論文 / 著書情報
Article / Book Information

題目(和文)	単一文書要約の高度化に関する研究
Title(English)	
著者(和文)	菊池悠太
Author(English)	Yuta Kikuchi
出典(和文)	学位:博士（工学）, 学位授与機関:東京工業大学, 報告番号:甲第10377号, 授与年月日:2016年12月31日, 学位の種別:課程博士, 審査員:高村 大也,新田 克己,寺野 隆雄,奥村 学,小野 功
Citation(English)	Degree:, Conferring organization: Tokyo Institute of Technology, Report number:甲第10377号, Conferred date:2016/12/31, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

単一文書要約の高度化に関する研究

東京工業大学
大学院総合理工学研究科
知能システム科学専攻
指導教員 高村 大也

菊池悠太

2016 年 9 月

目次

図目次

表目次

第 1 章

序論

1.1　研究の背景

文書要約とはその名の通り，入力として与えられた文書に対し，その内容を簡潔にまとめた要約を作成するタスクである．人手により作られた要約は，新聞記事の見出し，小説のあらすじ，科学技術論文の抄録など，さまざまな形で我々の生活においてすでに用いられている．ウェブ上では，検索エンジンの出力するスニペットや EC サイトにおいて製品レビューにしばしば付与される評価レートも要約といえる．また，本研究では文書の要約のみを取り扱うが，文書に限らず何らかの情報ソースがあった時に，その内容から重要な情報のみを取り出し簡潔な形でユーザに提示する操作全般を要約ということができる．これはたとえば映画の予告やスポーツ映像のダイジェスト映像などが該当する．また，映像作品のあらすじなど，入力となるオリジナルの情報とその要約が必ずしも同じ媒体ではない場合も存在する．このように，何かしらの情報の要約は，情報全体を精査することが難しい場合にその重要な部分のみを把握したいときや，そもそもその情報全体を見る必要があるか否かの判断を下すために，非常に有用である．しかしながら，人手による要約の作成は比較的コストの高い作業であることや，要約の対象となりうる情報が年々増え続けている現状を鑑みると，機械による要約の自動生成は非常に重要な役割を担っているといえる．とくに，近年のインターネットの発達に伴いウェブ上にアップされる文書は日々増加しており，すでに個人がアクセス可能な量としては膨大すぎる文書が存在している．そのような状況において，文書において重要で読者にとって必要な情報に効率的にアクセスすることが可能となるための技術の重要性は高まっている．

以降は，本研究の対象である文書要約，すなわち文書を自動的に要約する研究課題について説明する．文書要約は自然言語処理における応用研究の題材の一つとしてすでに半世紀以上も取り組まれている [77]．文書要約における標準的な設定では，要約器への入力となるのは文書と要約長である．入力は，後述するように，一つの文書であることもあれ

ば，関連した複数の文書で構成される文書集合のこともある[*1]．伝統的に，入力となる文書 (集合) は，原文書 (集合) と呼ばれる．要約長は，要約器が生成するべき要約の長さを指定するものである．要約器は，指定された要約長に柔軟に対応し，許された長さに応じた適切な要約を作成することが要求される．指定された要約長の，入力文書の長さに対する割合は圧縮率と呼ばれ，一般にこの数値が高いほど簡単で，低いほど限られた長さに適切な情報を詰め込む必要があるため難しい設定である．また，評価や訓練の際に用いられる人手によって作成された要約は参照要約と呼ばれる．

文書要約の問題を大きく二つに分けると，入力される文書がただひとつである単一文書要約と，同じトピックについて書かれた複数の関連文書の集合である複数文書要約が存在する．単一文書要約は 1950 年台から研究が行われ，単一の報道記事やエッセイ，科学技術論文などが主な対象として扱われてきた．一方で，2000 年台初頭から近年までの文書要約の歴史は複数文書要約の歴史であったといえる．複数文書要約は近年のインターネットの発達に伴い大量の文書が収集可能になったことにより盛んに研究が行われるようになった．たとえば，複数の報道機関が同じ事件や事柄について記述した報道記事の集合，ある製品について異なる複数のユーザが書いたレビュー文書の集合など，相互に関連を持った多くの文書を集約して，その中から重要な情報を要約として抽出する需要はインターネットの発展に伴って急速に高まっていった．近年発達している Twitter をはじめとする Social Network Service (SNS) では，現実のあるイベントに対して複数のユーザが感想や実況を投稿することがしばしばあり，その模様を時系列順にまとめ上げる研究も，複数文書要約の一種として取り組まれている [113, 59]．実際，2001 年から 2007 年まで行われた，文書要約を共通課題として設定した評価型国際会議である Document Understanding Conference (DUC) では，はじめの 2 年間は単一文書要約と複数文書要約の両者が対象タスクとして設定されていたものの，2003 年以降は複数文書要約のみに絞られている．2008 年から 2011 年，2014 年は，DUC の後継とも言える Text Analysis Conference (TAC) でも文書要約が課題として設定されたが，そこに単一文書要約は含まれていない．その間，多くの研究者の興味を集めた複数文書要約は，ただ重要な情報を要約するだけではなく，要約するべき情報のキーワードを指定した上で要約を作成する query-oriented summarization (DUC2003), すでにある程度の情報を持っている前提で，追加入力された文書集合から新規の情報のみを要約に含める update summarization (TAC2008)，あらかじめ決められたいくつかのトピックを過不足なく要約に含めることが要求される guided summarization (TAC2010) など，多くの特殊な設定で DUC や TAC でも取り組まれてきた．なお．各タスクの後ろの括弧内に示した情報は，そのタスクが初めて共通課題として

[*1] また，ユーザからのクエリなど，タスクによって特有の追加的な入力が与えられることもある．

設定された会議と年度である[*2]．また，個々の研究者による新たな設定の複数文書要約タスクが提案されるなど，いまだ多くの関心が寄せられている．たとえば，読むべき情報に階層的な優先順位を作り，読者がより知りたい情報を深く掘り下げつつ読み進めていくことを可能とする hierarchical summarization [19] や，要約を読む途中で分からないキーワードが存在した場合に，そのキーワードについて追加的に要約を作成する query-chain focused summarization[*3][9] などが一例である．このように，インターネットの発達に伴いその必要性が高まった複数文書要約は，より実用性を考慮した様々な状況を想定した設定が提案され，現在でも盛んに研究が行われている．この傾向は将来的にも変わらず，複数文書要約は今後も盛んに研究が行われていくことが予想される．

これまで述べたように，インターネットの普及が複数文書要約の発展を急速に後押しし，相対的に要約研究者の焦点が単一文書要約というタスクから離れてしまったことは事実である．しかしながら，もちろん単一文書要約への需要がインターネットの発達に伴い減少したということは無く，両者は異なった使途や需要を持つ．いま目の前にある一つの文書を，他の文書と独立に要約したいという需要が消えることはない．実際に，日本[*4]やアメリカ[*5]などにおける一部の報道機関は各報道記事に対して人手による要約を付与している．また，本節の冒頭で述べた要約の例はまさに単一文書要約である．技術的な側面で考えても，要約という共通点は存在するものの，両者の問題はそれぞれ異なる特性を持っており，どちらも取り組むべき課題が多く残っている．すでに述べた通り，複数文書要約は，関連した内容の文書の集合を入力として受け取る．このときその文書集合における共通のテーマに関わる重要なキーワードは文書横断的に多発するため，要約するべき重要な箇所についての手がかりを得やすい．対して単一文書要約ではそのような手がかりが利用できないため，その点においては複数文書要約と比較して難しいといえる [90]．反面，複数文書要約において技術的に難しい点としては，関連した複数文書の中には重複する内容の文が多く存在することが考えられるため，出来上がる要約も冗長になってしまうおそれがある．そのため，重要な内容を含めることに加えて冗長性を排除することも考慮に入れなければならない．また，異なる文書から少しずつ文を抽出する場合，要約として提示する際にはそれらの文の順序を決定しなければならない．単一文書要約の場合には，基本的に元の文書の並びのまま文を整列させれば良いため，多くの場合この問題の影響は少ない．もちろん，単一文書要約と複数文書要約が共通してもつ，文書要約としての難しい点も存在する．たとえば，要約の使途や画面の表示領域など様々な要因により要求される要

