

厚生労働科学研究費補助金（臨床研究等 ICT 基盤構築研究事業）
分担研究報告書

退院サマリの自動生成研究における研究基盤の整備

研究分担者 奥村 貴史
(北見工業大学 工学部・大学院工学研究科 教授)

研究要旨

臨床医は、入院治療をしていた患者が退院する際、それまでに記載していた入院カルテから退院サマリを作成する。この退院サマリの作成を効率化することができれば、医師の診療負担を直接軽減することが出来ることに加えて、様々な副次的な効果が期待される。そこで本研究分担では、この入院カルテの自動要約に向けた研究基盤の整備に取り組んできた。

今年度は、昨年度までの研究をさらに発展させ、4つの課題に取り組んだ。まず、一連の研究には自由に研究利用できるカルテの存在が必要となる。そこで、ダミーカルテの収集を進め、100件の整備を目指して活動を進めた。また、収集したダミーカルテを対象として、機械的な処理を可能とするためのアノテーション作業に取り組んだ。その際、初年度に行ったアノテーション、昨年度に試みたアノテーションを踏まえ、さらなる改善を図った。次に、このアノテーション作業と平行して、退院サマリの分析モデルであり生成モデルである「CASE モデル」の改善に取り組んだ。また、退院サマリの自動生成処理を精度管理するうえで必要となる「理想の退院サマリ」の確保に向けた検討を行った。

研究活動の結果、ダミーカルテは、目標を超える108件を集めることが出来た。また、これらのカルテを対象としたアノテーションを進めると共に、アノテーションガイドラインを高品質化することが出来た。さらに、このアノテーション済みカルテを用いて実現する退院サマリの分析に向けたモデルとして、修正CASEモデルを提示することが出来た。理想のサマリに向けた検討では、高品質な退院サマリを低コストに実現するための作業仮説を整理すると共に、実証に向けた研究デザインを策定することが出来た。今後、これらの研究基盤をベースとして、退院サマリの自動生成に向けたさらなる研究の発展を図りたい。

A. 研究目的

医師が患者を診察した際には、その情報を遅滞なく診療録に記載する必要がある(医師法第二十四条)。これがいわゆるカルテであるが、カルテには大きく2つの種類がある。1つは「外来カルテ」であり、外来診療において継続的に記載されていく。もう1つが「入院カルテ」であり、患者が

入院する毎に作成され、入院中の経過が退院に至るまで記載されることになる。入院カルテは、回診毎に記載されることも少なくなく、外来カルテとは記載の密度が大きく異なるうえ、長期入院などにおいては分量も大きなものとなる。そこで医療機関は、入院患者が退院した時点で入院中の経過を要約した文書を作成する。これが、「退院サマリ」である。

この退院サマリの作成は医師にとって負担であるため、退院サマリの作成を自動化することができれば医師の勤務負担を軽減することができる。そこで本研究分担では、この入院カルテの自動要約に向けた研究基盤の整備に取り組んできた。初年度においては、自動要約を実現するための処理モデル(CASE モデル)の提案を行った。CASE モデルは、退院サマリ中に含まれるセンテンス(文)が、「元カルテに由来するか否か」、「抽象度が高いか低いか」という2軸により分類しようと仮定したモデルである(図 1)。その有用性を示すためには、そもそも実際の退院サマリが本当にこの2軸分類により効果的に分類されるセンテンスにより成り立つのかを、実際の退院サマリの分析を通じて実証する必要がある。そこで2年目に、このモデルの実証に向けた研究基盤の構築を進めた。まず、一連の実証に求められる「ダミーカルテ」の収集を行った。また、退院サマリに含まれるセンテンスの自動分類を実現する分類器の構築を目指した。さらに、当該モデルに基づいて退院サマリを生成するための、ユーザーインターフェースの開発を行った。

