F-037

語句関連性とベイズ推論による対話型法律相談支援システム
Words Correlation and Bayesian Inference based Interactive Legislation Consultation Support System

水木 龍太郎† 黄 潤和†
Ryutaro Mizuki Runhe Huang

1. はじめに

法律は毎年新規追加，一部改正，廃止など把握するには困難な勢いで変化していく．人々がある日唐突にトラブルに巻き込まれ自分の身を守るために弁護士との相談は問題の解決には有効であるが，人間を介する解決法であり，インターネット上での膨大な相談者に対して解決に必要な知識を持つ回答者は少なく，回答者への負担が大きい．そこで本論文ではより欲しい情報を絞り込む，法律に関する相談の解決を支援する対話型システムを提案する．既存の法律相談や法律に関するテキストをそれぞれに形態素解析を行い，特徴語とその依存関係を視覚化した語句関連グラフを文書毎に生成する．ユーザーとシステム間の双方向の対話で関連している語句を推測し提案を繰り返して相談内容の語句関連グラフ化を行う．そのグラフに近似しているグラフを持つ文書をベイズ推定により推測し，ユーザーに相談内容と関連している文書を提案する．このシステムによって，容易に相談を解決するための情報を手に入れることができ，円滑に問題を解決する手助けになる．

このシステムでは対話に対話型遺伝的アルゴリズムが使われる．対話型遺伝的アルゴリズムとは遺伝的アルゴリズムの遺伝的操作を基にして，評価部分を人間の主観による提示個体の評価に置き換えたものである．解の評価を人間が行うため，進化の速度は低下し，世代交代数もあまり多く設定できなくなる．とはいえ，対話型進化計算は発想支援システムとして捉えることができ，解が不完全な部分解であってもそれをヒントに人間がアイデアを発想する，という可能性も考えられる．本論文では，対話型遺伝的アルゴリズムによって導かれたキーワードを用いて法律に関する文書を提案する．

2 システムの構造

2.1 システム全体の流れ

本論文で扱うシステムの全体図は図 1 のようになる．まず事前準備として Web 上の法律相談や法律に関するテキストを形態素解析で単語に分解し，単語から語句関連グラフを生成してデータベースに記憶しておく．ユーザーに年齢，職業，相談を 1 つの文章で入力してもらい，その文章も形態素解析で単語に分割する．その単語をデータベースに照合して発見された場合，その単語に関連するキーワードを探索してユーザーに提案して評価してもらい，要らないキーワードは順次交換を行いキーワード群を生成する．ユーザーとの対話により必要なキーワード群を決定したら，キーワード群からベイズ推論を用いて相談内容に合致する語句関連グラフを提案し，ユーザーの相談内容に役立ったのであれば良い評価としてデータベースにフィードバックすることでシステムの一連の流れとする．

† 法政大学情報科学部 Faculty of Conputer and Information Sciences, Hosei University

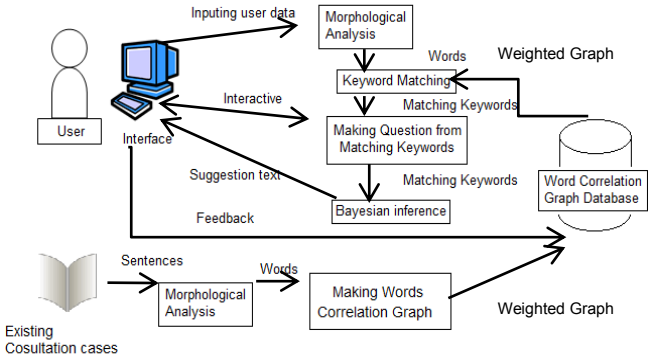


図 1.システム全体図

2.2 語句関連グラフ

本論文で提案するシステムは事前に Web 上から法律に関する文書や法律相談サイトでの相談文書を学習データとして収集し，形態素解析を用いて名詞や形容詞を抽出し，以下の式に従って語句関連グラフを生成する． 1 つの文書について語句関連グラフは複数のノードの繋がりとして表され，1 つのノードはノード間の繋がりを示す出現依存度とノードの大きさを表す吸引力の 2 つの要素を持ち，図 2 のように表すことが出来る．出現依存度 td(t,t')と吸引力 attr(t)は以下の式で計算される．

$$\mathrm{td}(\mathrm{t},\mathrm{t}') = \frac{sentences(t \cap t')}{sentences(t)}, \quad (1)$$

$$\mathrm{attr}(\mathrm{t}) = \sum_{t'' \in T} td(t'', t), \ (t \neq t'') \quad (2)$$

