

機械学習を用いたモバイル Web 通信品質の
測定評価に関する研究

著者

小林 史弥

電気通信大学大学院 情報システム学研究科

博士（工学）学位申請論文

2022年3月

機械学習を用いたモバイル Web 通信品質の
測定評価に関する研究

著者

小林 史弥

博士論文審査委員会

大坐 畠 智 准教授

大須賀 昭彦 教授

吉永 努 教授

小川 朋宏 准教授

山本 嶺 准教授

© Copyright 2022

by

Fumiya Kobayashi

All Rights Reserved

論文の英文要旨

T I T L E	A Study on Measurement and Evaluation of Mobile Web Communication Quality Using Machine Learning
N A M E	Fumiya Kobayashi

Smartphone users experience a decrease in QoE (Quality of Experience) due to latency during web browsing.

Therefore, we proposed a method for detecting QoE degradation due to web transfer latency delay based on the results of throughput/RTT (Round Trip Time) measurements at various locations and times. Throughput was measured by exchanging a large number of packets with a small volume, and RTT was measured by pinging a large number of packets to capture the instantaneous degradation of communication quality, and QoE was evaluated by whether the transfer latency was significantly delayed compared to normal at multiple Web sites. By applying machine learning to the throughput/RTT and QoE data collected by this measurement method, we proposed a method for creating, managing, and operating a decision maker that can determine QoE based on throughput/RTT measurement results.

機械学習を用いたモバイル Web 通信品質の

測定評価に関する研究

小林 史弥

概要

スマートフォンの普及に伴って、ユーザはあらゆる場所・時刻で大容量通信を気軽に利用するようになってきている。つまり、トラフィック量や接続数が集中する場所が、ユーザとともに移動し続けている。これにより、アクセスネットワークの一時的な輻輳を原因とする、QoE (Quality of Experience: ユーザ体感品質) の著しい低下が発生している。しかしながら、実サービス上でユーザがどのようなサービス (Web ブラウジング, 映像配信サービス等) を利用しているときに、どのような QoE 低下事象 (Web 転送待ち時間の増大, 再生の中断等) が発生しているのか、ユーザ動態の調査は必ずしも十分ではない。また、ユーザ動態がつかめたとしても、その不満を解消するためには QoE を低下させるほどのアクセスネットワークの輻輳が発生している場所・時刻を検出する手法が必要となる。

本論文では、ユーザ動態調査により QoE 低下の支配的サービスとその事象を明らかにした上、そのサービスの QoE がアクセスネットワーク要因で低下している場所・時刻を検出するための手法を提案する。ユーザ動態を調査するため、モニターユーザのスマートフォンに調査専用のアプリケーションをインストールし、QoE が低下したときの利用サービスや状況などのデータを収集した。これにより、Web ブラウジングにおける転送待ち時間の著しい遅延が、最頻出の QoE 低下事象であることを明らかにした。そして、その QoE 低下事象を検出するためのスループット/RTT (Round Trip Time: 往復遅延時間) 計測手法を考案した。従来のスループ

ット/RTT 計測手法が、ユーザの協力を得てさまざまな場所・時刻のスループット/RTT の計測に成功していることから、このような計測データから QoE 低下の有無を判定できれば有用であることに着目した。しかし、従来のスループット/RTT 計測は定常状態における平均値を評価することを目的としている。これでは瞬間的な品質劣化を捉えることができない。そこで、瞬間的な品質劣化を捉えられるように、小さなデータ転送を繰り返してスループットを計測する、間隔を空けて多数回の RTT を計測するという計測手法の提案を行った。また、Web 転送待ち時間は平均値ではなく、複数サイトで平常時より Web 転送待ち時間が著しく遅延している場合に QoE 低下ありとして扱った。Web 転送待ち時間もスループット/RTT と同じく通信セッションごとの分散が大きいため、複数回の計測結果から QoE 低下の有無を分類した。また、単一サイトの反復計測ではなく、複数サイトを計測対象とすることでサイト依存性を排除した。そして、スループット/RTT と QoE の関係を機械学習の一種である SVM (Support Vector Machine) を適用し学習した。これによって、RTT のみで QoE が低下した場所・時刻を高精度に検出できる判定器を作成できることを明らかにした。さらに、この提案手法を実サービスで運用するために、計測の負荷に直結するデータ量と判定精度の関係を分析した。結果として、RTT を 8 回と Web 転送待ち時間を 9 サイト計測したデータから十分な精度を有する判定器を作成することができた。

目次

概要	iii
図の一覧	v
表の一覧	vi
略語の一覧	vii
1 序論	1
1.1 背景	1
1.2 目的	2
1.3 貢献	4
1.4 論文の構成	5
2 関連技術	6
2.1 QoS	6
2.2 QoE	8
2.2.1 QoEの主観品質評価法	9
2.2.2 QoEの客観品質評価法	13
2.3 SVM	15
2.3.1 混同行列	15
2.3.2 評価メトリック	16
2.3.3 評価方法	18
2.4 QoE推定に関する先行研究	19
2.5 機械学習を用いたQoE推定手法	20
2.6 先行検討の課題	22
3 モバイル通信サービスの動態調査	24
3.1 概観	24
3.2 提案手法	24
3.3 結果	27
3.4 議論	29
3.5 結論	30

4	機械学習による判定器作成手法	32
4.1	概観.....	32
4.2	課題.....	33
4.3	提案手法.....	34
4.4	実施手順.....	38
4.4.1	Web 転送待ち時間計測の手順.....	41
4.4.2	スループット/RTT 計測の手順.....	42
4.4.3	Web 転送待ち時間遅延の判定手法.....	43
4.4.4	スループット/RTT の定量化手法.....	45
4.4.5	判定器の作成	45
4.5	評価実験.....	47
4.6	議論.....	49
4.7	結論.....	51
5	管理運用のための設計	53
5.1	特徴量.....	53
5.2	目的変数.....	57
5.3	議論.....	59
5.4	結論.....	62
6	結論	63
6.1	まとめ.....	63
6.2	貢献.....	63
6.3	今後の展望.....	64
	謝辞	66
	参考文献	67
	関連論文（学位論文を構成する論文）	76
	著者略歴.....	77

図の一覧

図 1 QoS と QoE の概念図	6
図 2 職業及び時間帯別の QoE 低下率	29
図 3 同一セット内の 2 つの Web サイトの転送待ち時間実測例	36
図 4 提案手法の全体像 (学習フェーズ, 判定フェーズ)	37
図 5 計測の手順	39
図 6 Web 転送待ち時間の累積分布関数の実測例	44
図 7 1 セット内に含まれる Ping の往復回数と判定精度の関係 ($21\text{pps}\times s$ 回)	54
図 8 1 セット内に含まれる Ping の往復回数と判定精度の関係 ($t\text{ pps}\times 4$ 回)	55
図 9 訓練用 Web サイト数の判定精度への影響	58

表の一覧

表 1 QoE 構成要素に関する主な勧告	8
表 2 代表的な主観品質評価法	10
表 3 ACR 法による評点と評定語の例	11
表 4 代表的な客観品質評価法	14
表 5 混同行列	16
表 6 先行検討と提案手法の対比	22
表 7 時間帯別の通信品質 (1 時間の平均値).....	29
表 8 実験条件	39
表 9 計測項目の設定	40
表 10 入力データの例	46
表 11 各判定器のパラメータと判定精度	48
表 12 (i)の混同行列	48
表 13 (ii)の混同行列	48
表 14 (iii)の混同行列	48
表 15 2 pps 以下における計測順と判定精度の関係.....	56
表 16 通信量の最適化効果シミュレーションの結果	59

略語の一覧

RTT: Round Trip Time

QoS: Quality of Service

QoE: Quality of Experience

MOS: Mean Opinion Score

SVM: Support Vector Machine

FP: False Positive

FN: False Negative

TP: True Positive

TN: True Negative

AUC: Area under the curve

ROC: Receiver Operating Characteristic

CDF: Cumulative Distribution Function

PLT: Page Load Time

SI: Speed Index

VoD: Video on Demand

1 序論

1.1 背景

スマートフォンの普及によるモバイル通信量の増大により [1], モバイル通信サービスの品質の重要性が高まっている。モバイル通信は固定通信と比較して通信品質が劣化する場所・時刻の変動が激しい。これは、モバイル通信ではユーザ端末が移動するために、ユーザが集中した場所・時刻において接続要求やトラヒックも集中し、通信品質が一時的に劣化するためである。このような事象が発生した場所・時刻を適切に把握できれば、通信事業者は通信品質劣化の原因を調査し、設備増強などの対策を検討できる。

これまで、通信品質を調査するためにドライブテストや定点観測にてスループットや RTT (Round Trip Time: 往復遅延時間) が計測されてきた [2]。ドライブテストとは、主に通信事業者が計測用のスマートフォンを車載し、各所でスループット/RTT などの通信品質を計測することを指す。任意の場所・時刻で、ユーザ端末で実測されるデータを収集できるという利点がある。定点観測は、計測機器を一定の場所に常設してデータを計測させ続けることを指す。その場所におけるデータを長期間にわたって計測できるという利点がある。

しかし、このような調査法だけでは通信品質を捉えるには不十分である。ドライブテストは多くの人的リソースが必要であり、定点観測は計測機器を常設するための場所を確保するなどのコストが必要である。そのため計測できる場所・時刻が限られ、全国さまざまな場所で劣化と平常化を細かく繰り返すモバイル通信品質の時空間変動を捉えるには十分ではない。

そこで、簡易にスループット/RTT を計測するアプリケーションを配信し、不特定多数のユーザによる計測結果を利用することで多数の場所・時刻での調査を可

能にする取り組みがある [3]-[6]. このようなアプリケーションは、通信事業者のみならずユーザにとってもメリットがあるため広く利用されている. たとえば, ユーザが契約中の通信サービスの品質を計測し, 他ユーザの計測結果と比較することでより QoS (Quality of Service) が良好な通信サービスへの乗り換えを検討するなどの使い方が挙げられる.

しかし, 従来のスループット/RTT 計測アプリケーションはスループット/RTT の実測を目的としているため, 通信の定常状態を計測している. 大容量の通信を行ったときの転送量を転送時間で除算してスループットを計算しており, RTT はその通信セッションで計測しているため, 瞬間的な品質劣化を捉えることはできない. Web ブラウジングなどはスループット/RTT の瞬間的な劣化によって引き起こされる転送待ち時間の遅延によって QoE が低下する [7]-[8]. そのため, 従来のスループット/RTT の計測手法では, 計測結果から QoE (Quality of Experience: ユーザ体感品質) の低下有無を把握することはできない.

1.2 目的

本論文の目的は, スループット/RTT の瞬間的な劣化を捉えることで, その計測結果から QoE が低下している場所・時刻を推定するための品質調査法を明らかにすることである. まず, ユーザ動態を調査し, QoE を低下させることが多いサービスを特定する. そして, そのサービスについて, 簡易に計測可能なスループット/RTT の計測の結果から, 各場所・時刻の QoE 低下の有無を判定する手法を提案する. そのために, 定常状態ではなく瞬間的な劣化を捉えられるスループット/RTT 計測手法を提案する. 最後に, 提案手法の管理運用手法を提案する.

まず, ユーザ動態の調査法について述べる. QoE を構成する要素はサービスごとに異なり, それに伴って必要とされる QoS の程度も異なる [9]-[21]. そこで, モバイル通信サービスのユーザが QoS に対して不満を持ちやすいサービスを理解す

ることが必要となる。しかし、あらゆるサービスに対して場所・時刻を問わず十分なスループット/RTTを提供することは現実的ではない。そこで QoE 低下の主要因となっているサービスと事象を明らかにする。たとえば、大容量転送を必要とする映像配信サービスにおける再生の中断や、小容量だが高速転送を必要とする Web ブラウジングにおける転送待ち時間の遅延などが想定される。これについて実ユーザを対象とした調査で明らかにする。

次に、スループット/RTT の計測結果から、各場所・時刻の QoE 低下の有無を判定する手法について述べる。スループット/RTT は通信セッションごとに瞬間的に大きく変動するため、従来の大容量転送時のスループット/RTT を計測する手法で収集したデータからは QoE を推定できない [7]-[8], [33]。VoD (Video on Demand) などのサービスでは大容量転送時のスループットが QoE と強い相関がある可能性はあるが、Web ブラウジングなどのサービスでは小容量転送時の瞬間的な品質劣化が QoE と強く関係することが考えられるためである。そこで、スループット/RTT は瞬間的な品質劣化を捉えるため、小容量の転送を繰り返して計測する手法を提案する。この提案手法で計測したスループット/RTT と QoE の関係を学習し、スループット/RTT の計測結果のみから QoE が低下している場所・時刻を検出できることを明らかにする。

最後に、提案手法の管理運用手法について述べる。スループット/RTT 計測のデータ量と判定精度のバランスを調査した結果を示し、管理運用に資する知見を提示する。品質調査における計測作業の負荷は小さいほど管理運用上は有利である。一方で、データ量が減るほど判定は困難になる。両者のバランスを示すことで、実運用時の設計を検討できるようにする。

1.3 貢献

本論文の貢献は以下の3点である。

第1に、ユーザ動態を調査する手法を示した。実験参加者のスマートフォンに、通信やアプリケーション利用のログを収集するアプリケーション（以下、エージェント）をインストールさせ、QoE低下を体感したときの状況を調査した。その結果、会社員が通勤時間帯や昼食時のWebブラウジングにおいてWeb転送待ち時間が著しく遅延したことによるQoE低下が多発していることを明らかにした。

第2に、専用アプリケーションを用いて計測したスループット/RTTから、その場所・時刻におけるQoE低下の有無を判定する手法を提案した。スループット/RTT計測時の通信セッションと、QoEが低下したときの通信セッションが異なるため、両者の関係は学習しづらい。これは、QoE低下事象が発生しているような混雑した場所・時刻においても、通信セッションごとにスループット/RTTの分散が大きいためである。さらに、定常状態の平均スループット/RTTでは、瞬間的な品質劣化の情報が失われ、QoE低下を検知できない [7]-[8], [33]。そこで、スループット/RTT計測手法を改善した。スループットについては、小さなファイル転送を繰り返して多数回計測し平均値を採用した。RTTについては、間隔を空けて多数回のPingを発出してRTTを計測し平均値を採用した。通信の過渡的な状態におけるスループットや、間隔を空けた複数回のRTTによって瞬間的な品質劣化を捉え、通信セッションに依らずその場所・時刻におけるQoE低下を検知できるようにした。また、目的変数の捉え方についても改善した。WebブラウジングにおけるQoEは、各サイトの転送待ち時間そのものではなく、複数サイトで平常時よりも転送待ち時間が遅延しているときに低下する [34]。そこで、複数のWebサイトの転送待ち時間をそれぞれ多数回計測し、その中央値を平常時の転送待ち時間とみなしてそこからの乖離が大きいサイトが複数確認された場所・時刻をQoE低下事象の発生ありとして目的変数とした。そして、瞬間的な劣化を捉えたスループ

ット/RTT の計測結果と、複数サイトで Web 転送待ち時間が著しい遅延しているかどうかの関係について SVM (Support Vector Machine) を適用することで [35]-[37], ひとたび両者の関係を学習すれば, 以後はスループット/RTT のみを計測し Web 転送待ち時間の著しい劣化事象の有無を判定する判定器が作成される手法とした. これにより, スループット/RTT 計測用アプリケーションの強みを生かして, さまざまな場所・時刻におけるデータの収集を可能にした.

第 3 に, 上記提案手法において, 学習や判定に使用するデータの計測量と判定精度のバランスを確認する手法を提案した. 特徴量となるスループット/RTT の計測量を変えて Web 転送待ち時間遅延の有無の判定精度に及ぼす影響を調査する. このとき, 特徴量のデータの間引き方を変えることで計測間隔を変えたことの判定精度への影響を調査できることを確認した. 次いで, 目的変数となる Web サイトの選定方法を検討するための手法を提案した. Web 転送待ち時間は通信セッションごとに大きく変動するため, その分散を捉えることが重要である. そこで, 各 Web サイトの転送待ち時間を多数回計測したときの分散や中央値の大小に基づいて, 学習データへの取り込み可否を判断することによって判定精度を高める方法を調査できることを確認した. 結果として, 特徴量は約 2.5 分にわたって 8 回分の RTT, 目的変数は中央値が大きい 9 サイトとすれば十分な判定精度を得られた.

1.4 論文の構成

本論文は以下のように構成される. 2 章で QoS と QoE, 機械学習に関連する技術を述べ, 関連研究のサーベイにより本研究の位置付けを明確にする. 3 章でモバイル通信のユーザ動態を調査する手法を説明する. ここで, Web 転送待ち時間の遅延が QoE 低下の主要因となっていることを明らかにし, 4 章で QoS の計測結果から Web 転送待ち時間遅延の有無を判定する手法の実施手順を説明する. 5 章で提案手法の管理運用に資する知見を述べる. 6 章で本論文をまとめる.