3年目は、上記研究のさらなる進展に取り組んだ。ダミーカルテは、一定数揃わなければ有効活用することができない。そこで、研究協力者の医師を募り100件の突破を目指して活動を進めた。また、このカルテを研究活用していくうえで、カルテに含まれる文章に対して、機械処理が可能となるようにその解釈情報を付与していく必要がある。この「アノテーション」を、カルテに含まれる全てのセンテンスに対して行うことには多くの手間が掛かるが、今までに付与したアノテーション情報に技術的な課題が生じていることが明らかとなった。そこで、アノテーションの基準を再考したうえで、全データを対象に再度のアノテーションを行った。さらに、これらの検討を通じて、CASE モデルそのものにも課題が明らかとなったことから、CASE モデルの修正を合わせて行った。

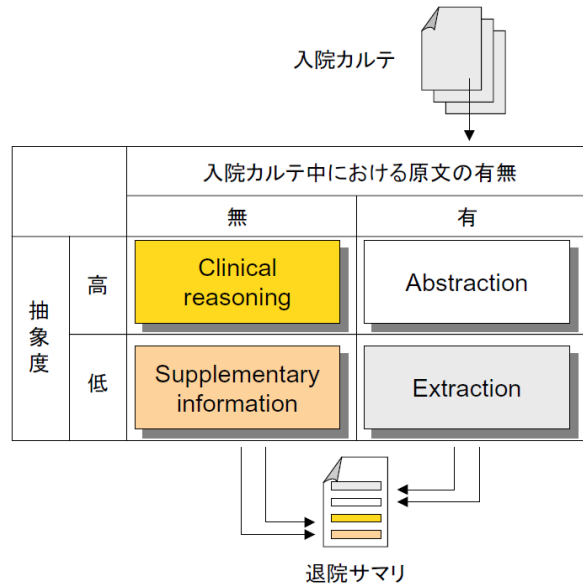


図 1 CASE モデル

本研究班では、上記のように、退院サマリの自動生成研究における研究基盤の整備を進めてきた。これによって、退院サマリの自動生成研究に関する各種の研究が進むことが期待される。一方で、退院サマリを自動生成したとしても、その精度が高まらない限りは有用な技術とすることができない。そのためには、生成した要約を客観的に評価できる必要がある。もし、個々の入院カルテに対して「理想的なサマリ」を用意することができれば、今後、生成したサマリを理想的なサマリとの差異によって評価することが可能となりうる。さらに、理想的な要約を処理系に学習させることで、さらなる精度向上に繋がる可能性もある。そこで、研究基盤整備の一環として、入院カルテから理想的な退院サマリを生成する手法についての検討を行った。

B . 研究方法

ダミーカルテ収集

ダミーカルテの収集には、多くの手間が掛かる。カルテを記載するには、医師の協力が必要となるが、多くの医師は、ひたすらデータを生成するような単純作業に興

味を示さない。また、医師の雇用には大きなコストが掛かる。医療用自然言語処理研究の発展が望まれているが、研究分野の停滞の一端は、この、研究利用できるデータの確保に多大な手間を要する点がある。この隘路を突破するためには、本研究の価値を理解し、データ生成に関わって頂ける研究協力医師を見つける必要がある。そこで、医師向けセミナー等にて研究状況を発表すると共に協力を呼びかけ、反応を示して下さった医師と詳細なやり取りを通じて研究協力頂く活動を続けた。医師の負担を少しでも減ずるため、提出データにおける課題の修正や欠損の補填は、別途、看護師の研究協力者へと依頼し、医師にはその作業結果の確認を依頼する形で作業を進めた。提出頂いたダミーカルテは、こうして内容をチェックしたうえで、その後の研究利用に適した状態へとするため、年齢、性別、主病名等の定型情報を整理し、XMLと呼ばれるフォーマットへと整理した。

カルテのアノテーション

また、協力医師による尽力によりダミーカルテを数多く用意することができたが、このカルテを研究活用していくうえでは、カルテに含まれる文章に対して、その解釈情報を付与していく「アノテーション」作業が必要となる。昨年までの研究においては、CASEモデルの実証に向けて、カルテ中の全てのセンテンスに対して「事実度」の情報を付与していた。ここで事実度とは、CASEモデルが想定する抽象度と逆の関係にある指標とする。そして、各センテンスに、評価者の判断に基づく「事実度の高低」に関するラベルと、そのセンテンスが肯定か否定かのいずれに関するものかの「極性情報」、その言及が確実なものか不確実なものかに関する「確信度」のラベルを付与していた。しかし、このラベルは、各センテンスを構文解析したうえで節レベルで評価したものでないことから、接続詞によって、1センテンス中に複数の極性や複数の事実度が混在する問題が生じて

