

トピックモデルおよび分類器学習を用いたノウハウサイトの同定

李 佳奇[†] 趙 辰[†] 林 友超[†] 丁 易[†] 川畑 修人[†]春日 孝秀^{††} 宇津呂武仁^{†††} 河田 容英^{††††}[†] 筑波大学 大学院システム情報工学研究科 〒305-8573 茨城県つくば市天王台 1-1-1^{††} 筑波大学 理工学群 工学システム学類 〒305-8573 茨城県つくば市天王台 1-1-1^{†††} 筑波大学 システム情報系 知能機能工学域 〒305-8573 茨城県つくば市天王台 1-1-1^{††††} (株) ログワークス 〒151-0053 東京都渋谷区代々木 1-3-15 天翔代々木ビル 6F

あらまし 本論文の前身研究では、ウェブページ集合に対してトピックモデルを適用した結果に対して、複数のトピックにまたがって出現するドメインはノウハウサイトのドメインである可能性が高いという仮説を立て、実際に、収集された候補サイトの半数が有用なノウハウサイトであるという結果を得ている。このことをふまえて、本論文では、ノウハウ知識を含むサイトの候補として収集されたサイト群に対して分類器学習手法を適用することによって、ノウハウ知識を多く掲載したノウハウサイト、および、それ以外の一般サイトを自動識別する手法を提案する。この手法においては、ノウハウ知識を含むサイトの候補として収集されたサイト群に対して、各サイトのドメインの URL 素性、各ウェブページ検索時に用いられたサジェストについての素性、トピックモデルを適用した結果から得られる素性、等を用いて SVM を適用することによって、ノウハウサイトの選定において一定以上の適合率、再現率を達成した。

キーワード ノウハウサイト、検索エンジン・サジェスト、トピックモデル、話題分布、収集・集約、分類器学習

Identifying Know-How Sites based on a Topic Model and Classifier Learning

Jiaqi LI[†], Chen ZHAO[†], Youchao LIN[†], Ding YI[†], Shuto KAWABATA[†], Takahide KASUGA^{††},Takehito UTSURO^{†††}, and Yasuhide KAWADA^{††††}[†] Grad. Sch. of Systems and Information Engineering, University of Tsukuba, Tsukuba 305-8573 Japan^{††} College of Engineering Systems, School of Science and Engineering, University of Tsukuba, Tsukuba 305-8573 Japan^{†††} Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba 305-8573 Japan^{††††} Logworks Co., Ltd. Tokyo 151-0053, Japan

1. はじめに

現在、インターネットの普及やソーシャルメディアの隆盛により、多くのユーザはウェブページから日常生活に役立つ知識を得ている。様々な知識をまとめたウェブサイトが多数存在する。Yahoo!知恵袋等の質問回答サイトでは、「就活の準備方法」、「結婚式スケジュールガイド」など、ユーザにとって、日常生活の手助けになるノウハウ知識が掲載されている。しかし、このような質問回答サイトやその他の多様なウェブサイトにおいては、多種多様な内容および質の情報が散在しているため、ユーザにとって役立つノウハウ知識を集約して提示することが必要

とされている。ここで、本論文では、検索者が詳細な情報を検索したい対象を「クエリ・フォーカス」と呼ぶ。そして、このような要求を満たすことを目的として、[5] では、指定されたクエリ・フォーカスに対して網羅的に検索エンジン・サジェストを収集し、クエリ・フォーカスとサジェストの AND 検索においてウェブページを収集した結果の文書集合に対してトピックモデルを適用し、そこからノウハウ知識の候補を多く収集し集約・俯瞰している。さらに、[3] においては、[5] において収集したウェブページ集合に対してトピックモデルを適用した結果に対して、複数のトピックにまたがって出現するドメインはノウハウサイトのドメインである可能性が高いという仮説を立