2 関連技術

QoS と QoE の関係を図 1 に示す。QoS は端末間に存在する通信の品質であり、QoE は端末やユーザーの主観まで含めた品質であると定義されている [11]-[20]。QoE は主観品質評価法を基礎として研究されてきた [38]-[40]。しかし、サービス提供中のように、QoE をユーザ自身に主観品質評価させることが適切でない場合などに、機械的に評価できる QoS の指標から QoE を推定できれば有用であることから、両者の関係に関する研究が数多く存在する [41]-[50]。以下、それぞれの評価法やその関係の学習方法について述べる。

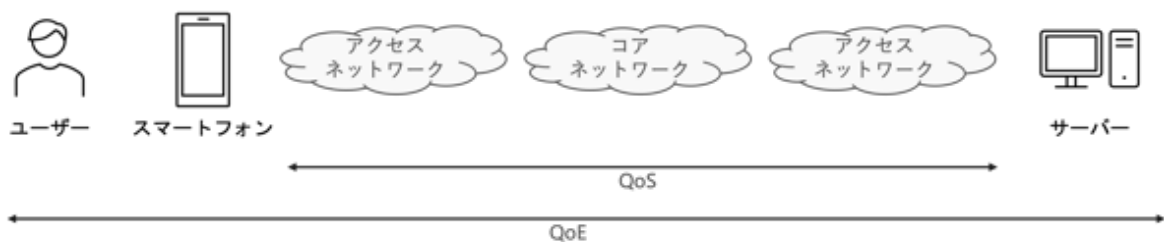


図 1 QoS と QoE の概念図

2.1 QoS

QoS の指標としては、End-to-End のスループット/RTT が用いられる [21]-[30], [33], [50]。計測方法としては、サービス利用中の通信セッションにおいてパケットをキャプチャし解析する手法と [21]-[30], [50]、サービス利用とは別の通信セッションにおいて一定サイズの計測用パケットを送受信する手法 [33] に大別できる。前者はその通信セッションにおけるスループット/RTT を扱えるため、QoE との関係分析がしやすい。さらに、パケット損失率やジッタなど多くの種類のデータを利活用でき精緻な分析が可能になる。しかし、収集コストが高い。後者は収集コストが低い、サービス利用中の通信セッションではないため、スループット

/RTT 以外の情報を利活用できず QoE との関係分析が容易ではない。それぞれに利点があるため、目的によって使い分けられる。評価メトリックとしては、大容量転送時の転送量を転送時間で除算してスループットとして扱い、RTT はその通信セッション確立時の値が用いられるなどが挙げられる。評価方法としては、場所・時刻ごとにスループット/RTT の平均値を比較するなど、ユーザが居住地におけるスループット/RTT の高い通信事業者を選定するなどに用いられている。

スループット/RTT 計測は、アプリケーションとして公開することでさまざまな場所・時刻の情報を収集できる [3]-[6], [51]-[55]。その 1 つである Speedtest.net は [6]、一般ユーザがスマートフォンのブラウザや専用アプリケーションを用いて計測専用のサーバとパケットを交換してスループット/RTT を計測できるサービスである。計測結果は匿名化されて集計され、場所・時刻ごとに品質のよい通信事業者などが公開されている。ユーザは自身が契約している通信事業者の QoS の実測値を知ることができるため、任意でも積極的に計測を行う。このように、スループット/RTT 計測アプリケーションという形で公開してデータを収集することによって、さまざまな場所・時刻の QoS を評価することが可能になる。また、この方式であればコアネットワークのデータにアクセスする必要がなく、通信の秘密に関わらない。しかし、スループット/RTT 計測サービスでは Web 転送待ち時間などのサービス利用時の QoE とのマッピングは行っていない。そのため、たとえ他の地域やキャリアと比較してスループット/RTT が劣っていたとしても、Web 利用には問題がないことも想定される。

従来のスループット/RTT の計測手法には、通信品質の瞬間的な劣化を捉えられない点に課題がある [7]-[8]。QoE 低下を引き起こすのは、定常状態におけるスループット/RTT ではなく、瞬間的で大きな劣化である。従来は大容量転送時のスループットと、そのセッションの 1 回分の RTT が計測値として用いられているため、瞬間的で大きな劣化を捉えられておらず、品質調査として不十分である可能性が指摘されている。

2.2 QoE

QoE とは、QoS の外側に存在するさまざまな要因の影響までも考慮したユーザの体感品質のことである。モバイル通信品質に対する QoE は、スループット/RTT などの純粋なネットワーク計測の結果に対して形成されるわけではない。モバイル通信の上で IP 電話や Web ブラウジングなどのサービスを利用し、通信の遅延、切断によってサービス利用が阻害されたときに、QoE の低下として知覚されるものである。

QoE の構成要素はサービスごとに大きく異なる [9]-[20]。そのため、サービスごとに勧告が存在する。表 1 に、ITU-T におけるサービスごとの QoE 構成要素に関する勧告を示す。

表 1 QoE 構成要素に関する主な勧告

サービス	ITU-T 勧告
IP 電話	G.107
Web ブラウジング	G.1031
映像配信	P.1202, P.1204
360 度映像	G.1035, P.360-VR
ゲーム	G.1032, P.809, G.1072

たとえば、IP 電話では声が発話者の口から電話機と通信を經由し受話者の耳へ到達するまでに存在する QoE 構成要素を示す次の式が規定されている [12]。QoS では扱われなかった利便性要因までも考慮していることが分かる。

品質評価値 (R 値)

$$= \text{雑音感} - \text{音量感} - \text{遅延} \cdot \text{エコー感} - \text{歪} \cdot \text{途切れ感} + \text{利便性要因}$$

また、Web ブラウジングでは QoE 構成要素として以下が規定されている [20].

- 1 ユーザ：不満を覚えやすい性質など. 一般に, 十分な数の評価者を用意することで評価への影響を低減する.
- 2 コンテキスト：場所, 緊急性等.
- 3 システム
 - 3.1 Web サーバ：CPU, メモリ等
 - 3.2 コンテンツ：要素の数, サイズ等
 - 3.3 クライアント：OS, ブラウザ, リソース読み込み手順等
 - 3.4 ネットワーク：レスポンスタイム等

この中で通信品質が担保できるのは 3.4 ネットワークのみである. そのため, 通信サービスの品質向上を目的とする場合には通信と関係がない QoE 構成要素についてはその影響を除去するか, 推定モデルの中で扱うことが求められる.

ネットワーク以外の QoE 要因への対応はサービスごとに異なる. 前述のとおり, IP 電話では利便性というコンテキストに近い要素がパラメータに取り入れられている [12]. また, 文献 [56]はゲームではコンテキストが QoE に強く影響するとして, 推定モデルに取り込んでいる. 一方, 文献 [57]は Web ブラウジングではコンテキストは QoE に影響を与えないとしている.

2.2.1 QoE の主観品質評価法

QoE がユーザの体感品質である以上, 主観品質評価が基本となる. 主観品質評価法は評価条件を揃えられる実験室実験を中心に発展し, その知見をフィールド調査に活用する取り組みが行われている. 以下, 順に説明する.

2.2.1.1 実験室実験

QoE の実験室実験における主観品質評価法は多数の勧告で定められている [38]-[40]. 代表的な主観品質評価法を表 2 に示す.

表 2 代表的な主観品質評価法

基準 尺度	絶対評価	相対評価
カテゴリー 尺度	(1) ACR 法 (Absolute Category Rating) (2) ACR-HR 法 (Absolute category rating with hidden reference)	(4) DCR 法 (Degradation Category Rating) (5) PC 法 (Pair Comparison)
連続尺度	(3) SSCQE 法 (Single Stimulus Continuous Quality Evaluation)	(6) DSCQS 法 (Double Stimulus Continuous Quality Scale) (7) SAMVIQ 法 (Subjective Assessment Methodology for Video Quality)

大きく、絶対評価か相対評価かという軸と、カテゴリー尺度か連続尺度かという軸で分類される。絶対評価は、動画などの評価用の刺激 1 つ 1 つに対してその体感品質を評価者に評価させる。相対評価は、2 つ以上の評価用の刺激を比較していずれが良いか、どの程度良いかなどを評価者に評価させる。一般に、相対評価のほうが評価コストは高まるが微細な劣化の影響を明らかにしやすいとされる。カテゴリー尺度は 5 段階から 11 段階程度で品質の良し悪しを評価させる。連続尺度はスライダーなどを使用して、たとえば 1~100 などの広い範囲を 1 刻みで評価させる。一般に、連続尺度のほうが評価コストは高まるが、より感覚的に評価がで

きる点が利点となる。このように、品質評価の目的によってさまざまな主観品質評価法が使い分けられている。

主観品質評価法の中でもとりわけ、ITU-T Recommendation P.910 [38] で規定される ACR (Absolute Category Rating) 法は単一刺激法 (Single Stimulus Method) とも呼ばれ、もっとも広く用いられる主観品質評価法の1つであるため、詳解する。評価者に、動画などの評価用の刺激を与え、その体感品質を表 3 に示すような尺度で評価させる。複数の評価者による評点の分散から MOS (Mean Opinion Score: 平均オピニオン評点) を計算する。MOS の活用方法は実験の目的に応じて調整されるが、たとえば MOS ≥ 3.5 を満足させるようにサービス品質を設計するなどが挙げられる。動画であればビットレートやフレームレートを調整することで MOS ≥ 3.5 を実現することになる。MOS は、評価者の属性に影響を受ける。わずかな品質劣化に敏感な評価者が参加していれば評点は下がる。人種や性別、年齢などによっても評点が厳しいまたは甘いなどの傾向があることも確認されている。これらの対策の1つとして、評価者の数を十分にすることで、評価者の属性の影響を最小化することが必要であると規定されている。最低でも 15 人の評価者が必要とされている。

表 3 ACR 法による評点と評定語の例

評点	評定語
5	非常に良い
4	良い
3	普通
2	悪い
1	非常に悪い

勧告を活用し、具体的なサービス利用における QoE の特性を分析した研究も多い [58]-[59]。たとえば、Web ブラウジングにおいては Web 転送待ち時間が 2~6

秒の間に MOS が落ち込んでいき [58], 10 秒に達すると MOS が 2.5 を下回り, すなわち許容できない程度まで落ち込むことが示されている [59].

2.2.1.2 フィールド調査

QoE はユーザによって決定される以上, ユーザが置かれた状況によって大きく変化する. 実験室実験では, 状況を統制することによってわずかな劣化に対しても厳しく評価することを可能にしてきた. 一方で, 外出中にも利用されるモバイル通信サービスにおける実験室とフィールドでの QoE 特性の違いに関する研究が進められている [32], [34], [60].

文献 [60] は, モバイル通信サービスにおいては実験室とフィールドで QoE が異なることを指摘している. モニターユーザを対象に, 各種アプリケーション利用直後に, その QoE を専用のスマートフォンアプリケーションを使用して評価させた. ACR 法を模擬して表 3 に示す評点を用いて評価させるとともに, QoE が低下した場合にはその理由 (Web 転送待ち時間の遅延, 動画ストリーミングの停止など) を記述させた. この方法により, フィールドでどのような QoE 劣化現象が発生したのかを明らかにすることができる. さらに, 実験後にインタビューを実施した. その結果, QoE は 5 段階評価等ではなく, QoE 劣化があったというバイナリで評価すべきであることが示された. また, QoE の低下をユーザが知覚するほどの事象はまれにしか発生しないため, その事象のみを把握すればよいとされた. 実際, 文献 [60] の実験では評点 1 (非常に悪い) がほとんどであることもこの主張を支持している. 実験室実験では Web 転送待ち時間のわずかな遅延なども知覚し評点を与えることが可能であるが, フィールドではわずかな品質劣化には寛容になる. そのため, 実験室実験のみではモバイル通信サービスの QoE を向上させるための取り組みとしては不十分であることになる. また, このことはユーザ自身も無自覚であり, 実験を経てユーザに認識されることになったことがインタビュー

一で確認されている。このように、エージェントを用いた実験ではユーザが無自覚な事象をも明らかにすることができる。

文献 [32], [34] も、フィールドにおいては QoE として MOS が適さないことを示している。Web ブラウジングにおいて、ユーザは単一の Web サイトにおける転送待ち時間の遅延には寛容であり、複数の Web サイトで平常時よりも著しく転送待ち時間が遅延しているときに QoE の低下を感じるということが明らかにされている。

2.2.2 QoE の客観品質評価法

主観品質評価法に対して、客観品質評価法と呼ばれる技術が存在する。QoE を活用すべき場面はサービスの設計段階だけではない。設計上は $MOS \geq 3.5$ を担保するようにしたとしても、実際のサービス提供中には QoS の劣化やユーザ端末の性能不足などによって $MOS \geq 3.5$ を満足しなくなることがある。このような QoE 低下事象を捉えて対応を行うことで、サービスに対するユーザ満足度を向上させていくことが重要である。しかし、前述のとおり QoE は主観品質評価が基本であり、サービス提供中に評価することは容易ではない。そこで、スループット/RTT のようにネットワーク上で自動的に計測できる指標から QoE を推定する技術が有用となる。このような技術のことを客観品質評価法と呼ぶ。

代表的な客観品質評価法を表 4 に示す。客観品質評価法はマルチメディア分野における検討が多くなされてきたため、その勧告が多数存在する [41]-[49]。

一般に、客観品質評価法においては、通信経路上でのパケット損失やパケット遅延ジッタなどを自動的に計測し、そのデータに基づいて主観品質評価値を推定する。これによって主観品質評価実験を行うことなく多数のユーザの実利用環境における QoE を常時監視でき、QoE の低下をユーザからの申告を待たずに検知し対策を検討できる。

表 4 代表的な客観品質評価法

モデル 項目	メディアレ イヤモデル	パケット レイヤモ デル	ビットス トリーム モデル	ハイブリ ッドモデ ル	プランニング モデル
入力	映像画素信 号	パケット のヘッダ 情報	ペイロー ド情報	左記情報 の組み合 わせ	ネットワーク や端末の品質 設計/管理パ ラメータ
出力	コンテンツ個別品質				コンテンツ平 均品質
主な用途	符号化品質 管理, 端末 機器の性能 比較	エンド端 末組み込 み型のユ ーザ品質 管理, 品 質劣化要 因の切り 分け	コンテンツ属性を加味 した品質管理 (利用シ ーンや取得可能な品質 情報に応じてモデルを 使い分ける)	机上のサービ ス品質設計・ 管理	
勧告	J.144/BT.1683 J.247/BT.1866 J.341/BT.1907 J.246/BT.1867 J.249 J.342/BT.1908	P.1201	P.1202	-	G.1070

2.3 SVM

本論文の目的には、機械学習の中でも SVM が有用である [35]-[37]。機械学習とは、データから反復的に学習し、パターンを見つけ出す手法を指す。スループット/RTT と QoE 低下の有無の関係を機械的に学習する手法の 1 つとして、SVM が該当する。

SVM は、複数の特徴量候補とバイナリで与えられる目的変数について実測値を与えることで、その関係を機械的に学習し、以後は特徴量の実測値のみを与えることで目的変数を推定する手法である。学習に使用できるデータが少ない場合に特に有効である。QoE 低下事象の発生がそもそもまれであることから、その測定は容易ではなく、学習に使用できるデータが少ないため、SVM は本論文の問題に適する。

以下、SVM について説明する。

2.3.1 混同行列

SVM では正解データの陽性/陰性と判定結果の陽性/陰性の組み合わせが 4 通り発生する。この組み合わせの数を表にしたものを、混同行列 (Confusion Matrix) と呼ぶ。混同行列を表 5 に示す。

1. True Positives (TP): 陽性と判定すべきときに陽性と判定した数
2. True Negatives (TN): 陰性と判定すべきときに陰性と判定した数
3. False Positive (FP): 陰性と判定すべきときに陽性と判定した数
4. False Negative (FN): 陽性と判定すべきときに陰性と判定した数

TP と TN は判定結果が正しいことを示し、この数値が大きいほど判定器の精度が高いということになる。

表 5 混同行列

		判定結果	
		陰性	陽性
正解	陰性	True Negative (TN)	False Positive (FP)
	陽性	False Negative (FN)	True Positive (TP)

2.3.2 評価メトリック

表 5から，以下に示すように複数の評価メトリックを計算できる．

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (1)$$

$$sensitivity = \frac{TP}{(TP + FN)} \quad (2)$$

$$specificity = \frac{TN}{(TN + FP)} \quad (3)$$

$$G - Mean = sensitivity \times specificity \quad (4)$$

$$False\ Positive\ Rate = \frac{FP}{(FP + TN)} \quad (5)$$

$$True\ Positive\ Rate = \frac{TP}{(TP + FN)} \quad (6)$$

以下, それぞれについて説明する.

2.3.2.1 Accuracy

Accuracy は式 (1) で求められる. 式 (1) より, TP か TN のいずれかが極端に高ければ, もう一方が低くとも Accuracy は高く評価されることが分かる.

2.3.2.2 G-Mean

G-Mean [61] は式 (2)–(4) で求められる. 式 (2)–(4) より, TP か TN のいずれかが極端に高くても, もう一方が低ければ G-Mean は低く評価されることが分かる. ゆえに, G-Mean は TP と TN の双方が高いことを以て良好な判定器と判断したい場合に利用される.

2.3.2.3 AUC

AUC (Area under the curve) [62] を説明するため, 式 (5) の偽陽性率と式 (6) の真陽性率について説明する. 偽陽性率は実際には負であるうち誤って正と判定する比率であり, 真陽性率は実際に正であるうち正しく正と判定する比率である. SVM では偽陽性率を高めるほど真陽性率が高まる. 偽陽性率と真陽性率の関係図が ROC (Receiver Operating Characteristic; 受信者動作特性) 曲線であり, AUC は ROC 曲線の下での面積を示す値である. すなわち, AUC が高いほど偽陽性率が低い時点で真陽性率が高いことを示す.