$sentences(t)$は文書中における語 t が含まれる文の数であり，$sentences(t \cap t')$は t と t'を同時に含む文の数である．吸引力は語 t が文書中の他の語を引き付ける力とし，他の語から語 t に対する出現依存度の総和として定義する．

各語間の出現依存度を指標とすることにより，文書に出現する二つの語の間には(i)双方向に強い依存度を持つ場合，(ii)一方向にのみ強い依存を持つ場合，(iii)双方向に弱い依存度を持つ場合の 3 つが考えられる．本システムではそれらに対して以下の行為を行う．(i)の場合ではそれらの語を密接な関係とみなして 1 つのノードに統合し，それらの吸引力も統合する．(ii)の場合は依存するノードから依存されるノードへリンクを張る．(iii)の場合では出現依存関係は無いとみなす．そうすることで，文書の内容をキーワードとそれらの依存関係を持ったグラフとして表している(図 2)．また，システムは全ての語句関連グラフを総合した独自の特徴語グラフを持ち，学習データ全体での語句の関連性データを持つ．例として法律相談サイトである弁護士ドットコムのある相談が掲載されているページをグラフ化す

Copyright © 2012 by
The Instiute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.

る(図 2). このケースは個人情報の漏洩についての相談であり，「個人情報」や「メールアドレス」と言った単語が頻出している．「メールアドレス」や「住所」といった単語が出る文章には「個人情報」という単語があり，一方的な依存関係が存在したのでリンクが張られた．また，「住所」と「電話番号」は同じ文章に常に含まれたため，同一ノードにまとまり，吸引力が統合された(図 3). 語句関連グラフを用いることによって，キーワードの関連性を表しつつ抽出できるため，キーワード提案の際に連想される用語を順に取り出すことが可能になる．

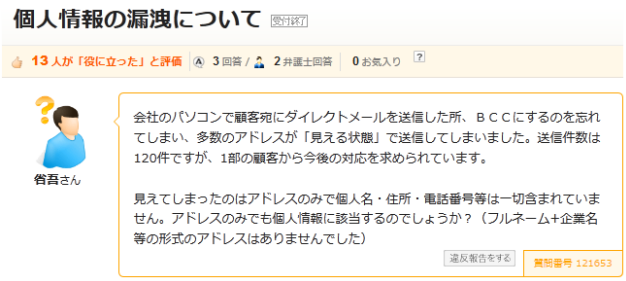


図 2 .学習用相談ケース

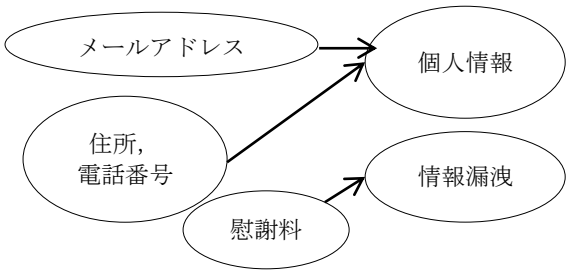


図 3 .語句関連グラフ化の一例

2.3 対話形式

本システムではユーザーに職業，年齢，相談したい内容を 1 文で入力して貰い，文章に形態素解析を行った結果抽出される名詞からユーザーの相談内容に関係の有りそうなキーワードを特徴語グラフから複数探索して提案する．具体的には，キーワードを持ったノードを特徴語グラフから見つけ出し，そのノードにリンクが張られているノードを吸引力の高い順に選択してキーワードを選出する．ユーザーは提案されたキーワード群から関係の有りそうなキーワードをできる限り選択しシステムに送信する．システムは選ばれたキーワードに関連の強いキーワードを選ばれなかったキーワードの代わりに改めて用意し，また稀に突然変異としてランダムに単語を選択したものを新しいキーワード群として生成して提案する．選択されたキーワードが十分な数が集まるまでこの対話を繰り返す．例えばユーザーが「詐欺で悩んでいる」と入力した場合，「詐欺」に反応して「オークション」「ワンクリック」などと言った単語を提案し，「オークション」が選ばれたらまたそれに関連する「出品者」「落札者」といった単語が並ぶようになる．