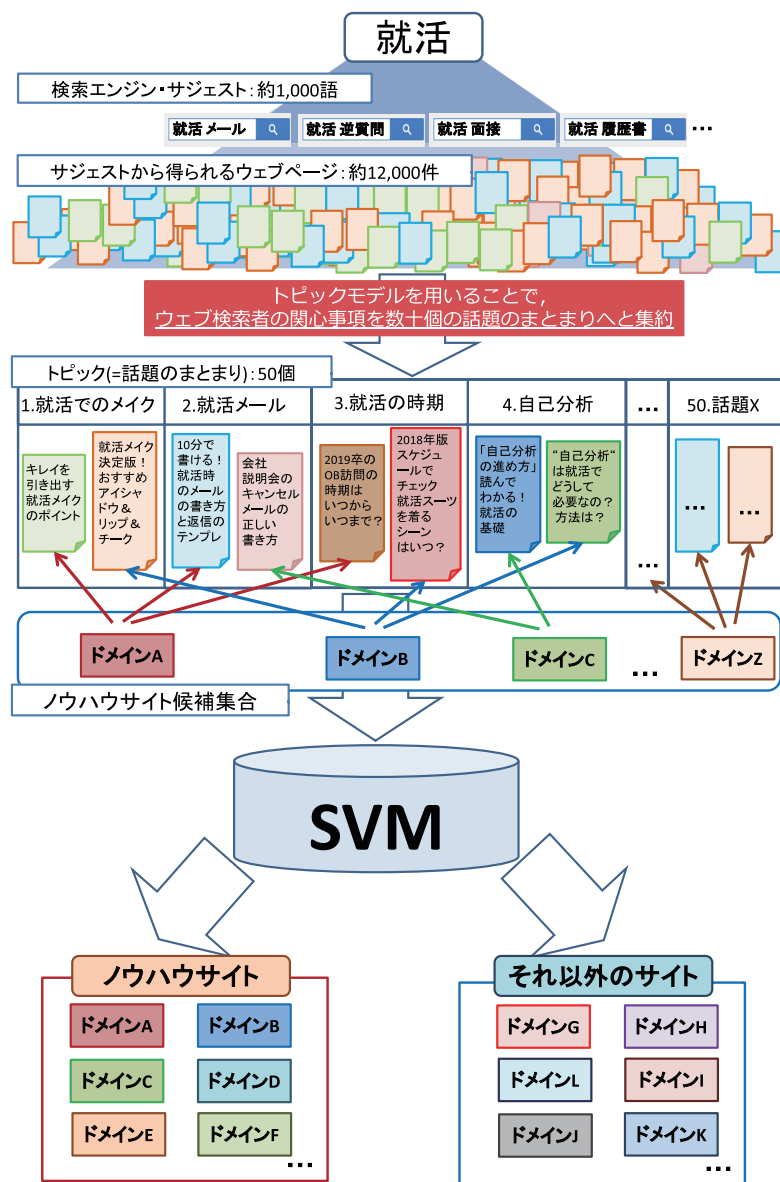


図 1 ノウハウサイト自動選定の流れ

表 1 ノウハウサイト候補のドメインに対する評価基準

ドメインそのものがノウハウ知識を提示する個別ページへのリンクを一覧するページである			A 群
ドメインそのものがノウハウ知識を提示する個別ページへのリンクを一覧するページではない	ノウハウ知識を提示する個別ページへのリンクを一覧するページが存在する	ドメインのトップからノウハウ知識を提示する個別ページへのリンクを一覧するページに容易に辿り着ける	B 群
		ドメインのトップからノウハウ知識を提示する個別ページへのリンクを一覧するページには容易に辿り着けない	C 群
	ノウハウ知識を提示する個別ページへのリンクを一覧するページが存在しない	ノウハウ知識を提示する個別ページが存在する	D 群
		ノウハウ知識を提示する個別ページが存在しない	E 群

て、実際に、収集された候補サイトの半数が有用なノウハウサイトであるという結果を得ている。

これに対して、本論文では、図 1 の全体像に示すように、[3] においてノウハウ知識を含むサイトの候補として収集されたサイト群に対して分類器学習手法を適用することによって、ノウ

ハウ知識を多く掲載したノウハウサイト、および、それ以外の一般サイトを自動識別する手法を提案する。この手法においては、ノウハウ知識を含むサイトの候補として収集されたサイト群に対して、各サイトのドメインの URL 素性、各ウェブページ検索時に用いられたサジェストについての素性、トピックモ

