DEIM Forum 2019 C1-2

ノウハウ収集のための中国語ウェブサイトの同定と分析

牛　文彬† 大川　遥平† 川畑　修人† 趙　辰† 聶　添†
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 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　本論文では，中国語を対象とし，中国でよく使用される検索エンジン「百度」を用いて特定のクエリ・フォーカスについて多数のノウハウを記述したウェブサイトを収集する手法を提案する．本手法においては，まず，特定のクエリ・フォーカスに関する検索エンジン・サジェストを用いて，関連ウェブページを網羅的に収集する．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，複数トピックにまたがって出現するドメインをノウハウ知識を含むサイトの候補とする．そして，選定された候補ウェブサイトを「ノウハウサイト」,「QA サイト」,「商用サイト」,「事例サイト」の四つに分類し，各種類サイトのノウハウ分布を分析する．
キーワード　ノウハウ収集, 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

Detecting and Analysing Chinese Web Sites for Collecting Know-How Knowledge

Wenbin NIU†, Yohei OHKAWA†, Shuto KAWABATA†, Chen ZHAO†, Tian NIE†, Takehito UTSURO††, and Yasuhide KAWADA†††

† Grad. Sch. of Systems and Information 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

現在社会においては，多くの人がウェブ情報を利用し，日常生活に役に立つ知識を容易に入手することができる．ウェブ利用者は，有用な情報が掲載されるウェブサイトから様々な知識を得ることにより豊かな生活を過ごせるが，インターネット上では情報や知識が氾濫しているため，大量の情報を集約して把握することは容易ではない．そのため，ユーザにとって必要な情報のみを提示することを目的とするウェブ情報の集約が不可欠である．

文献 [7] においては，日本語検索エンジンを対象として，ノウハウを多く掲載する「ノウハウサイト」を自動同定する手法を提案している．文献 [7] においては，まず，既存の検索エンジンによってノウハウサイトが適切に検索される割合を評価した．その結果，ノウハウが得られるサイトが検索される割合は，最も高精度であったクエリ・フォーカス (検索者が詳細な情報を検索したい対象．本論文における以降の例ではクエリ・フォーカスを「結婚」とする) を検索クエリとした場合でも，検索結果の上位 5 位以内において 60%程度であり，平均的には，上位 50 位以内において，20〜40%程度であった．それに対して，文献 [7] の手法では，特定のクエリ・フォーカスに着目して検索エンジン・サジェストを収集し，それらのサジェスト集合を用いて収集したウェブページ集合を対象として分類器学習 (注1) を適用することにより，検索エンジンよりも高い精度でノウハウが得られるサイトを自動識別することができた．

以上をふまえて，本論文では，文献 [7] の手法の考え方を中国語に適用することを目的とする．具体的には，中国における使用頻度が最も高いウェブ検索エンジン「百度」(注2) を情報源とし，特定のクエリ・フォーカスに着目し，ノウハウを多く掲載する中国語ウェブサイトの収集と分析を行った．本論文にお

(注1)：素性としては，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の素性，および，doc2vec [3] を用いた素性を用いた．
(注2)：https://www.baidu.com/