2.3.3 評価方法

2.3.3.1 クロスバリデーション

クロスバリデーションとは、学習データを任意の数に分割し一部のデータで判定器を作成して、残りのデータで判定精度を評価する手法である [53]. 分割されたデータのうち判定器の作成に使用するデータを次々と変えて、分割した数だけの回数分、判定器の作成と精度評価が行われる。各回の精度の平均値が最も高くなる判定器を採用することで、判定器の妥当性を高められる。

2.3.3.2 グリッドサーチ

グリッドサーチとは、判定器の精度を向上させるため、すべてのパラメータの組み合わせを試す方法である [63]. スループットと RTT の両方を判定に使う場合、スループットのみを判定に使う場合、RTT のみを判定に使う場合を試し、最も評価精度の良いものを探索する。SVM では各パラメータの寄与度を算出できないため、このような総当たりで比較する。特徴量は多いほど判定精度がよいとは限らないため、このような調査を行う。

このとき、判定精度が最も高くなるハイパーパラメータを導出する必要がある。ハイパーパラメータとは、誤判定の許容範囲を決定するための値であり、RBF kernel を用いる場合は C と γ の 2 変数がある。 C は SVM が解く 2 次計画問題の式の中で、 γ は RBF kernel の式の中で以下のとおり表される。

$$\min_{\beta} \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^N \xi_i$$

$$K(x, x') = \exp(-\gamma \|x - x'\|^2)$$

C が小さいほど誤分類を許容し、 γ が小さいほど単純な決定境界を作る。 C と γ の組み合わせごとに判定器が作成され、各判定器の判定精度が導出される。

2.3.3.3 不均衡データのサンプリング

計測したデータの中に劣化なしのデータが多く、劣化ありのデータが極端に少ない場合がある。これは、不均衡データと言われる。不均衡データに対してランダムにサンプリングして学習を行うと、常に劣化なしと判定する判定器が作成される。劣化なしと判定したほうが正解する確率が高いためである。そこで、不均衡データから劣化ありのデータを模擬的に作り出し、均衡データとして学習する手法が存在する [64]-[65]。これにより、実世界では不均衡データしか収集できない場合であっても均衡データとして学習を行い、精度の高い判定器を作成できる。本論文の研究対象はこれに該当する。日常的には劣化のないデータがほとんどであり、まれに検出される劣化のあるデータを捉えて判定器を作成する必要がある。

2.4 QoE 推定に関する先行研究

文献 [33]では、スループット/RTT 計測専用のサーバとスマートフォンの間での計測結果のみから、スループット/RTT 計測時とは別の通信セッションとなるサービス利用中の QoE を推定する手法が提案されている。前節ではいずれの文献もサービス利用中のパケットデータを解析することで多種のデータを扱い、精緻な分析を可能にしている。一方、サービス利用中の通信セッションにおけるパケットデータは使用せず、スループット/RTT からその場所・時刻においてサービスを利用したときの QoE を推定できれば、データの収集コストを抑え、さまざまな場所・時刻の QoE 推定が可能になる。文献 [33] では、エージェントを用いて各種サービス利用における QoE の低下が報告されたとき、その場所・時刻でのスループ

ット/RTT を複数回計測し、その関係を学習した。スループット/RTT の最悪値と QoE に一定の関係があることが示唆されたが、スループットと RTT の最悪値から QoE の低下有無を判別するには至らなかった。これは、スループット/RTT の計測手法が定常状態を計測するものとなっていたためと考えられる。

文献 [9]-[11], [57], [66]では、推定モデルの汎用化範囲が示されている。文献[9]-[11]では、サービスをまたぐ汎用化は困難であり、サービスごとに開発する必要性が主張されている。この根拠として、映像配信サービスでは再生の中断、Web ブラウジングでは Web 転送待ち時間が QoE に影響するなど、QoE の構成要素が異なるために同じ推定モデルが使用できなかったことが示されている。一方、文献 [57] では、Web ブラウジングにおいては、アプリケーションや OS、利用状況に関してはある程度の汎用性が認められている。すなわち、ニュースアプリなど、ブラウザエンジンを使用しているアプリケーション利用においてはブラウザと同じ QoE 推定モデルが適用できる。また、Android 端末で収集したデータに基づいて生成した推定モデルは iOS 端末やタブレット端末利用における QoE 推定にも適用できる。そして、文献 [66]では、混雑駅での学習結果は移動中の電車内にも適用できることが示されている。

文献 [9]ではまた、推定モデルが変化する要因を明確にできない以上、推定モデルの生成は常時行うべきであることが主張されている。この点、文献 [67]では QoS の実測値をオンラインで収集し、Hadoop [68] を使用して機械学習をオンラインで常時自動実行する仕組みを提案し、頻繁に推定モデルをアップデートすることを推奨している。

2.5 機械学習を用いた QoE 推定手法

機械学習を用いることで、パケットデータから得られる多種のデータと QoE の関係を学習する手法が提案されている。

文献 [23]-[24]では映像配信サービスについて機械学習を用いて MOS を推定する手法が示されている。特徴量の候補として QoS (スループット, パケット損失率, パケットの到着間隔), システムの状態 (バッファリング/解像度変更等), ユーザ行動 (シーク/最大化/停止等) が提案されている。

文献 [25]では VoIP サービスについて機械学習を用いて MOS を推定する手法が示されている。機械学習を用いることで E-model を超える精度を実現できたことが報告されている。

文献 [26]では 360 度映像配信サービスについて機械学習を用いて MOS を推定する手法が提案されている。特徴量の候補として QP (Quantization Parameter: 量子化パラメータ) や解像度などに加えて, ユーザの性別や興味のあるコンテンツカテゴリ, VR (Virtual Reality) への熟練度なども用いられ, MOS の代わりに VR 酔いを目的変数とした分析も行われた。その結果, ユーザは興味があるコンテンツについては酔いにくいことが示された。

文献 [27]-[32] では Web ブラウジングについて機械学習を用いて MOS を推定する手法が示されている。特徴量の候補として PLT (Page Load Time: ページ全体の読み込み時間), SI (Speed Index: スクロールしないで閲覧する事のできる画面領域の読み込み時間), RTT, Web サイトの情報 (カテゴリ, 動的コンテンツの有無, 通信量など) が提案されている。このとき, 推定モデルが学習に用いられた Web サイトに依存することが確認され, 学習に用いる Web サイトについて, 動的コンテンツの多寡や通信量などに配慮が必要であることが示唆されている。また, スマートフォン上でのレンダリング時間がデータに混入するとスマートフォンの性能に影響を受けるため, レンダリング時間を除いた通信に関わる Web 転送待ち時間のみを扱うことが望ましいことも確認されている。

なお, 機械学習の手法としては SVM のほかに決定木分析など, 多数の手法が各文献で利用されている。いずれの手法を使うかについては確定的な根拠や制限はなく, 推定精度が出ればよいとされている [22]-[32]。

2.6 先行検討の課題

ここまで、モバイル通信サービスにおける QoS と QoE の関係調査に関する先行検討を示した。表 6 に、先行検討と本論文における提案手法の関係をまとめる。

表 6 先行検討と提案手法の対比

先行検討 \ 項目	パケット解析 不要	瞬間的な劣化	フィールド QoE	機械学習による 関係分析
[33], [60]	Yes	No	Yes	No
[8], [22]-[32]	No	Yes	No	Yes
提案	Yes	Yes	Yes	Yes

まず本論文の目的を整理する。さまざまな場所・時刻の QoE を把握するためには、サービス利用中の通信セッションのパケットデータを扱うことはできない。そこで、計測専用のサーバとスマートフォン間でパケット交換して計測されるスループット/RTT を特徴量とする必要がある。このとき、スループット/RTT は瞬間的な品質劣化を捉えられる形式で計測される必要がある。また、目的変数としては MOS ではなく、サービスごとに異なるフィールドでの QoE とする必要がある。たとえば Web ブラウジングでは複数のサイトで著しい転送待ち時間の遅延が発生しているときを QoE 低下ありとして検出する必要がある。上記の特徴量と目的変数の関係について機械学習を適用し、特徴量を常時計測して判定器にかけることでさまざまな場所・時刻の QoE 低下を検知するシステムを実現できることが目指す姿となる。

文献 [33], [60] は、サービス利用中のパケットを用いず、単純なスループット/RTT 計測結果から、フィールドでの QoE をバイナリで捉えて推定することを提案

している。しかし、瞬間的な劣化を捉えられておらず、フィールドでの QoE を推定するモデルの確立には至らなかった。

文献[8],[22]-[32]は、スループット/RTT と QoE の関係を機械学習によって捉えられることを示している。特に文献[8] は、スループットの瞬間的な劣化を捉えることの重要性を強調している。しかし、これらの検討ではサービス利用中の通信セッションについてパケットデータを利用できることを前提としている。そのため、インプットされる情報として常に瞬間的な劣化が捉えられており、本論文の目的であるサービス利用中の通信セッションのパケット情報を扱えない条件には当てはまらない。また、QoE を MOS で捉えているという課題がある。フィールドにおいて QoE は MOS ではなく、サービスごとの QoE 低下事象の発生有無を捉えることが重要である[9]-[11], [32], [34]。たとえば Web ブラウジングにおいては、1 つの Web サイトの転送待ち時間から MOS を推定することは QoE の捉え方としては適切ではなく、複数の Web サイトで Web 転送待ち時間の遅延を体感したときを QoE 低下と捉えることが重要であることが示唆されているが [34]、QoE の低下を捉える手法は明らかとなっていない。

また、既存研究では QoE 推定の対象とすべきサービスの特定に関する検討は十分ではない。文献[9]-[11] より、QoE 推定モデルはサービスごとに検討が必要であることが分かっている。この点、ユーザによる各サービスの利用頻度やその品質に対する満足度を調査する手法が示されており[60]、かつ映像配信サービスや Web ブラウジングの利用頻度が高いことからそれらを対象とした研究が多い。しかし、モバイル通信サービスにおける QoE 低下事象を多く引き起こしているサービスの特定には至っていない。まず、モバイル通信サービスに対して不満を覚えるきっかけとなっているサービスを特定し、その QoE を瞬間的な品質劣化が加味されたスループット/RTT から推定する手法が必要である。

3 モバイル通信サービスの動態調査

3.1 概観

3章では、モバイル通信サービスに対して不満を覚えるきっかけとなるサービスを特定する手法を提案する。

そのために、モバイル通信サービスのユーザ動態を調査する必要がある [33], [60]. 3章では、先行検討を参考にモバイル通信においてユーザ動態を調査し、QoE低下が多く体感されているサービスを明らかにする方法とその調査結果を示す。

3.2 提案手法

従来、QoEを調査するためにはACR法などの主観品質評価が必要とされてきた [38]-[40].

しかし、文献 [60]がエージェントを利用して、サービス利用が阻害されるほどの著しい品質劣化が発生したときにのみユーザに報告させることで、実フィールドにおけるQoE低下事象を効率よく抽出できることが示されている。

本論文においてもエージェントを用いた調査を行う。エージェントは、通常のスマートフォンアプリケーションの形式で配布され、バックグラウンドでパケットキャプチャやOSのシステムログ（CPU/メモリ使用率、アプリケーションの利用時刻等）取得を行うと同時に、ユーザによって起動された際に、QoE低下報告を行うかを確認し、ユーザによって承認された場合にQoE低下が報告された時刻を記録する。3章の目的はユーザ動態の調査であるので、主観品質評価実験よりもエージェントを用いたフィールドでのデータ収集が有効である。なお、フィール

ドで広く動態調査を行うためにはアンケート調査を行うことも考えられる。しかし、アンケート調査はユーザ自身が十分に知覚している事象を対象に感想などを求めるには最適であるが、無自覚な事象について調査するには不適である。モバイル通信サービスに対して不満を覚える事象についてユーザ自身が必ずしもすべて知覚しているとは考えられないため、エージェントを用いた実験を行うことによって無自覚であった事象についても明らかにすることが重要である。

エージェントを用いた実験の利点を説明する。通信品質やアプリケーションの利用情報などをコアネットワークやアクセスネットワークではなくユーザ端末で収集することになる。これにより、誰にでも同様の調査ができるという再現性の高さがまず挙げられる。また、実際にユーザ端末上で観測された事実であることから、QoEに近い情報であるとも言える。他の方法としては、プロキシサーバなどでトラフィックを監視し、利用アプリケーションを推定するなどの取り組みも存在するが[69]、エージェントを用いた実験ではユーザ端末で観測することで推定を必要とせず、OSから渡される確かな情報に基づいて考察が可能であることもエージェントを用いた調査法の強みと言える。文献[33], [60]はエージェントの開発研究であるため、ユーザがモバイル通信においてQoE低下を多く体感しているサービスについて調査したものではないが、上記の利点を鑑みてエージェントを用いた調査が有効であると考えられる。

実際に、ユーザがモバイル通信においてQoE低下が多く体感しているサービスを明らかにすることを目的として、エージェントを用いた実験を行った。その際のエージェントの機能は以下のとおりである。

- 1 QoE低下の報告機能：実験参加者は、サービス利用中にQoE低下を体感した際、エージェントを起動しその旨を報告する。文献[33], [60]を参考に、ACR法等の5段階評価ではなく、報告をするということは耐え難いQoE低下があったことを表すものとして扱う。このとき、対象のサービス（Webブラウジング等）や、周辺状況（通勤電車、自宅等）を報告する。

2 通信品質測定機能

2.1 パケットキャプチャにより，サービス利用中の通信セッションの時間，データ量，転送時間を記録し，スループットを計算する．

2.2 QoE 低下の報告時に計測専用サーバへ ping を送信し，RTT を記録する．

3 その他

3.1 各データは実験参加者を識別する番号を持つため，QoE 低下の報告回数等とあらかじめ記録した年齢や性別などのユーザ属性の関係分析が可能である．

3.2 各サービスの利用回数を記録する．

実験条件は以下のとおりである．実験参加者の数については，マルチメディアサービスの QoE 評価法を定める P.910 が 15 人を最低条件としていることを参考に [38]，それを十分上回る数を募った．今回の実験は主観品質評価実験ではなく，対象サービスもマルチメディアサービスに限ったものではないものの，P.910 は評価者の属性の影響を最小化するために多くの検討がなされた結果を受けていることから，参考にできると判断した．

1 実験期間：20 日

2 実験参加者：44 人

2.1 性別：男性 6 人，女性 25 人，未回答 13 人

2.2 年齢：20 代 8 人，30 代 7 人，40 代 11 人，50 代 5 人，未回答 13 人

2.3 職業：主夫/主婦 12 人，学生 7 人，会社員 20 人，フリーランス 5 人

2.4 その他：全員東京在住かつ Android スマートフォン使用歴 3 か月以上，報告の信頼性を担保するために謝礼を設定し，実験前説明会と誓約書の作成を行った

評価メトリックとして、以下を利用した。

QoE低下の報告回数

$$\text{QoE低下率} = \frac{\text{QoE低下の報告回数}}{\text{サービス利用回数}}$$

$$\text{スループット} : \frac{\text{サービス1回の利用に関わる全セッションの転送量合計}}{\text{サービス1回の利用の終了時刻 - 開始時刻}}$$

RTT : QoE低下報告のたびに1 pps 分のRTT

QoE低下率は比較的ユーザ特性の影響を受けずに、サービスや通信品質の特性を捉えやすいと考えた。QoE低下の報告回数はサービス利用回数に依存するため、これだけではスマートフォンの利用頻度というユーザ特性を表してしまう可能性がある。

スループットは、今回の実験ではパケットキャプチャを利用した。キャプチャデータには、利用アプリケーションの情報が印加されているため、各アプリケーション利用時のスループットを正確に計算できた。なお、TCPのスロースタート[70]-[71]によって、セッションが短い場合には通信品質に問題がなくてもスループットが小さく見える場合があるが、本実験でQoE低下が報告されたアプリケーション利用についてはそのような事象が確認されるほどセッションが極端に短いケースは観測されなかった。

3.3 結果

本実験におけるその他の調査結果を図2に示す。図は、職業ごと、1時間単位のQoE低下率を示している。最大でも0.8%程度と、QoE低下事象の発生はまれであるとした文献[60]と同じ結果となった。会社員については8時、12時、17時前

後に QoE 低下事象が増加している。朝夕の通勤ラッシュと昼の休憩時間に QoE 低下事象が多いことが分かる。会社員以外はまだらであり、特徴が見られない。会社員以外は 8 時, 12 時, 17 時前後に QoE 低下率が高いわけではないので, コアネットワークやサーバの問題で QoE が低下しているとは考えづらい。会社員は朝夕に通勤電車, 昼の休憩時間にオフィス街に集中しており, 一斉にスマートフォンを利用したことでアクセスネットワークの一時的な輻輳が発生し, それによって QoE が低下したと考えられる。

さらに, 表 7 に各時間帯のスループットと RTT を示す。いずれも 1 時間単位の平均値である。図 2 から, 8-9 時の QoE 低下率は他の時間帯よりも 2 倍以上高いことが読み取れる。一方, 表 7 から, 8-9 時の通信品質はスループット/RTT いずれの観点でも他の時間帯よりも 2 倍以上良好という結果であったことが読み取れる。スループット/RTT の平均値では, QoE 低下率を推定できないことがこの結果からも分かる。

なお, QoE 低下時のサービスは Web ブラウジングが主であった。Web ブラウジング, 動画, チャット, メール, IP 電話, その他 (自由記述) の選択肢の中で Web ブラウジングが 9 割と圧倒的多数であった。