いた。また、当該データを元にして試行的な分類器の作成を行ってみたものの、高い分類精度を出すことができずにいた。そこで、アノテーションの基準を再考したうえで、全データを対象に再度のアノテーションを試みた。

CASEモデル検討

本研究において当初提案したCASEモデルは、退院サマリ中の各センテンスを、「元カルテに由来するか否か」、「抽象度が高いか低いか」という2軸により分類した。このモデルには全センテンスを2×2表のいずれかに落とし込むことが出来る分類モデルであることに加えて、4つに分類されるセンテンスをいかに生成するかという生成モデルを兼ねている。たとえば、Clinical Reasoning文は、元のカルテに由来せず事実度の低いセンテンスを指す。このカテゴリの文は元カルテからは生成しえないが、カルテ作成者の他のカルテが同一診療科の他の医師のカルテより類似したセンテンスを拾える可能性がある。Assessment文は、当該センテンスに近い文を元のカルテに見出しうる、事実度の低い文であり、元のカルテに記載されている事実度の高い文、すなわち患者病態等を評価した文であることになる。したがって、要約に際した手法としては、Extractive(抜粋的)な要約ではなく、Abstractive(抽象化的)な要約により生成されることが示唆される。しかし、入院カルテにおいても、SOAPモデル(カルテ記載においてSubjective, Objective, Assessment, Planの各カテゴリ毎に整理する手法)により記載している限り、Assessment部分やPlan部分には事実度の低い評価や将来計画が記載されることになる。これらは、退院サマリにおいて、Extractiveに抽出され記載される可能性があることから、CASEモデルの前提が崩れかねない状況が生じていた。そこで、CASEモデルを拡張し、入院カルテと退院サマリの関係を整理したうえで、必要な検討を行った。

理想のサマリ研究

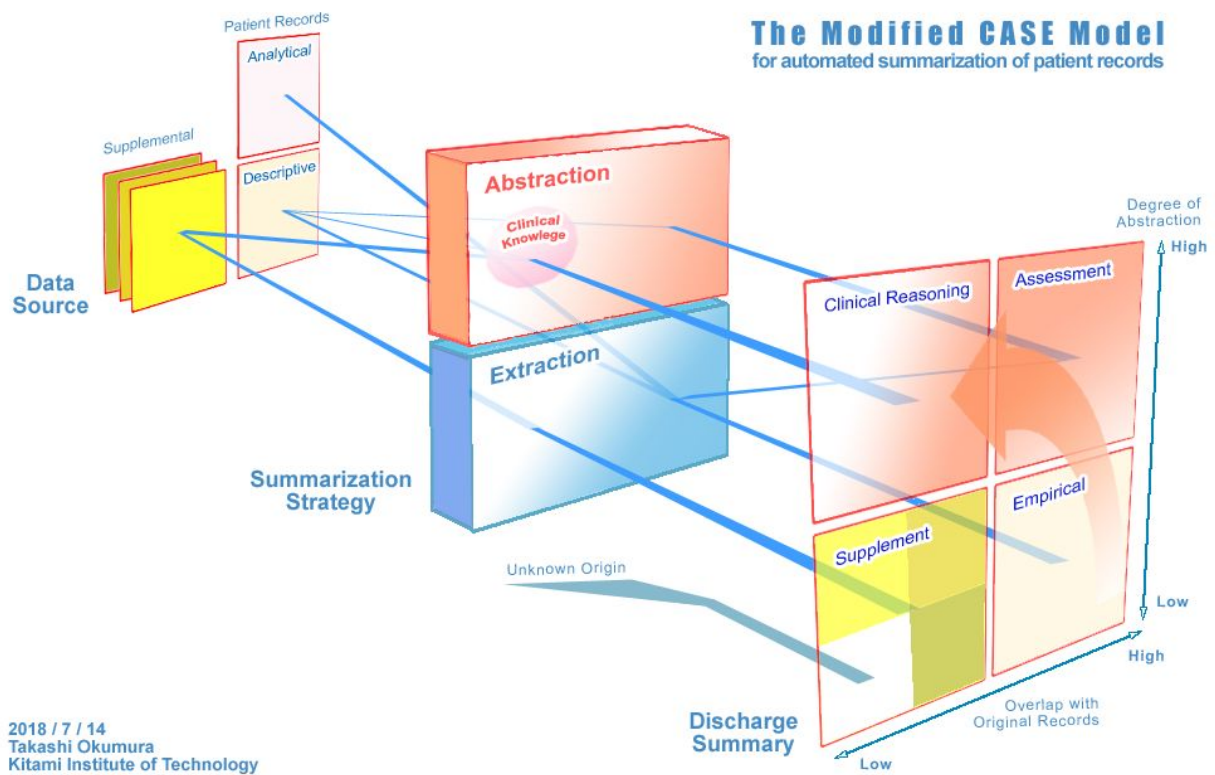
個々の入院カルテに対して「理想的なサマリ」を用意するためには、何が理想的なサマリなのかの定義をする必要がある。ただし、言及すべき事項を整理することをもって理想のサマリの定義とすると、疾患によってカルテに記載すべき内容も異なるため、疾患毎に理想の内容を定義する必要が生じ作業量が膨大となる。また、たとえばそのような定義を行ったとしても、患者には個体差があることに加えて、それぞれ合併症や同時に罹患している他の疾患がある。さらに、入院日数も異なれば、そもそもその入院カルテの質や量にも違いがある。したがって、理想のサマリを内容によって定義しようとする、そもそも「理想の入院カルテ」を用意する必要が生じ、医療現場の負担軽減を目指してきた本研究の方針にも反する結果となってしまう。医療現場の負担軽減を図るためには、当面の間は、現場において記載された現状の入院カルテを対象として理想的な退院サマリを生成する必要がある。