[*2] 各タスクが共通課題として継続された年数については，タスク毎に異なるため割愛する．

[*3] このとき，追加的な要約はすでに出力した要約との差分のみを出力することが要求されるため，本タスクは query-oriented summarization と update summarization を組み合わせたような要約タスクであるといえる．

[*4] http://news.livedoor.com

[*5] http://www.cnn.com

約長は大きく変化するため，要約長の違いに対して柔軟な要約器を作成することは重要である．また，原文書の文意を変化させないためには照応関係の解決や文書の論理構造を崩さないような考慮が必要である．文の部分的な削除や言い換えなど，オリジナルの文を書き換える処理が入る場合は，文法的に誤りのある非文の生成を避けることは，重要であるものの，非常に難しい課題の一つである．そもそも，文書を文に適切に分割する境界も自明とはいえず，用いる自動解析器によって文境界そのものが一致しないという問題も存在する．このように，単一文書要約と複数文書要約には共通の解くべき課題も存在するが，それぞれ特有の異なった課題も存在するため，相互に補完しつつ両方のタスクで研究を進めることが重要である．

本研究では，単一文書要約の高度化に向けていくつかの観点で研究を行う．

単一文書要約は，これまで述べたように複数文書要約とは異なる需要や技術的課題点が存在している．また，近年の技術の蓄積や大規模なデータの出現など，単一文書要約をより高度化させるための材料が揃いつつあり，一部の研究者の焦点が単一文書要約へ再び集まり始めているという事実もある．そのため今後は単一文書要約も再び盛り上がりを見せていくことが期待される．次節以降で，これまでの文書要約研究における枠組みや要素技術の変遷について，本研究で取り組む研究と併せて述べる．

また，本研究では具体的には二つの異なる側面に焦点を当てて研究を行うが，高度化の指標としてはどちらも ROUGE 値という自動評価指標を用いる．詳細は 3.2.2 節で述べるが，ROUGE(Recall-Oriented Understudy for Gisting Evaluation)[73] は現在の文書要約の研究において最も用いられている自動評価指標であり，人手による正解の要約の表層的な内容をどの程度網羅できたかを測る指標である．ROUGE は現在の文書要約研究において最も標準的に用いられている自動評価尺度であり，人手によって作成された要約の内容が，現文書のなかで要約に含めるべき重要な情報であるとみなすことで，文書要約における必須要件の一つである「現文書の重要な内容を適切に含んだ要約」の評価を行っている．

1.2　これまでの文書要約の研究動向

抽出型要約 (extractive summarization) は現在の文書要約研究において最も広く用いられるアプローチである．このアプローチは，文書をある言語単位（文，節，単語など）の集合とみなし，その部分集合を選択することで要約を生成する．これに対し，原文書に含まれない表現も用いつつ要約を行うアプローチは非抽出型要約[*6]と呼ばれる．そのため，非抽出型要約には高度な自然言語理解や自然言語生成という非常に難しい技術が必要である．本論文では抽出型要約のみに焦点を当てるが，非抽出型要約も重要な技術である上

[*6] あるいは概要型要約 (abstractive summarization).

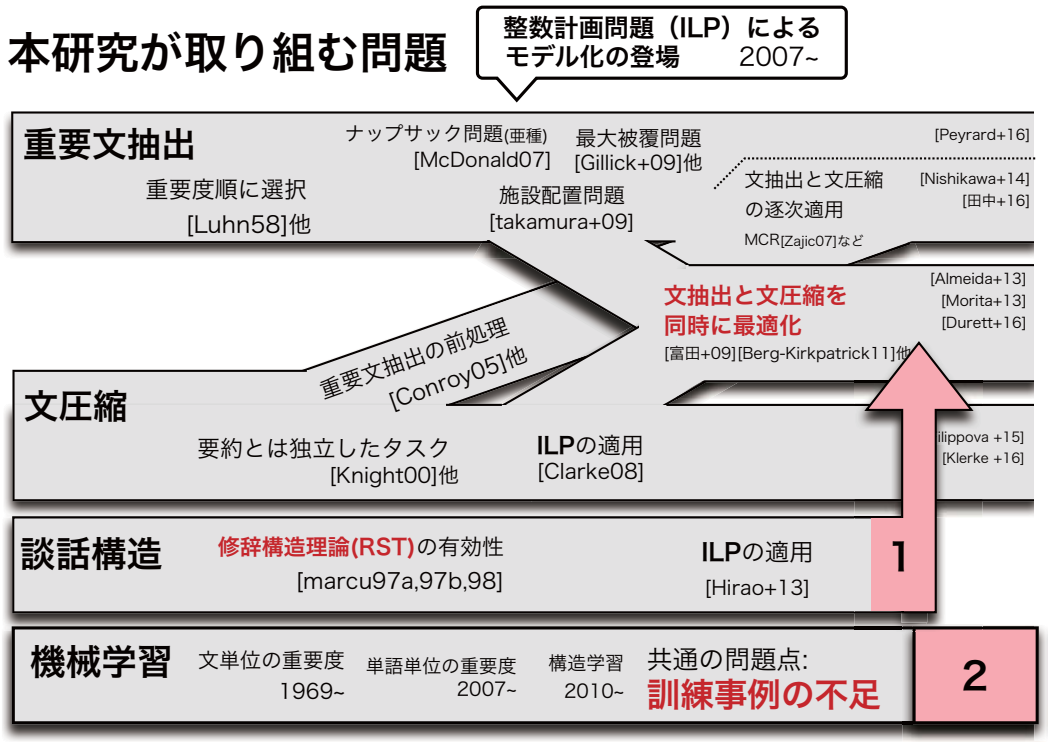

図 1.1　本研究と関係の深い文書要約研究における研究動向．本研究と特に関わりの深いトピックは赤字で示されている．また，本研究で取り組む問題はそれぞれ番号で示された右下の赤く塗りつぶされた領域である．

に，近年のニューラルネットワークや深層学習の発展に伴い注目が集まってきたといえるトピックであるため，5 章の今後の課題においてとりあげる．

文書要約の初期の頃から近年に至るまでの研究動向について，本研究で取り組む課題に関係の深いトピックを中心に時系列でまとめたものを図 1.1 に示す．抽出型要約のなかでも，文を抽出単位とする文抽出型要約は，文書要約研究の黎明期から使われ続けている最も長い歴史を持ったアプローチである [77]．文抽出型の要約アプローチは，原文書を文に分割し，様々な特徴に基づき各文に要約としての重要度を付与する．その後重要度に基づき要約に含める文を選択していくというアプローチである．文の重要度を決定する手がかりとして，文書中における単語の出現頻度や文の位置など，多くの特徴が有効に働くことが報告されてきた [77, 30, 95, 84]．初期の研究では，これらの手がかりに人手で重みを与えることで文の重要度を計算してきた [77, 30]．その後，これら手がかりの重みを自動で決定するためにナイーブベイズ分類器 [60, 4] や決定木学習 [80, 71, 136, 144]，Support Vector Machine[139] などの機械学習手法が適用された．これらの研究における共通の問題点としては，訓練データの不足が挙げられるがそれについては 1.3.2 節で詳しく述べる．

重み付けられた文の重要度に基づき，どのようにして実際に要約文を選択するかという

問題について，初期に取られた方法は，単に重要度順に文をランキングし，上位の文から順に選択していくという方法である．単一文書要約においては，原文書そのものが冗長である場合を除けば基本的に重要な文から順番に選択することができる．しかしながら，すでに述べた通り複数文書要約においては文書間で重複した表現が多数存在することが自然であるため，単に重要度順に選択すると冗長な要約が簡単に出来上がってしまう．そのため，一つの文を要約に選択するたびに，すでに選択済みの文との類似度を考慮して重要度を計算し直す手法 [13] も早い時期から提案されているが，文の選択自体は同様に計算しなおしたランキングの上位から選択するものである．