提案手法の仮説: 複数トピックにまたがって出現するドメイン=ノウハウサイトのドメインと見なす

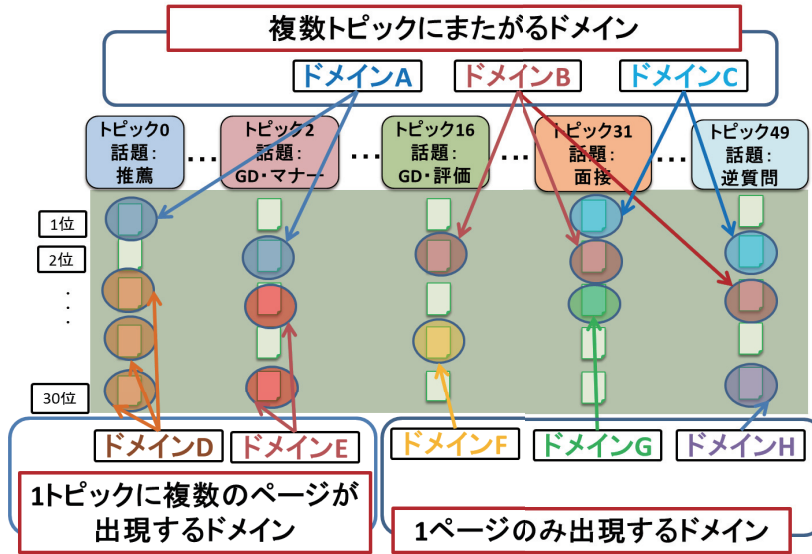


図2 トピックモデルにおけるサイトの分布に基づくノウハウサイト候補集合の選定

表2 各クエリ・フォーカスにおけるサジェスト数, ウェブページ数, および, 各トピック上位30件のウェブページから収集されたノウハウサイト候補の数

クエリ・フォーカス	サジェスト数	ウェブページ数	ノウハウサイト候補の数 (正例/負例)
就活	926	12,078	82 (51/31)
結婚	959	13,256	85 (40/45)
花粉症	850	9,745	102 (45/57)
マンション	958	13,742	89 (57/32)
虫歯	838	9,462	92 (45/47)
食中毒	815	7,573	98 (28/70)
合計	5,346	65,856	548 (266/282)

デルを適用した結果から得られる素性, 等を用いて SVM を適用することによって, ノウハウサイトの選定において一定以上の適合率, 再現率を達成した。

2. 検索エンジン・サジェストを用いたウェブページの収集

本論文では, 分析の対象であるクエリ・フォーカス「就活」, 「花粉症」, 「結婚」, 「マンション」, 「虫歯」, および, 「食中毒」について, 検索エンジン・サジェストを利用してウェブページを収集する。そのためにまず, Google 検索エンジンを用いて, 一クエリ・フォーカスにおいて, 約 100 通りの文字列を指定し, 最大約 1,000 語のサジェストを収集する。クエリ・フォーカス「花粉症」, 「結婚」, 「就活」, 「マンション」, 「虫歯」, および, 「食中毒」において, 収集されたサジェストの数を表 2 に示す。ここで, クエリ・フォーカスに対して収集されたサジェストの集合 $\mathbb{S}$ として, あるサジェスト $s \in \mathbb{S}$ に対して, クエリ・フォー

カスとの AND 検索によって上位 N 件以内に検索されるウェブページ p の集合を $P(s, N)$ とする。本論文では, $N = 20$ とする。各クエリ・フォーカスに対して収集されるウェブページの集合 P_w を次式で定義する。

$$P_w = \bigcup_{s \in \mathbb{S}} P(s, N)$$

ウェブページの収集においては, Google Custom Search API^(注1) を用いた。各ウェブページ p を検索する際に指定されたサジェスト s を集めた集合 $\mathbb{S}(p)$ を次式で定義する。

$$\mathbb{S}(p) = \left\{ s \in \mathbb{S} \mid p \in P(s, N) \right\}$$

3. トピックモデルを用いたウェブページ集合の集約

前節で収集したウェブページ集合 P_w に対してトピックモデルを適用することにより, ウェブページ集合の集約を行う。

本論文では, トピックモデルとして潜在的ディリクレ配分法 (LDA; Latent Dirichlet Allocation) [1] を用いる。LDA を用いてトピックモデルを推定する際には, 語 w の集合を V として, 語 $w (w \in V)$ の列により表現されたウェブページの集合およびトピック数 K (本論文の評価実験においては $K = 50$ とする) を指定して, 各トピック z_n ($n = 1, \dots, K$) における語 w の確率分布 $P(w|z_n)$ ($w \in V$), および, 各ウェブページ p におけるトピック z_n の確率分布 $P(z_n|d)$ ($n = 1, \dots, K$) を推定する。