そこで、我々は、理想のサマリの定義として、何が書かれているかの内容に関する定義は行わず、代わりに、「入院カルテと退院サマリを第3者の医師に添削させ、その結果をさらに別の医師に添削させると

いう作業を繰り返し、修正が行われなくなったサマリ」を、理想のカルテであると定義した。これにより、疾患の種別や患者側の既往歴や個人差、入院日数の違いなどに関わらず、理想的なカルテを客観的な形で定義することが実現する。

一方、このような操作的な定義の場合、本当に「安定」した退院サマリが得られるのかという懸念が生じる。たとえば、添削者によって、とある記述が削除された後に、別の添削者によって同じ記述が加筆された場合、退院サマリは「振動」することになる。あるいは、添削者によって加筆や削除するポイントが異なる場合、退院サマリは収束せず、逆に「発散」することになる。糖尿病の教育入院やクリニカルパスに基づく入院の場合には、カルテ記載も定型的になる可能性が高いため、安定サマリが得られやすい一方、診断困難症例の検査目的での入院などでは発散しやすい等、入院の種類によってもこの傾向に差がでることも考えられる。

そこで、この「理想のサマリ」を操作的に定義することにより、本当に理想のサマリを得ることができるかの検証を進めた。具体的には、上記の検証を行うための研究プロトコル策定に加えて、医師を対象として検証実験を行うための検討を進めた。



2018 / 7 / 14
Takashi Okumura
Kitami Institute of Technology

図 2 修正 CASE モデル

C . 研究結果

ダミーカルテ収集

ダミーカルテの収集と編纂には、全ての過程に医療従事者の関与が求められることに加えて、作業者の錬度が重要となることから、必然的に高コストかつ時間を要することになる。本研究分担では、2年強に渡る広報活動と収集活動の結果、入院カルテ・退院サマリのペアを含むダミーカルテを108件、収集することができた。これらのファイルは、電子カルテよりSS-MIX2形式にて抽出したカルテデータと等価に扱うことができるよう、XML化したうえで、入院日、退院日、主疾患等のいくつかのメタデータを付与した形となっている。データの品質は高く、一見したところ普通のカルテ情報と遜色がない水準とすることができている。

