

NTCIR-13
MedWeb コーパス・ガイドライン
Medical Natural Language Processing for Web Document



Version 2.0 (2017/07/04)

Organizers

Eiji Aramaki (Nara Institute of Science and Technology)
Shoko Wakamiya (Nara Institute of Science and Technology)
Mizuki Morita (The University of Okayama)
Yoshinobu Kano (Shizuoka University)
Tomoko Ohkuma (Fuji Xerox)

1 分類の手順

対象となる症状は以下の通りである。

| | | | |
|---------|------|-----|---------|
| インフルエンザ | 風邪 | 花粉症 | 下痢 |
| 頭痛 | 咳・たん | 熱 | 鼻水・鼻づまり |

¹該当症状が、²今現在、³本人とその周囲にあるかどうかを判別し分類した。全症状に¹共通の基準、²時間に関する基準、³空間に関する基準を設けた。また、症状ごとに、罹患性や原因性を加味しながら 症状別の判断基準を設けた。それら判断基準から症状ありと判断されたものを「ポジティブ (p, positive)」, それ以外を「ネガティブ (n, negative)」とした。

なお、本ガイドラインでは、分かりやすさのために、ある症状の判定結果がポジティブ (p, positive) の場合には (症状+)、ネガティブ (n, negative) の場合には (症状-) の形式のラベルとともに発言例を示す。また、「インフルエンザ」「咳・痰」「鼻水・鼻づまり」に関しては、スペースの都合上、それぞれ略した症状名を用いている (「インフル」「咳」「鼻水」)。

2 大部分の症状に共通の基準と症状別の例外

2.1 大部分の症状に共通の基準

ラベル付けでは、実用性を重視し、過去の情報（数週間前の情報など）や重篤でない症状（飲み過ぎによる頭痛など）はネガティブとするなど、医学的重要性を考慮する必要がある。このため、以下の3つの基準を原則とする。

【症状（の事実性）の適合】 まず、疾患に罹患したという事実が確認される必要がある。原則として、単に疾患名を含んでいるだけの話題としての発言は除く（ネガティブとみなす）。詳細は2.3節を参照いただきたい。

【時制の適合】 次に、過去の情報は、サーベイランスという観点からは意味がないと考える。ただし、ある種の疾患に関しては、24時間より前であっても、臨床的に重要視される場合がある。例えば、既存のインフルエンザのサーベイランスでは一週間単位で情報をまとめて配信しているため、これよりも細かい時間粒度であれば意味を持つ。どれくらい以前の情報を除くかであるが、原則24時間以内を目安として、当日の情報、または、その前日までの情報あれば、これをポジティブとみなす。詳細は2.4節を参照いただきたい。

【空間の適合】 最後に、罹患者の位置情報が特定される必要がある。発言者が罹患者本人である場合は、発言者の位置情報が発症場所になる。伝聞などの場合は、発言者の近傍にいると推測される場合はポジティブとし、そうでない場合はネガティブとする。近傍判定は同一県内かどうかを基準とする。詳細は2.5節を参照いただきたい。

以下の表に、これらをまとめたものを記す。

| | |
|----|-------------------------------------|
| 症状 | a. 症状を直接的に表現している場合 |
| | b. 症状を間接的に表現している場合 |
| | c. 否定されているが症状の事実が認められる場合 |
| | d. 回復過程の場合 |
| | e. 原因不明の場合 |
| | f. 症状の疑いのみが認められる場合 |
| | g. 症状がなくなった場合 |
| | h. 予想や過程のモダリティ表現が含まれる場合 |
| | i. 罹患とは関係がない話題の場合 |
| | j. 人以外の症状の場合 |
| | k. 比喩表現・ジョークの場合 |
| 時制 | a. 当日の症状である場合 |
| | b. 前日の症状である場合 |
| | c. 過去（一昨日以前）の症状である場合 |
| | d. ごく近い過去の症状と推測される場合 |
| | e. 継続または頻繁に起きている症状で、今現在も症状としてみられる場合 |
| | f. いつの時点の症状か不明な場合 |
| 空間 | a. 本人と同じ地域にいると分かる場合 |
| | b. 本人と同じ地域にいる可能性が高い場合 |
| | c. 集団の症状の場合 |
| | d. 不特定多数への呼びかけ、近傍にいない人の症状の場合 |
| | e. 特定の相手の症状の場合 |