2007 年，McDonald らは複数文書要約を組合せ最適化として定式化した．具体的には，文をアイテム，文長をアイテムの重量，要約長をナップサック容量とみなす 0-1 ナップサック問題をベースに，目的関数に独自の変更を加えた整数計画問題 (Integer Linear Programming; ILP) としてモデル化する手法を提案した [86]．以降，様々な ILP の定式化が提案され，実際にこれらの手法は要約文書の情報の網羅性の指標となる自動評価手法である ROUGE (Recall-Oriented Understudy for Gisting Evaluation)[73] 値の向上に大いに貢献してきた [86, 31, 110, 111]．ILP の登場が要約研究に与えた影響は大きく，近年報告されている抽出型要約研究のほとんどは，ILP を用いたものであることから，ILP の登場は文書要約研究におけるひとつのパラダイムシフトであったといえる [2, 97, 88, 38, 45, 52, 103, 28]．このとき，目的関数において用いる文の重要度は，前述のように機械学習で推定するものもあれば，頻度などのヒューリスティックに基づき決定されるものの両方が存在する．機械学習を用いる場合は，すでに述べた問題点である訓練コーパスの不足という問題は存在するものの，今後十分な量の訓練事例が利用可能になるにつれて，より活用されていくことが予想される．

これまでの説明では，抽出単位の変化には焦点を当てず，文抽出に限定して説明してきた．文抽出型要約への ILP の導入によって，高い網羅性を持った要約の生成が可能となった一方で，要約手法が持つ要約長に対する柔軟性も，情報の網羅性と密接な関係をもつようになった．文を抽出単位とすることには，生成された要約の文法性が保証されるという利点がある．しかし，高い圧縮率，すなわち原文書の長さと比較して非常に短い長さの要約文書が求められている場合，文を抽出単位とすると十分な量の情報を要約文書に含めることが出来ず，情報の網羅性が低くなってしまうという問題があった．この問題に対し，文抽出と文圧縮を組み合わせるアプローチが存在する．文圧縮とは，主に単語や句の削除により，対象となる文からより短い文を生成する手法である．文圧縮のアイディア自体は 2000 年前後から提唱されていたものである [56]．当初は文圧縮はそれ単体でひとつのタスクであり，その後すぐに文書要約の要素技術として使用するアイディアが提唱されたが，両者を単純に組み合わせるだけでは必ずしも最終的な結果がよくなるわけではなかった [72, 116]．ILP が登場すると，文抽出型の文書要約だけでなく，単体としての文圧縮も

ILP で定式化する研究も現れはじめた [20]．その後，ILP 以前のように重要文抽出と文圧縮を独立したパイプラインとして逐次適用するのではなく，両者を同時に解くように ILP の定式化を行う手法が研究されはじめた [143, 10]．これによる利点は，高い情報の網羅性と要約長への柔軟性を持った要約が生成できることであり，その後は多くの手法がこれらの手法をベースとしたものになっている [2, 52, 28]．それらの手法では，構文解析器により獲得した文の木構造を利用し，その木構造の根付き部分木を実行可能解とする制約を設けることで実現している．文の構造の表現方法として，依存構造と句構造の二種類が標準的に用いられる．このように，ILP の出現は要約研究に大きな転換をもたらし，文抽出と文圧縮を同時に最適化する手法は今日において最も盛んに研究が行われている手法の一つになっている．

ILP は，文書要約研究に大きな影響を与えた新しい枠組みである．一方で，文書要約の研究自体は約半世紀の歴史があり，その間には多くの知見が蓄積されている．その一つの例として文書の談話構造を利用することの有用性がある．談話構造は，基本的に文書内におけるテキストセグメント[*7]同士の関係を意味する．Marcu らは文書の談話構造を規定する理論の一つである修辞構造理論 (Rhetorical Structure Theory; RST) が，文書要約の手がかりとして有効であることを報告した [82, 83, 84]．RST は文書における二つのテキストセグメント同士の関係性を二分木の形で階層的に表現する．このとき二つのセグメント間には非対称な役割があるため，その非対称性を利用して各テキストセグメントの重要度を計算した．要約システムに必要とされる側面はいくつかあるが，そのうちの一つである一貫性 (coherence)[46, 81] に対して談話構造が有効に働くことが示された．要約が原文書の談話構造を保持していない場合，原文書の意図と異なる解釈を誘発する文書が生成されてしまうおそれがある．すなわち，原文書と似た談話構造を持つように要約文書を生成することは，要約を生成するために重要な要素である[*8]．RST を要約に利用した初期の研究は，RST における談話の最小単位である Elementary Discourse Unit (EDU) を重要度に基づき順序付けし，高いものから要約に選ぶという手法であり，十分な情報の網羅性が保証されないという問題があった．Hirao らは，EDU の抽出に基づく要約生成を ILP により定式化することで，その問題を解決した [45]．彼らは，従来の RST における標準的な表現方法である談話構造木をそのまま用いることの問題点を指摘し，EDU の依存構造木 (DEP-DT) に変換し，依存構造木の刈り込みにより要約を生成する木制約付きナップサック問題 [50] として要約を定式化した．彼らの手法は，ILP という新しい枠組みに，RST という古くから存在した知見を違和感なく組み込む方法を示し，RST が新たな枠組みの中でも有効に働くことを示したという点で貢献が大きい．また，彼らの研究は，単一文書

[*7] 文，句，節など，意味を持った複数単語の連なり，あるいはそれらの組み合わせ．

[*8] 原文書は常に一貫性を持った文書であることを仮定している．

要約に ILP を適用した初めての事例である．これまで複数文書要約ではその有効性が確かめられてきた ILP による定式化が，単一文書要約においても変わらず有効であることが示された．

ここまで，今日における文書要約における研究動向を述べた．抽出型要約の抽出単位としては，文からはじまり，近年ではより高い圧縮率を実現するために文圧縮などの技術で文の一部のみを要約に組み込む手法が盛んに研究されていることについて述べた．また，要約の生成手段として，古典的には抽出単位の重要度の高いものから選択していくものであったが，今日においては要約を ILP として定式化し大域最適解を求めることで情報の網羅性を高める枠組みが支配的な手段となっていることについて述べた．また，ILP の登場以前から蓄積されていた知見やアイディアを，ILP という新しい枠組みに適用する試みの一例として，RST による研究事例を紹介した．抽出単位の重要度を決定する方法として，機械学習を利用する試みは古くからあるものの，大量の訓練データの不足という問題が常に存在してきたことを述べた．

本研究ではこれらを踏まえ，二つのことに取り組む．一つは，文書要約への RST の利用という側面を更に発展させるために，現在の文書要約研究において最も精度が良いとされている文抽出と文圧縮を同時に最適化するモデルに RST の情報を違和感なく組み込むための手法の提案である．これは，過去に蓄積された要約に対する知見を最新の手法に組み込むための取り組みであるとも言える．もう一つは，文書要約における機械学習の活用を加速させるため，従来活用されてこなかった大規模な要約資源 (報道記事とその人手による要約の対を大量に収録したコーパス) を訓練データとして効率的に活用する手法の提案である．

1.3　本研究の目的

1.3.1　文間の依存関係を考慮した文抽出と文圧縮の同時最適化手法

RST は，これまで文書要約に有効である知見がすでに報告 [84, 25] されており，Hirao らによってその知見が ILP による定式化においても変わらず有効に働くことが示された [45]．しかしながら，RST を抽出型要約に組み込む従来の手法の問題点は，その抽出粒度にある．従来手法は抽出の単位として，RST において規定される談話の最小単位である EDU を用いて要約の生成を行ってきたが，それが要約において必ずしも最適な単位であるとは限らない．