カルテのアノテーション

新たなアノテーション方式の設計に際しては、旧方式よりも少しでもシステムティ

ックな分類が実現するよう工夫を凝らした。まず、センテンスが言及している内容を、カルテに頻出する話題から整理した8つのラベルとして付与した。具体的には、「患者自身に関する(医学的)情報」、「患者以外の何者かに関する情報」、「誰か、何かの行動に関する文」、「誰か、何かの対応に関する文」、「患者の治療に関する言及」、「患者の診断」、「何かの推定」、「可能性に関する言及」について、当てはまるものをマルチラベル式に付与するものとした。これらのうち、前5者は事実度が高い内容であり、後3者は事実度合いが低い内容と考えられる。そこで、センテンスに付与されるラベルが前5者のみの場合には、事実度が高いセンテンスとして扱い、「d: descriptive」とのラベルを付与した。また、後3者のいずれかが含まれるセンテンスの場合は、事実度が低い、すなわち、抽象度が高いセンテンスとして、「a: abstractive」のラベルを付与した。ただし、このルールは単純明快ではあるものの、時折例外が生じるため、a、dのラベルについては目視確認のうえ付与した。そのうえで、このデータを用い、汎用性の高い言語表現モデルであるBERT

を用いてセンテンスの a/d 分類を試みた。利用したデータセットにおいて、descriptive なセンテンス数は abstractive なセンテンスの 5 倍ほどの分量があるため (d:a = 0.83:0.17)、全ての予測を d としても 83%は正解できてしまうことになる。そこで、a センテンス側の予測性能の向上を目指したところ、再現率(実際に a であるもののうち、a であると正しく予測されたものの割合: recall) 0.70、適合率(正と予測したデータのうち、実際に正であるものの割合: precision) 0.70 の性能を得ることができた。これは、決して高い性能ではないが、単純な手法に基づく予測性能であり、今後のチューニングによってさらなる性能向上が十分に期待できる状況となっている。

修正 CASE モデルの検討

次に、CASE モデルの考察を重ね、アップデート版として修正 CASE モデル(図 2)を策定した。その結果、当初の CASE モデルの弱点であった、退院サマリ中のセンテンスと元の入院カルテ中のセンテンスとの対応付けを行うと共に、そのそれぞれの生成物たるセンテンスの生成メカニズムの検証に道を拓くことができた。たとえば、このモデルでは、退院サマリ中の Assessment 文の由来として、入院カルテに含まれる記述(Descriptive 文)からの Abstractive な要約の他に、もともとのカルテに含まれる分析的な記述(Analytical 文)を Extractive に要約することで生成されることが示されている。また、センテンスの由来を元の入院カルテに見出すことができない場合、たとえば、Clinical Reasoning 文の由来として、入院カルテから臨床推論を経て記載されるケース、その他の情報源から臨床推論を経て記載されるケースがあることを示している。また、事実度の高い Supplement 文としては、その

他の情報源から Extractive に要約されるケースの他に、その他の未知の情報源からの情報が含まれる点が示されている。これらは、退院サマリ中のセンテンスの分類モデルであると同時に、それぞれのセンテンスをいかに(自動)生成しうるかを示す生成モデルとなっており、その完成度を高めることが出来た。

理想のサマリ研究

自動要約に関する研究分野に、要約研究用のデータセットは存在する。しかし、そもそもそのデータセットに含まれるトレーニング用の要約をいかに高品質に生成するかについては、体系だった研究が見当たらない。エキスパート間のコンセンサスを持って gold standard とするにせよ、誰かが最初の要約作業を行いその要約を修正・添削するのか、複数のエキスパートが平行で要約を行いその結果を用いて討議するのか、品質やコストにとってどれが正しい戦略であるかは自明な問いではない。文献検索を行ってみても、どうも先行研究はみあたらないことから、研究の蓄積が浅い分野であろうことが推察された。近年におけるニューラルネットワークの技術革新によって、高品質なトレーニングデータさえ用意できれば高品質なニューラル要約が可能となる可能性が高い。もし、高品質な要約を低コストに作成できる方法論が確立すれば、退院サマリの要約だけでなく、自動要約研究自体に大きな貢献を果たすことができよう。

