

論拠推定のための主観的推論知識の学習

稲垣 隆史

石塚 満

東京大学大学院情報理工学系研究科

1 はじめに

現在、推論知識獲得の研究が盛んに行われている。推論ルールの知識を獲得することで、ある事象の後に起こる出来事を予測したり、逆に結果から原因を推定することが可能となる。これらの研究は、誰の目にも正しい“常識”を獲得することに主眼を置いている。対して本研究が目指すのは、人によって真偽の異なるような知識、すなわち主観的推論知識を獲得することである。

主観的推論知識は、主に議論などで結論を導く論拠として現れる。様々な論拠を知識として獲得することで、新しい議論的となるような主張がなされたとき、それに対してあり得る様々な考え方を推定することができ、議論における視野狭窄や議論漏れを防ぐことができると考えられる。しかし、主観的に書かれた文章は比較的くだけた表現や構文が多く使われるため、既存の研究と同様の方法では精度の良い抽出はしにくいと考えられる。そこで本研究では、議論コーパスと文書要約技術を用いた論拠の獲得手法を提案する。

2 議論コーパスからの推論知識抽出

2.1 議論サイトの利用

本研究が扱う主観的知識はその性質上、獲得のために Wikipedia などの主観を含みにくいコーパスを利用することができない。したがって、人間が自由に書いた文章から知識を抽出することになるが、その場合、くだけた表現やノイズの多い文章を扱わなければならない。「because」や「since」などの語彙パターンや言語的情報のみを使用しての論拠抽出の精度は非常に低くなると考えられる。そこで、コーパスとして議論サイトを利用することを提案する。ここでいう議論サイトとは、議題に対して利用者が賛成もしくは反対の立場をともに意見を投稿することのできるサービスとする。コーパスとして議論サイトを利用することの利点は、推論知識抽出を言語処理的に行う際の、精度を下げる要素を一部緩和できることにある。まず、議題はそのまま推論知識の結論として抽出することができる。また、利用者によって書き込まれた意見は議題に賛成する、あるいは反対する理由を表すものと考えられるので、明示された立場とともに、議題に対する論拠として抽出することができる。

2.2 文書要約を用いた論拠文判別

議論サイトに投稿される書き込みには、冗長な例示や繰り返しも多い。そこで、投稿された複数文の書き込みの中から論拠として必要十分な文を抽出することを考える。ここで、書き込みのうち重要度の高い文に論拠が示されていると仮定すると、このタスクは文書要約技術を用いて解くことができると考えられる。本研究では、複数文からなる書き込みの系列に、各文が論拠を表すかどうかを意味が重複しないようにラベル付けするタスクに取り組む。系列ラベリング問題には条件付確率場 (CRF) がよく利用されており、その手法はデータ x が与えられたときのラベル割り当て y の確率を

$$P(y|x) = \frac{1}{Z(x)} \exp(\mathbf{w} \cdot \Phi(x, y)) \quad (1)$$

Learning Subjective Knowledge for Opinion Reasoning
Takashi INAGAKI, Mitsuru ISHIZUKA, Graduate School of Information Science and Technology, University of Tokyo

表 1: 議論サイトに書き込まれた書き込みの例。意味の重複が多く、論拠として分かりづらい。

議題	Should suicide become a legal human right?
書き込み	I agree that suicide should be a legal human right, because everyone should be able to have control of their body. When someone wants to die, they should be able to make that decision for themselves and for their body.(中略)People should not be able to impress their own moral or religious views on others to prevent them from ending their life.

とモデル化し、条件付確率が最大となるような割り当てを出力するものである。ここで $\frac{1}{Z(x)}$ は正規化のための定数、 Φ は素性関数である。本手法では CRF の素性に、単一文に関するものとして、文の長さ、書き込み中の文の位置、大文字の割合、議題との類似度といった情報の他、「because」「since」などの論拠を特徴付けるような単語を用いた。また、2つの文にまたがる素性として、2つの文の類似度を取り入れた。2文 x, y 間の類似度として、

$$\text{sim}(x, y) = \sum_{(w_1, w_2) \in M(x, y)} \frac{s(w_1, w_2)}{|x| + |y|} \quad (2)$$

を設定した。ここで $M(x, y)$ は文 x および y に含まれる単語の任意の組の集合である。 $s(w_1, w_2)$ には概念辞書 WordNet¹ の path 類似度、Leacock-Chodorow 類似度、Wu-Palmer 類似度を使用した。CRF の実装として、mallet²を使用した。

3 実験

3.1 データセット

議論サイト Debate.org³ 内で行われている 1835 の議論から、議題および書き込みを抽出し、そのうちの 200 の書き込みを無作為に選択し、2人のアノテータにより、その書き込みの各文について、必要十分な論拠であるかどうかのラベル付けを行った。そして、2人のアノテーションが一致した書き込みと、片方がもう片方を含むような書き込みにおける共通部分の文を、論拠として必要十分な文の正解データセットとした(データセット A とする)。その結果、119 の書き込み、628 の文が得られた。そして、比較手法用のデータとして、データセット A のデータについて、文単位で論拠として有効な文かどうか、意味内容の重複を許すラベル付けを行った(データセット B とする)。

3.2 評価方法

データセット A に対して提案手法を適用し、論拠抽出の性能を評価した。比較手法として、データセット A に対して単一文の素性のみを用いたサポートベクトルマシン (SVM) による識別結果を評価した。また、データセット B に対し

