Supplementary Materials for "Why Question Answering using Sentiment Analysis and Word Classes"

Jong-Hoon Oh* **Kentaro Torisawa†** **Chikara Hashimoto ‡**
Takuya Kawada§ **Stijn De Saeger¶** **Jun'ichi Kazama‖** **Yiou Wang****
Information Analysis Laboratory
Universal Communication Research Institute
National Institute of Information and Communications Technology (NICT)
{*rovellia,†torisawa,‡ch,§tkawada,¶stijn,‖kazama,**wangyiou}@nict.go.jp

1 Causal Relation Acquisition for Causal Relation Features

(顔ダニ, にきび) (face mite, acne)	(アクネ菌, 吹出物) (Propionibacterium acnes, rash)
(シラミ, 皮膚病) (pediculus, skin disease)	(バイオフィルム, 歯周病) (biofilm, periodontal disease)
(環境汚染物質, 花粉症) (pollutant, hay fever)	(活性酵素, 肌トラブル) (active enzyme, skin problems)
(食品成分, アレルギー) (food constituent, allergy)	(刺激成分, 乾燥肌) (stimulus, dry skin)
(アクネ菌, にきび) (propionibacterium acnes, acne)	(ダニ, ダニアレルギー) (acari, allergy for acari)
(酸, 虫歯) (acid, cavity)	(化合物, 尿臭) (chemical compound, uraroma)
(過酸化脂質, しみ) (lipoperoxide, liver spot)	(メラニン, くすみ) (melanin, smell)
(メラニン色素, そばかす) (melanin pigment, freckles)	(アクネ菌, ニキビ) (propionibacterium acnes, acne)
(メラニン, そばかす) (melanin, freckles)	(たんぱく質, ニオイ) (protein, dry skin)
(悪玉菌, 口臭) (bad bacteria, halitosis)	(チロシナーゼ ソバカス) (tyrosinase, freckles)

Table 1: Top-20 causal-relation instances. Here we represent the instances as (A, B), where A is a word representing a cause of B.

De Saeger et al. (2009) proposed a weakly supervised method that uses class-dependent pattern induction for knowledge acquisition. It takes as input a number of seed patterns, such as "X causes Y" for the acquisition of causal relation instances, and learns a large amount of paraphrase patterns to extract relation instances from a corpus. These patterns are called class-dependent paraphrase patterns, as they place semantic word class restrictions on the noun pairs they may extract, like "X:chemical causes Y:disease." These class restrictions enable to distinguish between multiple senses of frequent but highly ambiguous patterns. For instance, given a class independent pattern "X causes Y" as seed, if we restrict X and Y in "Y from X" to the classes of chemicals and diseases (as in "cancer from cadmium"), the class dependent pattern "Y:disease from X:chemical" becomes a valid paraphrase of "X causes Y." Note that, other class restrictions of the same pattern (e.g., "Y:products from X:company", as in "iPhone from Apple") may not yield a valid paraphrase of "X causes Y." To obtain word classes they use a large-scale word clustering algorithm (Kazama and Torisawa, 2008), and rank each instance in the corpus according to a score based on the semantic similarity between the seed patterns

Freq.	Pattern	Freq.	Pattern
6215	Bの原因となるA (A is a cause of B)	1253	Bを防ぐA (A preventing B)
2665	Bを起こすA (A brings about B)	1225	Bを伴うA (A accompanies B)
2521	AはBを引き起こす (A causes B)	1216	AはBの原因になる (A is a cause of B)
2509	Bを引き起こすA (A causing B)	1033	Bの原因であるA A being a cause of B
1902	AがBを引き起こす (A causes B)	1003	AによりBする (doing B by A)
1707	AがBを起こす (A causes B)	953	Aに起因するB (B arising from A)
1271	AからくるB (A came from B)	883	AはBを起こす (A brings about B)
1269	AはBの原因となる (A is a cause of B)	861	Aに伴うB (B accompanied by A)

Table 2: Causal relation patterns used for acquiring the top-100,000 causal relation instances. Here A is a word representing a cause of B. Freq. represents the number of causal relation instances acquired by this pattern.