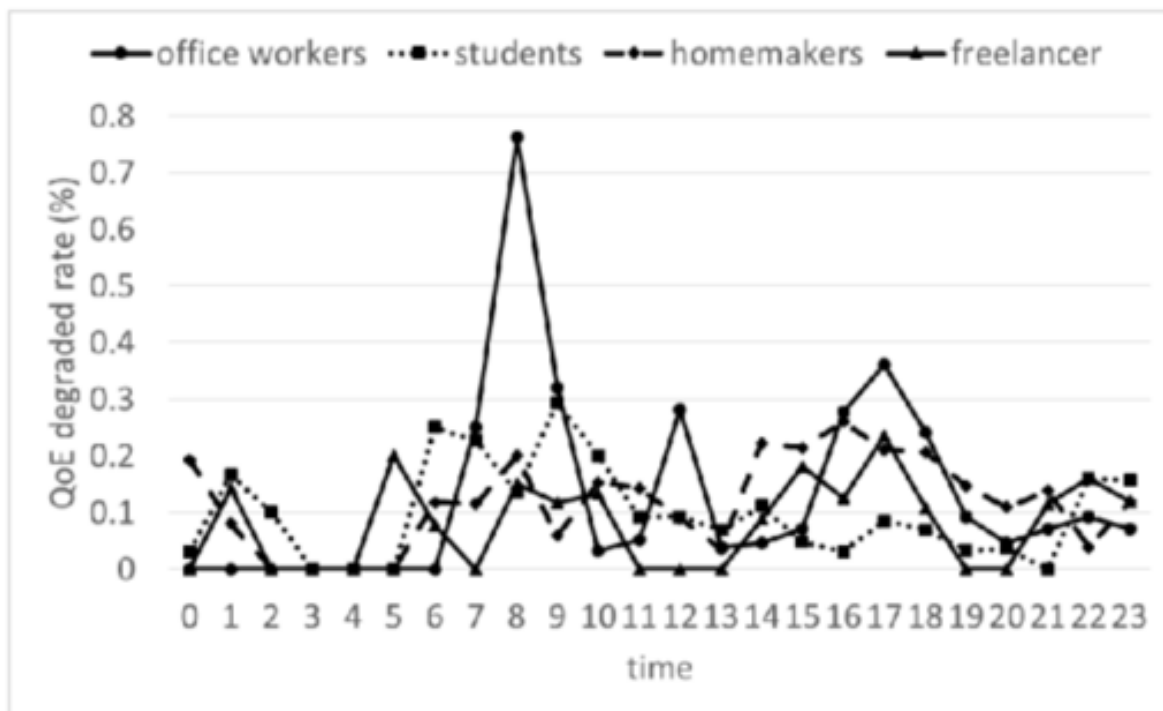


図 2 職業及び時間帯別の QoE 低下率

表 7 時間帯別の通信品質 (1 時間の平均値)

項目 \ 時間帯	早朝(5-6 時)	通勤時間帯(8-9 時)	夕方(15-16 時)
スループット	139 Kbit / s	503 Kbit / s	109 Kbit / s
RTT	472 ms	199 ms	398 ms

3.4 議論

まず、スループット/RTT の平均値が QoE 低下率と比例しない原因について考察する。実験におけるスループット/RTT は各サービス利用中のパケット解析によるものであり、瞬間的な品質劣化を捉えることを目的としたものではなかった。そのため、スループット/RTT と QoE 低下率とが無関係なデータとして集計されてしまったと考えられる。

次に、Web ブラウジングについて QoE 低下報告が多かった原因を考察する。文献[66]より、Web ブラウジングのほか、メール、チャットや映像配信サービスなどの利用頻度も高い。メール、チャットについては転送量が小さいため QoE 低下につながりにくかったと予想できる。一方、映像配信サービスは転送量が大きく QoE 低下が予想されるが、その報告は少なかった。この理由を考察する。映像配信サービスの主な QoE 低下要因は再生の中断である[23]-[24], [72]。特に VoD はクライアントに 10 秒程度のバッファがあるため、瞬間的なスループット/RTT の劣化では再生の中断には至らない。そのため、バッファが小さい Web ブラウジングのほうが瞬間的なスループット/RTT 劣化の影響が強く、QoE 低下報告数の多さにつながったと考えられる。

3.5 結論

エージェントを用いた計測によって、各サービスの利用ログ、スループットと RTT、QoE を計測でき、その結果から QoE 低下が頻発しているサービスが Web ブラウジングであることを特定できた。すなわち、ユーザ動態を調査し、モバイル通信サービスに対して不満を覚えるきっかけとなる主たるサービスを特定するという目的は達成できた。

しかし、エージェントを用いたデータ収集だけでは本論文の目的を達成できない。エージェントを活用したデータ収集は、実験室に評価者を拘束する ACR 法などを用いた主観品質評価実験と比べれば実験コストは低いものの、プライバシーに関わる実験となるため、評価者に対する謝礼の支払いの設定や収集されるデータや評価者の権利などについて誤解がないように対面での説明会を必要とする上、パケットキャプチャによる巨大なファイルの管理を必要とすることや、各サービスの利用頻度など個人情報に関わるデータを収集してしまうことから、Speedtest のような調査に比べれば実験コストは高い。そのため、さまざまな場所・時刻の

品質調査ができない。簡易に計測できるスループット/RTT から QoE を推定する手法が必要である。

4 機械学習による判定器作成手法

4.1 概観

3章では、通勤時間帯における Web 転送待ち時間の著しい劣化が、QoS に対する不満を覚える状況として顕著に現れた。

4章では、スループット/RTT の計測結果から瞬間的な品質劣化を検知して、フィールドにおける Web ブラウジングの QoE 低下事象の発生を捉える手法を提案する。さまざまな場所・時刻での QoE 評価を可能とするために Web ブラウジング中のパケットデータを用いないことが前提となる。そこで、場所と時刻を一定の空間と一定の時間幅で区切ってデータを扱うこととする [73]。以下、この区間のことをセットと呼称する。セット単位でスループット/RTT と Web 転送待ち時間遅延の関係を教師あり学習によって捉えることで、スループット/RTT の計測結果から Web 転送待ち時間劣化が発生したセットを判定する判定器を作成する。このとき、スループット/RTT は瞬間的な劣化を捉えられるようにするため、スループットは転送量を小さく多数回計測し、RTT は計測間隔を空けて多数回計測する。また、Web 転送待ち時間そのものや各転送待ち時間の MOS では QoE を捉えられないことから [34]、Web 転送待ち時間は多数のサイトで著しく待ち時間が遅延している状態を QoE 低下ありとみなす。

ユースケースとしては、スループット/RTT を常時計測し、判定結果を地図上にプロットするなどが考えられる。たとえば 1 km 四方、1 時間単位で地図をメッシュに区切って QoE 低下が発生しているメッシュを可視化する。また、Web 転送待ち時間も常時計測し、判定器の作成・更新に必要なデータセットが揃う都度、自動的に判定器を作成・更新することが考えられる。文献 [67] では分散処理技術を用いて高速に機械学習を完了させることを提案していた。これを Amazon Redshift

などのパブリッククラウドを活用して実現することにより、さらに低コストでのリアルタイム更新が可能になる [74].

続いて、前提となるスループット/RTT, Web 転送待ち時間の捉え方を説明する. スループット/RTT はサーバ, 通信事業者のコア網, 通信事業者設備とユーザ端末間のアクセス区間, 端末の各区間で劣化し得る. この中で, モバイルはユーザ端末が移動することからアクセス区間が一時的に混雑し通信品質が劣化する事象が発生しやすく, その発生を推定するための検討が多数存在する [3]-[6]. そのため, 本論文ではアクセス区間の通信品質劣化を対象とし, 他の区間での劣化を極力排除して検討する. 一方, Web ブラウジングにも様々な処理の段階があり, どこまでを Web 転送待ち時間として捉えるか定める必要がある. ユーザ端末から Web サーバにコンテンツ転送を要求し, コンテンツが転送され, コンテンツの内容がブラウザ上に描画される. この内, 描画処理は端末上での処理となり, 通信が関与しない. そのため, スループット/RTT と関係しづらい描画処理時間は対象外としなければ, QoE 推定モデルに悪影響が出る [32]. そこで, ユーザによる転送要求からユーザ端末への Web コンテンツの転送完了までを Web 転送待ち時間として扱う.

4章は以下のように構成される. 4.2節でスループット/RTT の計測結果から Web 転送待ち時間が遅延したセットを判定するにあたっての課題を整理する. 4.3節で課題解決のための手法を提案する. 4.4節で実際の実験を用いて実施手順を説明する. 4.5節で提案手法の判定精度を含む実験結果を紹介する. 4.6節で結論を述べる.

4.2 課題

本節では, セット単位で, スループット/RTT の計測結果から Web 転送待ち時間遅延の有無を判定する手法を検討するにあたっての課題を整理する. Web ブラ

ウジング中の通信セッションを解析せずとも、同一セット内のスループット/RTT計測結果から Web 転送待ち時間を推定できる可能性については多数の報告があるものの [75]-[78], スループット/RTT の計測結果から Web 転送待ち時間遅延の有無を判定する手法の確立には至っていない。

第 1 の課題として、スループットの計測手法が挙げられる。従来は定常状態における通信品質の実力値を把握することが目的に、大容量ファイルを転送する手法が採用されている [3]-[6]。しかし、この方法では瞬間的なスループットの落ち込みを捉えることができない。瞬間的な劣化を捉えることで QoE 推定に活用できることが示されており、これに対応した計測手法が必要である [7]-[8]。

第 2 の課題として、RTT の計測手法が挙げられる。通常、Ping は軽量なため、QoE 低下が発生するほど QoS が劣化しているセットであってもほとんどの場合には短時間で完了してしまう [79]。RTT が著しく劣化するというまれな事象を捉えるための計測手法が必要である。

第 3 の課題として、Web 転送待ち時間の計測手法が挙げられる。Web サイトは接続の都度、広告などのコンテンツが変わることから、そもそも QoS に劣化がないセットで期待される Web 転送待ち時間そのものが変動する。すなわち、Web 転送待ち時間を秒単位等で捉えたとしても、その要因が切り分けられない。さらに、Web サーバやユーザ端末など、QoS とは別の要因による遅延も起こり得る。この性質を踏まえた上で、アクセス区間の劣化に起因する Web 転送待ち時間遅延が発生していることを捉えるための計測手法が必要である。

4.3 提案手法

4.2 節の課題に対応した計測手法や評価メトリック及びその関係を学習する方法を提案する。

第1の課題に対応して、スループットは小容量ファイルを多数回計測し、その平均値を評価メトリックとする。これによって、セッション確立後に大容量転送を行ったことによる高スループットを記録するのではなく、セッション確立が速やかに行われたかどうかを評価可能にする。

第2の課題に対応して、RTTは間隔を空けて多数回計測し、その平均値を評価メトリックとする。これによって、まれに観測されるRTTの著しい劣化を検出し、そのセットにおける通信品質の劣化状況を評価可能にする。

第3の課題に対応して、複数のWebサイトで転送待ち時間を計測し、その劣化度を評価メトリックとする。

劣化度について解説する。まず、各Webサイトの転送待ち時間を多数回計測し、その中央値を算出する。そして、各セットにおける各Webサイトの転送待ち時間（実測値）を、各Webサイトの中央値で除算した値を劣化度とする。劣化がまれな事象であることから、各Webサイトを多数回表示したときのWeb転送待ち時間の中央値は、そのWebサイトにおける平常時のWeb転送待ち時間とみなせる。この平常時の値と比較してWeb転送待ち時間が著しく長い事象を劣化として捉える。

しかし、1つのWebサイトの転送待ち時間が偶発的に著しく劣化することが考えられる。これを説明するため、図3にWeb転送待ち時間の不安定さを示す。図3は同一のセット・同一のユーザ端末で2つのWebサイトの転送待ち時間を連続で計測し、横軸をセットとしてサイトごとの転送待ち時間をプロットした図である。横軸が同じである青と橙の頂点は同一セットにおける各サイトの待ち時間を示す。黒の両矢印が示すとおり、同一セットでスループット/RTTが同程度と考えられるときでも、Webサイトごとの転送待ち時間に乖離が生じ、また、特定のサイトが遅延するとは限らず、不安定である。このとき、2つのWebサイトの平均値では黒の両矢印の大きさを捉えづらく、最悪値では1回の計測結果の影響が過大である。

そこで、複数の Web サイトで転送待ち時間が平常時から著しく劣化する事象を、QoE が低下したセットとして扱う。これを実現するため、劣化度が一定値以上の Web サイトが一定数以上存在する場合に QoE が低下したセットとみなす。

さらに、Web 転送待ち時間の計測対象サイトの要件を 2 点説明する。(1) Ajax や CSS 等の Web 構造が Web 転送待ち時間に影響するという報告があるため [27], [80]-[81], 計測対象サイトはそれぞれ異なる Web 構造を持つように設定することで、各サイトが持つ Web 構造の違いによる遅延の影響を抑える。(2) 計測対象とする Web サイトの数は、捉えたいセットの時間幅に計測が収まる程度に調整する必要がある。Web サイト数の目安については、5 章にて検証を行う。

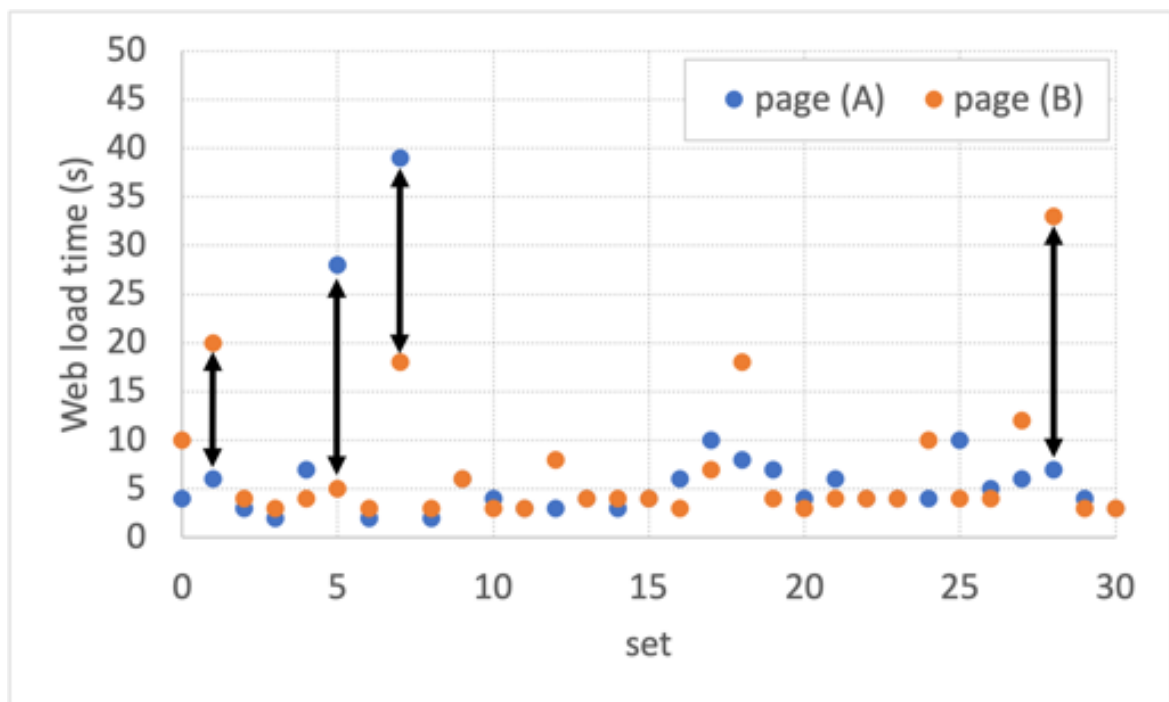


図 3 同一セット内の 2 つの Web サイトの転送待ち時間実測例

上記の計測手法と評価メトリックで得られたデータ間の関係を学習する方法として、SVM を利用する。1 セットの計測中に混雑具合等の周辺状況が変動しないように、計測にかかる時間に留意しながら計測シナリオを定め、1 セット内に

Web 転送待ち時間計測からスループット/RTT 計測を行う。セットごとのスループット/RTT の平均値を特徴量候補，Web 転送待ち時間の劣化有無のバイナリ情報を目的変数として SVM を適用する。図 4 に，SVM における学習フェーズと判定フェーズの概要を示す。

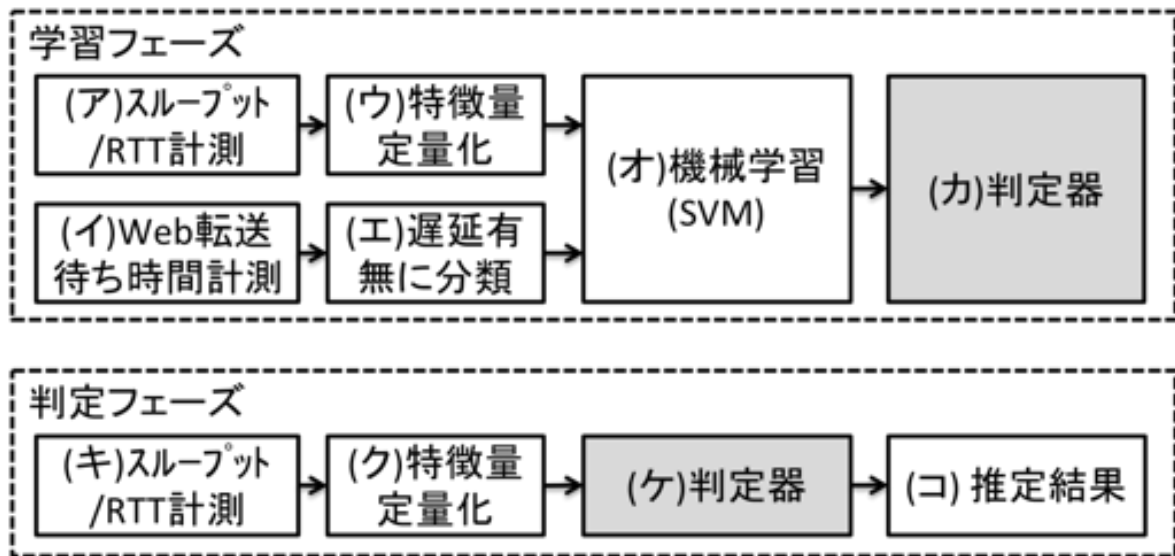


図 4 提案手法の全体像 (学習フェーズ，判定フェーズ)

学習フェーズについて解説する。前述の計測手法を用いて各セットの (ア) スループット/RTT と (イ) Web 転送待ち時間を連続で計測する。スループットは少量のデータ転送で複数回計測し，RTT は間隔を空けて多数回計測する。Web 転送待ち時間は構造が異なる複数のサイトを対象に計測する。さらに，前述の評価メトリックを定量化するため，(ウ) スループット/RTT は平均値を計算し，(エ) Web 転送待ち時間はあらかじめ設定した劣化度のしきい値に基づいて遅延あり/なしに分類する。そして，(オ) スループット/RTT の平均値を特徴量候補，Web 転送待ち時間の劣化有無を目的変数として SVM [35]-[37] を適用し，(カ) 判定器を出力する。SVM は多数の特徴量と 2 値の目的変数の関係を学習し，特徴量の実測値を入力すると目的変数の判定結果を出力する，判定器を作成するための教師あり学習の手

