

議論に基づく Web 推薦システム*

福本 太郎[†]

新潟大学大学院自然科学研究科[†]

澤村 一[‡]

新潟大学自然科学系[‡]

1 はじめに

インターネットの普及に伴い検索エンジンから得られる検索結果の数は爆発的に増え、ユーザが自分の見たい情報を取得することが難しくなっている。これまでに効率的な検索技術がいくつか開発されてきたが、それらは定量的な評価（質問のキーワードと一致した回数やサイトの人気度など）しか行えず、ユーザの好みのような定性的な評価に応じた検索結果を返すことはできない。さらに、質問に対して大量の結果が検索エンジンから返ってきたとき、ユーザは自分でその膨大なデータの中から求める情報を選択しなければいけない。この時、特殊構文を用い検索結果を絞り込むこともできるが、一般的なユーザにとってこれを扱うことは用意なことではない。この問題を解決するシステムとして Chesñevar らが ArgueNet[1] を提案した。ArgueNet は従来の検索エンジンと議論フレームワークを統合したシステムであり、検索エンジンから得た検索結果に対してユーザの選好基準にとって適切であるかどうか議論することで、検索結果を適切なもの、判断がつかないもの、適切でないものの3つに分類する。構文的、定量的な評価だけでなく、ユーザの好みのような定性的な評価基準にしたがって分類でき、検索結果をさらに自分の求める情報かどうか評価する煩わしさからユーザを助ける。しかし ArgueNet では、我々人間が扱うような不確実な知識を表現できないため分類が大まかすぎる。そのためにユーザの選好基準に対して適切であるものが、判断がつかないものに含まれてしまうことがある。そこで本研究では我々の提案する LMA[2] と検索エンジンを統合した LMA-ArgueNet を提案する。LMA は我々の提案する議論フレームワークで、これを導入することで曖昧な知識を基に議論させることが可能であり、ArgueNet よりさらに細かな分類が可能となる。

2 LMA-ArgueNet

本システムでは EALP[2] を用いてユーザに選好基準を記述してもらい、論証の交換することで議論を実現し、議論結果を用いて検索結果を自動分類する。表 1 に選好基準の記述例を示す¹。これはクラシックギタリストを検索する際のユーザの選好基準の例である。(i) は好きなギタリストについて書かれたサイトは適切であると読む。(ii) は巨匠について書かれたサイトは適切でないわけがないと読む。(iii) は信頼するギタリストについて書かれたサイトは適切であると読む。(iv) は日本語で書かれたサイトでなければ適切でないと読む。(v) はそのギタリストの演奏を聞いたことがあるかどうかわからなくても、ハバナコンクールで受賞しているならば信頼すると読む。(vi) は John が巨匠であると読み(vii) は David の演奏を聞いたことがあるかどうかわからないと読む。

論証とは規則の有限列である。例えば表 2 のサイト s_2 について、以下のような論証が作成される。 $[\sim rel(s_2) : f \leftarrow website(s_2, John) : t \ \& \ greatmaster(John) : t, greatmaster(John) : t \leftarrow]$ これはサイト s_2 がユーザの選好基準にとって適切であることを主張するための論証である。論証の集合から正当化された論証を決定し、これを用いて検索結果进行分类する。ここで正当化の理解を容易にするために議論木を導入する。議論木とは、すべての枝が議論となる木である。ただし、すべての提議 $move_i = (提案者, Arg_i)$ に対し、その子は Arg_i を打破する論証である。論証 Arg を根とする議論木のすべての枝の終端が提案者であったとき論証 Arg は正当化されたという。以下に検索結果分類のための議論の流れを示す。

- Step 1 ユーザが検索エンジンに質問し、検索結果 $L = [s_1, \dots, s_i, \dots, s_n]$ を得る。L の要素 s_i は、情報 $info(s_i)$ (ファイル名等の特徴情報) によって特徴づけられている。
- Step 2 各 $info(s_i)$ から特徴情報を抽出し、EALP facts を作成する。
- Step 3 ユーザの選好基準 KB と $facts$ の和集合を使用し、各 s_i について $rel(s_i) : \mu$ と $\sim rel(s_i) : \mu$ を議題として議論を展開する。

*LMA-ArgueNet: Argumentation-based Web Recommender System

[†]Taro Fukumoto (fukumoto@cs.ie.niigata-u.ac.jp), Graduate School of Science and Technology, Niigata University

[‡]Hajime Sawamura (sawamura@ie.niigata-u.ac.jp), Institute of Natural Science and Technology, Academic Assembly, Niigata University

¹EALP では任意の真理値を注釈として使用できるが、紙面の都合上、本論文では真理値 $FOUR = \{t, f, \perp, \bot\}$ のみ言及する。この真理値は古典論理的な真偽に加え矛盾と未定義を表すことができる。

表 1: ユーザの評価基準

(i) $rel(S):t$	$\leftarrow website(S, N):t \& like(N):t$
(ii) $\sim rel(S):f$	$\leftarrow website(S, N):t \& greatmaster(N):t$
(iii) $rel(S):t$	$\leftarrow website(S, N):t \& trust(N):t$
(iv) $\sim rel(S):t$	$\leftarrow url(S, U):t \& \text{not domain}(U, "jp"):t$
(v) $trust(N):t$	$\leftarrow listen(N):t \& contest(habana, N):t$
(vi) $greatmaster(John):t$	$\leftarrow$, (vii) $listen(David):t \leftarrow$