QS1	Q: 海水をバケツですくっても、無色透明な液体なのに、宇宙や宇宙ステーションなど、遠く遠くから見ると青く見えるのははなぜですか？ (Seawater looks a colorless transparent liquid if it is in a bucket. But why does it appear blue when we see it from space or space stations?)
	Q: 毎日学校に行く時は自転車の空気は１ヶ月経っても抜けないのに、休みになり乗らなくなると急に空気が抜けるのはなぜですか？ (I am using a bicycle every day to commute to school and I have never suffered from deflated tires of my bicycle for a month. But why are tires of my bicycle deflated so fast over the weekend or if I do not ride my bicycle?)
QS2	Q: 佐藤栄作がノーベル平和賞を受賞したのはなぜですか？ (Why was the Nobel Peace Prize awarded to Eisaku Satō?)
	Q: ハリウッドが映画で有名なのはどうしてですか？ (Why is Hollywood so famous for movies?)
	Q: 松下電器産業が社名をパナソニックに変更したのはなぜですか？ (Why did Matsushita Electric Industrial Co., Ltd. change its company name into Panasonic?)

Table 3: Samples of questions in QS1 and QS2

and each class dependent pattern the instance co-occurs with.

We applied this method to our target corpus (600 million Japanese web pages) and used the top 100,000 causal relation instances. In our experiments we use these 100,000 instances along with 490 causal relation patterns that acquired more than 10 causal relation instances in the top-100,000 results as causal relation features for B-Ranker+CR. Table 1 lists the top-20 instances by their extraction score and Table 2 lists the top-20 patterns by their frequency in the acquisition of the top-100,000 causal relation instances.

2 Questions in QS1 and QS2

Table 3 lists samples of questions in QS1 (obtained from the Japanese version of *Yahoo! Answers*) and QS2 (created by our annotators). As described in our paper, questions in *Yahoo! Answers* are aimed at human readers and users often "set the stage" by giving lots of background information about their question. Table 3 shows some typical questions in this style from QS1. On the contrary, questions in QS2 more compact than those in QS1.

3 Correct Question-Answer Pairs in Our Test Sets

Table 4 and 5 show samples of correct question-answer pairs in our test set, where the samples in Table 4 and those in Table 5 are distinguished by the condition whether a question has a sentiment phrase. Sentiment phrases identified by *opinion extraction tool* are represented with underlined strings and their polarities are shown in red for negative and blue for positive. As shown in Table 5, polarities of sentiment phrases in a question and its answer sentences are often in agreement if both the question and its answer candidate have sentiment phrases. This follows our hypothesis concerned with sentiment orientation and we think this can explain the improvement using phrase-polarities as features. Note that *opinion extraction tool* does not always correctly identify sentiment phrases or their polarities. For example, it assigned a wrong phrase-polarity, positive, to sentiment phrase コアラの数が減ってきているのは (is the number of koala bears decreasing) of Q8 in Table 5. In this case, it is hard to expect that this wrong phrase-polarity can contribute to finding correct answers of Q8.

References

Stijn De Saeger, Kentaro Torisawa, Jun'ichi Kazama, Kow Kuroda, and Masaki Murata. 2009. Large scale relation acquisition using class dependent patterns. In *Proc. of ICDM 2009*, pages 764–769.