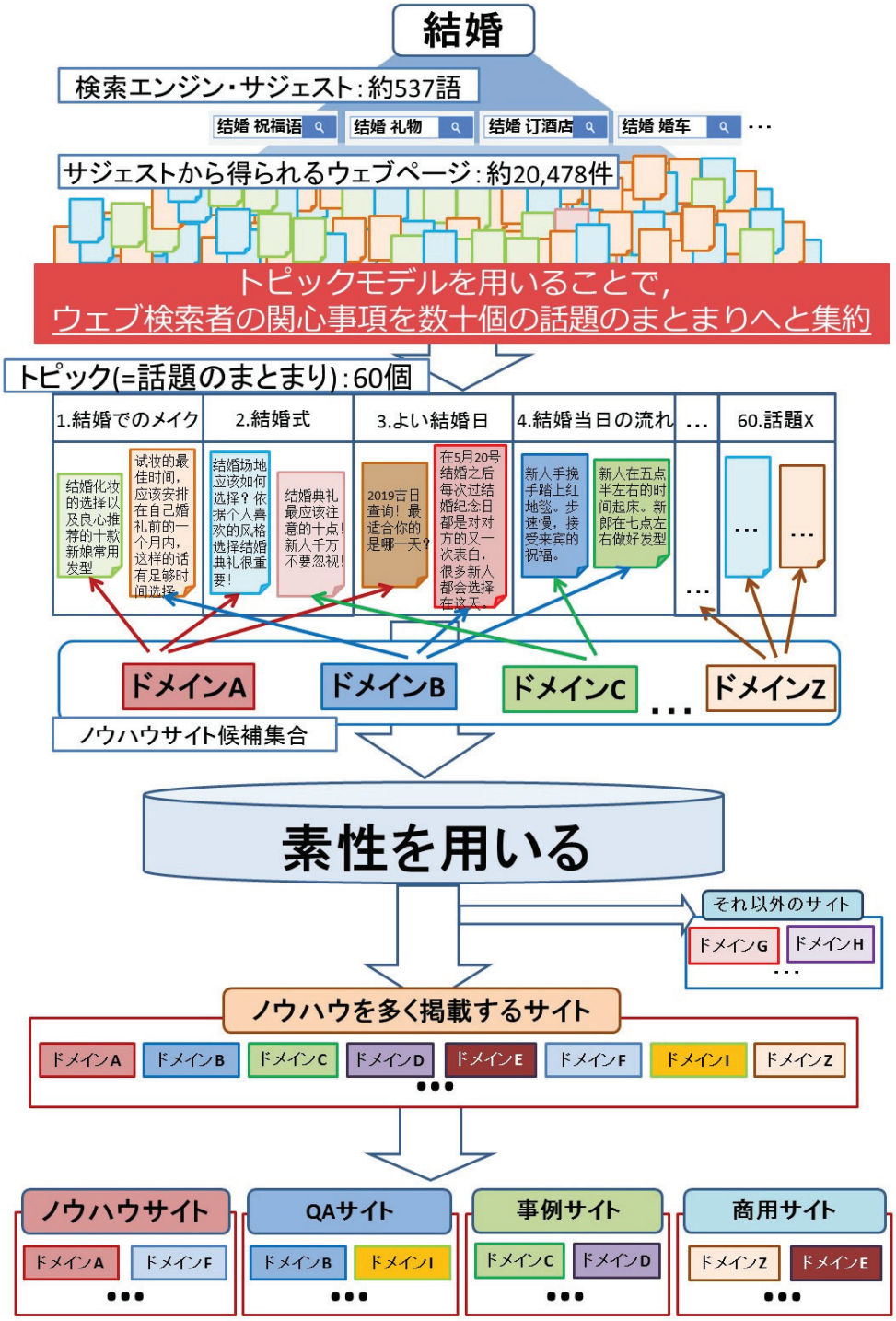


図 1　ノウハウを多く掲載する中国語サイト収集と分析の流れ

ける全体的流れを図 1 に示す．

まず，特定のクエリ・フォーカス (図 1 の例では「結婚」) に対してサジェストを収集する．そして，収集されたサジェストを用いて「クエリ・フォーカス AND サジェスト」の形のクエリを作成し，これを用いてウェブページの収集を行う．その後，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，ウェブページ集合の話題を集約する．次に複数のトピックにまたがってウェブページを持つサイトにおいては，ノウハウが多く掲載されていると仮定し，ノウハウを多く掲載する中国語ウェブサイトを選定する．そして，ノウハウを多く掲載する中国語ウェブサイトを「ノウハウサイト」,「QA サイト」,「商用サイト」,「事例サイト」の四種類に分類し，各種類のサイトに

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

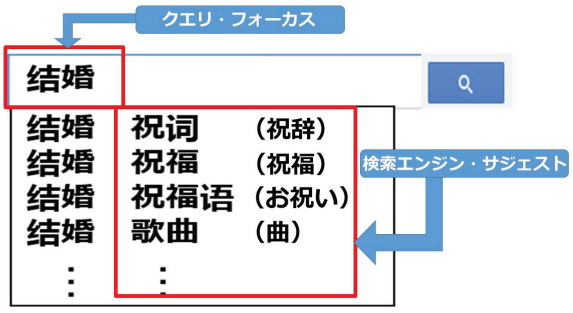


図 2　中国語検索エンジン「百度」における検索エンジン・サジェストの例

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「ノウハウサイト」の例（「wed114 結婚ネット」，http://www.wed114.cn/）