表 2: Web サイトの特徴情報 $info(S)$

s_1	$website(s_1, Dai):t \leftarrow$ $contest(Tokyo, Dai):t \leftarrow$ $url(s_1, "http://www.music1.jp"):t \leftarrow$ $domain("http://www.music1.jp", "jp"):t \leftarrow$
s_2	$website(s_2, John):t \leftarrow$ $contest(unknown, John):t \leftarrow$ $url(s_2, "http://www.music2.uk"):t \leftarrow$ $domain("http://www.music2.uk", "uk"):t \leftarrow$
s_3	$website(s_3, David):t \leftarrow$ $contest(habana, David):t \leftarrow$ $url(s_3, "http://www.music3.uk"):t \leftarrow$ $domain("http://www.music3.uk", "uk"):t \leftarrow$

Step 4 図 1 に従い検索結果 L の要素を分類し、新たな検索結果リスト L' を作成する。

LMA-ArgueNet のフレームワークを図 2 に示す。4 値では検索結果 L の要素を図 1 に示すように 6 つの集合に分類する。6 つはそれぞれ、 S^r : 適切なサイト、 S^{pr} : たぶん適切なサイト、 S^{ri} : どちらとも言えるサイト、 S^u : どちらかわからないサイト S^{pi} : たぶん適切でないサイト、 S^i : 適切でないサイトであり、この順番にユーザへ示す。 $S^{pr}(S^{pi})$ は現時点では適切である (適切でない) と判断するだけの情報が不足していることを示し、 S^u は情報が足らなくどちらとも判断できないサイトを示す。これらは将来ユーザによって評価基準が更新されるか、もしくはサイトが更新されることによって適切または適切でないサイトになる可能性がある。

簡単な例を用いて分類例を示す。「クラシックギタリスト」をキーワードとして検索エンジンを使用したところ表 2 のサイトがこの順番で返されたとする。表 1 の評価基準を用いて LMA-ArgueNet を使用すると、議論の結果、 s_1 は情報が足りないため S^u : へ。 s_2 は日本語で書かれている保証はないが巨匠のサイトであるため S^{pr} へ、 s_3 も日本語で書かれている保証はないがハバナコンテストで受賞しているギタリストなので S^r へ分類される。

3 まとめ

検索エンジンと LMA を統合することで、ユーザが不確実、不完全、曖昧な評価基準を記述できる新しい検索方法を与え、実行例を通してその効果を示した。将来の課

題としては情報が足りないため分類ができなかったサイトに対してユーザに評価基準の入力を促し個人に適応していくことや、ユーザインターフェースの改良がある。

ALGORITHM LMA-ArgueNet

INPUT Query q , EALP $\mathcal{P}$

OUTPUT List L' (Classified search results)

BEGIN

Let $L = [s_1, \dots, s_i, \dots, s_n]$ be the output of solving query q using a web search engine.

Initialize $S^r, S^{pr}, S^{ri}, S^u, S^{pi}$ and S^i as empty sets. $\{S^r, S^{pr}, S^{ri}, S^u, S^{pi}$ and S^i stands for the set of relevant, probably relevant, both relevant and irrelevant, undecided, probably irrelevant and irrelevant search results, resp.}

FOR EVERY $s_i \in L$

DO

Solving query $rel(s_i):t$ using P'

IF $rel(s_i):t$ is justified and $rel(s_i):f$ is not justified THEN add s_i to S^r .

ELSE IF $\sim rel(s_i):f$ is justified and $rel(s_i):t$ is not justified THEN add s_i to S^{pr} .

ELSE IF $rel(s_i):t$ is justified THEN add s_i to S^{ri} .

ELSE IF $\sim rel(s_i):t$ is justified and $rel(s_i):f$ is not justified THEN add s_i to S^{pi} .

ELSE IF $rel(s_i):f$ is justified and $rel(s_i):t$ is not justified THEN add s_i to S^i .

ELSE add s_i to S^u .

Return $L' = [S^r, S^{pr}, S^{ri}, S^u, S^{pi}, S^i]$

END

図 1: LMA-ArgueNet 分類アルゴリズム

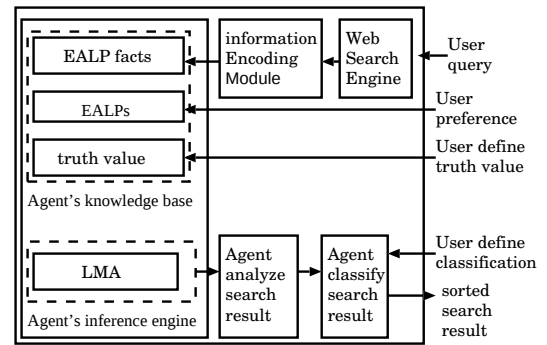


図 2: LMA-ArgueNet フレームワーク

参考文献

- [1] C. I. Chesñevar and A. G. Maguitman. *ArgueNet: An argument-based recommender system for solving web search queries*. In Procs. of the 2nd. International IEEE Conference on Intelligent Systems, pp282-287 (2004).
- [2] T. Takahashi and H. Sawamura. *A logic of multiple-valued argumentation*. In Procs. of AAMAS04, pp800-807 (2004).