法の一つである。これにより、スループット/RTT の平均値から、同一セットにおける Web 転送待ち時間遅延の有無を判定する判定器を作成する。

判定フェーズについて解説する。学習フェーズとは別のセットにおいて (キ) スループット/RTT 計測を行い、(ク) 平均値を計算する。その結果を (ケ) 学習フェーズで作成した判定器に入力することで (コ) 同一セットにおける Web 転送待ち時間遅延の有無が判定される。

学習フェーズで判定器を作成した後は判定フェーズが繰り返される。判定フェーズではスループット/RTT のみを計測し Web 転送待ち時間は計測しない。スループット/RTT を学習フェーズと同様の手法で計測し、作成済みの判定器に入力することで、そのセットにおける Web 転送待ち時間遅延の有無を判定する。

4.4 実施手順

本節では提案手法の実施手順を実例に基づき説明する。表 8 に実験条件を示す。

判定器の作成には Web 転送待ち時間が遅延している状況と、遅延していない状況の双方での計測結果が必要である。一般に遅延の発生率は低いため、ユーザ端末が集中し遅延が予想される場所・時刻において計測した。以下、訓練データとしてのスループット/RTT と Web 転送待ち時間を計測し、SVM によって判定器を作成するための手順を、図 5 を用いて説明する。なお、各計測項目の設定は表 9 のとおりとした。

表 8 実験条件

計測地	都内ターミナル駅付近 1 箇所
計測日時	平日 18-20 時, 20 日間
端末・通信環境	通信事業者 2 社(A, B), LTE, Android スマートフォン各 2 台 (全 4 台, 同一機種, OS バージョンは 4.4 (ビルド番号まで同一), CPU2.2GHz quad core, メモリ 2GB), TCP バージョンは CUBIC
セット単位	1 セット 30 分
計測したセット数	655 セット (1 日約 9 セット × 4 端末 × 18 日間)

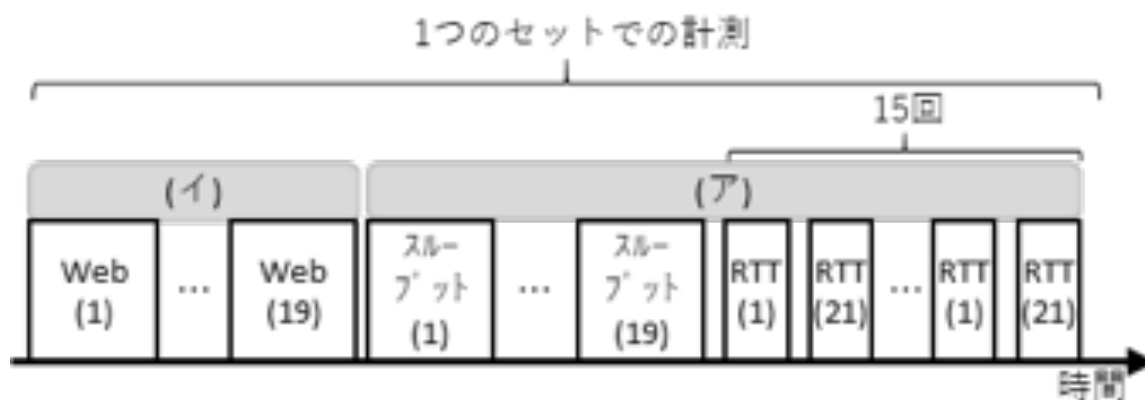


図 5 計測の手順

表 9 計測項目の設定

計測項目	設定
Web 転送待ち時間	<p>試行回数 19 回 重複なしの 19 サイト</p> <ul style="list-style-type: none"> ● EC (Electronic Commerce) サイト: 5 ● ポータルサイト: 5 ● マスメディア: 3 ● 検索: 1 ● ブログ 1 ● その他 4 <p>試行間隔 転送完了後 1 秒</p> <p>Time out 120 秒</p>
スループット	<p>大手 CDN から定型コンテンツをダウンロード</p> <ul style="list-style-type: none"> ● 転送量 2MB ● 転送回数 19 回 ● 転送間隔 20 秒 ● Time out 180 秒
RTT	<p>大手検索エンジンの公開サーバ宛</p> <ul style="list-style-type: none"> ● パケット長 64 bytes ● 1pps (packet per second) × 21 パケットを 20 秒おきに 15 回

まず、スマートフォン上で(イ) Web 転送待ち時間を複数回、複数サイトにて計測する。続けて、同一セットにおける(ア)スループットを複数回計測する。さらにその後、同一セットにおける(ア)RTTを複数回計測する。Web 転送待ち時間計測から RTT 計測までを1セットとする。多数のセットで計測を行い、セットごとにスループット/RTTの平均値と Web 転送待ち時間の遅延有無を示したデータを作成し、これを SVM に入力することで判定器を得る。

以下、4.3.1 節で(イ) Web 転送待ち時間の計測手法を、4.3.2 節で(ア)スループット/RTTの計測手法を説明する。4.3.3 節で Web 転送待ち時間遅延あり/なしの分類手法、4.3.4 節でスループット/RTTの平均値の計算手法を説明する。4.3.5 節で Web 転送待ち時間遅延の有無とスループット/RTTの平均値の関係を SVM で学習し判定器を得る手法を説明する。

4.4.1 Web 転送待ち時間計測の手順

図4の(イ)で示される Web 転送待ち時間の計測手順を説明する。ブラウザ標準 API である Navigation Timing API で規定されたタイムスタンプを用いて計測する [82]-[83]。コンテンツの転送時間を含む DOM (Document Object Model) の構築が完了した時刻を示す `domContentLoadedEventEnd` の時刻から、Web 転送を開始した時刻を示す `navigationStart` の時刻を差し引いた値を、Web 転送待ち時間として使用した。

次に、計測対象とする Web サイトの選定基準を説明する。第1, 第2の課題を考慮し、以下の観点で選定した。

- (1) 多様な Web 構造を持つ
- (2) 計測対象数は 5~20 サイト程度を目安とする

(1) は、多様な Web 構造を考慮するためである [27], [82]-[83]. Web サイトには、テキスト中心の Web サイトや、Ajax を活用した動的な Web サイトなどが存在する. もし、特定の Web 構造ばかりを評価対象とすると、Web 転送待ち時間とスループット/RTT の関係がその Web 構造に依存したものとなり、遅延の影響の出方に偏りが出る可能性がある.

(2) は、通勤ラッシュによる Web 転送待ち時間の遅延を想定し、セット内の人口密度が 30 分単位で大きく変化すると考え、1 セットの計測時間が 30 分程度となるように設定した. 今回は、その中で 19 サイトとした. これにより、計測時間が長引いて Web 転送待ち時間計測時とスループット/RTT 計測時とで混雑度合い等の状況そのものが変わってしまうことを避けられるようにした.

4.4.2 スループット/RTT 計測の手順

図 4 の (ア) で示されるスループット/RTT の計測手順を説明する. Web 転送待ち時間の計測が完了した直後、サーバからユーザ端末に約 2MB のファイルを 19 回、20 秒おきに転送し、各回のスループットを計測した. さらに、スループット計測の直後に Ping を 1pps で 21 サンプル、20 秒休みを 15 回繰り返して RTT を計測した. スループットと RTT で計測サーバは異なるが、今回はアクセス区間の劣化を想定するため、いずれもサーバ要因での品質劣化を防ぐ観点から Web サーバとは別の安定的に稼働していることが期待される CDN を利用した.

また、計測にあたっては文献 [84] が示す、スループット/RTT 計測において遵守すべき以下の事項に従った. これらに従わない場合、計測結果が揺らぎ、信用できないデータになる.

- (1) 計測直前から計測完了まで、端末の位置や手に持つ/机に置く等の環境を変えずに計測する

(2) 計測中は計測以外の送受信パケットが0またはわずかであること

本論文の実験では、駅近辺の机上にスマートフォンを設置し、バックグラウンドで計測シナリオ以外の通信が発生しないよう、不要なアプリケーションをあらかじめアンインストールした。また、計測中はパケットキャプチャを行い、計測以外の送受信パケットが0またはわずかであることを確認した。

4.4.3 Web 転送待ち時間遅延の判定手法

第1の課題を考慮して、Web 転送待ち時間データをそのまま利用するのではなく、各セットを Web 転送待ち時間遅延あり、なしに分類した。SVMを適用するために多数の学習データを計測する必要があることから、Web 転送待ち時間は自動的に転送・計測を繰り返すシステムを用いた。そのため、たとえば ACR 法[38]のような主観品質評価は行わず、Web 転送待ち時間に一定の基準を設けて遅延あり、なしに分類した。その手順は以下のとおりである。

- (1) 全計測データ (655 セット) から、Web サイトごとに転送待ち時間の中央値を算出する
- (2) 各セットにおける各 Web サイトの転送待ち時間計測結果の、(1) で求めた中央値からの劣化度を、以下の式で定量化する

$$\text{劣化度}(x) = \frac{\text{各セットにおける各Webサイトの計測値}}{\text{各Webサイトの中央値}}$$

- (3) 各セットにおいて、劣化度 x が 2 以上の Web サイトが 2 割以上存在するとき、そのセットを Web 転送待ち時間遅延ありと判定し、その他の場合を Web 転送待ち時間遅延なしと判定する

(1)~(3)の手順について説明する．一般に Web 転送待ち時間の劣化事象の発生頻度は小さいため，多数回の計測を行ったときの中央値は，その Web サイトにおける正常な転送待ち時間であると期待できる．図 6 に，実験で計測したある Web サイト (1 サイト) の転送待ち時間の累積分布関数 (CDF: Cumulative Distribution Function) を示す．

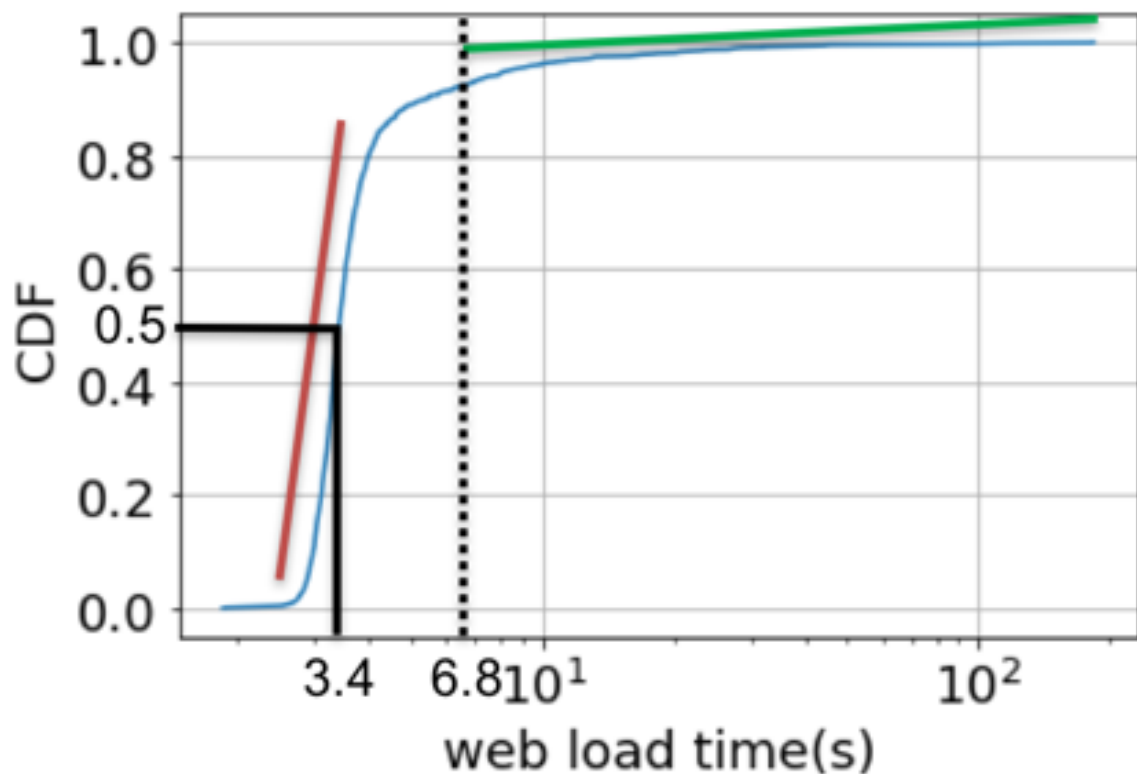


図 6 Web 転送待ち時間の累積分布関数の実測例

黒の実線は中央値 (3.4 秒) を示している．赤い線が示すように，中央値付近では勾配が大きいことから，中央値前後のデータが多いことが分かる．黒の破線は中央値の 2 倍の値 (6.8 秒) を示している．緑の線が示すように，中央値の 2 倍より大きい値では勾配が小さいことから，中央値の 2 倍以上のデータは少ないことが分かる．また，文献[58]-[59]より，Web 転送待ち時間が 6 秒以上では MOS も落ち込んでいることから，個別の Web サイトの待ち時間として中央値の 2 倍以上を遅

延ありとして扱うことには妥当性がある。そして、遅延ありの Web サイトが1つではなく複数含まれるということは、Web サーバ要因の劣化は排除されており、かつユーザがスループット/RTT に不満を感じる状態である。これは、文献 [34] が指摘した、複数サイトの転送待ち時間が遅延して初めて QoE が低下するという知見にも符合する。この捉え方を具体化するため、1つのセットの中に劣化度 x が大きい Web サイトが 19 サイト中 2 割 (3 サイト) 以上のとき、そのセットを Web 転送待ち時間が遅延したセットとして扱うこととした。

4.4.4 スループット/RTT の定量化手法

(ア) スループット/RTT の定量化手法を説明する。瞬間的な品質劣化を捉えるために、セットごとに、19 回分の小容量の転送におけるスループット (転送量 2MB ÷ 転送時間) 計測結果の平均値と、20 秒おきに 21pps × 15 回分の RTT 計測結果の平均値を計算する。

4.4.5 判定器の作成

判定器を作成する手法を説明する。Web 転送待ち時間計測からスループット/RTT 計測までを 1 セットとして、セットごとに Web 転送待ち時間遅延の有無と、スループット/RTT の平均値の関係を SVM で学習し、判定器を作成した。計測したデータの半数を学習用、残りの半数を評価用データとして用いた。SVM ではカーネル関数を選択する必要があるが、今回は RBF (Gaussian) kernel を使用した。表 10 に、セットごとの Web 転送待ち時間遅延有無とスループット/RTT の平均値を並べたデータの例を示す。遅延有無という目的変数に対し 2 つの特徴量の候補を用意した。特徴量が多いほど判定精度がよいとは限らないため、スループット

/RTT の両方を特徴量とする場合，片方のみを特徴量とする場合とで SVM を実行し，判定精度を比較する．

表 10 入力データの例

セット	Web 転送待ち時間(1:遅延)	スループット平均値(Mbps)	RTT 平均値(ms)
0	0	7.6	0.23
1	1	1.8	0.61
...
655	0	3.1	0.40

手順としては以下のようになる．

- i. 特徴量としてスループット/RTT それぞれの平均値を採用し，ハイパーパラメータを変動させて判定器を作成する． C と γ の組み合わせの数だけ判定器が作成される．その中で，G-Mean が最大となる判定器を選出する．
- ii. 特徴量としてスループットの平均値を採用し，ハイパーパラメータを変動させて判定器を作成する．その中で，G-Mean が最大となる判定器を選出する．
- iii. 特徴量として RTT の平均値を採用し，ハイパーパラメータを変動させて判定器を作成する．その中で，G-Mean が最大となる判定器を選出する．

ハイパーパラメータの範囲は以下とする ($26 \times 31 = 806$ 通り)．

・ C : $2^{-10}, 2^{-9}, \dots, 2^{14}, 2^{15}$

・ γ : $2^{-20}, 2^{-19}, \dots, 2^9, 2^{10}$

上記の手順により，提案手法 (i), (ii), (iii) の判定器を作成する．

4.5 評価実験

計測結果のうち、Web 転送待ち時間の遅延ありのセットは 127 セット (19%) であった。まず (i) の判定器を作成した。18 日分のデータを学習データとテストデータにそれぞれ 9 日ずつ、8 日ずつ、...3 日ずつと分割し、最も G-Mean が高くなる判定器を採用した。学習データとテストデータは可能な限り連続するデータを採用した。(i) で 9 月上旬の 4 日間を学習データ、同月中～下旬の 4 日間をテストデータとしたときの判定器が、G-Mean が最大となり選出された。(ii), (iii) の判定器は、同じデータの粒度で作成した。表 11 に各判定器のハイパーパラメータと、テストデータに適用したときの Accuracy, G-Mean, AUC を示す。表 5 より、(i) と (iii) が(ii) よりも Accuracy, G-Mean, AUC が高いことを確認した。また、(i) は (iii) よりも Accuracy が 0.01 高い一方、AUC が 0.06 低く両者は同程度の判定精度と見ることが出来る。さらに、表 12～表 14 に(i)～(iii) の混同行列をそれぞれ示す。これらの結果より、遅延なしについては、(i)～(iii) とともに遅延なしの検出数を示す TP が 104 以上であり、すべての遅延なし ($TP + FN = 112$) の 8 割以上と高い精度で検出できた。一方、遅延ありについては、遅延ありの検出数を示す TN が (i), (iii) が 20 で同数だが (ii) は 15 と低く、すべての遅延あり ($FP + TN = 30$) のうち検出できた比率が 5 割～約 7 割と差が開いた。