そこで、入院カルテを対象とした要約については一旦保留し、より一般的な問題へと条件を緩和して、理想の要約に関する検討を進めた。とりわけ、入院カルテを対象とした検証は被験者(医師)の確保のためのコストが高いことから、同じ構造を有する他の問題にシフトし基礎的検討を加えることは、研究戦略としても合理的なものと

考えられた。

対象としてまず選択したのは、「就活に関するブログ」の要約タスクである。就活ブログは、ある程度の長さがある時系列なテキストと考えられる。また、複数面接のステップや最終的な結果等、イベントに関する記載が含まれることに加えて、活動を通じた成功要因や失敗要因についての考察が含まれる点で、入院カルテに類似した構成となっている。さらに、就職活動は大学生のほとんどに関わるイベントであり、他人の就活より学べる点も多いことから、大学内での被験者確保に有利なテキストと言える。そこで、今回の理想のサマリ作成の予備実験として、研究協力者に依頼して就活ブログを要約し、その他の被験者に見せ、さらに添削させるという試みを行った。

予想された。そこで、より簡便なタスクとして、「Wikipedia に含まれる会社紹介から適切なサイズの社史部分を抜粋したデータセットを作成し、その要約を試みる」タスクへと修正を加えるものとした。また、この予備実験における挙動を分析することで、「理想のサマリ」を客観的に評価する手法に加えて、要約作業をより低コストに抑える手法について、研究デザインを定めることができた。

D . 考察

ダミーカルテ収集

自然言語処理分野の研究論文においては、ダミーデータを用いて研究を行う場合、利

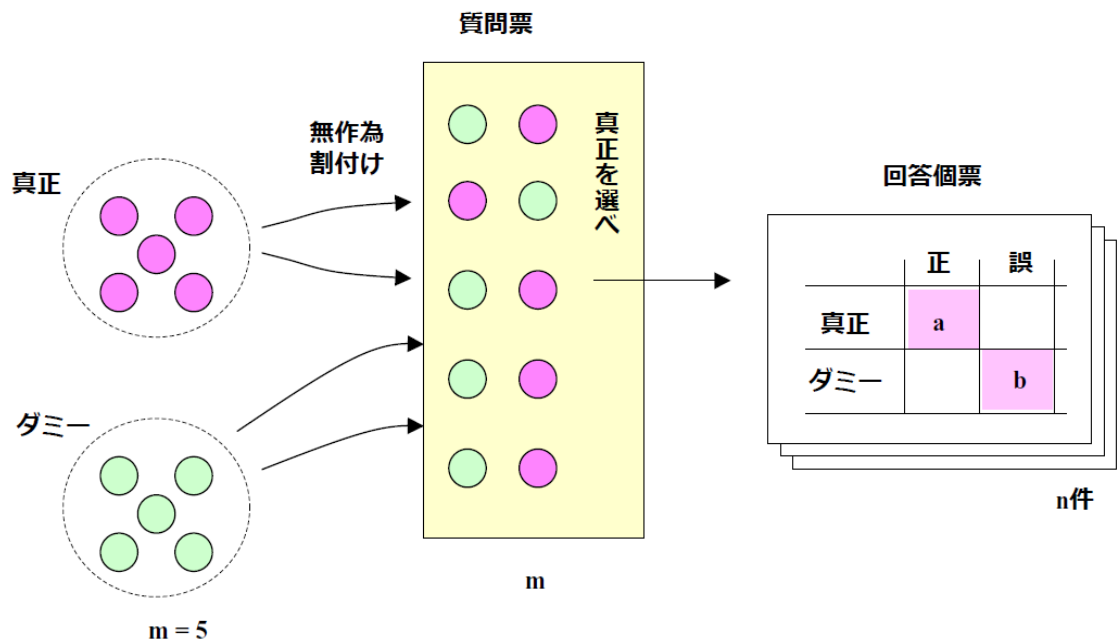


図3 ダミーカルテの真正性検証

試行の結果、まず、最初の要約に時間が掛かり、被験者への負担が予想外に大きいことが明らかとなった。これを第1世代の要約と称するとすると、第2世代以降の要約は相対的に負担が低くなるが、それでも、今後の研究に際した被験者の確保に困難が