本論文では, ウェブページ p におけるトピックの確率分布

(注1): <https://developers.google.com/custom-search/>

表 3 SVM の素性

分類	定義
URL 素性	f_1 : ドメインの URL において, 組織種・地域コード “com”, “org”, “jp”, “net”, “co.jp” の 5 種の各々の有無を個別に二値素性 ($f_{11}, \dots, f_{15}$) として用いる
	f_2 : ドメインの URL におけるプロトコルがセキュア (Secure) か (HTTPS) 否か (HTTP) の二値素性
サジェスト素性	f_3 : 当該ドメインを検索する際に指定されたサジェストの延べ観測数
	f_4 : 当該ドメインを検索する際に指定されたサジェストの異なり観測数
ウェブページ素性	f_5 : 当該ドメインにおいて収集されたウェブページ数
トピックモデル素性	f_6 : 当該ドメインのウェブページが出現するトピック数
複合素性	$f_7 = f_3 - f_4$: サジェストの延べ観測数 - 異なり観測数
	$f_8 = f_7 / f_5$: (サジェストの延べ観測数 - 異なり観測数) / (収集されたウェブページ数)
	$f_9 = f_5 - f_6$: 収集されたウェブページ数 - ウェブページが出現するトピック数
	$f_{10} = f_5 / f_6$: 収集されたウェブページ数 / ウェブページが出現するトピック数

において確率が最大となるトピックをウェブページ p に割り当てる. その結果, 全ウェブページの集合 P_w に対して, トピック $z_n (n = 1, \dots, K)$ が割り当てられたウェブページの集合 $P_w(z_n)$ は次式で表される.

$$P_w(z_n) = \left\{ d \in P_w \mid z_n = \underset{z_u (u=1, \dots, K)}{\operatorname{argmax}} P(z_u | d) \right\}$$

4. 評価用サイト集合

本節では, ノウハウサイト自動同定の対象として用いる評価用サイト集合の収集方法について述べる. 本節の収集方法の基本的な考え方は, [3] において我々が提案したノウハウサイト候補の自動収集手法に基づく. [3] においては, 前節でウェブページ集合 P_w に対してトピックモデルを適用した結果のうち, 特に, 各トピック上位 30 位以内のウェブページのドメインに対して, 表 1 に示す基準のもとで, 各ドメインがノウハウサイトとみなせるか否かの判定を行った. そこでは, 表 1 の基準のうち, A 群, B 群, C 群に属するドメインがノウハウサイトであるという立場に立った上で, 特に, 複数のトピックにまたがって出現するドメイン (図 2) においては, 相対的にノウハウサイトが多く含まれる, という仮説を立て, この仮説が有効であることを示した.

本論文でも, [3] における仮説「トピックモデルを適用した結果において, 各トピック上位 30 位のウェブページのドメインのうち, 複数のトピックにまたがって出現するドメイン中にノウハウサイトが多く含まれる」をふまえて, ウェブページ集合 P_w に対してトピックモデルを適用した結果において, 各トピック上位 30 位のウェブページのドメインのうち, 複数のトピックにまたがって出現するドメインを収集し, ノウハウサイト自動同定の対象として用いる評価用サイト集合とする.

本論文では, 特に, 分析対象のクエリ・フォーカスとして, 「就活」, 「花粉症」, 「結婚」, 「マンション」, 「虫歯」, および, 「食中毒」を用いる. これらのクエリ・フォーカスに対して収集されたサジェスト数, ウェブページ数, およびサイト数 (ノウハウサイト候補数) を表 2 に示す. また, これらのノウハウサ

イト候補に対して, 表 1 に示す基準にしたがい, A 群 ~ E 群の判定を行った後, A 群, B 群, C 群のサイトをノウハウサイトとみなして正例とし, その他のサイトを負例として集計した結果もあわせて示す.