2.2 症状別の例外

上記の原則には例外が存在する。例えば、「頭痛」の症状では、疾患を原因としない（「飲み過ぎ」など）発言が多い。疾患サーベイランスとしては、これらの発言はノイズとみなされるべきである。さらに、疾患ごとに多様な例外がある。前述の例では「飲み過ぎ」の他に、「服薬」「刺激物」「精神的」「乗り物酔」「つわり」「生理痛」などによる症状も除かれるべきであろう。このような例外を次の表にまとめる。

| | 症状 | | | |
|----------------|-------------------------------|-------------|---------------------------------|----------|
| | 症状疑い表現を含めるか ※ ¹ | 単語のみ表現を含めるか | 例外 | |
| | | | 症状とみなす | 症状とみなさない |
| インフルエンザ | | | | |
| 風邪 | ○ | ○ | | |
| 花粉症 | ○ | ○ | | |
| 下痢 | ○ | ○ | 食べ過ぎ 消化不良 飲酒 服薬 刺激物 | |
| 頭痛 | ○ | ○ | | 視覚嗅覚 |
| 咳・たん | ○ | ○ | 飲酒 刺激物 | |
| 熱 | ○ | ○ 微熱のみ | 花粉症 注射の副作用 | |
| 鼻水・鼻づまり | ○ | | 花粉症 | 気温変化 |

※¹「っぽい」「気がする」「かもしれない」「可能性がある」「前兆」「しそう」「予感」等の疑い表現が含まれるケース。ただし、発言のコンテキストも含めて判断するため、疑い表現の有無によりポジティブやネガティブの判定結果が一意に決まるわけではない。

2.3 症状の事実性の基準

a. 症状を直接的に表現している場合

症状を直接的に表現している場合が該当する。症状の強さの有無は問わず、軽い症状も含める。

| | |
|-------------------------------|--------|
| ううむ、腹の痛みと下痢はあいかわらずだ… | (下痢+) |
| 朝喉が痛くなるのは花粉症で口呼吸になってることがあるようだ | (花粉症+) |

b. 症状を間接的に表現している場合

通院、投薬、他症状併発などから症状の事実が間接的に推測される場合が該当する。具体的に対策している場合（対策をしようとしている場合も含む）は、ポジティブとみなす。花粉症やインフルエンザは症状に罹患する以前から予防のために対策する場合も多いため、具体的な行動を伴うかを精査する。

| | |
|-------------------------|-------------|
| 風邪薬を飲んで今から寝ます。 | (風邪+) |
| 食べたくもないラーメンを胃に入れ頭痛薬を飲む | (頭痛+) |
| 頭痛対策として好きな音楽を聴くこともある。 | (頭痛+) |
| 花粉症に効果があると噂の眼鏡を買ったぞ | (花粉症+) |
| 花粉症に処方するめっちゃまく薬誰かつくってー！ | (花粉症+, 鼻水+) |
| インフル対策に外出時は、マスクをしてるよ。 | (インフル-) |
| 花粉症対策にはヨーグルトがいいみたいだよー。 | (花粉症-) |

c. 否定されているが症状の事実が認められる場合

文面では症状が否定されていても、それと類する症状が伺える場合はポジティブとみなす。多くの症状について、一般の患者が十分な知識を持っておらず、見過ごしてしまう可能性を考慮する。診断や自己判断も含める。

| | |
|---|----------------------|
| にしても、頭が痛いwww風邪・・・じゃないよな、あんなに食べたし | (頭痛+, 風邪+) |
| 時期的にはヒノキ？目の痒みもあるので、単純に風邪と言い切れないのよね | (花粉症+, 風邪+) |
| 花粉症だと思ってたら、微熱が出てきて風邪だと判明！間違えることもあるんですね。 | (花粉症+, 風邪+, 熱+, 鼻水+) |
| 医者間違えて花粉症だって診断しやがった。 | (花粉症+, 風邪+, 鼻水+) |

d. 回復過程の場合

症状から回復する過程においても、症状は弱いとはいえ罹患しているといえるため、完全に症状がなくなっていない場合は、ポジティブとみなす。ただし、症状が完全になくなっている場合はネガティブとみなす。

| | |
|--------------------------------------|-----------|
| 頭痛吐き気状態から回復したっぽい | (頭痛+) |
| 週末をまるまる熱出して寝てた。週明けの今日、鼻水が少々出るけどほぼ復活。 | (熱-, 鼻水+) |

e. 原因不明の場合

| | |
|-----------------|-------|
| 最近すぐ頭が痛くなる。なんでだ | (頭痛+) |
|-----------------|-------|

f. 症状の疑いのみが認められる場合

患者が十分な知識を持っておらず、症状・疾患を看過してしまう可能性を考慮し、症状・疾患の疑いのみが認められる場合においてもポジティブとみなす（表現例：「っぽい」「らしい」「似た」「のような」「気がする」「かもしれない」「ぞみ」「可能性がある」「なのかな」「前兆」「しそう」「予感」）。