表 11 各判定器のパラメータと判定精度

パラメータ/判定精度 特徴量	C	γ	Accuracy	G-Mean	AUC
(i) スループット/RTT	2^4	2^{-10}	0.78	0.83	0.81
(ii) スループット	2^{14}	2^7	0.67	0.58	0.49
(iii) RTT	2^5	2^{-10}	0.79	0.83	0.81

表 12 (i)の混同行列

	predicted positives	predicted negatives
real positives	TP: 112	FN: 10
real negatives	FP: 10	TN: 20

表 13 (ii)の混同行列

	predicted positives	predicted negatives
real positives	TP: 104	FN: 18
real negatives	FP: 15	TN: 15

表 14 (iii)の混同行列

	predicted positives	predicted negatives
real positives	TP: 112	FN: 10
real negatives	FP: 10	TN: 20

上記より、判定精度の観点では (i) と (iii) のいずれかを採用すべきであるが、計測コストの観点を加えると (iii) のほうが有利となる。(i) に必要なスループット計測は RTT 計測と比較して計測時間、通信量の観点で計測コストが高くなるためである。

なお、今回の実験結果では特徴量を RTT の平均値のみとする (iii) を採用できるが、場所等によって適切な特徴量が変動する可能性があるため、判定器を作成する都度、特徴量を選択する必要がある。

4.6 議論

まず、特徴量について議論する。瞬間的な品質劣化を捉えるために、スループットは転送量を小さく回数を多くしたが特徴量として採用されなかった。これは、スループットでは転送量を小さくしてもなお瞬間的な劣化を捉えられなかったと言える。一方、RTT を間隔に留意して多数回計測しその平均値を用いることで、より高精度に QoE 低下を検出できた。この計測手法については、本論文の提案手法に則って QoE 判定器を作成する場合に一般的に適用できると考えられる。すなわち、今後、瞬間的な通信品質の劣化を捉えるにはスループット計測は不要であり、間隔に着目した RTT 計測を行う必要があるということになる。

次に、提案手法の適用範囲について議論する。今回は Web ブラウジングを対象としており、QoE 推定モデルはサービスごとに異なることから[9]-[11]、対象サービスは Web ブラウジングに限定される。しかし、文献[57], [66], [72]で、以下によっては Web ブラウジングの QoE 推定モデルが異なることが明らかにされていることから、この範囲において提案手法は汎用性があると考えられる。

- 端末種別 (スマートフォン/タブレット)
- OS (Android/iOS 等)

- アプリケーション (Chrome/Safari 等)
- コンテキスト

特に、アプリケーションを問わないということは、ブラウザというカテゴリでリリースされているアプリケーションでなくとも、HTML レンダリングエンジンを活用しているニュースアプリケーションなどの QoE も推定できるものと考えられる。また、Web 転送待ち時間からは端末上での描画処理を除いていることから、HTML レンダリングエンジンの違いが与える影響もほぼなく、普遍性があると考えられる。

判定器の地理的な適用領域については、基地局ごとに通信環境が異なることから、判定器は基地局ごとに作成が必要であることが想定される。しかし、判定器作成時と異なる場所で QoE を判定できるかどうかについては今後の検討が必要である。

判定器の更新頻度については、今後の検討が必要である。判定器に影響する要因を明確にできないため、基地局単位で、必要なデータセットが揃う都度判定器の更新が行われるように、判定器の更新作業を自動化することが望ましいと考えられる。

また、通信に関わる技術進歩に影響を受けて判定器の作り直しが必要になることが想定される。たとえば、通信品質劣化時の通信セッションの切断と再接続が圧倒的に高速になれば、RTT の瞬間的な遅延の QoE への影響度合いが変わると考えられる。

しかし、間隔を空けて RTT を計測することで通信品質の瞬間的な劣化を捉えられることには普遍性があり、特徴量と目的変数の計測手法は長期にわたって通用することが期待できる。

今後、瞬間的な通信品質の劣化を捉えることの重要性はますます高まっていくものと予想される。映像配信サービスも VoD からライブストリーミング、リアル

タイムストリーミングと、よりリアルタイム性が高くクライアントのバッファが小さい配信形態も普及すると、瞬間的な劣化が QoE に与える影響が大きくなると考えられる。映像配信サービス以外でも、IoT 機器の遠隔制御や車々間通信など、広帯域大容量通信は必要としないが瞬間的な劣化が QoE に与える大きいサービスの普及が予想される。そのため、瞬間的な劣化を捉える手法の重要性は増大していくものと考えられる。

4.7 結論

4章では、簡易に計測可能なスループット/RTT の分布を特徴量として、Web 転送待ち時間の分布を遅延有無の 2 値で判定する手法を提案し、その有効性を示した。

第 1 の課題に対応し、瞬間的な劣化を捉えるためにスループット計測の転送量を 2 MB と小さくした上、計測を 19 回繰り返してその平均値を特徴量候補としたところ、この特徴量候補は使用する必要がないという実験結果となった。これは、2MB と小さな転送量であっても瞬間的な品質劣化は埋没してしまい、瞬間的な劣化を捉えられなかったためと考えられる。

第 2 の課題に対応し、RTT は間隔を空けて $21 \text{ pps} \times 15 \text{ 回} = 315 \text{ 回}$ もの多数回を計測し、その平均値を特徴量候補としたことで、QoE 低下の有無を判定できた。これは、RTT という単純な値であっても、間隔を空けて平均値を取ることで瞬間的な劣化を捉え、どの Web サイトを表示しても遅いと感じるようなアクセスネットワークの劣化状況と対応することを示していると考えられる。

第 1, 第 2 の課題への対応結果を併せると、今後、本論文の提案手法に基づいて QoE 判定器を作成する場合には、特徴量候補としては多数回の RTT のみを用いればよく、スループットを計測する必要はない。

第3の課題に対応し、複数のWebサイトの転送待ち時間が著しく遅延しているときをQoE低下ありと定義したことで、RTT平均値からのQoE低下有無の判定が可能になった。この評価メトリックでは、Webサイトのコンテンツや構造に依存しない劣化事象を扱っており、サーバ側の要因ではなくアクセスネットワークの一時的な輻輳を原因とする遅延に絞って判定できていることになる。

5章では、より実用性を高めるために管理運用のための設計について議論する。

5 管理運用のための設計

本論文の提案手法は，判定器を作成するための教師データの必要量と，判定器の作成が必要となる頻度によって管理運用の容易性が大きく異なる．判定精度と計測量のバランスを取ることで，計測による計測者や通信事業者設備への負担を軽減できる．4章の実験では，判定器の作成コストは4時間 × 4日間の計測期間を要する．また，計測に必要な通信量は表8と表9より，4日間で約2.2GB/台である．これを，管理運用のためにどこまで削減できるかを論ずる．

5.1節では，特徴量であるスループット/RTT計測の計測回数や転送量を増減し判定精度への影響を評価する．5.2節では，目的変数であるWeb転送待ち時間の計測対象となるWebサイトを変化させ判定精度への影響を評価する．5.3節では，4章と5章の結果を踏まえて管理運用の設計について議論し，5.4にて結論を述べる．

5.1 特徴量

4章の実験では，RTTを21pps × 15回 = 315pps分のデータを計測していた．通信品質の分布の偏りを捉えるために必要なRTTの計測量を確認するため，このデータの一部を教師データとしたときの判定精度を検証する．計測量が減ったときの判定精度の低下幅が小さければ，より小さな計測負荷でWeb転送待ち時間の遅延を判定できることになる．

まず，図7にRTTを21pps × 1~15回分としたときの判定精度を示す．橙色がG-Mean，青色がAccuracy，灰色がAUCを示す．図7において横軸が3以下のときはG-MeanとAUCのいずれかが0.5以下と低く，横軸が4以上でいずれの指標

も 0.8 程度で安定することから、21pps×4 回で十分な精度を得られると判断した。

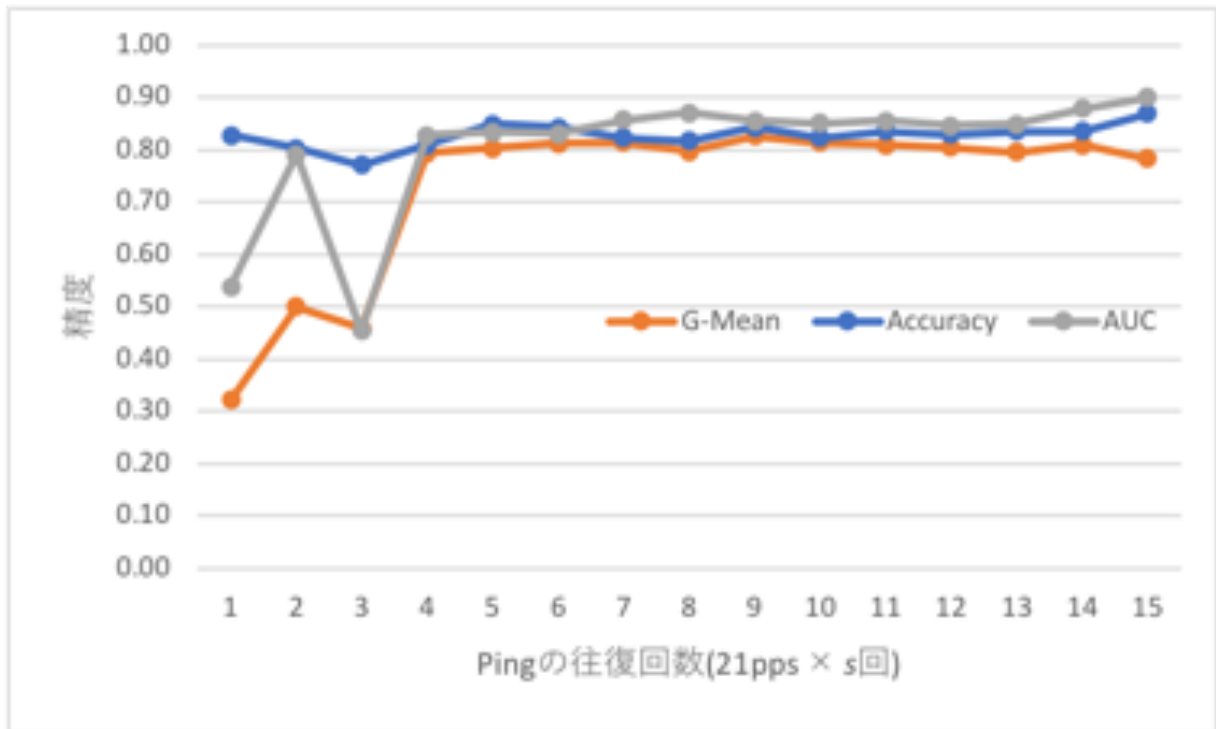


図 7 1 セット内に含まれる Ping の往復回数と判定精度の関係 (21pps×s 回)

次に、図 8 に RTT を 1~21 pps × 4 回としたときの判定精度を示す。このとき、単純に計測した順番にデータを採用する場合に加えて、間隔を変えてデータを採用した場合について調査を行った。図の、間隔 0 と示される折れ線は左から、計測した順番にデータを 1 つ採用した場合、2 つ採用した場合の判定精度をそれぞれ示している。間隔 1 と示される折れ線は左から、1 回目と 3 回目の計測データを採用した場合、1 回目と 3 回目、5 回目の計測データを採用した場合の判定精度を示している。これによって、RTT の計測間隔をどの程度設けることで品質の瞬間的な劣化を捉えやすいのかを検証している。図では判定精度として G-Mean を示している。G-Mean は 0.64 から 0.82 の間で変動している。データ個数が少ないときでも判定精度が高い場合が多いことから、データ個数が少なくても計測の間隔が

離れていれば同じセットの通信品質の瞬間的な劣化を捉えられていると考えられる。表 15 に、データ個数を 2 pps 以下としたときの判定精度を、G-Mean の降順で示す。1 番目と 10 番目以降のデータを採用すれば G-Mean と AUC が 0.75 以上、Accuracy が 0.8 以上となっている。RTT はおよそ 1 秒に 1 回計測していることから、10 秒間隔で RTT を 2 回計測し、それを 4 度繰り返せば十分な判定精度になるといえる。

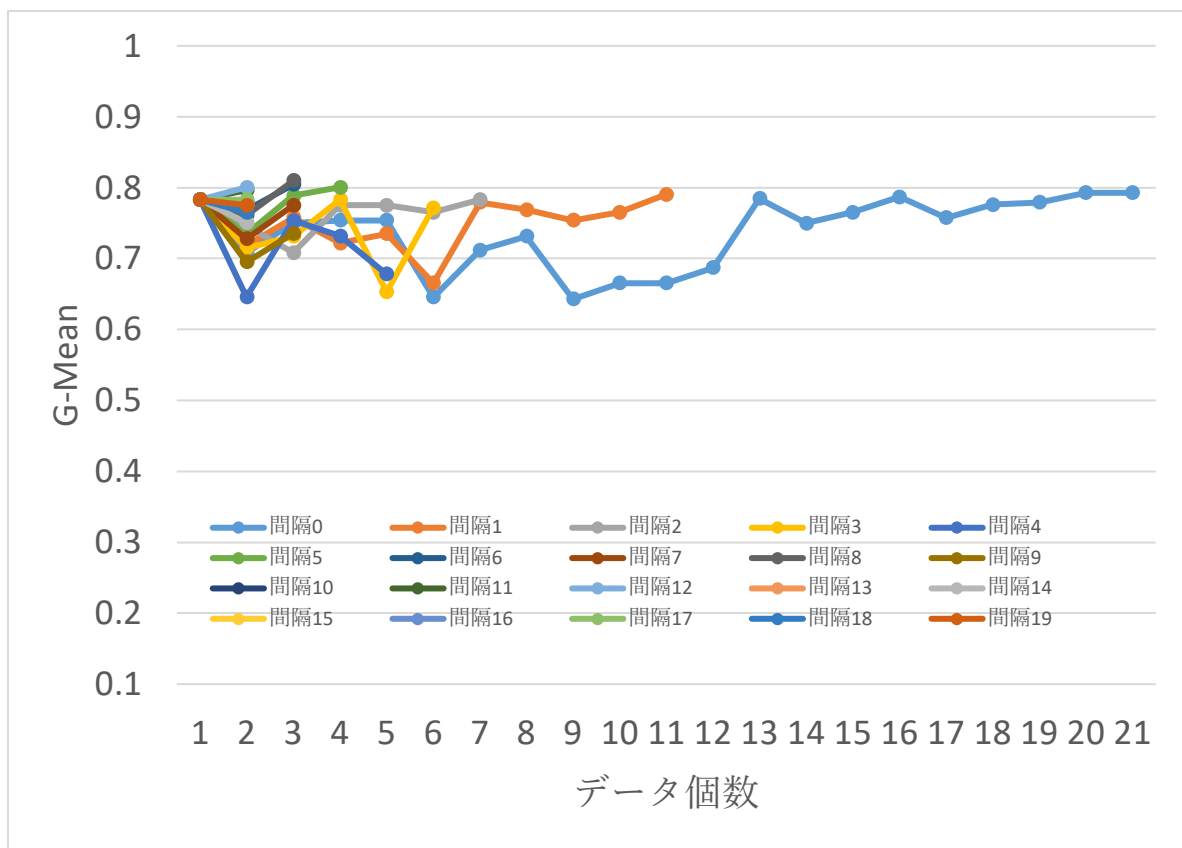


図 8 1 セット内に含まれる Ping の往復回数と判定精度の関係 (t pps×4 回)

表 15 2 pps 以下における計測順と判定精度の関係

学習データ (計測順)	G-Mean	Accuracy	AUC
1, 14	0.80	0.82	0.82
1, 13	0.80	0.82	0.79
1	0.78	0.82	0.78
1, 19	0.78	0.82	0.81
1, 15	0.78	0.81	0.80
1, 18	0.78	0.81	0.79
1, 12	0.78	0.80	0.83
1, 17	0.78	0.80	0.81
1, 21	0.78	0.80	0.77
1, 8	0.77	0.84	0.81
1, 20	0.77	0.81	0.77
1, 10	0.76	0.80	0.78
1, 16	0.75	0.81	0.80
1, 4	0.75	0.80	0.78
1, 7	0.73	0.81	0.80
1, 9	0.73	0.80	0.82
1, 3	0.72	0.81	0.77
1, 5	0.72	0.80	0.78
1, 2	0.71	0.79	0.78
1, 11	0.70	0.80	0.80
1, 6	0.65	0.81	0.87

5.2 目的変数

続いて、訓練データの収集において計測すべき Web サイトについて考察するための分析を行った。前述のとおり、判定器が特定の Web サイトに依存しないように複数サイトを対象とすべきである。その上で、どのような Web サイトを何サイト程度採用すべきか調査した。4 章では 19 サイトの Web 転送待ち時間の分布の偏りから各セットを Web 転送待ち時間の劣化あり/なしに分類し、判定器を作成していた。本項では計測対象となる Web サイトを減じたときの判定精度への影響を確認する。

図 9 に、訓練に用いる Web サイトの数と種類を変化させたときの判定精度を示す。4 サイト以上で劣化度 x が 2 以上の場合に遅延ありセットと定めることから、計測対象が 8 サイト以下になると遅延ありセットが存在しない場合があるため、9 サイトを下限とした。Accuracy, G-Mean, AUC のいずれも同様の傾向を示したため、G-Mean に着目した。Web 転送待ち時間のばらつきを捉えるため、Web 転送待ち時間の分散または中央値が大きいもの/中間のもの/小さいものから順に一定数の Web サイトを選んで精度との関係を調査した。図 9 より、中央値が大きい (橙の実線) 順に選出した場合、9 サイトでも G-Mean が 0.8 近く、サイト数が増えると精度が下がる傾向にある。9~12 サイトであれば中央値が大きい順に選出した場合、最も G-Mean が高くなる。逆に、中央値が小さい (青の実線) 順に選出した場合、最も G-Mean が低くなる。