たとえば，EDU は文よりは短いもののそれなりの長さを持ったテキストセグメントであるため，そのまま抽出単位とする場合は要約長に対する柔軟性の面で問題が生じる．

本研究の目的は，文抽出と文圧縮の同時最適化モデルに，一貫性のある要約の作成に有

効である談話構造を違和感のない形で組み込むことで高い情報の網羅性と要約長への柔軟性を持ち，原文書の談話構造を保った要約が生成できる要約手法を開発することである．これまで，文書要約に談話構造を加える試みと，文抽出と文圧縮の同時モデルは，どちらも文書要約において重要な要素であるにもかかわらず，独立に研究されてきた．その大きな要因の一つは，両者の扱う抽出粒度の違いである．前者は EDU であり，後者の抽出粒度は文（圧縮され短くなった文も含む）である．抽出単位を文や EDU というそれなりの長さのテキストスパンにすると，ある要約長制約に対し，選択可能なテキストスパンの組合せは自ずと限られ，情報の網羅性を向上させることが困難な場合がある．我々は，文間の依存関係に基づく木構造と単語間の依存関係に基づく木構造が入れ子となった**入れ子依存木**を提案し，その木構造に基いて要約を生成することでこの問題に取り組む．

1.3.2　要約器の効果的な訓練のための大規模要約資源の活用手法

他の様々なタスクと同様に，文書要約においても機械学習が何らかの形で用いられる場面は増えてきている [130, 64, 2, 44, 93, 112, 105]. しかしながら，文書要約において機械学習を用いる上での共通の問題は，訓練事例（要約-原文書対）の不足である．これは，単純な分類問題と比較し人手による要約作成はコストが高いことに由来している．そのため，従来のほとんどの研究は数百事例という小規模なデータによる学習を余儀なくされており，他の主要なタスクと比較しても十分とはいえない[*9]．この現状は，複雑な手法の開発や有効な素性の構築，人手による詳細な分析などを困難にしている．

現在，このような現状の打破に寄与すると最も期待できるコーパスとして，New York Times Annotated Corpus (NYTAC)[102] が存在する．NYTAC を構成する約 180 万記事のうち，およそ 65 万もの記事に，専門家による要約が付与されている．これは，潜在的には単一文書要約において最も大規模な訓練データとみなすことができる．しかしながら，これまで NYTAC の人手要約を何らかの形で利用した研究 [47, 127] は存在するが，直接的に大規模な訓練事例として利用したという報告はなされていない．NYTAC を直接的に訓練事例として利用する上で留意すべき[*10]は，そこに含まれる人手要約が特定の目的のもと統制されて作成および整備されたわけではないという点である．文書要約にはその作成方法や使途に複数の種類が存在しており，それぞれ特性や必要となる技術が異なるが，NYTAC においてはそれらが明示的に区別されることなく混在している．そのため，特定の要約器，例えば現在最も標準的なアプローチである文抽出に基づく要約器を訓練する際

[*9] 例えば，機械翻訳における古典的なコーパスの一つである Hansard コーパス [21] には約 130 万もの事例が収録されている．

[*10] これは同時に，従来 NYTAC が要約器の大規模訓練データとして用いられて来なかった理由であると考えられる．

に，全ての事例を機械的に利用することが必ずしも有効であるかは疑問である．

本研究における我々の目的は，NYTAC を単一文書要約における大規模な訓練事例として利用する上での効果的な手法を示すことである．訓練のどの段階でどのような形で NYTAC を利用するべきか，また全ての事例を用いるべきか何らかの基準でその部分集合に限って利用するべきか，実験を通して明らかにする．評価の際は，特定の評価セットに依存した結果を避けるために，単一文書要約において評価に用いられている複数の評価セット（ターゲットデータ）を用意する．

ここで我々が扱う問題は，ドメイン適応 [69] の一つのケースだと考えることができる．ドメイン適応では，両ドメイン間で学習されたパラメータの線形補間や適用元ドメインにより訓練された分類器を素性として利用するなど， シンプルでありながら強力ないくつかの手法が存在しており [24]，今回のケースにも容易に適用が可能である．さらに，ドメインにあわせた事例選択 (instance selection, instance weighting) も同様にドメイン適応の方法として知られている [6, 124]．本研究では，ドメイン適応分野の研究を参考にいくつかの標準的な手法を用意し比較することで， NYTAC を利用するにあたり有効な方法を確かめる．

1.4　本論文の構成

本章以降の構成を示す．2 章では，本研究に関連する従来の研究について紹介する．従来提案されてきた ILP による定式化の例や文圧縮の方法，談話構造の利用などいくつかの技術のほか，文書要約における機械学習の利用事例を中心に説明する．

3 章では，従来高い精度が確認されている文抽出と文圧縮の同時最適化モデルに，新たな構造的制約として修辞構造に基づく文と文の間の依存関係を組み込む新たなモデルを提案する．提案手法を従来の同時最適化モデルや木制約付きナップサック問題による要約手法と比較評価したところ，文書要約の自動評価指標である ROUGE において最高精度が得られることを確認した．また，高い圧縮率が要求される要約設定へのさらなる柔軟性を獲得するため，文圧縮の制約を緩和するよう拡張を加え，その有効性を示す．

4 章では，ターゲットとなる少数の整備された評価データに加え，大規模であるが整備のされていない要約資源である NYTAC がある状況で，後者を有効に要約器の訓練に活用するための手法を提案する．ドメイン適応の分野から標準的な 5 つの手法を取り上げ，さらに訓練する要約器の特性に併せた事例のフィルタリングをオプションとして用意することで，それらの組み合わせが訓練に与える影響を確かめる．実験の結果，要約器を NYTAC で一度訓練したあと，そのパラメータを初期値として所望のターゲットデータで追加的に訓練する手法が有効に働くことが分かった．加えて，使用する要約器の特性に併せた事例選択を事前に行うことで一部の評価データセットでは更に精度が大きく向上する

ことを確認した.

最後に，5 章において本研究のまとめと今後の展望について述べる.

第 2 章

関連研究

1 章において，本研究の目的と共にこれまでの文書要約における研究動向について振り返った．本章では，本研究において具体的に取り組む二つの問題に焦点を当てて，関連研究を紹介する．すなわち，組み合わせ最適化問題による要約のモデル化と文書要約における機械学習の利用の二つのトピックである．

2.1 文間の依存関係を考慮した文抽出と文圧縮の同時最適化手法

McDonald は，複数文書要約をナップサック問題の亜種 [86] として定式化した．ナップサック問題は，価値と重量を持つアイテムの集合とナップサックの容量が与えられたときに，ナップサック容量を超えない範囲の重量で価値が最大となるようにアイテムの部分集合を選択する整数計画問題の一種である．ここで文をアイテム，文長をアイテムの重量，要約長をナップサックの容量とみなすことで文書要約をナップサック問題として違和感なく定式化できる．McDonald の手法と単純なナップサック問題との違いは，ナップサック問題の目的関数は選択したアイテム (文) の価値の総和であるのに対し，彼の目的関数が Carbonell らによって提案された Maximum Marginal Relevance (MMR)[13, 39] に基づくものになっている点である．MMR は，文書要約における重要文のランキングや情報検索における文書のランキングにおいて，入力集合中に類似した事例が多く存在する場合に，結果のランキングに似たような事例が多く含まれ冗長になってしまうことを避けるために提案された手法である．具体的には MMR を要約に適用する際は，McDonald による以下の式 [86] の第二項のように，すでに要約として選択された文集合との類似度をペナルティとして導入することで同じような事例が再び結果に出現することを防ぐ:

$$\sum_i \alpha_i Rel(i) - \sum_{i<j} \alpha_{ij} Red(i,j) \tag{2.1}$$

ここで，α_i は i 番目の文が要約に含まれた場合に 1 となる決定変数であり，$Rel(i)$ は i 番目の文を選んだ場合に得られる利得である．α_{ij} は，i 番目の文と j 番目の文をどちらも要約に含めた場合に 1 となる決定変数であり，$Red(i,j)$ はその二つの文を同時に要約に含めた場合に生じる冗長度合いである．$Rel(i)$ には単語頻度や文の位置など，従来の要約における重要文の手がかりがそのまま用いられ，$Red(i,j)$ としては文の類似度が用いられる．MMR による要約手法の欠点のひとつは，文の選択自体は貪欲法に基づくものであったため．解の最適性が保証されない点であった．McDonald の手法は，MMR に基づいた目的関数を持つ ILP として定式化することで厳密解を求めることを可能にし，実際に実験によって ROUGE 値が向上することを確認した．