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「QA サイト」の例（「百度知道」，https://zhidao.baidu.com/）

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「商用サイト」の例（「58 同城」，http://www.58.com/）

おけるトピック分布を分析する．さらに，ノウハウを多く掲載するサイト群に対して，文献 [7] と同等の素性を用いることにより，ノウハウを多く掲載する中国語ウェブサイトの自動同定精度の評価を行う．

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

2.1　検索エンジン・サジェスト

各検索エンジン会社においては，ウェブ利用者による検索ログが蓄積されている．多くのウェブ利用者が検索したクエリのうち，特に高い関心が持たれた語を抽出することにより，検索エンジン・サジェストとして提示するサービスを提供している．ここで，検索エンジン・サジェストとして提示された語は，クエリ・フォーカスに対して，AND 検索の形で二つ目以降に入力した語を情報源として抽出したものである．本論文では，検索エンジン・サジェストにはウェブ利用者の関心事項が反映されていると考えて，検索エンジン・サジェストの収集を行った．特に，中国でよく使用される検索エンジン「百度」を用いて，クエリ・フォーカス「結婚」を対象に，検索エンジン・サジェストの収集を行った．

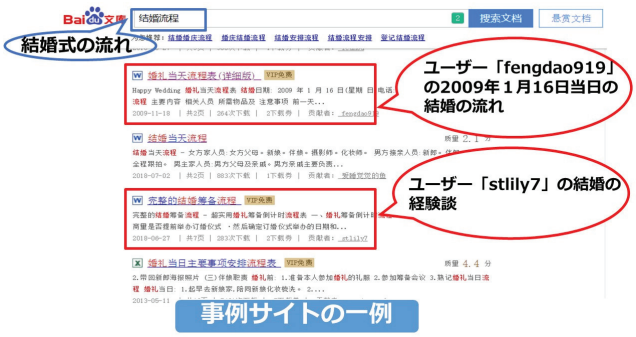


図 6　A 群「ドメインそのものがノウハウ知識を提示する」の「事例サイト」の例 (「百度文庫」, https://wenku.baidu.com/)

2.2　検索エンジン・サジェストの収集

中国語の「百度」検索エンジンに対して，一つのクエリ・フォーカスに対して 300 通りの文字列を指定し，最大 3,000 語のサジェストを収集することができる．300 通りの文字列とは，中国語のピン音の部首であり，例えば検索欄に「結婚」と入力すると,「祝福 (祝福)」,「戸口 (戸籍)」などがサジェストとして掲示される．クエリ・フォーカス「結婚」に対してそれらのサジェストを収集することにより，537 語のサジェストが収集された (表 2).

2.3　ウェブページの収集

ウェブページの収集においては，クエリ・フォーカス，および，収集したサジェストを指定 (「クエリ・フォーカス AND サジェスト」の AND 条件での検索語を指定する) して，中国語検索エンジン「百度」を用いることにより，中国語のウェブページを対象として収集を行った．一つの第二検索語あたり，ウェブ検索結果の上位 20 ページを収集し，重複ページおよび開けないページを削除した結果，検索結果として，20,478 件のウェブページが収集された (表 2).