2.4 ベイズ推定による文書提案

まずユーザーから情報を入力された段階で選ばれたキーワード群から最尤法を用いてそれぞれの語句関連グラフが相談に合致する確率を計算する．大抵の場合，それだけで提案候補を決定するほどの確率にはならない．しかし，3.3 で説明した対話を繰り返すことで必要と判断されたキーワードと不必要と判断されたキーワードが判明していく．そこでベイズ推定を用いて合致確率を再計算することによって提案候補の確率を対話毎に動的に変動させる．確率は次の計算式で求められる．

$$P(A|X) = \frac{P(X|A)P(A)}{\sum_B P(X_B)P(B)} \quad (3)$$

P(A|X)は対話後の文書 A が提案される確率，P(A)は対話前の文書 A が提案される確率であり，P(X|A)は尤度である．

一定の対話回数，もしくは合致確率が十分高い候補が現れた場合，その候補文書を提案する．その文書がユーザーを満足させられるものでは無かった場合はキーワードを再収集するが，キーワードが限界まで追加され新しく追加されなかった場合は合致確率が高く算出された順に提案を行う．例えば，「オークション」「詐欺」というキーワードだけではオークション詐欺の相談までしか絞れないが，「落札者」「送料」までのキーワードが判明すれば 送料問題で悩んでいるというより詳細に相談内容に近い文書を絞り込むことができる．

2.5 フィードバックによる強化学習

提案された文書が有用とユーザーに判断された場合，ユーザーの情報及び入力に使われた単語を提案された文書の語句関連グラフにユーザーの情報を語句関連グラフに追加する．既存単語は単語の出現回数に 1 を足し，新出単語は新規ノードとして追加する．それにより有用なキーワードは，よりユーザーに提案されやすくなるように学習される．

3.評価

本論では形態素解析で抽出する品詞は名詞のみとしたが抽出する品詞を変更するとシステムの信頼度に影響を与える可能性がある．例えば名詞に加えて形容詞を抽出してキーワードとして使用した場合，個人の嗜好が表れやすくなるが，互いに相反するような感情が同じ文書に出現すると提案確率の上昇を阻害する可能性がある．また，流行の相談に対応するためには学習データは頻繁に更新する必要がある．

4.結論

本論文では，語句関連グラフやベイズ推定を用いた対話型提案システムを提案した．このシステムが実現化すれば，一般の人々にとってキーワードが分からなくてもより簡易な操作で相談に合う文書を提案することが出来る．規模の小さい相談であれば自己解決力の上昇も見込める．また，このシステムは学習型だが既に存在する文書を学習データとして利用できるので他分野での利用も考えられる．

参考文献

[1] 井上 博行, 袁 丹, 岩谷 香栄 , “対話型進化計算による配色作成支援システム”, 日本知能情報ファジィ学会誌, Vol.21, No.5 (2009)
[2] 小野,本村,麻生, “ベイジアンネットによる映画コンテンツの推薦方式の検討”,電子情報通信学会技術報告,vol104,no.NC2004-66,p55-60,2004
[3] 森 治憲, “ベイズ法における事前分布の持つ情報の評価法”,日本統計学会誌,第 40 巻,第 1 号,2010 年 9 月,p1-22
[4] Bickel, Peter J. and Doksum, Kjell A. (2001). *Mathematical Statistics, Volume 1: Basic and Selected Topics* (Second (updated printing 2007) ed.). Pearson Prentice–Hall. ISBN 0-13-850363-X.

Copyright © 2012 by
The Instiute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.