用したテキストが本物と品質等において変わりが無いことを証明しなければ評価が下がるケースがある。そこで、この後の研究利用に際しては、まず、「本物であることが証明されているカルテ」とダミーカルテとを医師が区別できないことを示すことが望

ましい。前者については、同意に基づいて、研究協力者より直接提供を受けることで用意が可能と考えられる。後者については、今回収集したダミーカルテからランダムサンプルすることで用意しうるものと考えられる。

研究デザインとしては、無作為に生成した真正データとダミーデータのペアからなる質問票を作成し、これを被験者に提示する手法によって、等質性の検証が実現する(図3)。もし、真正データとダミーデータが判別可能であれば、真正を正とし、ダミーを誤とした解答が多数となる。一方、この解答がランダムとなるのであれば、両データは目視にて判別不能であり、ダミーデータの品質が真正カルテに等しいことを示すことができる。今後、データと被験者の確保を行い、実際の検証を行いたい。

カルテのアノテーション

上記のように品質管理を行ったダミーカルテについて、今年度、初年度に付与したアノテーションを白紙に戻して、さらに別のアノテーションを行った。その結果、自動分類においても性能向上が認められたが、アノテーションのガイドラインとしてはさらなるブラッシュアップの余地がある。

たとえば、現状では「治療」に言及するセンテンスについて、事実度が高いものと分類している。しかし、治療には、「治療した」という事実の記載だけでなく、今後の治療計画に関するものがあり、現在のラベルはそれらを混在させている可能性がある。また、「誰か、何かの行動に関する文」、「誰か、何かの対応に関する文」という区分についても、「行動」の主体が、患者なのか医療者なのかが不明確となっている。

これらを整理すると、事実度が高いセンテンスについては、主語ないし動作主が患者であるかそれ以外であるかという軸(述

語の表す動作の主体)と、「描写」か「行動」かという述語種別の軸によって、2×2に分類することが、アノテーションガイドラインとしても明確化するのではないかと考えられる。たとえば、「その後空咳も出てきた」という記述において、主語は「空咳」であるものの、動作主は患者となる。ただし、カルテ記載において、この動作主は必ずしも明示的には記載されず、とりわけ患者の場合は少なからず省略されることになる。医療用自然言語処理においては、この動作主の検出というタスク自体に価値がある可能性があるため、今後、再度のアノテーションによりデータ整備を目指したい。

また、事実度の低い、抽象度が相対的に高いセンテンスについても、改善の余地がある。とりわけ、最初のアノテーションにおいて付与した Probable、Certain といったモダリティ情報について、今回のアノテーションでは活用できていない。今後、これらを統合したデータを作成することにより、データセットの価値をさらに高めることが望ましい。

修正 CASE モデルの実証

今回提案した「修正 CASE モデル」は、概念的なモデルであるが、大量の電子カルテがあれば、その妥当性を統計的に検証しうる。

退院サマリ中に含まれるセンテンスの要約手法として、Extractive な要約由来のセンテンスが多いことが分かれば、入院カルテの自動要約に向けたハードルは一気に下がる。また、入院カルテに由来しないセンテンスについて、我々の仮説が示しているように、他の入院カルテに含まれる割合が明らかになることによって、入院カルテの自動要約のハードルはさらに下がることになる。由来が明らかとなれば、元テキストのどういう位置や文脈のテキストが退院サ

マリに出現しやすいか等、多くの統計データを取得していくことが可能となる。

問題は、「テキストに由来しないセンテンスの生成」と、「医学知識に基づいた abstractive な要約の実現」となる。前者については、そもそもそのカテゴリのセンテンスとして、主治医しか知りえない情報なのか、オーダリングシステム中のどこかに存在しうる情報なのか、それぞれがどれくらいの比率で存在するのかという点が重要となる。このいずれにおいても、現在は研究の蓄積が乏しいが、我々の研究によって明らかとなっていくことが期待される。医学知識に基づいた abstractive な要約は技術的な難易度が高いが、こちらも、入院カルテと退院サマリの膨大なペアデータを用いた機械学習により、ニューラル機械翻訳技術と同様にトレーニングデータの物量である程度解決できる可能性がある。