| | |
|--------------|-------|
| 風邪引く前兆出てるナウ | (風邪+) |
| すごい肩こり。風邪かも？ | (風邪+) |

g. 症状がなくなった場合

完全に症状がなくなった場合はネガティブとする（表現例：「治った」「おさまった」「止んだ」「克服」「復活」「完全回復」）。薬による治癒表現は一時的な治癒である可能性や服薬中又は服薬直後であることからポジティブとする。

| | |
|----------------------------------|--------|
| なんかがんばったら鼻づまり止んだ。 | (鼻水-) |
| 薬飲んだら咳とまった！ | (咳・痰+) |
| あ、下痢症状は治まりました。 | (下痢-) |
| 頭痛や気だるさも水飲んで一晩寝たらなおるよ！って事でなおったよ！ | (頭痛-) |

h. 予想や過程のモダリティ表現が含まれる場合

「つもり」「しかける」「したら」「予定」「だったら」などのモダリティ表現が含まれており、文脈として症状の事実性が認められない場合はネガティブとみなす。

なんかだるい・・・もしこれで咳と熱が出たら●●を疑うかも。校内で感染者が出て休校中なわけだしw (熱-)

風邪じゃありませんように *mb* (風邪-)

i. 罹患とは関係がない話題の場合

罹患とは関係なく、単なる話題として症状・疾患表現が出現した場合は、ネガティブとみなす。ニュース、一般論、広告などが該当する。

鼻水が固体化し始めたら風邪寛解の合図 (鼻水-, 風邪-)

News: 【新型インフル】大阪の感染者83%が38度以上の高熱 - MSN□□ニュース <http://□□> (インフル-)

j. 人以外の症状の場合

罹患とは関係なく、単なる話題として、症状・疾患表現が出現した場合は、ネガティブとみなす。ニュース、一般論、広告などが該当する。

今日は子犬さんが風邪気味なので早めに帰ってあったかいものを作ってあげたいと思います。... (風邪-)

朝からパンが熱暴走...例の電源が高熱になって落ちるヤツ。 (熱-)

k. 比喻表現・ジョークの場合

誤表記を含む。

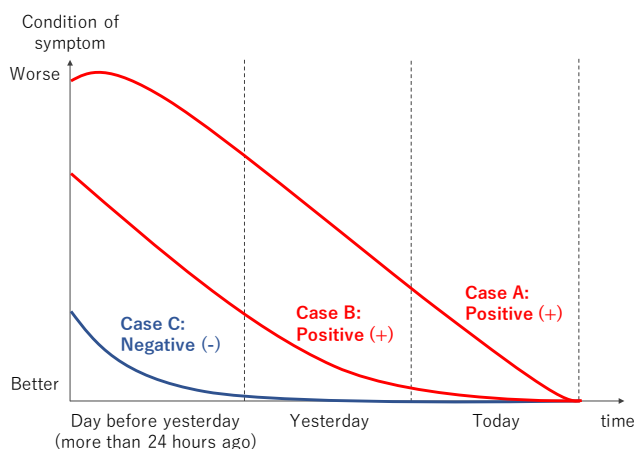
絶賛発熱中 (熱-)

台風並みに風邪が強いな (風邪-)

2.4 時制の適合

各種症状の出現は重篤度や個々による差はあるが、回復期に入ると緩やかに症状が治まっていくことが一般的である。そのため前日の症状であっても今現在わずかに残っている可能性や、治癒直後であることも考慮し、当日（24時間以内）あるいは前日の症状と分かるものは積極的に抽出する。

以下の図は症状の推移図であり、y軸は症状の度合い、x軸は時間である。ケースAとケースBはそれぞれ当日（24時間以内）と前日に症状が少なからず存在しているとしてポジティブとする。ケースCは24時間以上前の症状としてネガティブとする。



a. 当日の症状である場合

当日の症状であれば、24時間以内に症状が存在しているとして、今現在症状がなくてもポジティブとする。

| | |
|---|------|
| 朝微熱→研究室ミーティング欠席→熱さが→受け取れないはずの宅急便受け取り | (熱+) |
| 高熱が出てこれはもうダメだと思ったら4時間後に平熱に戻ったので仕事しに来たでござるの巻 | (熱+) |

b. 前日の症状である場合

前日の症状であれば、24時間以内に少なからず症状が存在しているとして、今現在症状がなくてもポジティブとする。

| | |
|---|-------|
| 昨日からの風邪が治らない。 | (風邪+) |
| 食欲がない以外は昨日熱があったのが不思議なくらい元気なんだけどな。平熱に戻ったし。 | (熱+) |