McDonald により文書要約を組合せ最適化問題の枠組みで定式化することの有効性が示されて以降，多くの新しい要約の定式化が提案された．高村らや Gillick らは，複数文書要約を最大被覆問題として定式化した [110, 38]．これは，与えられた要約長を満たす範囲で選択された文の集合によって，原文書集合中の概念単位（ユニグラムやバイグラム）の集合を出来る限り被覆するよう要約を生成するものである．各概念単位には被覆した場合の利得が付与され，同じ概念単位を複数回被覆，すなわち生成した要約中に同じ概念単位が複数出現しても利得は加算されない．そのため最大被覆問題は，複数文書要約において問題となる要約の冗長性を避けることが可能となっており，複数文書要約のモデルとして非常に有効であることが示された．

Takamura らは，複数文書要約を施設配置問題として定式化した [111]．原文書集合内の全ての文が，要約中の文のいずれかにより出来る限り表現されるように要約の生成を行う．Takamura らは，ある文が別の文を被覆した場合の利得を定義し，原文書集合中の文が要約として選択した文集合のいずれかの文に割り当てられた時に利得の総和が最大になるよう要約文の選択を行う．施設配置問題も，最大被覆問題と同様に，出来上がる要約の冗長性を排除する上で非常に有効な手段である．

これまで述べたようにさまざまな形式の要約手法が提案されたが，それらの主たる目的のひとつには冗長性の排除があった．それは，複数文書要約における入力である多くの関連文書の集合がもつ冗長性に起因するものであり，入力がひとつの文書により構成される単一文書要約では，単純な 01 ナップサック問題に近い形式の手法も用いられる．

Hirao らは，ナップサック問題においてアイテム間の依存関係を制約として追加した木ナップサック問題として単一文書要約を定式化した [45]．依存関係において上位の文を選択しなければ，その下位の文を選択できない．Hirao らはこの依存関係に，RST による談話構造を利用した．これにより，要約において原文書の談話構造を保存した要約生成が可能となり，単純なナップサック問題を有意に上回ることが確認された．

Nishikawa らは隠れマルコフモデルにナップサック制約を加えた隠れ半マルコフモデルとして要約を定式化した [93]．彼らは，文と文の間のテキスト結束性 [79] をモデルに組

み込み，要約に選択された文の重要度と，要約内で隣接する文と文のつながりの良さの両方を最大化する問題として要約を定式化した．

上記の手法は，文を抽出単位とした手法である．現在の抽出型要約の主流である文抽出，文圧縮の同時モデルでは，与えられた文書 (群) から，文を単語間依存木として表現し，その根付き部分木を刈り込む，すなわち単語を削除することで要約を生成する．また，これを ILP として定式化する研究が盛んに行われている [2, 97, 88, 38]．しかし，文から抽出する単語列を単語依存木の根付き部分木に限ると高圧縮な要約設定において情報の網羅性が低下するおそれがある[*1]．さらに，これらの同時モデルでは文書が持つ談話構造を考慮しないため，情報の網羅性が高く自動評価指標の ROUGE において高い値を得ることができるが，一貫性に欠けた要約が生成されてしまうという欠点がある．

また，同時モデルにおける文圧縮の手段として，文を依存構造木ではなく句構造木として表現し，その部分木を抽出する手法も提案されている [62, 120, 10]．句構造木は連続した単語列の文法的な役割を階層構造として表現した木であるため，この木を刈り込む際には連続した単語列（句）を同時に削除することが多くなる．よって，依存構造木を用いた刈り込みと比較すると要約長に柔軟な刈り込みが難しい．

一方，一貫性をもった要約を生成する手法として RST を利用した手法が提案されている．Daumé III and Marcu は，RST を利用した Noisy-channel モデルに基づく手法を提案した [25]．彼らの手法は一貫性を持った要約の生成が可能であるが，情報の網羅性という観点で最適解が得られるとは限らない．また適切な確率を計算するために大量のコーパスを必要とする上に，計算量の問題で長い文書に適用できないという欠点があった．Marcu は，RST の構造を利用して EDU の順位を決定し，ランキング上位の EDU を要約として抽出した [84]．Uzeda らは，Marcu の手法を含む合計 6 つの手法を組み合わせる手法を提案し，オリジナルの手法との比較評価を行った [115]. 3.2.3 節では彼らの報告にある数値も参考値として載せている．

Marcu らの手法を含む EDU のランキングに基づく手法は，MMR 同様に解の最適性が保証されない保証されないという欠点がある．Hirao らはこれを解決するため，前述したように，EDU の依存木を構築し，その依存関係に基づいて EDU を選択する問題を木制約付きナプサック問題として定式化した [45]．これらの手法は RST におけるテキストの最小単位である EDU をそのまま抽出単位としていたが，EDU が文書要約においても適切な抽出単位であるかについては，要約長に対する柔軟性の面で疑問が残る．EDU は文よりも短いとはいえそれなりの長さを持ったテキストスパンである．そのため，要約に含める情報の組み合わせの自由度は比較的低く，かつ EDU のようなテキストスパンを対象と

[*1] ただし，単一の文に対し明示的に圧縮率が与えられる文圧縮タスクだけに限れば，文に複数の根を与える手法も存在する [34].

した構文解析器がないため，文圧縮のような技術が適用できない．これに関しては 3.2.3 節で評価実験結果をふまえて考察を行う．

Hirao らの手法は提案手法に最も強く関連している．両者の違いは，Hirao らの手法が EDU をノードとする依存木から EDU を選択する要約手法であるのに対し，提案手法は文間の依存関係と単語間の依存関係が入れ子構造を成す木から単語を選択する要約手法であるという点である．

また，文重要度の決定に貢献する特徴を調べた文献 [76] でも，RST の有効性が示されている．

これまで，文書の（大域的な）談話構造を利用した要約手法について紹介したが，隣接した文同士のつながりを評価し，文の局所的な並びを最適にすることに取り組む研究も存在する [94, 93, 18]．これらの方法では，修辞構造解析器を必要としないため，論理構造が明確でなく自動解析の精度が期待できない文書においては有効である．一方で，文書の大域的な談話構造を考慮した要約生成はできない可能性がある．

RST を要約に組み込む研究の多くは RST で定義される修辞構造の構造木をそのまま利用したものが多かった [84, 25] が，Hirao ら [45] は，RST の談話構造木をそのまま用いることの問題点を指摘し，EDU の依存構造木 (DEP-DT) に変換し，依存構造木の刈り込みにより要約を生成する木制約付きナップサック問題 [50] として要約を定式化した．

2.2　要約器の効果的な訓練のための大規模要約資源の活用手法

他の様々なタスクと同様に，文書要約においても様々な形で機械学習が利用されている．機械学習を利用した最も初期の研究は Kupiec らによるものである [60]．Kupiec らは，ナイーブベイズ分類器を用いて文が要約に含まれるかどうかの確率を予測する分類器を訓練した．手がかりとして，位置情報，単語の出現頻度，手がかり表現など，それまでの研究により有効性が確認された特徴を用いている．このとき，訓練データとしては要約そのものを利用するのではなく，重要文を人手で選択することで構築された重要文集合を利用している．その後，ナイーブベイズ分類器 [114, 4] の他に，決定木学習 [80, 71, 136, 144, 5] や対数線形モデル [96]，Support Vector Machine[44] などが文の重要度を予測するために利用されてきた．

また，重要度を付与する単位を文ではなく文中の単語にする研究も存在する．Yih ら [130] は文ではなく単語の重要度をロジスティック回帰により予測する分類器を構築し，単語の予測確率の総和を要約全体のスコアとして，要約のデコードを行った．Hong らは，NYTAC に含まれる人手による要約の一部を分析し，そこから文書中のキーワードを表す