分析対象のクエリ・フォーカスのうち, 「就活」, 「結婚」, および, 「花粉症」における A 群のノウハウサイトの例を, それぞれ, 図 3, 図 4, および, 図 5 に示す.

5. 素 性

本論文の分類器学習において用いた素性の一覧を表 3 に示す. なお, 5.1 節 ~ 5.4 節の素性を単位素性と呼び, それらの単位素性を複数個組み合わせた素性を複合素性 (5.5 節) と呼ぶ.

5.1 URL 素性

ノウハウサイト候補のドメインの URL において, 組織種・地域コード “com”, “org”, “jp”, “net”, および, “co.jp” の 5 種の各々の有無を個別に二値素性 ($f_{11}, \dots, f_{15}$) として用いる. また, ドメインの URL におけるプロトコルがセキュア (Secure) か (HTTPS), 否か (HTTP) を二値素性として用いる.

5.2 サジェスト素性

評価対象のノウハウサイト候補の集合を T , 評価対象の一つのサイトを $t (t \in T)$ とし, サイト t 中のウェブページ集合を $P(t)$ とする. ここで, 各ウェブページ $p (p \in P(t))$ を検索する際に指定されたサジェストを集めた集合 $S(p)$ の和集合を次式

$$S(t) = \bigcup_{p \in P(t)} S(p)$$

で定義する. そして, サイト t におけるサジェストの異なり観測数

$$|S(t)|$$

をノウハウサイト自動選定の素性の一つとして用いる. 同様に, サイト t の各ウェブページ $p (p \in P(t))$ を検索する際に指定されたサジェストの延べ数

$$\sum_{p \in P(t)} |S(p)|$$



図 3 A 群「ドメインそのものがノウハウ知識を提示する」のサイトの例
(「賢者の就活」, <http://kenjasyukatsu.com/>)



図 4 A 群「ドメインそのものがノウハウ知識を提示する」のサイトの例
(「結婚レシピ」, <http://www.wedding-recipe.com/>)

をサイト t に対して求め、サイト t におけるサジェストの述べ観測数として、ノウハウサイト自動選定の素性の一つとして用いる。

5.3 ウェブページ素性

サイト t 中のウェブページの集合 $P(t)$ 中のウェブページ数

$$|P(t)|$$

をノウハウサイト自動選定における複合素性において用いる。

5.4 トピックモデル素性

トピックモデルを適用した結果においては、ウェブページ $p \in P(t)$ に対して、ウェブページ p におけるトピックの確率分布において確率が最大となるトピック $z(p)$ をウェブページ p に割り当てる。

$$z(p) = \underset{z_u (u=1, \dots, K)}{\operatorname{argmax}} P(z_u | p)$$

ここで、サイト $t \in T$ に属する各ウェブページ p のトピック $z(p)$ の集合を次式

$$z(t) = \bigcup_{p \in P(t)} \{z(p)\}$$

で定義する。そして、サイト t におけるトピックの異なり観測数

$$|z(t)|$$

をノウハウサイト自動選定における複合素性において用いる。

5.5 複合素性

以下の素性を用いる。

(1) サジェストの延べ観測数 - 異なり観測数

$$\sum_{p \in P(t)} |S(p)| - |S(t)|$$



図 5 A 群「ドメインそのものがノウハウ知識を提示する」のサイトの例
 (「花粉症対策と症状対策」, <http://www.kafun-taisaku1.com/>)

(2) (サジェストの延べ観測数 - 異なり観測数) / (収集されたウェブページ数)

$$\frac{\sum_{p \in P(t)} |S(p)| - |S(t)|}{|P(t)|}$$

(3) 収集されたウェブページ数 - ウェブページが出現するトピック数

$$|P(t)| - |z(t)|$$

(4) 収集されたウェブページ数 / ウェブページが出現するトピック数

$$\frac{|P(t)|}{|z(t)|}$$

6. 評価

6.1 評価手順

本論文では、LIBSVM^(注2)を用いて評価実験を行った。カーネル関数としてRBFカーネルを用い、交差検定の手順として、全6個のクエリ・フォーカスのうち、5個分のクエリ・フォーカスのサイトで訓練をして、残った1個分のクエリ・フォーカスで評価をする、という6分割交差検定を行った。SVMのハイパーパラメータの調整は、グリッドサーチによって行った。