3.　トピックモデルの適用による候補ウェブサイトの選定

3.1　トピックモデル

検索エンジン・サジェストを指定して検索エンジンを用いた検索を行うことにより，多くのウェブページ収集することができる．しかし，収集されるウェブページの話題の多くは重複し冗長であるため，収集されたウェブページ集合の話題の集約が必要である．そこで，本論文では，収集されたウェブページ集合に対して，トピックモデルの 1 つである潜在的ディリクレ配分法 (LDA; Latent Dirichlet Allocation) [1] を適用する．LDA のツールとしては，GibbsLDA++(注3) を用いる．トピック数 K を 10 から 80 の範囲で変化させてトピック推定を行った後，話題集約の性能が最もよいトピック数を人手で決めた．その結果，本論文の分析において用いたクエリ・フォーカス「結婚」においては，$K = 60$ を採用した．LDA のハイパーパラメータ α および β の値については，GibbsLDA++のデフォルト値で

(注3)：http://gibbslda.sourceforge.net/

表 3　ノウハウを多く掲載する中国語ウェブサイトの種類

ノウハウサイト	QA サイト	商用サイト	事例サイト	合計
15	14	16	20	65

ある $\alpha = 60/K$，$\beta = 0.1$ とし，Gibbs サンプリングの反復回数を 2,000 に設定した．

3.2　ウェブページに対するトピックの割り当て

ウェブページ集合に対して LDA を適用することにより，各トピック z_n $(n = 1, \ldots, K)$ における語 w の確率分布 $P(w|z_n)$ $(w \in V)$，および，各文書 d におけるトピック z_n の確率分布 $P(z_n|d)$ $(n = 1, \ldots, K)$ を得ることができる．そこで，本論文では，次式によって，トピック z_n に対してウェブページ集合 $D(z_n)$ を割り当てる．

$$D(z_n) = \left\{ d \in D \,\middle|\, z_n = \underset{z_u\ (u=1,\ldots,K)}{\operatorname{argmax}} P(z_u|d) \right\}$$

3.3　候補ウェブサイトの選定

本節では，ノウハウを多く含む中国語ウェブサイトの候補サイトの選定手順について述べる．

本論文では，特に，文献 [4] の手法を用いて，ノウハウを多く含む中国語ウェブサイトの候補サイトを選定した．文献 [4] においては，収集されたウェブページ集合にトピックモデルを適用した結果において，各トピックの確率順上位 30 件のウェブページのドメインに対して，表 1 の基準に基づき，各ドメインがノウハウを含むサイトとなるか否かの判定を行った．表 1 に示す基準では，サイトのうち，ノウハウが含まれるか，および，どの程度ノウハウが存在しているかの二つの条件で A 群，B 群，C 群，D 群，E 群に分類する．そのうち，A 群，B 群，C 群のサイトはノウハウを一覧するページを含み，D 群，E 群はノウハウを一覧するページを含まないと定義する．特に，文献 [4] においては，複数のトピックにまたがって出現するドメイン (図 7) においては，ノウハウサイトが多く含まれるという仮説を用いた．その結果，仮説を満たすサイトの半分がノウハウサイトであるという結果となり，仮説の有効性が示された．

文献 [4] の結果をふまえ，本論文では，文献 [4] の仮説を採用し．各トピックの確率上位 30 件のウェブページのドメインのうち，複数のトピックに存在するドメインを収集して，候補ウェブサイト集合を作成した．クエリ・フォーカス「結婚」を用いて収集された検索エンジン・サジェスト数，ウェブページ数，および，複数トピックにまたがるドメイン数を表 2 に示す．そして，候補ウェブサイト集合に対して表 1 の基準に基づき，A 群 ~E 群の判定を行った結果，表 2 に示す A 群，B 群，C 群のサイト数となった．

4.　ノウハウを多く掲載する中国語ウェブサイトの分類

前節の手順により収集された候補ウェブサイトに対して，各サイトの特徴をふまえてウェブサイトの分類を行った．本論文では，特に，クエリ・フォーカスを「結婚」に着目し，それに関するノウハウ収集のためのウェブサイトを

表 2　クエリ・フォーカス，サジェスト数，ウェブページ数，および，トピック数

クエリ フォーカス	サジェスト数	ウェブページ数	トピック数	複数トピックにまたがるドメイン数	A 群	B 群	C 群	A・B・C 群合計
結婚	537	20,478	60	556	21	16	28	65