重要な単語の予測をロジスティック回帰により行った．また，単語と文両方の重要度を併用する手法も存在する．Wan ら [119] は，単語同士の類似度グラフと文同士の類似度グラフを別々に構築した上で，各単語から，その単語が出現した文にもエッジを張ることで，単語-単語間，文-文間，単語-文間の三種類のエッジが存在するグラフを構築した．その後，単語の重要度を予測するときは文の重要度を，文の重要度を予測するときに単語の重要度をそれぞれ用いつつ再帰的に両者の重要度の更新を行った．

ILP の登場以降は，その定式化の性質上，単語やバイグラムといった文を構成する文よりも小さな単位に対する重要度が用いられることが多くなっている．例えば，要約を最大被覆問題として定式化する手法は，要約として選択した文に含まれる概念単位（主にユニグラムやバイグラム）ごとの重要度を考慮し，要約として被覆された概念単位集合の重要度の総和を要約のスコアにしていた．そのため，最大被覆問題による要約モデルにおいて機械学習を利用する場合，分類器も単語単位の重要度予測を行うことが自然である．Takamura ら [112] は，バイグラムを概念単位として用いる最大被覆問題に基づく要約モデルにおいて，単語の重要度を構造化 SVM を用いて予測した．素性自体は従来の研究において用いられているものだが，彼らの手法は文書要約において構造学習を適用した初めての研究である．Sipos ら [105] は，単語単位の重要度を用いる最大被覆に基づいたモデルに加え，文間の類似度を用いる MMR に基づいたモデルにも構造学習による学習を行い両者の比較を行った．Li ら [64] は，ある文を選択したことにより新たに被覆した要約中のバイグラムの数をゲインとし，要約を作成し終えた時点におけるゲインの総和が最大になることを目的関数に設定した要約手法を提案した．ゲインを計算するために，原文書に出現した各バイグラムが要約中で何回出現するか，その出現頻度を予測する分類器を構築した．Berg-kirkpatrik ら [10] や Almeida ら [2] は，バイグラムを単位とする最大被覆問題に拡張を加え，文の句構造木の部分木を削除した場合に得られる利得も考慮することで文圧縮も同時に考慮した最適化問題として定式化した．このとき，バイグラムの被覆による利得と部分木の削除による利得の両方に対して特徴量を設計し，その重みを構造学習により最適化した．このように，ILP を利用した場合は文抽出と文圧縮の同時最適化モデルにおいては単語を単位とした学習が用いられる [28]．もちろん，ILP による要約手法の中でも Nishikawa ら [93] や Kikuchi ら [54] などのように，ナップサック問題に基づく要約手法では文単位の重要度を予測する手法が主流である．

1 章でも述べた通り，文書要約に機械学習を適用する場合における共通の問題は，訓練データの不足である．DUC2001 では，参加者に 567 の原文書-参照要約対が提供された．機械翻訳において古典的なコーパスの一つである Hansard コーパス [21] には約 130 万もの事例が収録されていたことを考えると，非常に小さいサイズであるといえる．Ziff-Davis コーパスは約 7,000 の原文書-参照要約対が収録されており，一部の研究で要約の分析に用いられてきた [85, 23] が，要約器の訓練のためにはあまり活用されることはな

かった．直近の研究では Nishikawa ら [93] が約 13,000 事例の日本語報道記事とその要約対を用いた学習を行い，文書要約においても訓練事例数の増加が性能向上に寄与することを報告しており，訓練データの充実は更なる文書要約研究の発展に不可欠であると言える．また，一部の報道機関は，ウェブ版のホームページに人手による要約とみなせる情報が付与された記事を公開している．Svore らは，アメリカ合衆国の報道機関である Cable News Network (CNN)[*2]の各報道記事に付与されている人手による 3 文の story highlight を収集し，ニューラルネットワークに基づく要約モデルの訓練を行った [109]．日本語においては，田中らが Livedoor News[*3]において各報道記事に付与された 3 行の要約を利用して要約器の訓練を行った．また，Genest らは人手で付された要約そのものではなく，TAC2008 および 2009 において参加者のシステム要約に付与された人手の評価値を利用した [1]．事例数は約 4000 であり，原文書に対して任意の要約が与えられた時にそのふさわしさを予測する要約のスコア関数を学習し，そのスコアが人手コストの高い評価指標である Pyramid スコアと高い相関を獲得した．

本研究では，NYTAC を大規模な訓練事例として活用する手法について考える．NYTAC の要約データを部分的に活用する試みはいくつかの先行研究に見受けられる．いくつかの手法は，NYTAC の要約によって訓練された言語モデルを利用している [63, 47]．Li and Nenkova [68] は NYTAC を用いて文の特異性を推定している．また Yang and Nenkova [127] は，文書の第一段落がその文書全体にとって重要な情報を含んでいるかを判定するために NYTAC を利用している．本研究は NYTAC を直接的に訓練データとしているが，これらの研究は NYTAC を補助的な情報として利用している．本研究以外で NYTAC を直接訓練事例として利用しているもう一つの例が Durrett らによるものである [28]. 彼らは，Kikuchi ら [54] による要約手法にいくつかの変更と照応関係に関する新たなる制約を加え，構造化 SVM を用いた学習を行った．彼らは大規模な NYTAC のサイズを活かして複雑なモデルの学習を行ったが，訓練事例は先行研究が人手でフィルタリングした NYTAC の一部を利用していた．我々の研究では，用いる要約器の特徴に合わせて自動的に事例のフィルタリングを行うことで訓練の効率を高めることが目的であり，両者において目的とする貢献が異なる．

[*2] http://edition.cnn.com/

[*3] http://news.livedoor.com/

第 3 章

文間の依存関係を考慮した文抽出と文圧縮の同時最適化手法

3.1　入れ子依存木の刈り込みによる要約文書生成

本研究の目的は，要約文書の一貫性と情報の網羅性が高く，かつ要約長に柔軟な要約手法を提案することである．具体的には，文間の依存関係に基づく木構造と単語間の依存関係に基づく木構造が入れ子となった**入れ子依存木**を提案し，その木構造に基いて要約を生成する手法を提案する．提案手法について，図 3.1 に示す例で説明する．本研究で提案する入れ子依存木は，文書を文間の依存関係で表した**文間依存木**で表現する．文間依存木のノードは文であり，文同士の依存関係をノード間のエッジとして表現する．各文内では，文が単語間の依存関係に基づいた**単語間依存木**で表現されている．単語間依存木のノードは単語であり，単語同士の依存関係をノード間のエッジとして表現する．このように，文間依存木の各ノードを単語間依存木とすることで，入れ子依存木を構築する．そして，この入れ子依存木を刈り込む，つまり単語の削除による要約生成を ILP として定式化する．生成された要約は，文間依存木という観点では必ず文の根付き部分木となっており，その部分木内の各文内，すなわち単語間依存木の観点では単語の部分木となっている．ここで，文間依存木からは必ず木全体の根ノードを含んだ根付き部分木が抽出されているのに対し，単語間依存木はそうでないものも存在することに注意されたい (図 3.1 中の $*$)．従来，文圧縮を文書要約に組み込む研究では，単語間依存木の場合も必ず根付き部分木が選択されていたが，限られた長さで重要な情報のみを要約に含めることを考えると，単語の根付き部分木という制約が情報の網羅性の向上の妨げとなる可能性がある．そこで提案手法では，根付きに限らない任意の部分木を抽出するために，部分木の親を文中の任意の単語に設定できるよう拡張を加える．