本論文の評価においては、各素性の有効性を評価するため、 $f_{7 \sim 10}$ の複合素性を含め、 $f_{1 \sim 6}$ の単位素性の様々な組み合わせを評価した結果、 $f_{1 \sim 4, 7 \sim 10}$ の8素性の組み合わせが最も高い性能を示した^(注3)。また、判定結果の推定信頼度を出力した上で、

信頼度に対して下限値を設け、高信頼度な判定結果に限定した範囲での適合率、再現率の評価を行った。参照用ノウハウサイト集合(正例集合)を R 、信頼度 $conf$ が下限値 c 以上でノウハウサイト(正例)と判定されたサイト集合を $S(conf \geq c)$ として、再現率($conf \geq c$)、および、適合率($conf \geq c$)を次式で定義し、この値をプロットすることにより評価を行う。

$$\text{再現率}(conf \geq c) = \frac{|R \cap S(conf \geq c)|}{|R|}$$

$$\text{適合率}(conf \geq c) = \frac{|R \cap S(conf \geq c)|}{|S(conf \geq c)|}$$

6.2 評価結果

図6の評価結果から分かるように、正例が少ないクエリ・フォーカス「食中毒」では性能が悪かった。それに対して、正例、負例のバランスが良い「食中毒」以外のクエリ・フォーカスでは相対的に高い性能を示した。

7. ノウハウ分布の調査と分析

参照用ノウハウサイトに対して、素性としてトピックモデル素性を用いて、各サイトの持つトピック数の降順の上位10サイトを対象として、全50トピックのうち「就活」に関するノウハウとの関連が強い32トピックにおけるウェブページの分布を示したものを表4に示す。このように、各ノウハウサイトにおいては、「就活」に関連するノウハウとの関連が強い32トピック全体にわたって、広範囲の話題が掲載されていることが分かる。

値素性のとり方を評価した。その結果、左端を固定し、右端が異なる約40~100個程度の範囲をとる二値素性を用いた場合に最も高い性能となったため、この素性設定を用いた。その結果、 f_3 は89個、 f_4 は111個、 f_7 は35個、 f_8 は111個、 f_9 は47個、 f_{10} は80個の二値素性となった。