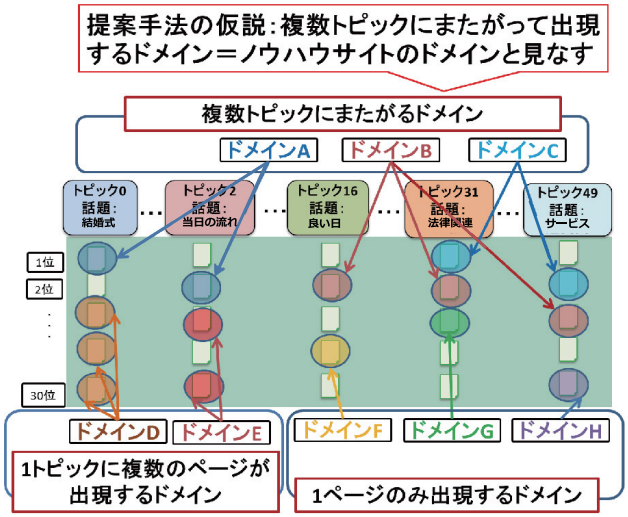


図 7　トピックモデルにおけるサイトの分布に基づく候補ウェブサイトの選定

（1）　ノウハウサイト
（2）　QA サイト
（3）　商用サイト
（4）　事例サイト

の四種類に分類した．

「ノウハウサイト」とは，多くのノウハウ知識を蓄積し，それらの知識をまとめて提示するウェブサイトである．例えば，ウェブサイト「wed114 結婚ネット」(図 3)(注4) は「結婚」専門のノウハウサイトであり，このサイトでは「結婚」に関する広範囲の知識が提示されており，閲覧者は多方面のノウハウ知識を得ることができる．

「QA サイト」は，ウェブ閲覧者が，自身の実情をふまえた質問を投稿し，他の閲覧者から回答を受けるという形式によって運営されているサイトである．「結婚」にとどまらず，あらゆる話題についての代表的な中国語 QA サイトとして，中国「百度知道」(図 4)(注5) が知られている．

「商用サイト」は多数の商品やサービスを掲載するサイトである．例えばサイト内の検索欄でクエリ・フォーカス「結婚」を検索することにより，結婚に関する商品・サービスが検索されるだけでなく，「結婚」に関する注意事項等のノウハウもまとめて閲覧することができる．「結婚」にとどまらず，あらゆる話題についての代表的な中国語「商用サイト」として，「58 同城」(図 5)(注6) がよく知られている．

(注4)：http://www.wed114.cn/
(注5)：https://zhidao.baidu.com/
(注6)：http://www.58.com/

「事例サイト」は，クエリ・フォーカスに関する過去の事例を掲載することにより，各閲覧者の利用状況に応じて，参考となる事例を通してノウハウや知識を提示するサイトである．「事例サイト」において結婚式の流れに関するノウハウや知識を知りたい場合には，「結婚式の流れ」に関する過去の事例を参照することが一般的である．「結婚」にとどまらず，あらゆる話題についての代表的な中国語「事例サイト」として，「百度文庫」(図 6)(注7) が挙げられる．

本論文においてクエリ・フォーカス「結婚」を対象として収集した候補ウェブサイト集合における上記四種類のサイトの数の内訳を表 3 に示す．

5. 素　　性

中国側においてノウハウを多く掲載するウェブサイトを同定するための素性として，表 4 に示すサジェスト素性，ウェブページ素性，トピックモデル素性，および，検索ボリューム素性を用いる．

5.1 サジェスト素性

本論文では，検索エンジン・サジェストを用いてウェブページの収集を行っているため，各ウェブページには検索エンジン・サジェストが対応付けられる．そして，各ウェブサイトには，検索エンジン・サジェストが対応付けられたウェブページの集合が掲載されていることから，結果的に，各ウェブサイトには検索エンジン・サジェストの集合が対応付けられることになる．そこで，各ウェブサイトのサジェスト素性として，各ウェブサイトに対応付けられるサジェスト集合中のサジェストの種類数，および，ウェブページ中の延べ観測頻度を用いる．

具体的には，4. 節で選定したサイト集合を T とし，T 中の一つのサイトを $t(\in T)$ とする．そして，サイト t に含まれるウェブページの集合を $P(t)$ とし，各ウェブページ $p(\in P(t))$ に対応付けられるサジェストの集合を $S(p)$ として，サイト t に対応付けられるサジェストの集合 $S(t)$ を次式で表す．