Q1	なぜボジョレヌーボーの解禁日は全世界で同じ日なのですか？ (Why is Beaujolais Nouveau of the year released on the same day all over the world?)
A1	.... ボジョレヌーボーは世界的に人気が高いためメーカーが十分出来上がっていないにもかかわらず、出荷するようになってしまったので１９８４年から解禁日を１１月の第３木曜日に決められて、この解禁日までワインとしてきちんと造り込むことになった、という経緯があります。 ... (... Because Beaujolais Nouveau has been the most popular wine in the world, the manufacture rushed to ship the wine even though its process was not finished. For this reason, selling Beaujolais Nouveau of the year has been disallowed before the third Thursday of November since 1984. ...)
Q2	宮崎駿監督が映画の主要なキャストに職業声優の起用を避ける傾向があるのはなぜですか？ (Why does Hayao Miyazaki, a Japanese film director, have tendency not to cast professional voice actors for his film?)
A2	...『紅の豚』以降、自分の作品にプロの声優を殆ど起用せず、俳優、女優に依頼する事が多い。その理由は、かつて属した東映動画の草創期の長編作品が、 俳優、女優を起用していたことへの原点回帰であると言われている。... (... After producing the Porco Rosso, he has casted actors or actresses as a voice actor rather than professional voice actors for his work. It is known that this tendency comes from his experience when he made long films at the beginning of his career in Toei Animation. ...)
Q3	台風とか渦巻きの方向は決まってて北半球と南半球で逆だそうですがなぜこのような現象がおきるんですか？ (Why are the spiral directions of cyclones different according to Hemispheres of the Earth?)
A3	大規模なサイクロンの地上付近の回転方向は、北半球では反時計回り、南半球では時計回りと常に決まっています。 .... このような回転方向は、「コリオリの力」とよばれる、地球上の気体に働く見かけの力によって決まっています。 The spiral direction of huge cyclones is clockwise in the Northern Hemisphere, while counter clockwise in the Southern Hemisphere. ... This is due to the Coriolis force that affects atmospheric dynamics on the Earth.
Q4	ガムを噛みながらチョコ食べちゃうときがあるのですが、なぜガムが溶けてしまうのでしょうか？ (Why does gum base become melted when eating chocolate while chewing gums?)
A4	... ガムベースは、チクルなどの植物性樹脂やドイツで開発された酢酸ビニル樹脂に、弾力性を出すポリイソブチレンなどを加えて作られる。 酢酸ビニル樹脂は脂溶性。 一方チョコレートは油脂を含んでいる。 一緒に食べることで、ガムの組織がバラバラになり溶けていくということになる。... (... Gum base is made of polyvinyl acetate, polyisobutylene and plant resins such as chicle. Here, polyvinyl acetate is fat-soluble. On the contrary, chocolate has fats. If we eat chocolates when chewing gums, the fat-soluble polyvinyl acetate in gum base reacts with the fats in chocolates and, as a result, the structure of the gum is broken. ...)

Table 4: Samples of correct question-answer pairs in our test set, where questions have no sentiment phrase

Q5	なぜ輸入品の値段が上昇するとインフレ懸念が強まるのですか？ (Why does the risk of inflation increase if prices of imported goods rise?)
A5	... コスト・プッシュ・インフレは賃金・原材料費・地代等のコスト上昇が生産価格や販売価格に影響を与えて生じるインフレである。輸入品の高騰や海外インフレが原因となる場合もあります。... (...Cost-push inflation is a type of inflation caused by substantial increases in production and selling cost such as costs for wage, materials and rent. A sudden rise in prices of imported goods is one reason of the cost-push inflation.)
Q6	なぜ、アレルギー性鼻炎になるのですか？ (Why does allergic rhinitis occur?)
A6	...アレルギー性鼻炎は鼻に入ってきた抗原を排除しようとして、鼻粘膜が過剰に反応して引き起こされます。 ... (... Allergic rhinitis occurs when a human body responses allergens inhaled by an individual in order to eliminate the allergens which are detected by a sensitized immune system as harmful to the body.)
Q7	佐藤栄作がノーベル平和賞を受賞したのはなぜですか？ (Why was the Nobel Peace Prize awarded to Eisaku Satō?)
A7	.. 佐藤栄作元総理はその後ノーベル平和賞を受賞した。 受賞理由は外交交渉で沖縄返還を実現した事、そして日本の非核三原則政策を打ち立てた事であるという。.... (... After that the Nobel Peace Prize was awarded to Eisaku Satō, the former prime minister of Japan. This award is due to his two achievements: Japan's attainment of the reversion to Japan of the Okinawa Islands by diplomatic negotiations and settlement of the three non-nuclear principles in Japan. ...)
Q8	オーストラリアでコアラの数が減ってきているのはなぜですか？ (Why is the number of koala bears decreasing in Australia?)
A8	... 20世紀になって、オーストラリアのコアラの数は300万匹から8万匹にまで減少しています。 ハンティングや山火事、都市開発などが、このような悲しい結果を招いているのです。.. (... In the 20th century, the number of koala bears in Australia has decreased from 3 millions to 80 thousands. Hunting, wildfire and development lead to this tragedy. ...)

Table 5: Samples of correct question-answer pairs in our test set, where questions have at least one sentiment phrase. Note that opinion extraction tool assigned a wrong polarity to コアラの数が減ってきているのは in Q8.