(注2) : <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

(注3) : 各素性については、素性のとり値の分布に応じて、重複のない多数の範囲をとる二値素性、および、重複のある多数の範囲をとる二値素性等、様々な二

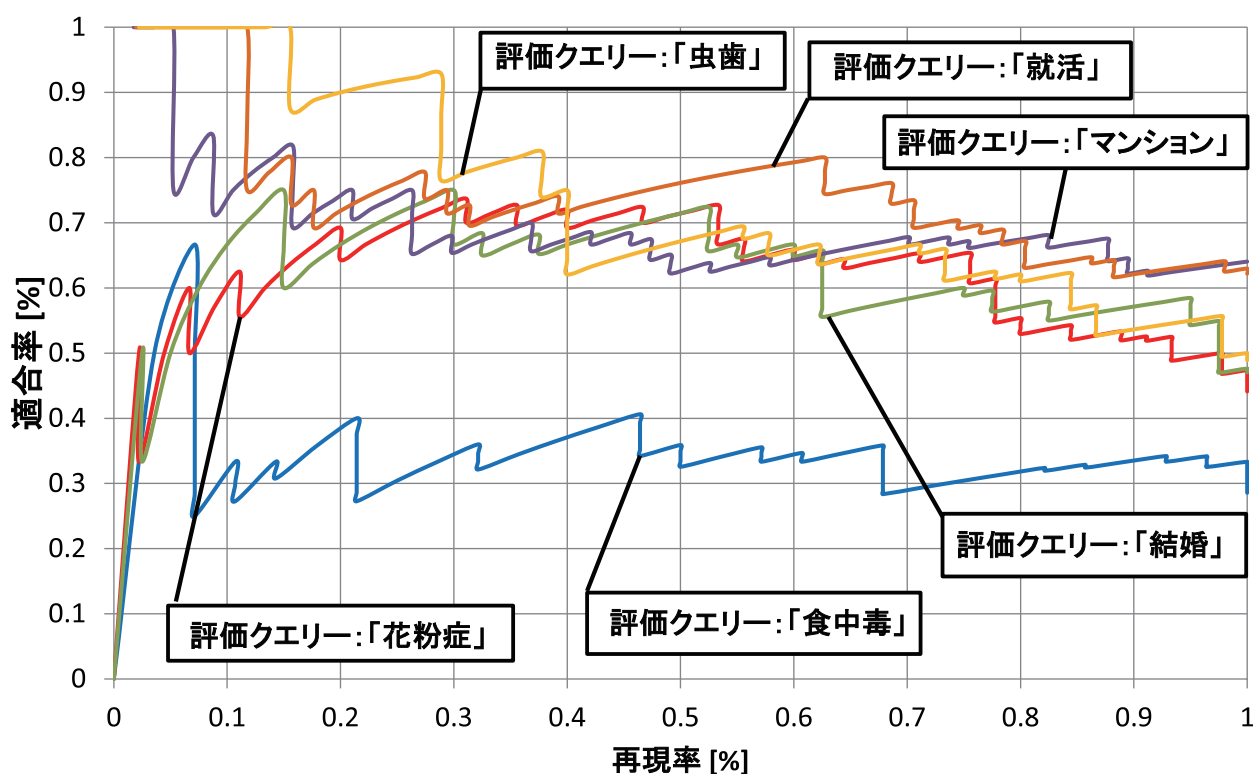


図6 ノウハウサイト自動同定の評価結果

8. 関連研究

本論文のノウハウサイト同定タスクに関連して、先行研究として、ウェブページの分類に関する手法がいくつか提案されている。例えば、[6]では、大学サイト中のウェブページを種類別に分類するタスクにおいて、ページタイトル、本文テキスト、および、被リンク元のアンカーテキストを素性として用いる手法を提案している。一方、本論文のノウハウサイト同定タスクにおいては、識別対象がウェブページではなくサイトの単位であり、かつ、素性として用いる情報も、サイト中のウェブページにおける話題の分布、ウェブ検索時のサジェストの延べ数・異なり数、サイト中のウェブページ数等、ウェブページ分類において用いられる素性とは大きく異なっている。

また、[2]では、クエリを実現するためのサブタスクの収集方式において、行為を表す動詞表現の形式を用いる方式を提案している。2014年12月に開催されたNTCIR-11^(注4)においては、[2]とほぼ同様の仕様でTask Mining Taskも実施された。本研究においても、Task Mining Taskで用いられたクエリリストおよび評価手順[4]のもとで、本論文の提案手法を適用する方式の可能性を検討する必要があると考えられる。ただし、Task Mining Taskのタスク設定では、クエリを実現するためのサブタスク群を動詞表現の形式で出力することにとどまっており、ノウハウ知識そのものは収集の対象として扱われていない。

これに対して、本研究において収集の対象となるのは、ウェブページ群の形式で表現されたノウハウサイトであり、この点

において上記の関連研究とは大きく異なっている。

9. おわりに

本論文では、一つのクエリ・フォーカスについて、ウェブから自動収集されたウェブページ集合に対して、トピックモデルを適用して集約を行い、その後、分類器学習を適用することにより、ノウハウサイトを自動選定する手法を提案した。今後の課題としては、特定のクエリ・フォーカス、および、クエリ・フォーカスに対するサジェストを用いて収集した大規模ウェブページ集合に対するトピックモデル適用結果を経由せず、特定のサイトから得られる情報のみを素性として分類器学習を適用することにより、ノウハウサイト自動同定を実現することが挙げられる。

文 献

- [1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, Vol. 3, pp. 993–1022, 2003.
- [2] 加藤龍, 大島裕明, 山本岳洋, 加藤誠, 田中克己. タスクの汎化と特化に着目したWebからのタスク検索. 第6回DEIMフォーラム論文集, 2014.
- [3] 李佳奇, 趙辰, 林友超, 馬場瑞穂, 宇津呂武仁, 河田容英, 神門典子. トピックモデルにおける話題分布特性を手がかりとするノウハウサイトの収集. 第9回DEIMフォーラム論文集, 2017.
- [4] Y. Liu, R. Song, M. Zhang, Z. Dou, T. Yamamoto, M. Kato, H. Ohshima, and K. Zhou. Overview of the NTCIR-11 IMine task. In *Proc. 11th NTCIR Workshop Meeting*, pp. 8–23, 2014.
- [5] 守谷一朗, 井上祐輔, 今田貴和, 轟添, 宇津呂武仁, 河田容英, 神門典子. 質問回答事例および検索エンジン・サジェストを用いたノウハウ知識の相補的収集. 第7回DEIMフォーラム論文集, 2015.