いずれにおいても、今回提案した修正 CASE モデルが基点となって統計取得と研究の発展が見込まれることになる。今後、まずは実データを用いたモデルの実証を目指したい。

理想のサマリ研究

「理想の退院サマリ」の確保に向けては、今年度、客観的な基準に基づく理想的なサマリの作成に向けた仮説の構築と、その仮説の実証に向けた研究デザインの策定を行うことができた。この実証に際しては、退院カルテそのものを用いた研究のコストが高いことから、条件を緩和し、より一般性の高いテキストとして Wikipedia に含まれる企業の社史を対象とした要約タスクとなっている。これにより、文書の要約研究に向けた高品質な要約文書を低コスト、高品質に実現しうることを実証することができれば、次のステップとして、医師を対象とした退院サマリの高品質な整備手法を検証

することができる。最終的なゴールである高品質な退院サマリの確保にはまだ道程が遠いものの、作業としては明確化したことから、実現に漕ぎ着けたい。

E . 結論

本研究分担では、退院サマリの自動生成研究の発展に向けて、研究基盤の整備に取り組んだ。そのために、総合に関連した 4 つの課題に取り組み、i) 108 件のダミーカルテ収集、ii) ダミーカルテに対するアノテーションとガイドライン策定、iii) サマリの分析と自動生成に向けたセンテンス分類モデルの構築、iv) 理想の退院サマリの作成に向けた予備実験と研究デザインの策定を行うことができた。

今後、ダミーカルテを対象としたアノテーションの更なるブラッシュアップにより、センテンス分類モデルの実証に向けた分類器の精度管理用データが用意できることになる。この分類器が実用精度となった段階で、実際の入院カルテ・退院サマリを対象とした網羅的な分析を開始することが出来る。この分析は、退院サマリがいかに作成されているかの統計取得を図るものであり、退院サマリの自動生成を実現する研究アプローチを検討するうえで決定的な価値を有することになる。理想のサマリの検討は、この自動生成をニューラル機械翻訳技術の応用により実現する際、生成したサマリを理想的なサマリとの差異によって評価する評価系の実現に大きく貢献する。さらに、高品質なトレーニングデータを大量に用意することができれば、さらなる精度向上に繋がることが期待される。

上記の方向性により、退院カルテの自動生成研究の進展が見込まれる。さらに、自動生成したサマリを提示し、「訂正タスク化」することで、継続的な精度向上が望め

るとともに、医療用自然言語処理の課題の一つである曖昧性解消のためのデータ収集が低コストに実現することになる。また、入院カルテ・退院サマリのペアデータからは、コストの掛かるコーパス整備なしに、様々な副産物が期待される。たとえば、個々の医師やチーム毎に個人辞書を整備することで、「表記ゆれ」の問題を解消することができる。さらに、医療用 NLP の最大の課題であったカルテにおける表記の「用言化」に対し、サマリ訂正タスクを通じた言い換え辞書の整備が実現する可能性がある。

このように、退院サマリの自動作成研究が、技術的な難度が高い医療用自然言語処理技術に対して技術的なブレイクスルーをもたらすことが期待される。また、退院サマリの自動生成技術は、実現に向けて医療現場からの期待も高い技術である。今後、コンピュータがカルテ記載を理解できるようになれば、医療用人工知能の発展を通じて、診断や治療において医療従事者を支援する様々な機能が実現する。そのためにも、今後の研究を通じて、さらなる研究の発展を図りたい。

F．研究発表

1．論文発表

なし

2．学会発表

なし

G．謝辞

理化学研究所 革新知能統合研究センター 松本裕治先生、首都大学東京システムデザイン学部 小町守先生、安道健一郎さんには、退院サマリの自動要約アプローチに向けたセンテンス分類に際して多くの助力を頂きました。また、田鎖麻衣さん、大阪大学医学部 宮本紘子さんには、ダミーカルテ整理とアノテーションにご尽力下さいました。北陸先端科学技術大学院大学 浅井拓也さん、北見工業大学 荒田真輝さん、斎藤健斗さん、寺下俊さんは、理想のサマリ研究に貢献して下さいました。また、ダミーカルテの整備に多くの医師の先生方にお力をお借り致しました。この場をお借りして、感謝を申し上げます。