

## P2P コンテンツ共有におけるユーザの集合知適用に関する一検討 A Contents Selection System by User's Collective Intelligence on P2P Networks

岡田 聡† 大島 浩太‡ 寺田 松昭‡  
Satoshi Okada Kohta Ohshima Matsuaki Terada

### 1. 背景

近年、計算機の高性能化とブロードバンド環境の普及に伴い、拡張性が高く、サーバの負荷を分散可能な P2P (Peer-to-Peer) の技術を用いたさまざまなサービスが誕生した。P2P を利用したサービスとして、Gnutella[1] や Winny[2] などのファイル共有ソフトが多くのユーザに利用されている。

しかし、流通するコンテンツ数の増大と内容の多様化により、ユーザが欲しいコンテンツを取得することが困難になっている。目的のコンテンツを得るまで取得を繰り返す行為は、手間を増大させるだけでなく、通信トラフィックの増加も招いている。また、意図しないコンテンツのダウンロードによる、スパムやウィルスの被害も増加している。これらの問題は、コンテンツの検索時にユーザに対してコンテンツの内容に関する十分な情報が提供されていないことが一つの原因であると考えられる。

本稿では、これらの問題を解決するためのシステムの提案と試作について述べる。

### 2. 提案システム

提案システムは、P2P コンテンツ共有に参加するユーザの集合知を利用することで、流通する各コンテンツを選別することを目的とする (図1)。

本稿では、各ユーザが提供する各コンテンツに対する評価情報の集積を集合知と定義する。集合知の内容はコンテンツの推薦や感想、スパムやウィルスの警告など多岐に渡る。これらのユーザの集合知を検索時に提供することで、P2P ネットワークの利用者に容易にコンテンツの選別をさせることが狙いである。

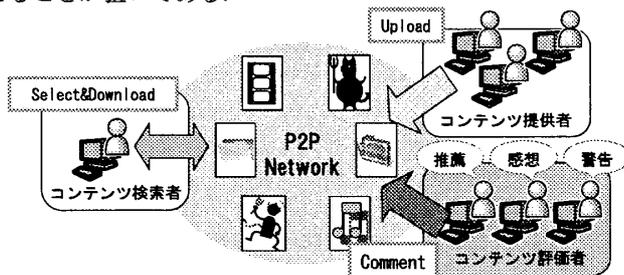


図1 提案システムイメージ図

提案システムでは、あるユーザがアップロードしたコンテンツに対して、その他のユーザが評価としてコメントを行う。その評価は、コンテンツを検索したユーザに提供される。本システムの特長を以下に示す。

#### (1) 第三者によるコンテンツの評価

コンテンツの選別を行うためには、選別の根拠となる多

くの情報が必要である。また、それらの情報はある一個人による主観的なものではなく、客観的なものでなければならない。提案システムでは、第三者である複数のユーザが評価を行うことで、複数の視点からの客観的な評価を形成することができる。

#### (2) コンテンツの詳細情報提供

ユーザが提供する評価はコンテンツが本物か捏造かといった情報だけでなく、音声や画像の品質や具体的な内容にまで及ぶ。そのため、各コンテンツの評価を参照することでコンテンツ同士を比較することが可能となる。例えば、内容が類似したコンテンツを検索する場合に、より高品質なコンテンツの選択や、より目的に合致したコンテンツの選択が可能になる。

### 3. システムの課題

提案システムには以下のような課題がある。

#### (1) コンテンツの選別が可能な検索方式

コンテンツの選別は大きく二つに分けられる。提案システムでは以下の二つの選別を実現する検索方式にする必要がある。

##### a. 悪質なコンテンツ取得の回避

悪質なコンテンツとは、スパムやウィルス感染したコンテンツ、捏造コンテンツを指す。これらは全てのユーザにとってダウンロードを回避すべきコンテンツである。

##### b. ユーザの目的や嗜好にそったコンテンツの選別

コンテンツを検索する際の目的は様ではない。また、コンテンツに対する満足度も各ユーザの嗜好に大きく依存する。そのため、あるユーザにとっては有用なコンテンツでも、別のユーザにとっては回避すべきコンテンツとなる可能性は大いに存在する。

#### (2) コンテンツとその評価の管理方式

提案システムでは、管理するコンテンツ及びコンテンツの評価の数が膨大になると予想される。そのため、拡張性があり、管理コストの低い方式にする必要がある。

また、ユーザの集合知を形成するためには、情報の中立性を確保する必要がある。そのため、提供された各コンテンツの評価は特定の人物によって検閲できない構成にすることが望ましい。

### 4. コンテンツ及び評価の検索

提案システムでは、以下の二つのデータを新たに作成する。

#### (1) メタデータ

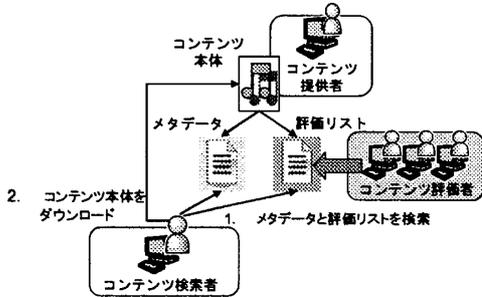
コンテンツの要約情報である。コンテンツ名や拡張子、コンテンツサイズ、作成日時などが含まれる。