このように，要約としての一貫性と要約長への柔軟性を獲得するために文書を入れ子依

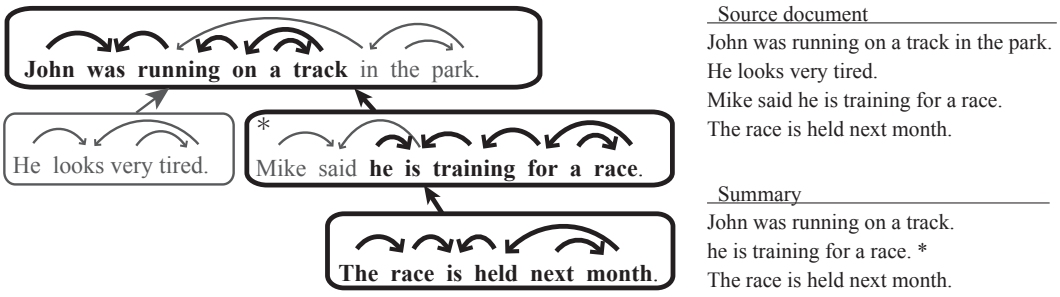

図 3.1　提案手法の概要．原文書は二種類の依存木に基づく入れ子依存木として表現される．提案手法は，文間依存木からは根付き部分木，その各ノードは単語間依存木の部分木となっているように単語を選択することで要約を生成する．

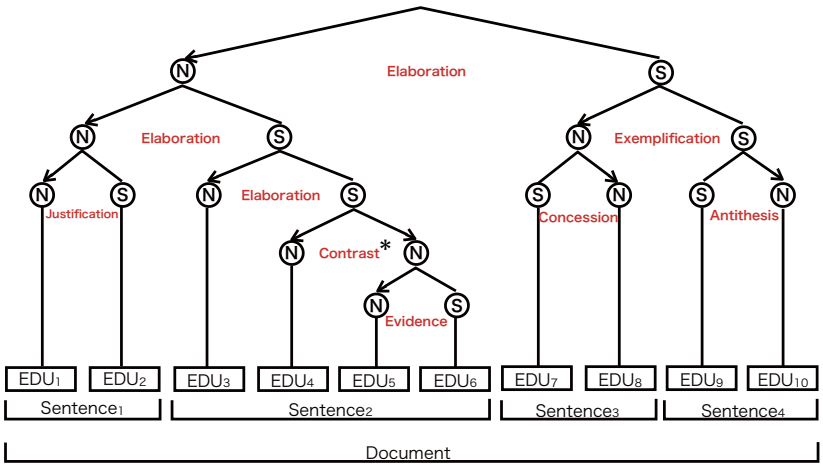

図 3.2　RST による文書の表現．EDU_{1-10} は具体的なテキストになっており，おおよそ節に相当する．隣り合ったテキストスパンは再帰的に修辞関係により関連付けられており，最終的に文書全体で木構造を構成する．N と S はそれぞれ核と衛星に対応し，間に書かれたラベルがそれらの間の修辞関係である．

存木として表現し，入れ子依存木から要約文書を生成する問題を ILP として定式化することで高い情報の網羅性を持った要約生成を行う．本節では，入れ子依存木の構築についての詳細と，ILP での定式化について説明する．

3.1.1　修辞構造理論

修辞構造理論 (Rhetorical Structure Theory; RST)[81] は，文書の談話構造を表現するために提案された理論である．文書を EDU に分割し，連続した EDU 同士（あるいは，複数の EDU をつなぎあわせたテキストスパン）を修辞関係で関連付けることで，談話構造木を構築する．構築される木は，終端ノードが EDU，非終端ノードが子ノード間の修辞

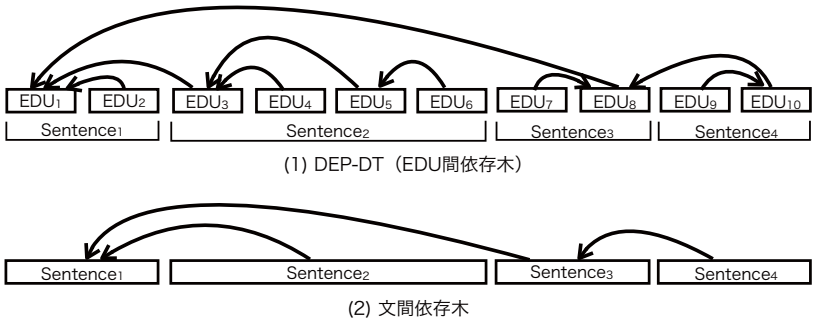

図 3.3　依存構造に変換された文書の修辞構造．(1)Hirao らによる EDU 間の依存関係を表した DEP-DT であり，図 3.2 の木構造を変換したもの．なお，multinucleus の関係を持つテキストスパンを成す EDU_4 と EDU_5 はそれぞれ独立に一つ上の核である EDU_3 に依存している．(2) 本研究で用いる文間依存木．Hirao らの DEP-DT を変換したもの．

関係をラベルに持つ木構造で表現される．図 3.2 に，談話構造木の例を示す．図において二つの非終端ノードの間に書かれているラベルがそのノード間の修辞関係である．具体的には例示，補足，背景などの関係により，テキストスパン同士がどのような関係にあるかを表現する．今回用いたコーパスでは合計で 89 種類の修辞関係が存在した．また，修辞関係と共に各テキストスパンには核 (Nucleus) か衛星 (Satellite) のいずれかのラベルが付与される．核はその修辞関係において中心的な役割を担い，衛星は補助的な役割を持つ．例えば補足という修辞関係では，補足される方のテキストスパンが核であり，その内容を具体的に補足したテキストスパンが衛星となる．図においては，各非終端ノードの N と S がそれぞれ核と衛星を表している．なお，図 3.2 の $*$ のように，複数の核からなる多核 (multinucleus) という性質を持った修辞関係も存在する．

3.1.2　入れ子依存木の構築

Hirao ら [45] は RST の木構造を変換することで依存構造に基づく DEP-DT を構築した．DEP-DT は，EDU 間の依存関係を直接表現しており，この依存木を刈り込むことで一貫性を保った要約が生成できる．図 3.3 に，図 3.2 の木構造から変換された DEP-DT と本研究で用いる文間依存木を示す．Hirao らの用いた DEP-DT は，EDU をノードとするものであったが，本研究で取り扱うモデルは文間の関係をとらえたものであるため，これを文をノードとした依存木へと変換する．具体的には，同じ文に属する EDU 集合をまとめ，文内の親となっている EDU の依存先を，その文の依存先として採用する．依存先の EDU は他の文に属しているため，この変換規則により，文間の依存関係を持った木構造（文間依存木）を取得することができる．次に文間依存木の各ノードとなる文に対し依存

構造解析を行い，単語間の依存構造（単語間依存木）を獲得する．以上の処理により，文書が文の係り受け木，各文内では単語の係り受け木により構成された**入れ子依存木**を構築する．本研究では，この入れ子依存木から不要な単語を刈り込むことで要約文書を生成する．

3.1.3　整数計画問題による定式化

入れ子依存木からの単語の削除による要約文書の生成は，整数計画問題として定式化できる．具体的には，ある目的関数のもと，文間依存木の根付き部分木，根付き部分木中の各文は単語間依存木の（任意の）部分木となるように単語を選択することで要約を生成する．提案手法は次の整数計画問題で定式化できる．