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

そして，サイト t が持つサジェストの種類数

$$|S(t)|$$

を素性として用いる．さらにサイト t に含まれるウェブページが検索される際に用いられたサジェストの延べ頻度

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

も素性として用いる．

(注7)：https://wenku.baidu.com/

表 4 素　　性

素性名	定義
サジェスト素性	当該ドメインのページを検索する際に指定されたサジェストの延べ頻度
	当該ドメインのページを検索する際に指定されたサジェストの種類数
ウェブページ素性	当該ドメインに対して収集されたウェブページ数
トピックモデル素性	当該ドメインのウェブページが出現するトピック数
検索ボリューム素性	当該ドメインのページを検索する際に指定されたサジェストの うち直近一ヶ月の検索数が最大となるものの検索数
	当該ドメインのページを検索する際に指定されたサジェスト 集合における直近一ヶ月の検索数の平均値

5.2 ウェブページ素性

当該ドメイン t に含まれるウェブページ集合 $P(t)$ のページ数

$$\left|P(t)\right|$$

を一つの素性として用いる．

5.3 トピックモデル素性

ウェブページ集合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 $p(\in P(t))$ に対してトピックの確率分布が得られる．そこで，確率値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．

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

ここで，各サイト $t(\in T)$ に対して，t に含まれるウェブページ p に割り当てられたトピック $z(p)$ を集めた集合 $z(t)$ を次式

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

で定義する．そして，サイト t に対応付けられたトピック集合 $z(t)$ 中のトピックの種類数

$$\left|z(t)\right|$$

を素性として用いる．

5.4 検索ボリューム素性

検索エンジン・サジェストの検索ボリュームを素性として用いる．検索エンジン・サジェストの検索ボリュームとは，検索エンジンにおける一定期間でのサジェストの検索数である．検索エンジン会社はサジェストが検索された頻度によってウェブ利用者の関心事項および関心度合いを分析することができる．本論文では，中国の検索エンジン「百度」によるツール「百度指数」(注8) により，一つのサジェストあたり直近一ヶ月の検索数を収集する．そして，一つのサイト t に対応付けられるサジェストの集合 $S(t)$ における直近一ヶ月の検索数の平均値および最大値を素性として用いる．

6. 評　　価

6.1 評価手順

表 2 に示す「複数トピックにまたがるドメイン数」556 ドメインを対象として，「ノウハウを含むサイト」か否かの予測を行い，その予測結果の正誤を評価する．「ノウハウを含むサイト」の参照用集合 R は，表 2 における A・B・C 群合計 65 サイト (表 3 の合計 65 サイト) から構成される．「ノウハウを含むサイト」の予測手法としては，前節で述べた各素性を用い，各素性の値の降順に評価対象サイトを順位付けする．具体的には，素性の値 $conf$ が下限値 c 以上となる評価対象サイトの集合を $T(conf \geqq c)$ として，$T(conf \geqq c)$ 中における参照用集合 R の要素の含有数を求め，再現率 $(conf \geqq c)$，および，適合率 $(conf \geqq c)$ を次式で定義し，評価を行う．

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

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

6.2 評価結果

図 8 に評価結果を示す．相対的には，検索ボリューム素性 (最大値・平均値) 以外の素性を用いた場合に高い性能となった．また，検索ボリューム素性においては，最大値を用いた場合の方が平均値よりも高い性能となった．