¹<http://wordnet.princeton.edu/>

²<http://mallet.cs.umass.edu/>

³<http://www.debate.org/opinions/>

てSVMによる文単位の論拠識別の後、類似度の高い文を群平均法を用いてクラスタリングし、SVMのスコアが高い1文のみを抽出した結果も評価した。その際、クラスタの類似度には式(2)を使用し、関数 $s(x,y)$ にはWordNetのpath類似度を用いた、クラスタを併合する閾値は、F値の最も高くなる0.8とした。クラスタの代表ベクトルには、SVMにおけるスコアの最も高いものを選択した。評価は文単位でのPrecision, Recall, F値の他に、書き込み単位で正しく抽出できた正解率を比較した。SVMの実装はLIBSVM⁴を使用し、CRFによる提案手法、SVMいずれも5分割交差検定で学習および識別を行った。

3.3 結果

表 2: 論拠抽出の評価

	提案手法	SVM	SVM+Clustering
文 Precision	.555	.337	.250
文 Recall	.418	.385	.549
文 F 値	.477	.359	.343
書き込み正解率	.416	.269	.126

CRF, 及び比較手法による論拠文抽出の結果を表2に示す。CRFによって行った論拠文抽出の結果は、文単位のPrecisionでSVMによる手法と比べて22パーセント、SVMとクラスタリングを併用した手法と比べて30パーセントも良いものとなった。また、書き込み単位の正解率で見ても、CRFによる論拠文抽出が比較手法を大きく超える結果となった。文単位の論拠識別と、意味の重複する文の除去を別々に行うよりも、提案手法における文間の類似度という形で、まとめて判別の方が良い結果となることが示された。

提案手法のCRFによる学習では、文の長さ、単語「because」を含むかどうか、文が先頭にあるかどうかという素性に、高い重みがつけられた。また、2つの文の類似度が高い場合、どちらかの文のみを残し、もう一方は抽出しない方向に重みが学習され、これらの素性が、必要十分な論拠抽出に有効な素性であるということが示された。

提案手法によって正しく抽出できた、また適切に抽出できなかった論拠の例を表3および表4に示す。適切に抽出できたものは、ほとんどが書き込みの先頭1文のみという形であったが、この例では2文目を適切に抽出できている。これは、文の長さの素性によって、より内容の充実した文に高いスコアを与えることができたからだと考えられる。適切に抽出できなかったものは、先頭にあるという素性に影響されて、適切な文を選択できなかった例である。抽出例を見ると、先に述べたいくつかの素性に影響を強く受け過ぎている傾向があることが分かる。書き込み単位の正解率が0.416と、高い数字とはいえないことと共に、より有効な素性の導入する必要性を示唆している。

4 関連研究

利用者が自由に書き込める形式のサイトを要約する研究には、Chan[1]らの研究がある。この研究は、QAサイトの書き込みをCRFを用いて要約しており、文の長さや位置、文間の類似度などの本研究でも用いた素性の他、投稿者のレートなどテキスト以外の情報を多く素性として取り入れているが、本研究で扱っているようなテキスト中の単語などの情報は素性として用いていない。

Wang[2]らの研究では、意見を要約するための素性として、トピックとの関連度を用いている。また、文が主観的か客観的かを識別するために、感情分析の研究を用いているが、本研究で扱っている主観的意見とは、トピックにこだわることなく様々な視点から考えた意見であるという点や、必ずしも感情を含むものではないという点で異なっている。

⁴<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

表 3: 適切に抽出された推論ルールの例

議題	Is multiculturalism a better social model than assimilation?
書き込み	Multiculturalism is watering down the individual differences and uniqueness found in different nations. Multiculturalism is not a better social model because it waters down the uniqueness and individual differences that a nation possesses. It is much better for people to assimilate into the chosen society they are joining is better than trying to find of multicultural mix that takes away the individuality of the nation.
抽出された文	Multiculturalism is not a better social model because it waters down the uniqueness and individual differences that a nation possesses.

表 4: 適切に抽出されなかった推論ルールの例

議題	Should breeds of dogs that are considered dangerous, like the Pit Bull, be banned?
書き込み	Pitbull killings are getting out of hand. It might be the owners, but for some reason, pitbull killings of other animals and children is getting out of hand. I'm not sure they should be banned but, they need to be put away when they attack other animals or humans.
正解の文	It might be the owners, but for some reason, pitbull killings of other animals and children is getting out of hand.
抽出された文	Pitbull killings are getting out of hand.

5 おわりに

本研究では、コーパスとして議論サイトを利用し、文書要約技術を用いた主観的推論知識の獲得手法について述べた。また、実験によって提案手法の論拠抽出性能を評価し、比較手法より良い結果が得られることを示した。論拠推定を行うためには、本研究で用いた手法に加えて、入力された議題に対して可能性のある論拠を提示する必要がある。未知の議題にも対応できる形で知識を獲得するための手法としては、概念体系であるオントロジーの利用が考えられる。また、未知の議題に対して獲得した知識を適用して論拠を出力するには、議題間の類似度を適切に設定することが必要である。これらのタスクについては、論拠抽出の性能向上と合わせて今後の課題としたい。

参考文献

- [1] Wen Chan, Xiangdong Zhou, Wei Wang and Tat-Seng Chua Community Answer Summarization for Multi-Sentence Question with Group L_1 Regularization In *Proc. of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 582-591, 2012.
- [2] Dong Wang and Yang Liu A Pilot Study of Opinion Summarization in Conversations In *Proc. of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 331-339, 2011.