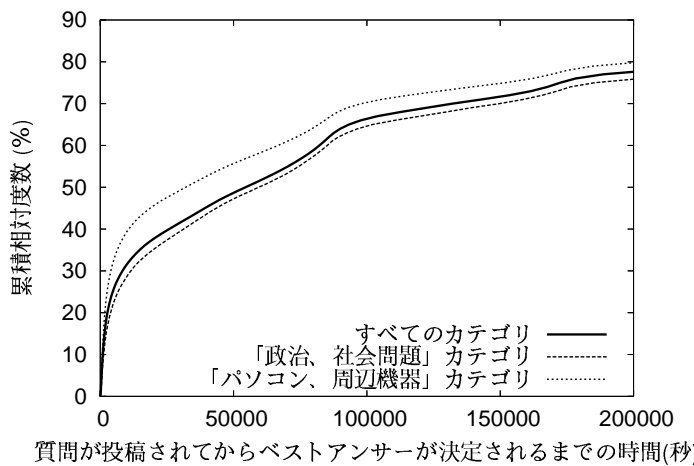
累積相対度数	すべてのカテゴリ	政治、社会問題	パソコン、周辺機器
10	66	69	80
20	92	94	112
30	120	119	147
40	154	150	191
50	202	187	250
60	281	244	339
70	440	337	495
80	881	526	843
90	3085	1154	2338
100	2401084	2075985	2210490

図5 質問が投稿されてから最初の回答が投稿されるまでの時間(秒)



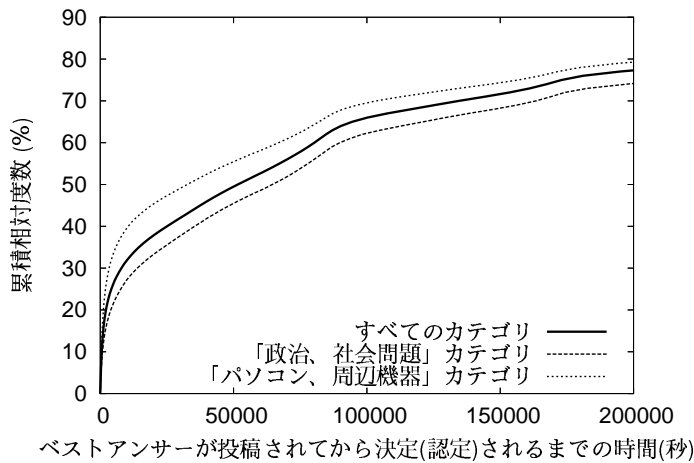
累積 相対度数	すべての カテゴリ	政治、 社会問題	パソコン、 周辺機器
10	65	118	43
20	161	308	99
30	323	624	182
40	630	1174	318
50	1278	2211	549
60	2832	4424	983
70	6804	9925	1968
80	18509	27041	4902
90	53885	77565	17947
100	3166955	3166955	2531847

図 6 2 件以上の回答が投稿された質問で、回答間の投稿時間差 (秒)



累積 相対度数	すべての カテゴリ	政治、 社会問題	パソコン、 周辺機器
10	948	1270	780
20	2694	3890	1730
30	8342	11495	3979
40	25841	31273	10687
50	54489	59400	32046
60	82179	85164	66710
70	132147	149711	97882
80	244547	260467	204075
90	513413	600089	484023
100	3711642	2509193	2502099

図 7 質問が投稿されてからベストアンサーが決定されるまでの時間 (秒)



累積 相対度数	すべての カテゴリ	政治、 社会問題	パソコン、 周辺機器
10	554	848	411
20	2149	3971	1179
30	7936	13914	3342
40	25049	35375	10505
50	51593	64505	32610
60	80275	89851	66908
70	134539	163454	103501
80	244975	284711	215799
90	510712	605736	493338
100	3801020	3680915	3374757

図 8 ベストアンサーが投稿されてから決定 (認定) されるまでの時間 (秒)