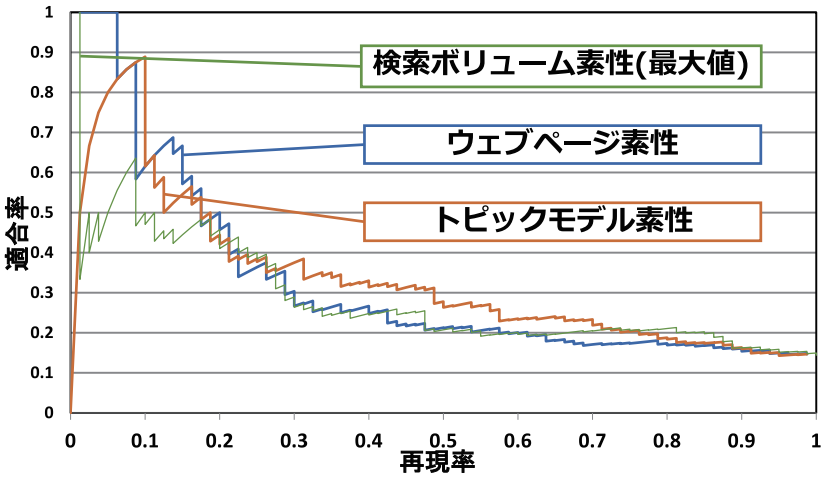
7. 中国語ウェブサイトにおけるノウハウ分布の分析

ノウハウを多く掲載する四種類の中国語ウェブサイトに対して，全トピック (トピック数 $K = 60$) のうち，クエリ・フォーカス「結婚」に関するノウハウを掲載するページを多く含むトピックを人手で選定した結果，28 個のトピックが選定された．トピック数の降順 1~6 位のサイトにおけるそれらの 28 個のトピックの分布を表 5 に示す．中国語ウェブサイトにおけるノウハウ知識は，主に「ノウハウサイト」,「QA サイト」,「商用サイト」,「事例サイト」の四種類のサイトに分布しており，多様な種類のサイトにおいて「結婚」に関するノウハウが掲載されることが分かる．

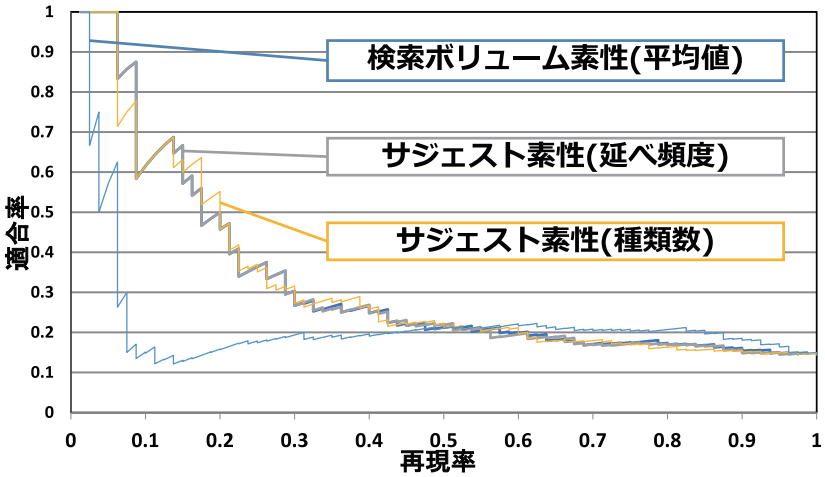
8. 関 連 研 究

本論文の先行研究として，文献 [2] では，検索エンジン・サジェストを用いて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することにより，各トピックのうち，サジェストが多く含まれるウェブページを提示することにより，多様な話題のウェブページ集合を提示する機構を実現した．そして，

(注8)：http://index.baidu.com/



(a) ウェブページ素性・トピックモデル素性・検索ボリューム素性 (最大値) の評価結果



(b) サジェスト素性 (延べ頻度・種類数)・検索ボリューム素性 (平均値) の評価結果

図 8　ノウハウを多く掲載する中国語ウェブサイト同定の評価結果 (各素性ごと)

文献 [6] においては，文献 [2] に基づき，トピックモデルだけでなく分散表現も併用し，検索エンジン・サジェストの集約における粒度を詳細化する手法を提案した．

ノウハウ収集・集約に関して，文献 [5] では，特定のクエリ・フォーカスに着目して LDA を適用し，各トピックのうち確率が高い文書の内容を分析することにより，クエリ・フォーカスに対するノウハウ知識を幅広く収集し，ノウハウ知識の候補の集約・俯瞰を実現した．文献 [7] では，文献 [4] で提案された手法で収集したサイト集合に対して，分類器の一つである SVM を適用することによって，ノウハウが多く含まれるノウハウサイト，および，ノウハウサイト以外の一般サイトを自動的に識別する手法を提案した．