#### (2) 評価リスト

コンテンツに対する各ユーザの評価情報が含まれる。

†東京農工大学大学院 工学府 情報工学専攻  
‡東京農工大学大学院 共生科学技術研究院

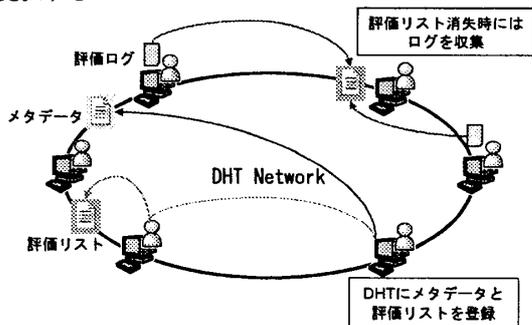
提案システムでは、ユーザが検索する対象をメタデータ、評価リスト、コンテンツ本体の三つに分類する。メタデータと評価リストはアップロード時に各コンテンツ毎に一つずつ作成される。ユーザの評価は評価リストに登録される。検索時にはメタデータと評価リストが提供され、ユーザはこれらをもとに選別を行う(図2)。また、集められた各ユーザの評価は、判断材料を増やすために検索時には全て提供される。



### 5. コンテンツ及び評価の管理

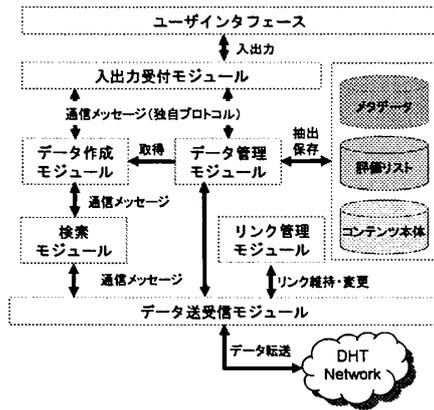
メタデータと評価リストは DHT (Distributed Hash Table) で管理する(図3)。DHTを用いることでネットワークに負荷をかけずに、高速かつ確実に検索することが可能になる。また、管理コストをユーザ間で分散できる。DHTにおけるメタデータと評価リストの配置場所は、ハッシュ関数によりランダムに決定される。メタデータと評価リストの配置場所は各ユーザから隠蔽される。これにより、情報の検閲を防ぐ。DHTのアルゴリズムとして chord[3]を用いる。

また、ノードの故障や DHT 参加・離脱時のデータ転送ミスなどによる評価リストの消失に対応するため、各ノードが評価ログとして自身が登録したコンテンツの評価を保存する。消失時には要求に応じて該当するコンテンツの評価を提供する。

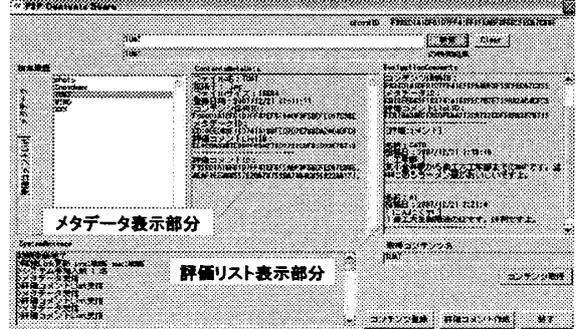


### 6. 実装

提案システムのソフトウェア構成を図4に示す。検索モジュールでは、入力となるデータとハッシュ関数を用いて、メタデータと評価リストの登録場所となるハッシュ値を作成する。使用したアルゴリズムは SHA-1[4]である。また、リンク管理モジュールでは、受信したメッセージから DHT における successor, predecessor, ルーティングテーブル等の管理を行う。データ管理モジュールは、通信メッセージに基づいて、受信したデータの保存や他のユーザからの要求に応じてデータの抽出などを行う。



実行画面では、検索によって取得したメタデータと評価リストが表示される(図5)。取得したデータを利用して、ユーザはコンテンツ本体をダウンロードするかどうか選択することが可能である。また、検索したデータは保存されるため、各コンテンツの評価を比較することも可能である。



### 7. まとめ

本稿では、P2P コンテンツ共有においてユーザの集合知を適用した場合の有効性を述べた。ユーザが各コンテンツを評価し、それらをユーザ間で共有できるシステムの提案と試作を行った。本システムにより、P2P ネットワークに流通する多種多様なコンテンツの選別が可能になることが期待できる。

今後の課題として、提案システムの有効性評価、及び大規模環境での性能評価が挙げられる。また、効率の良い複製やキャッシュの配置方法の検討、ユーザビリティを向上させるための、より効果的な選別方法の検討などが挙げられる。

### 参考文献

[1] Gnutella <http://www.gnutella.com>  
 [2] 金子勇著, アスキー書籍編集部 編集: "Winny の技術" ASCII(2005)  
 [3] Ion Stoica, Robert Morris, David Karger, M.Frans Kaashoek, Hari Balakrishnan: "Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications" ACM SIGCOMM 2001, San Diego, CA(2001)  
 [4] Request for Comments: 3174 <http://www.faqs.org/rfcs/rfc3174.html>