ここで、Web サイトの選定方法について考察する。今回の計測結果においては、計測負荷を抑える観点では、分散が中程度の 5 つの Web サイトを対象に訓練データを収集すればよいと言える。なお、分散が大きい Web サイトから選出する場合 (青色)、1 サイトでも G-Mean が 0.8 近い一方、Web サイト数が増えても G-Mean が高まるとは限らず不安定であり、分散が小さい Web サイトから選出する

場合 (灰色) は 11 サイト以下のときに G-Mean が低いため、分散が大きいもの/中間のもの/小さいものが偏りなく選出したほうが、精度が良い。

さらに、中央値が小さい Web サイトから選出した場合に G-Mean が低くなる原因について考察する。中央値が小さいということは、Web 転送量が少なく、それゆえスループット/RTT の影響が小さいと考えられる。各 Web サイトの中央値と通信量の相関係数は 0.71 であり、この考察を支持する。本論文の目的が、通信品質劣化による Web 転送待ち時間の遅延の発生を捉えることであることに鑑みれば、極端に Web 転送量が小さい Web サイトを訓練データとして選定することを避けることが望ましい。図 9 が示すように、中央値が大きい Web サイトを計測対象とすることで 9 サイトでも高い精度を得られることから、計測負荷削減にもつながる。

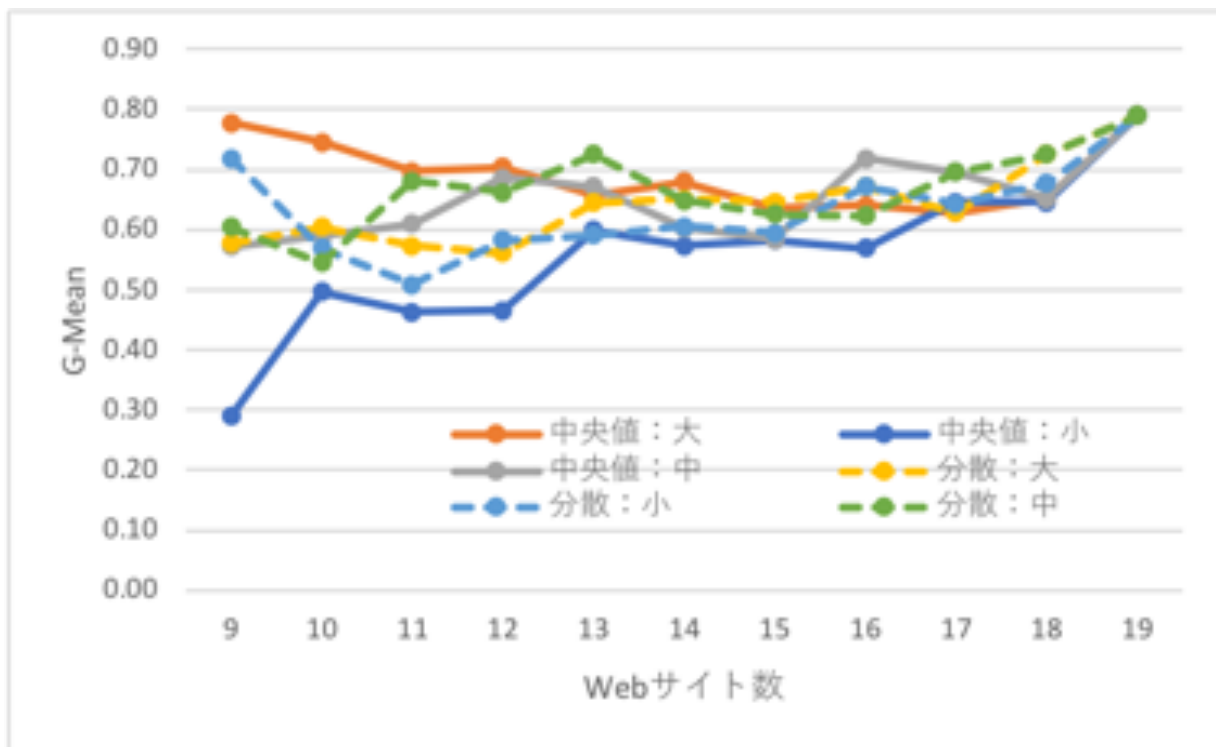


図 9 訓練用 Web サイト数の判定精度への影響

5.3 議論

まず，計測負荷を最小化した場合の判定器の作成コストをシミュレーションする．表 16 にその結果を示す．4 章の実験では約 2.2GB/台であった計測負荷は，5 章の結果から約 0.6 GB まで減少し，約 3 割まで負担を減らすことができた．学習フェーズに必要な教師データの計測の負荷が下がったことで，判定器の更新にかかるコストが低減できた．同時に，判定フェーズに必要な特徴量の計測についても，4 章の実験と比較して計測回数観点で 97%減とごく少数の RTT を計測すればよい．このことから，さまざまな場所・時刻のデータを広く収集するという目的に合致した判定器を作成することができたと言える．

表 16 通信量の最適化効果シミュレーションの結果

計測項目	4 章実験の通信量 (スマートフォン 1 台あたり)	最小化した通信量 (スマートフォン 1 台あたり)
Web 転送 待ち時間	約 304 MB (約 2 MB×19 サイト×8 セット)	約 144 MB (約 2 MB×9 サイト×8 セット)
スループット	約 240 MB (2 MB×15 回×8 セット)	0 MB
RTT	約 0.2 MB (64 bytes×21 パケット×15 回×8 セット)	約 0.004 MB (64 bytes×2 パケット×4 回×8 セット)
合計	約 544 MB (4 日間で約 2.2 GB)	約 144 MB (4 日間で約 0.6 GB)

ここからは、今後の管理運用のための設計を見据え、本論文で得られた知見の適用領域について議論する。

まず、計測量について議論する。5章では、特徴量はRTTを10秒間隔で2pps×4回、目的変数は中央値が大きい9サイトでQoE低下有無を高精度に判定できた。この計測量で場所・時刻を問わず高精度な判定器が作成できるかについては今後の確認が必要になる。計測量を削減すれば判定精度が低下する可能性があるため、今回の結果を以て計測量を確定してよいことにはならない。しかし、実験では品質劣化が発生しやすい通勤時間帯の混雑駅で学習しており、同様の条件では同程度の計測量で通用すると予想される。また、混雑していない場所・時刻においても、表14より、劣化ありだけでなく劣化なしを9割以上(112/122)正しく判定できていることから、同様に計測量も共通すると予想される。

続いて、判定器の作成粒度について考察する。4.6節でも取り上げたように、判定器に影響を与える要因の変動を予見・制御できない以上、判定器は長期間使い回すのではなく、オンラインで常時更新、比較検討されるべきものである[9]、[67]。特に、基地局の位置変更やモバイル通信/Web技術の進展(5G[85]やAMP(Accelerated Mobile Pages)[86]、QUIC[87]など)が判定器にいかなる影響を与えるかは予見できない。この考えに立てば、以下に示すような管理運用を行い、適宜最適な判定器が選択される状態を作り上げることが望ましい。

さらに、4章と5章の知見を総括して実運用における管理運用手法の例を示す。

以下の機能を実現することで管理運用コストを抑えつつ、通信品質の劣化を原因としてQoEが低下している場所、時刻を調査できる。

- 1 アクセス数の多いWebサイトについて、アクセスネットワークを含めて劣化要因がないと考えられる環境でその転送待ち時間の中央値を評価し、中央値が大きいWebサイトを計測対象として適するWebサイトと

して認定する。Webサイトの構造やコンテンツは変動し続けるため、上記の手順は自動化し、常時計測対象サイトが更新される仕組みを作ることが望ましい。

- 2 以下の機能を具備する Web ブラウザアプリケーションを配布する。アプリケーションのインストール時に、データ収集に関するユーザからの同意を得る。
 - 2.1 バックグラウンドで、専用サーバに対して 10 秒間隔で Ping を発出し、RTT を計測し続ける。計測データは順次データベースへ送信する。
 - 2.2 1で認定された Web サイトが表示されたとき、Navigation Timing API を用いて Web 転送待ち時間を計測する。計測データは順次データベースへ送信する。プライバシーを極力保護する観点から、各 Web サイトを示す URL すべては収集せずに、各サイトのトップページのデータのみを収集する。
 - 2.3 Web ブラウザアプリケーション利用中の位置情報を記録し、順次データベースへ送信する。
- 3 ユーザが同一地点に長時間留まった場合に、その場所・時刻の RTT と Web 転送待ち時間のデータが収集できる。同一地点かどうかの判断方法として、緯度経度が一定以内の変化であるかを確認するか、スマートフォンが接続しているアンテナが変わったかを確認するなどが考えられる。Web 転送待ち時間のデータを用いて判定器を作成するには、9 サイト程度のデータが必要となるため、判定器を作成できるほどデータ密度が高まることは多くはないことが予想されるが、人口密集地域で 1 ヶ月に 1 度程度の頻度で実現されればよい。一方、RTT は約 5 分で判定フェーズのみ学習データとしては十分な量が計測できる。日常的には、このデータを用いて判定を行えばよい。

- 4 劣化ありと判定される頻度が高い場所について設備増強計画の優先度を高めるなどが考えられる。さらに、設備を増強したあとに、その場所で Web 転送待ち時間の劣化が発生する頻度を追跡調査することで、設備増強のコスト対パフォーマンスを正確に把握することができる。

5.4 結論

5章では、学習フェーズにおける教師データ及び判定フェーズにおける入力データの計測負荷と判定精度の関係について議論した。その結果、学習フェーズにおいては計測量の観点で4章での実験の3割のデータで十分な判定精度を保てることが明らかとなった。同時に、判定フェーズにおいては計測回数の観点で4章での実験の3%のデータで十分判定できることが明らかとなった。

さらに、ここまでの知見を総合し、実運用の手法について議論した。提案手法においてフローが定義されており、人間による主観的な判断を必要としないことから、判定器の作成・更新や判定器へのデータ入力、判定結果に基づくアラートの生成までを自動化できる。すなわち、提案手法を用いることで全国規模での常時監視が可能であることを示した。従前はユーザから QoE の低下が報告された場所についてドライブテストなどを行う必要があった。この方法では、そもそもユーザにとって申告稼働が面倒なため QoE が低下していても報告されることなく不満だけを募らせ解約に至ることも考えられる。さらに、QoE 低下の報告を受けることができたとしても、ドライブテストなどの調査稼働が過大であることや、調査時に必ずしも QoE 低下事象が再現されるわけではないこと、アクセスネットワークの一時的な混雑を原因としない場合も含まれることなど、多数の要因によって十分なフォローができるとは限らなかった。提案手法によって全国規模で常時監視できれば、上記のような課題を解決することができる。

6 結論

6.1 まとめ

これまで、パケットデータと機械学習を用いて、スループット/RTTとQoEの関係を学習する手法が確立されてきた[8], [22]-[32]. パケットデータにはスループット/RTTの瞬間的な劣化に関する情報が含まれているため[8], Webブラウジングなどの瞬間的な劣化がQoEに強く影響するサービスにおいても有効であった[27]-[32].

一方, さまざまな場所・時刻でモバイル通信品質を調査するためには, パケットデータを用いない, 単純なスループット/RTT計測結果からQoEを推定する手法が必要とされている[3]-[6]. 従来は定常状態のスループット/RTTを計測していたが, これでは瞬間的な劣化を捉えられず, WebブラウジングなどのQoEを判定することはできない[33], [60].

そこで, 過渡状態のスループット/RTTをさまざまな場所・時刻で調査する手法を提案した. その結果, RTTを間隔に留意して複数回計測し, QoEとの関係について機械学習を適用することによって, WebブラウジングのQoEを判定できることが分かった.

6.2 貢献

本論文の貢献は以下のとおりである.

まず, ユーザ動態を調査してWeb転送待ち時間の著しい劣化がモバイル通信品質に対するQoEを低下させる主要因であることを明らかにした. また, 通勤ラッシュや昼のオフィス街など人の移動によってアクセスネットワークが一時的に混雑したことが原因と考えられる事象が多いことを明らかにした.

次に、簡易に計測可能なスループット/RTTの分布を特徴量として Web 転送待ち時間の分布を遅延有無の 2 値で判定する手法を提案した。スループット/RTT, Web 転送待ち時間のいずれについても、瞬間的な品質劣化を捉えることが重要であることに着目し、スループットは小容量の転送の反復、RTT は間隔を空けての計測の反復、Web 転送待ち時間はサイトごとに平常時の待ち時間からの乖離度によって捉えた。その結果、RTT のみで高精度に QoE 低下セットを検出できることを明らかにした。

続いて、その管理運用手法について議論した。提案手法により、特定の Web サイトのコンテンツや構造に依存せずに Web 転送待ち時間の劣化有無を推定することができる。また、目的変数を複数サイトの乖離度から分類される QoE 低下の有無として捉えることによって、説明変数は RTT という単純な指標で推定できる。それもわずか 8 パケット分の RTT で十分な判定精度を得られることを示した。さらに、訓練データの収集において計測すべき Web サイトについて考察し、極端に Web 転送量が小さい Web サイトを避けることが訓練データ計測の負荷削減及び判定精度向上の観点で有効である可能性を示した。

ここで、提案手法の普遍性についてまとめる。RTT の瞬間的な品質劣化を捉えるための計測手法と、Web 転送待ち時間の乖離度による QoE 低下有無の分類というフレームワークは普遍性がある。ただし、RTT の計測間隔や、5 章で議論した RTT の計測回数、Web サイト数などの計測データ量の削減結果については、場所などの影響を受けるかどうかについて今後の研究が必要である。

6.3 今後の展望

本論文では、スループット/RTT 専用の計測結果を入力とすることを前提に議論した。しかし、間隔を空けて計測した RTT のみから高精度に QoE 低下を検知できたことから、基地局などが収集している RTT のパッシブ計測データから QoE 低下

を推定できる技術を検討する必要性が明らかとなった。そもそも QoE 判定のための計測を必要とせず、基地局単位で常時計測している RTT を間隔に留意してサンプリングし、予め用意した判定器に入力することで、モバイル通信サービスのカバーエリア全域における QoE 低下を検出できる可能性が生まれた。

今後は、フィールドで調査した QoE 低下セットを正解データとして、これを通信事業者のパッシブ計測データから検知できるかを調査する予定である。これが期待どおり検知可能であれば、人口過密地域で一般ユーザのスマートフォンから、Web 転送待ち時間のデータとそのとき接続していた基地局の情報だけを自動で収集し続けられれば、後は該当セットにおける TCP などの通信中の RTT のパッシブ計測データについて間隔を変えて特徴量候補として機械学習を適用して判定器を常時アップデートできる。その判定器に、カバーエリア全域の RTT データを入力し続け、QoE 低下セットと判定された基地局のカバーエリアについて対策を行うなど、非常に効率的な品質向上施策の計画、実行及びその効果の評価が実現できる。

謝辞

長年にわたってご指導ご鞭撻いただいた指導教員の大坐畠智准教授に心より感謝いたします。社会人であることをはじめ様々な要因によって格別のご負担をおかけいたしました。懇切丁寧なご助言と温かいご支援によって、研究を完成させることができました。また、大坐畠先生とともに本研究について討論いただき、ご助言を頂戴した策力木格准教授に感謝いたします。そして、ご多忙のなか時間を割いて審査を引き受けてくださった電気通信大学の須賀昭彦教授、吉永努教授、小川朋宏准教授、山本嶺准教授には、研究の本質に関わるご指摘を賜った上、その内容について深く議論させていただいたことで成長することができました。ここに厚く感謝の意を表します。

また、日本電信電話株式会社にて研究のご指導をいただいた各位にお礼申し上げます。特に、高橋玲氏、野尻秀樹氏、岡本淳氏、川口銀河氏、小川秀貴氏には多大なご協力とご支援をいただきました。この場を借りて、お礼申し上げます。

最後に、家族に感謝します。妻 友香は、多くを引き受けて時間を与えてくれるだけでなく、ときに誰より厳しくご鞭撻くださり、そのおかげで成し遂げることができました。子 綜真は、知見を世に残す取り組みに対して大きな動機を与えてくれました。常に道を明るく照らしてくれた2人に、改めて感謝いたします。

参考文献

- [1] CISCO, "Cisco visual networking index: Forecast and methodology, 2015-2020," June 2016.
- [2] 原田 篤, 井上 祐樹, 栗田 大輔, 小原 辰徳, "世界主要ベンダとの 5G 伝送実験," NTT ドコモ テクニカル・ジャーナル, Vol. 23, No.4, pp. 59-68, Jan. 2016.
- [3] Federal Communications Commission, "Measuring Broadband America," Feb. 2021. (<https://www.fcc.gov/general/measuring-broadband-america>)
- [4] Q. Plessis, M. Suzuki, T. Kitahara, and S. Ano, "Group Mobility in Mobile Networks: Signaling Based Detection and Network Utilization Modeling," Proc. of Global Communications Conference (GLOBECOM), pp.1-7, Dec. 2016.
- [5] J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z. M. Mao, S. Sen, and O. Spatscheck, "An In-depth Study of LTE: Effect of Network Protocol and Application Behavior on Performance," Proc. of SIGCOMM Computer Communication Review, Vol. 43, Issue 4, pp.363-374, Oct. 2013.
- [6] Ookla, "Location Performance." (<http://www.speedtest.net/location-performance>)
- [7] P. Casas, A. Sackl, R. Schatz, L. Janowski, J. Turk and R. Irmer, "On the quest for new KPIs in mobile networks: The impact of throughput fluctuations on QoE," 2015 IEEE International Conference on Communication Workshop (ICCW), 2015, pp. 1705-1710, doi: 10.1109/ICCW.2015.7247426.
- [8] F. Kaup, F. Fischer and D. Hausheer, "Measuring and predicting cellular network quality on trains," 2017 International Conference on Networked Systems (NetSys), 2017, pp. 1-8, doi: 10.1109/NetSys.2017.7903960.
- [9] S. Aroussi and A. Mellouk, "Survey on machine learning-based QoE-QoS correlation models," 2014 International Conference on Computing, Management and Telecommunications (ComManTel), 2014, pp. 200-204, doi: 10.1109/ComManTel.2014.6825604.
- [10] M. Seufert, N. Wehner, V. Wieser, P. Casas and G. Capdehourat, "Mind the (QoE) Gap: On the Incompatibility of Web and Video QoE Models in the Wild," 2020 16th

- International Conference on Network and Service Management (CNSM), 2020, pp. 1-5, doi: 10.23919/CNSM50824.2020.9269041.
- [11]ITU-T Recommendation E.811, "Quality measurement in major events," Mar. 2017.
- [12]ITU-T Recommendation G.107, "The E-model: a computational model for use in transmission planning," June 2015.
- [13]ITU-T Recommendation P.1203, "Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport," Oct. 2017.
- [14]ITU-T Recommendation P.1204, "Video quality assessment of streaming services over reliable transport for resolutions up to 4K," Jan. 2020.
- [15]ITU-T Recommendation G.1035, "Influencing factors on quality of experience for virtual reality services," May 2020.
- [16]ITU-T Recommendation P.919, "Subjective test methodologies for 360° video on head-mounted displays," Oct. 2020.
- [17]ITU-T Recommendation G.1032, "Influence factors on gaming quality of experience," Oct. 2017.
- [18]ITU-T Recommendation P.809, "Subjective evaluation methods for gaming quality," June 2018.
- [19]ITU-T Recommendation G.1072, "Opinion model predicting gaming quality of experience for cloud gaming services," Jan. 2020.
- [20]ITU-T Recommendation G.1031 "QoE factors in web-browsing," Feb. 2012.
- [21]H. Koto, N. Fukumoto, S. Niida, H. Yokota, S. Arakawa, and M. Murata, "Users' reaction to network quality during web browsing on smartphones," Proc. of International Teletraffic Congress (ITC), pp.1-9, Oct. 2014.
- [22]P. Casas, M. Seufert, N. Wehner, A. Schwind and F. Wamser, "Enhancing Machine Learning Based QoE Prediction by Ensemble Models," 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), 2018, pp. 1642-1647, doi: 10.1109/ICDCS.2018.00186.