これに対して，本論文では，中国で最もよく使用される検索エンジン「百度」を情報源として，ノウハウを多く掲載する中国語ウェブサイトを四種類に分類した．

9. おわりに

本論文では，特定のクエリ・フォーカスに対して，検索エンジンを情報源として，ウェブから検索エンジン・サジェスト，および，ウェブページを収集した．そして，ウェブページ集合に対してトピックモデルを適用することにより，複数トピックにまたがるサイトが，ノウハウが多く掲載されるサイトであるとみなし，候補ウェブサイトの集合を選定した．その後，いくつかの手がかりを素性として評価することにより，ノウハウを多く掲載する中国語ウェブサイトの識別性能を評価した．今後の課題としては，ノウハウが多く含まれるサイトにおける特徴を素性として用いて，分類器によって，ノウハウを多く掲載する中国語ウェブサイトを自動同定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] 井上祐輔, 今田貴和, 陳磊, 徐凌寒, 宇津呂武仁, 河田容英. 検索エンジン・サジェストおよびトピックモデルを用いたウェブ検索結果の集約. 第 8 回 DEIM フォーラム論文集, 2016.

[3] Q. Le and T. Mikolov. Distributed representations of sentences and documents. In *Proc. 31st ICML*, pp. 1188–1196, 2014.

[4] 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.

表 5　自動収集された候補ウェブサイト群中のトピック分布

トピックID	内容	トピック数降順順位/A 群，B 群または C 群/サイト種類/サイト名					
		1 位	2 位	3 位	4 位	5 位	6 位
		A 群	C 群	C 群	B 群	A 群	A 群
		ノウハウサイト	QA サイト	事例サイト	商用サイト	ノウハウサイト	商用サイト
		wed114 結婚ネット	百度知道	百度文庫	58 同城	婚礼猫	婚博会
1	披露宴サービスの提供	○			○	○	○
2	電子招待状の作成	○			○		○
3	サービスを提供する販売店のコツ	○			○	○	○
4	結婚当日の流れ	○	○	○	○	○	
5	結婚に関する商品の販売	○			○	○	○
6	結婚式の計画	○	○	○		○	
7	結婚でのメイク，ネクタイ等の選び方	○	○	○	○	○	○
8	結婚式の日の選び方	○	○	○		○	
9	新婚旅行先の推薦	○			○		
10	結婚に関する法律の提示	○	○	○		○	
11	結婚前の準備リスト	○	○	○	○	○	○
12	結婚式当日の花婿のスピーチ原稿	○	○				
13	結婚の心理学	○					
14	結婚のお祝いの言葉	○	○	○		○	○
15	地域別結婚習慣の違い	○	○	○			
16	結婚相手の両親に会う際の礼儀・持参品	○	○	○	○		
17	結婚後の生活	○					
18	結婚式での雰囲気作り	○		○		○	
19	結婚用部屋の飾り方	○	○	○	○		
20	結婚式費用	○	○	○	○		
21	結婚式でのゲーム		○	○			
22	プロポーズの仕方	○	○	○		○	
23	婚姻届の書き方	○	○	○			
24	結婚写真の選び方	○	○		○		○
25	祝儀の金額	○	○	○	○	○	
26	花婿の両親が用意する現金の金額	○	○	○			
27	結婚式の贈り物の注意事項	○	○	○	○		○
28	祝儀袋の表紙に書く内容例	○	○	○	○	○	

[5] 守谷一朗, 井上祐輔, 今田貴和, 聶添, 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.

[6] T. Nie, Y. Ding, C. Zhao, Y. Lin, T. Utsuro, and Y. Kawada. Clustering search engine suggests by integrating a topic model and word embeddings. In *Proc. 18th SNPD*, pp. 581–586, 2017.

[7] Y. Ohkawa, S. Kawabata, C. Zhao, W. Niu, Y. Lin, T. Utsuro, and Y. Kawada. Identifying tips Web sites of a specific query based on search engine suggests and the topic distribution. In *Proc. 3rd ABCSS*, pp. 4347–4353, 2018.