c. 過去（一昨日以前）の症状である場合

一昨日以前（24時間より前）の症状であればネガティブとする。

| | |
|---|------|
| こないだサウナに入って一切汗が出なくてそのあと少し激しいスポーツをしたらその夜かなりの高熱を出したから今日はサウナに入る前にポカリスエットをたくさん飲んだ | (熱-) |
| 更に発表会当日に39度近い熱が出たのも思い出した。一番目にしてもらって、速攻帰ろうと思ったんだけど、誰も順番変わってくれなかった。 | (熱-) |

d. ごく近い過去の症状と推測できる場合

いつの症状であるか明示されていない場合でも、当日あるいは前日の症状であると推測される場合はポジティブとする。

| | |
|------------------|-------|
| 最近すぐ頭が痛くなる。なんでだ | (頭痛+) |
| 謎の高熱で3日ほど死んでました。 | (熱+) |

e. 継続または頻繁に起きている症状で、今現在も症状としてみられる場合

| | |
|-------------------------------|-----------|
| 冬の間はずっと風邪をひいてる気がしてきた。痰が黄色い。 | (風邪+, 咳+) |
| うう……。豚肉食ったら、なぜかいつも下痢をする・・・orz | (下痢+) |

f. いつの時点の症状が不明な場合

状況の具体性から推測して判断する。

| | |
|--|-----------------|
| 焼き肉の翌日は下痢 | (下痢+) |
| 今年のインフルエンザは咳系統らしいですよ！って熱が出て咳が出まくって倒れた俺が言ってみる | (インフル-, 熱-, 咳-) |

2.5. 空間の適合

a. 本人と同じ地域にいると分かる場合

都道府県あるいはそれよりも小さい単位を地域とする。ただし、国内のサーベイランスを対象としているため、本人と同じ地域にいる可能性が高い場合でも、国外の場合は考慮しない。

| | |
|---|---------|
| 電車の中に激しく咳している人がいる。風邪うつさないでね。 | (咳+) |
| やばいやばいやばいよ・・・!!近くで新型インフルエンザ発症者出たらしい、ああ、恐いよ・・・ | (インフル+) |
| でも学校閉鎖は・・・正直おいしいです^p^ | |

b. 本人と同じ地域にいる可能性が高い場合

家族など、本人と同じ地域にいる可能性が高い場合には、同居別居問わず周囲の症状とする。伝聞の場合も含む。ただし、国内のサーベイランスを対象としているため、本人と同じ地域にいる可能性が高い場合でも、国外の場合は考慮しない。

| | |
|---|-------------|
| 今、家族で風邪が流行!!親父を除いて!!うがい手洗い大事な!! | (風邪+) |
| 娘が発熱しました。昨日じいさんばあさんと豊川稲荷に行った時にもらった風邪か? | (熱+) |
| 兄がインフルになったらしい。 | (インフル+, 熱+) |
| 風邪なのに仕事とかなんなの?移すの? | (風邪+) |
| 咳込んでわざと風邪をうつそうと頑張るのやめてほしいわー。うつらないけど*Tw* | (咳+, 風邪+) |

c. 集団の症状の場合

集団の範囲を特定できて、かつ、その集団が同じ地域にいる可能性が高い場合(学校、クラブなど)には、周囲の症状としてポジティブとする。それ以外の場合(ブログ仲間など)には、空間の適合判定が困難であるためネガティブとする。

| | |
|--|-------------|
| 部活のメンバー全員で握手会に行く予定だけど、インフルが絶賛蔓延中だから不安だよ。 | (インフル+, 熱+) |
| みんな下痢って、まさか集団食中毒?? | (下痢-) |
| 周りに咳してる人が多いときはいつも以上に手洗いを徹底しないと。 | (咳・痰-) |
| 犬のブログを通じて知り合った人が鼻づまりで息ができず苦しいらしい。たかが鼻水だけの鼻風邪と侮れない。 | (鼻水-) |
| センター試験直前なのに、学校内で風邪が流行るなんて大変だなあ。 | (風邪+) |

d. 不特定多数への呼びかけ、近傍にいない人の症状の場合

呼びかけや集団の範囲を特定できない場合には、空間の適合判定が困難であるためネガティブとする。

| | |
|-------------------------------|-------|
| 風邪なみなさんは、耳鼻科へ行くといいです。全部解決します。 | (風邪-) |
| なんか風邪ひいてる人が多い気がする[mb] | (風邪-) |

e. 特定の相手の症状の場合

特定の人との会話(RTや@□□等)において、相手方の症状に対する励ましや警告が多々見受けられる。この場合の相手方を周囲の人と解釈する。明らかに見知らぬ人の発言だと思われる場合はネガティブとする。

| | |
|---|------|
| 毎日のように熱がどう薬がどう体調が悪いだの mixi で書いてる子が、病院で休みを潰したくないから行かないとか書いてて、流石にちょっとツッコミを入れることにきめた | (熱-) |
|---|------|