- [23] Irena Orsolíc, Dario Pevec, Mirko Suznjević, and Lea Skorin-Kapov, "A machine learning approach to classifying YouTube QoE based on encrypted network traffic," *Multimedia Tools and Applications*, Volume 76, Issue 21, pp.22267-22301, Nov. 2017.
- [24] Yue, Ting & Wang, Hongbo & Cheng, Shiduan & Shao, Jie. (2019). Deep Learning based QoE Evaluation for Internet Video. *Neurocomputing*. 386. 10.1016/j.neucom.2019.12.082.
- [25] P. Charonyktakis, M. Plakia, I. Tsamardinou and M. Papadopoulou, "On User-Centric Modular QoE Prediction for VoIP Based on Machine-Learning Algorithms," in *IEEE Transactions on Mobile Computing*, vol. 15, no. 6, pp. 1443-1456, 1 June 2016, doi: 10.1109/TMC.2015.2461216.
- [26] M. S. Anwar, J. Wang, W. Khan, A. Ullah, S. Ahmad and Z. Fei, "Subjective QoE of 360-Degree Virtual Reality Videos and Machine Learning Predictions," in *IEEE Access*, vol. 8, pp. 148084-148099, 2020, doi: 10.1109/ACCESS.2020.3015556.
- [27] Trevisan, Martino & Drago, Idilio & Mellia, Marco. (2018). PAIN: A Passive Web performance indicator for ISPs. *Computer Networks*. 149. 10.1016/j.comnet.2018.11.024.
- [28] Hora, Diego & Asrese, Alemnew & Christophides, Vassilis & Teixeira, Renata & Rossi, Dario. (2018). Narrowing the Gap Between QoS Metrics and Web QoE Using Above-the-fold Metrics. 10.1007/978-3-319-76481-8_3.
- [29] Antoine Saverimoutou, Bertrand Mathieu, Sandrine Vaton, "A 6-month analysis of factors impacting web browsing quality for QoE prediction," *Computer Networks*, Volume 164, 106905, Sep. 2019.
- [30] T. Hoßfeld, F. Metzger and D. Rossi, "Speed Index: Relating the Industrial Standard for User Perceived Web Performance to web QoE," 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), 2018, pp. 1-6, doi: 10.1109/QoMEX.2018.8463430.
- [31] Baraković, Sabina & Skorin-Kapov, Lea. (2017). Modelling the relationship between design/performance factors and perceptual features contributing to Quality of

- Experience for mobile Web browsing. *Computers in Human Behavior*. 74. 311-329.
10.1016/j.chb.2017.04.046.
- [32] Casas, Pedro & Wassermann, Sarah & Wehner, Nikolas & Seufert, Michael & Schüler, Joshua & Hossfeld, Tobias. (2021). Mobile Web and App QoE Monitoring for ISPs - from Encrypted Traffic to Speed Index through Machine Learning.
- [33] Hirochika Asai, Kensuke Fukuda, Hiroshi Esaki, "Social benchmarking of QoS & QoE in cellular data networks," IEEE International Communications Quality and Reliability (CQR) Workshop, Marco Island, FL, USA, May 2013.
- [34] Salutari, Flavia & Hora, Diego & Dubuc, Gilles & Rossi, Dario. (2019). A Large-scale Study of Wikipedia Users' Quality of Experience. WWW '19: The World Wide Web Conference. 3194-3200. 10.1145/3308558.3313467.
- [35] A.J. Smola, B. Schölkopf, "A tutorial on support vector regression," *Journal of Statistics and Computing*, Vol. 14, Issue 3, pp.199–222, 2004.
- [36] Deng J., Berg A.C., Li K., Fei-Fei L., "What Does Classifying More Than 10,000 Image Categories Tell Us?," *European Conference on Computer Vision, Lecture Notes in Computer Science*, vol. 6315, pp.71-84, 2010.
- [37] scikit-learn.org, "Choosing the right estimator," (https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html)
- [38] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," Apr. 2008.
- [39] ITU-R Recommendation BT.500-11, "Methodology for the Subjective Assessment of the Quality of Television Pictures," June 2002.
- [40] ITU-R Recommendation BT.1788, "Methodology for the subjective assessment of video quality in multimedia applications," Jan. 2007.
- [41] ITU-T Recommendation J.144, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," Mar. 2001.
- [42] ITU-T Recommendation J.247, "Objective perceptual multimedia video quality measurement in the presence of a full reference," Aug. 2008.

- [43]ITU-T Recommendation J.341, "Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference," Mar. 2016.
- [44]ITU-T Recommendation J.246, "Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference," Aug. 2008.
- [45]ITU-T Recommendation J.249, "Perceptual video quality measurement techniques for digital cable television in the presence of a reduced reference," Jan. 2010.
- [46]ITU-T Recommendation J.342, "Objective multimedia video quality measurement of HDTV for digital cable television in the presence of a reduced reference signal," Apr. 2011.
- [47]ITU-T Recommendation P.1201, "Parametric non-intrusive assessment of audiovisual media streaming quality," Oct. 2012.
- [48]ITU-T Recommendation P.1202, "Parametric non-intrusive bitstream assessment of video media streaming quality," Oct. 2012.
- [49]ITU-T Recommendation G.1070, "Opinion model for video-telephony applications," June 2018.
- [50]ITU-T Recommendation G.1030, "Estimating end-to-end performance in IP networks for data applications," Feb. 2014.
- [51]L. Anegekuh, L. Sun and E. Ifeachor, "A screening methodology for crowdsourcing video QoE evaluation," 2014 IEEE Global Communications Conference, 2014, pp. 1152-1157, doi: 10.1109/GLOCOM.2014.7036964.
- [52]Hossfeld, Tobias & Keimel, Christian & Hirth, Matthias & Gardlo, Bruno & Habigt, Julian & Dieopold, Klaus & Tran-Gia, Phuoc. (2014). Best Practices for QoE Crowdstesting: QoE Assessment With Crowdsourcing. *Multimedia, IEEE Transactions on*. 16. 541-558. 10.1109/TMM.2013.2291663.
- [53]Gardlo, Bruno & Egger-Lampl, Sebastian & Hossfeld, Tobias. (2015). Do Scale-Design and Training Matter for Video QoE Assessments through Crowdsourcing?. 15-20. 10.1145/2810188.2810193.

- [54] Hossfeld, Tobias & Keimel, Christian & Timmerer, Christian. (2014). Crowdsourcing Quality-of-Experience Assessments. *Computer*. 47. 98-102. 10.1109/MC.2014.245.
- [55] M. Varvello, J. Blackburn, D. Naylor, and K. Papagiannaki, "EYEORG: A Platform For Crowdsourcing Web Quality Of Experience Measurements," *Proc. of Conference on emerging Networking EXperiments and Technologies (CoNEXT)*, pp. 399-412, Dec. 2016.
- [56] M. Suznjevic, L. Skorin-Kapov and M. Matijasevic, "The impact of user, system, and context factors on gaming QoE: A case study involving MMORPGs," 2013 12th Annual Workshop on Network and Systems Support for Games (NetGames), 2013, pp. 1-6, doi: 10.1109/NetGames.2013.6820606.
- [57] Asrese, Alemnew & Walelgne, Ermias & Bajpai, Vaibhav & Lutu, Andra & Alay, Ozgu & Ott, Jörg. (2019). Measuring Web Quality of Experience in Cellular Networks.
- [58] Zhang, Xu & Sen, Siddhartha & Kurniawan, Daniar & Gunawi, Haryadi & Jiang, Junchen. (2019). E2E: embracing user heterogeneity to improve quality of experience on the web. *SIGCOMM '19: Proceedings of the ACM Special Interest Group on Data Communication*. 289-302. 10.1145/3341302.3342089.
- [59] Jahromi, Hamed & Delaney, Declan & Hines, Andrew. (2020). Beyond First Impressions: Estimating Quality of Experience for Interactive Web Applications. *IEEE Access*. PP. 1-1. 10.1109/ACCESS.2020.2979385.
- [60] Katarzyna Wac, Selim Ickin, Jin-Hyuk Hong, Lucjan Janowski, Markus Fiedler, and Anind K. Dey, "Studying the experience of mobile applications used in different contexts of daily life," In *Proceedings of the first ACM SIGCOMM workshop on Measurements up the stack (W-MUST '11)*, Association for Computing Machinery, pp. 7-12, Aug. 2011.
- [61] Tang Y, Zhang YQ, Chawla NV, Krasser S., "SVMs Modeling for Highly Imbalanced Classification," *IEEE Trans Syst Man Cybern B Cybern.*, Vol.39 No.1, pp.281-288, Feb. 2009.
- [62] T. Fawcett, "An introduction to ROC analysis," *Journal of Pattern Recognition Letters* Vol. 27 Issue 8, pp. 861-874, June 2006.

- [63] C. M. Bishop, "Pattern Recognition and Machine Learning," Springer Science+Business Media, LLC, 2006.
- [64] Y. Sun, A. Wong, M. Kamel, "Classification of imbalanced data: a review," International Journal of Pattern Recognition and Artificial Intelligence Vol. 23, No. 4, pp.687-719, Nov. 2011.
- [65] Imam, T., Ting, K. M., Kamruzzaman J., "z-SVM: An SVM for Improved Classification of Imbalanced Data," Proc. of AI 2006: Advances in Artificial Intelligence, pp.4-8, Dec. 2006.
- [66] Morrison, Alistair & Xiong, Xiaoyu & Higgs, Matthew & Bell, Marek & Chalmers, Matthew. (2018). A Large-Scale Study of iPhone App Launch Behaviour. 1-13. 10.1145/3173574.3173918.
- [67] Uzun, Mete & Abul, Osman. (2016). End-to-end internet speed analysis of mobile networks with mapReduce. 1-6. 10.1109/ISNCC.2016.7746114.
- [68] The Apache Software Foundation, "Apache Hadoop" (<http://hadoop.apache.org>)
- [69] T. Iwai, A. Nakao, "Adaptive mobile application identification through in-network machine learning," Proc. of 18th Asia-Pacific Network Operations and Management Symposium (APNOMS), pp.1-6 Oct. 2016.
- [70] M. Allman, V. Paxson, E. Blanton, "TCP Congestion Control," RFC 5681, 2009
- [71] V. Jacobson, "Congestion avoidance and control," Proc. ACM SIGCOMM'88, pp.314-329, 1988.
- [72] Hossfeld, Tobias & Egger-Lampl, Sebastian & Schatz, Raimund & Fiedler, Markus & Masuch, Kathrin & Lorentzen, Charlott. (2012). Initial Delay vs. Interruptions: Between the Devil and the Deep Blue Sea. 2012 4th International Workshop on Quality of Multimedia Experience, QoMEX 2012. 10.1109/QoMEX.2012.6263849.
- [73] J. Sommers, P. Barford, "Cell vs. WiFi: On the Performance of Metro Area Mobile Connections," Proc. of Internet Measurement Conference (IMC), pp. 301-314, Nov. 2012.
- [74] Amazon Web Services, Inc., "Amazon Redshift" (<https://aws.amazon.com/jp/redshift>)

- [75] L. Pierucci, "The quality of experience perspective toward 5G technology," *Journal of Wireless Communications*, Vol. 22, Issue 4, pp. 10-16, August 2015.
- [76] M. S. Alias, S. A. M. Rofie, S. M. M. Hassan, A. Shafii, and M. S. B. Ibrahim, "An experimental QoE performance evaluation of HTTP over Multipath TCP," *Proc. of Control System, Computing and Engineering (ICCSCE)*, pp. 320-325, Nov. 2016.
- [77] T. Uhl, J. Klink, and P. Bardowski, "New metric for world wide web service quality," *Journal of Telecommunications and Information Technology*, Vol. 2014, Issue 2, pp.50-58, 2014.
- [78] A. Sackl, S. Egger, and R. Schatz, "The influence of network quality fluctuations on Web QoE," *Proc. of Quality of Multimedia Experience (QoMEX)*, pp. 123-128, Sept. 2014.
- [79] Toshihiko Kato, Xiaofan Yan, Ryo Yamamoto, Satoshi Ohzahata, "A Study on Round-trip Time Estimation from Unidirectional Packet Traces Using Different TCP Congestion Control Algorithms," *International Journal On Advances in Networks and Services*, pp. 1-9 July 2019.
- [80] T. Everts, "Rules for Mobile Performance Optimization," *Communications of the ACM*, Vol. 56, Issue 8, pp. 52-59, Aug. 2013.
- [81] jQuery, "Ajax" (<http://api.jquery.com/category/ajax>)
- [82] T. Tominaga, K. Sato, N. Yoshimura, M. Masuda, H. Aoki, and T. Hayashi, "Web-Browsing QoE Estimation Model," *IEICE Transactions on Communications*, Vol. E100.B, No.10, pp.1837-1845, 2017.
- [83] W3C, "Navigation Timing," Dec. 2012. (<https://www.w3.org/TR/2012/REC-navigation-timing-20121217>)
- [84] A. Gember, A. Akella, J. Pang, A. Varshavsky and R. Caceres, "Obtaining In-Context Measurements of Cellular Network Performance," *Proc. of ACM Internet Measurement Conference*, pp. 287-300, Nov. 2012.
- [85] J. G. Andrews, "What Will 5G Be?," *IEEE Journal on Selected Areas in Communications*, Vol.32, Issue6, pp. 1065-1082, June 2014.
- [86] Google, "AMP(Accelerated Mobile Pages)." (<https://amp.dev>)

[87]Langley, Adam & Iyengar, Janardhan & Bailey, Jeff & Dorfman, Jeremy & Roskind, Jim & Kulik, Joanna & Westin, Patrik & Tenneti, Raman & Shade, Robbie & Hamilton, Ryan & Vasiliev, Victor & Riddoch, Alistair & Chang, Wan-Teh & Shi, Zhongyi & Wilk, Alyssa & Vicente, Antonio & Krasic, Charles & Zhang, Dan & Yang, Fan & Swett, Ian. (2017). The QUIC Transport Protocol: Design and Internet-Scale Deployment. 183-196. 10.1145/3098822.3098842.

関連論文（学位論文を構成する論文）

(1)

全著者名：小林史弥，川口銀河，岡本淳，大坐畠智，小川秀貴，高橋玲

論文題目：モバイル環境における通信品質指標を用いた機械学習による Web 転送遅延判定方式

印刷公表の方法及び時期：2020 年 6 月，電子情報通信学会論文誌 B, Vol.J103-B, No.6

説明：本文第 4 章を中心に関連する，機械学習による Web 転送遅延の判定方式を提案した論文

(2)

全著者名：Fumiya Kobayashi, Ginga Kawaguti, Hideki Nojiri, Akira Takahashi

論文題目：Crowdsourced QoE evaluation of mobile device usage with consideration on user dependence

印刷公表の方法及び時期：2015 年 8 月，17th Asia-Pacific Network Operations and Management Symposium (APNOMS)

説明：本文第 3 章を中心に関連する，モバイル環境における Web 転送遅延の重要性を調査した国際会議

著者略歴

小林 史弥

2005 年 4 月 同志社大学 文化情報学部 文化情報学科 入学

2009 年 3 月 同志社大学 文化情報学部 文化情報学科 卒業

2009 年 4 月 日本電信電話株式会社 入社

2012 年 10 月 電気通信大学 大学院 情報理工学研究科 情報学専攻
博士後期課程 入学

2015 年 7 月 東日本電信電話株式会社へ異動

2020 年 3 月 電気通信大学 大学院 情報理工学研究科 情報学専攻
博士後期課程 満期退学