(注4) : <http://research.nii.ac.jp/ntcir/ntcir-11/index-ja.html>

表 4 自動収集されたノウハウサイト群中のトピック分布

トピック ID	内容	トピック数: 1~10 位 / ドメインに対する評価基準: A 群 または B 群 / ノウハウサイト名									
		1 位	2 位	3 位	4 位	5 位	6 位	7 位	8 位	9 位	10 位
		B 群	B 群	B 群	B 群	B 群	A 群	A 群	A 群	B 群	A 群
		マイナビ	マイナビ学生の窓口	リクナビ	キャリアタス就活	JOBPASS 新卒	賢者の就活	外資就活ドットコム	就活ノート	Jobweb	履歴書 Do
1	就活でのスーツ、ネクタイ、コートなど服装	○	○	○		○	○	○	○	○	○
2	就活での TOEIC スコア、英語・外国語系学生の就活	○	○	○	○	○		○	○	○	
3	就活でのメイク	○	○	○	○						○
4	就活で印象が良くなる髪型の作り方	○	○		○	○	○		○		○
5	理系の就職活動の「学校推薦」を解説、研究内容の書き方	○	○	○	○	○	○	○		○	
6	就活での内定時期、内定辞退、「誓約書」など	○	○	○	○	○	○	○	○	○	○
7	就活前に考える理想のファーストキャリア、注目すべき職業など	○	○	○	○	○		○		○	
8	既卒の就活、大学卒業後就活のやり方など	○		○	○	○	○				○
9	就活メール、返信用封筒などのマナー			○							
10	就活の時期、選考解禁の時期など	○	○	○		○	○		○		
11	就活の選考過程、自己分析の方法など		○		○	○					
12	履歴書書き方、送付、履歴書のサイズなど	○	○	○		○	○	○	○	○	○
13	就活に必要な健康診断の項目、料金など	○	○	○	○	○					○
14	業界研究・企業研究の仕方	○	○	○	○	○	○	○	○	○	
15	就職偏差値ランキング、業界別倍率など	○		○	○						
16	証明写真撮影ポイント、履歴書の写真サイズなど		○		○	○			○	○	○
17	就活で約立つアプリなど	○		○	○			○		○	
18	就活前にやっておくべきこと・必要なものなど	○	○			○	○	○	○	○	
19	SPI, WEB テスト (玉手箱・TG-WEB) などの対策	○	○	○	○		○	○	○	○	○
20	長所と短所の見つけ方、自己分析のやり方など	○	○	○	○	○	○	○	○	○	○
21	就活でのゲーム業界、ゲーム会社など	○	○	○		○	○				○
22	就職面接のマナー、面接で緊張しないコツなど	○	○	○		○	○	○	○	○	○
23	グループディスカッションのコツ、進め方、対策など	○	○	○	○		○	○	○	○	
24	面接や説明会での質問コツ		○	○	○		○	○	○		○
25	就活が辛いときの気持ちの切り替え方、不安を乗り越える方法	○	○	○	○	○	○	○	○	○	○
26	就活の面接で嘘をつくらざるべきこと、面接で正しい答えかた	○	○			○	○	○	○	○	○
27	就活やインターンを安く乗り切る方法、アルバイトを両立させる方法	○	○	○	○			○	○	○	○
28	就活動教員採用試験の並行	○	○	○		○		○		○	○
29	エントリーの方法・エントリーシートの書き方	○	○	○	○		○		○		○
30	就活セミナー情報、合同企業説明会情報など	○		○	○	○					
31	業界を絞る方法・業界の選び方、企業研究方法	○	○	○	○	○	○	○	○	○	
32	ブラック企業の特徴と見分け方	○	○	○	○		○	○	○	○	○

[6] A. Sun, E.-P. Lim, and W.-K. Ng. Web classification using support vector machine. In *Proc. 4th WIDM*, pp. 96–99, 2002.