$$\text{max.} \quad \sum_{i}^{n}\sum_{j}^{m_i} w_{ij} z_{ij}$$

$$\text{s.t.} \quad \sum_{i}^{n}\sum_{j}^{m_i} z_{ij} \leq L; \tag{3.1}$$

$$x_i \geq z_{ij}; \qquad \forall i, j \tag{3.2}$$

$$x_{\text{parent}(i)} \geq x_i; \qquad \forall i \tag{3.3}$$

$$z_{\text{parent}(i,j)} + r_{ij} \geq z_{ij}; \quad \forall i, j \tag{3.4}$$

$$r_{ij} + z_{\text{parent}(i,j)} \leq 1; \quad \forall i, j \tag{3.5}$$

$$\sum_{j=0}^{m_i} r_{ij} = x_i; \qquad \forall i \tag{3.6}$$

$$r_{ij} \leq z_{ij}; \qquad \forall i, j \tag{3.7}$$

$$\sum_{j \notin R_c(i)} r_{ij} = 0; \qquad \forall i \tag{3.8}$$

$$r_{i\text{root}(i)} = z_{i\text{root}(i)}; \qquad \forall i \tag{3.9}$$

$$\sum_{j=0}^{m_i} z_{ij} \geq \min(\theta, m_i) x_i; \forall i \tag{3.10}$$

$$\sum_{j \in \text{sub}(i)} z_{ij} \geq x_i; \qquad \forall i \tag{3.11}$$

$$\sum_{j \in \text{obj}(i)} z_{ij} \geq x_i; \qquad \forall i \tag{3.12}$$

$$x_i \in \{0, 1\}; \qquad \forall i \tag{3.13}$$

$$z_{ij} \in \{0, 1\}; \qquad \forall i, j \tag{3.14}$$

$$r_i \in \{0, 1\}; \qquad \forall i \tag{3.15}$$

n は対象とする原文書に含まれる文の数であり，m_i は文 i の単語数である．w_{ij} は単語 ij（i 番目の文における j 番目の単語）の重みである．x_i は，文 i を要約に含めるときに 1 となる決定変数であり，z_{ij} は，単語 ij を要約に含めるときに 1 となる決定変数である．目的関数は，要約に含まれた単語の重みの総和であり，この関数を最大にするように単語を選択する．

式 (3.1) は，要約として選択される単語数が L 以下であることを保証するための制約式である．式 (3.2) は要約に含まれていない文中の単語を要約に含めてしまうことを防ぐための制約式である．式 (3.3) は文間依存木から文を選択する時に，その木構造を保つことを保証する．これは文間依存木からは必ず根を含む根付き部分木が選択されることを意味する．parent(i) は，文間依存木において文 i の親となる文のインデックスを返す関数である．

式 (3.4) から式 (3.9) には決定変数 r_{ij} が含まれている．r_{ij} は，単語 ij を根とした部分木を要約文書に含める場合に 1 となる．式 (3.4) は，単語間依存木から単語を選ぶ場合，その木構造を保つことを保証する．ただし，単語間依存木の根以外の単語 ij を根として部分木を抽出する場合は，その単語 ij には必ず親となる単語 $parent(i,j)$ が存在する．その状況では z_{ij} が 1 のまま $z_{parent(i,j)}$ を 0 にすることが許容されなければならず，それを可能とするのが左辺第二項の r_{ij} である．なお，$parent(i,j)$ は，文 i に対応する単語間依存木において，単語 ij の親となる単語のインデックスを返す関数である．ただし，このままでは，$z_{parent(i,j)}$ と r_{ij} のどちらも 1 である場合も許容されてしまうため，2 つの変数が同時に 1 となることを制限するための制約式 (3.5) を追加する．同様に，このままでは r_{ij} のみが 1 となっている場合も許容されてしまう．r_{ij} が 1 である場合はその単語 ij は必ず要約に含まれなければならないため，制約式 (3.7) を追加することで対処する．本研究では文 i が要約に含まれる場合 (x_i = 1) は，そこから抽出される部分木は高々一つであるとしている．そこで，式 (3.6) により，一つの文から複数の部分木 (の根) が生じることを制限している．また，文 i における単語間依存木の根に相当する単語 (root(i)) に関しては，根付き部分木を抽出する場合，すなわち $r_{\text{root}(i)}$ が 1 であるときのみ要約に含めることを保証する必要があり，そのため制約式 (3.9) を追加している．

冒頭で述べた通り，本研究では単語間依存木から部分木を抽出する際は，根となり得るのは文中の動詞と，単語間依存木全体の根となる単語に限っている．そこで，それ以外の単語が根となることを防ぐことを保証するため，制約式 (3.8) を追加する．ここで，$R_c(i)$ は文 i 中で根の候補となる単語，すなわち動詞のインデックス集合を返す関数である．式 (3.10) は，文（に対応する単語間依存木の部分木）を要約に含めるための最低の単語数を規定するための制約式である．これは，単語の削除により木を刈り込むという手法の性質上，極端に短く刈ってしまうと非文になる可能性が高くなることを防ぐ目的がある．また，要約を最適化問題としてモデル化しているので，目的関数を最大化するために

要約長の限界まで単語を選択しようとして，刈り込みを無制限に許容すると，極端な例では 1 単語からなる部分木を選択してしまうため，それを防ぐための制約である．式 (3.11) は，部分木を抽出する際は必ず一つ以上の主語を含むことを保証する制約である．同様に式 (3.12) は，目的語を一つ以上含むことを保証する制約である．ここで，$sub(i)$ と $obj(i)$ は，それぞれ文 i 中の単語のうち係り受けラベルが主語，目的語である単語のインデックス集合を返す関数である．

提案手法の文圧縮が単語間依存木の刈り込みに基づいている以上，その操作により非文が生成されてしまう可能性がある．ここで，非文となる部分木の生成を避けるための二種類の追加的な制約を導入する．一つ目の制約式は単語対に対するものである:

$$z_{ik} = z_{il}. \tag{3.16}$$

式 (3.16) は．単語 ik と単語 il は必ず同時に要約に含まれることを保証する．これは，片方だけを要約に含めてしまう場合に非文となってしまうような組に対して定義される．具体的には，係り受けタグが PMOD [*1], VC [*2]である単語とその親の単語，否定詞とその親の単語，係り受けタグが SUB あるいは OBJ である単語とその親となっている動詞，形容詞の比較級 (JJR) あるいは最上級 (JJS) とその親の単語，冠詞とその親の単語，“to” とその親の単語である．二つ目の制約式は単語列に対するものである:

$$\sum_{k \in s(i,j)} z_{ik} = |s(i,j)| z_{ij}. \tag{3.17}$$

式 (3.17) は単語の集合に対し，集合中のいずれかの単語を要約に含めるとき，集合中の他の全ての単語も要約に含めることを保証する制約である．具体的には，固有名詞列（品詞タグが PRP%, WP% あるいは POS のいずれかである単語列）や，所有格とその係り先の単語，その間に含まれる全ての単語列である．ここで，$s(i,j)$ は，単語 ij と，例に上げた関係にある単語インデックスの集合を返す関数である．

3.2　評価実験

3.2.1　実験設定

提案手法の有効性を示すために評価実験を行った．実験には RST Discourse Treebank (RST-DTB) [14] に含まれる要約評価用のテストセットを用いた．RST-DTB は Penn Treebank コーパスの一部の文書（Wall Street Journal から収集された 385 記事）からなるコーパスであり，RST に基づく木構造が人手で付与されている．さらにその内 30 記事について，人手で作成された要約文書（参照要約）が存在しており，それらの文書を

[*1] 前置詞とその子の単語の間の関係

[*2] 過去分詞形や現在進行形など，動詞が連続する時のそれらの動詞間の関係

評価用テストセットとした．実験に用いたすべての文書は，Penn Tokenizer によりトークンに区切った．要約システムの入力となる要約長は，参照要約の有するトークン数とした．テストセットに含まれる 30 記事の参照要約には，平均して原文書の 25% 程度の長さの **long** 要約と，平均して原文書の 10% 程度の長さの **short** 要約の二種類が存在する．本実験では両方のテストセットについて先行研究との比較を行う．評価尺度としては ROUGE (Recall-Oriented Understudy for Gisting Evaluation)[73] を用いる[*3]．ROUGE は，現在の文書要約において最も標準的に用いられている自動評価手法であり，次節で詳しく述べる．

抽出粒度の妥当性について検証するため，比較手法として EDU を単位とした木制約付きナップサック問題による要約手法 [45] と，文を単位とする木制約付きナップサック問題による要約手法を用意した．また，単一文書要約において強力なベースラインとなる LEAD 法との比較も行なった．LEAD 法は，要約長に達するまで文書の冒頭から抽出単位を選択していくことで要約文書を生成する手法である．本稿では EDU を抽出単位とする LEAD_{EDU} と，文を抽出単位とする LEAD_{snt} との比較を行う．さらに，提案手法において文（単語間依存木）から部分木を抽出する際に根付き部分木に制限する手法 (**根付き部分木抽出**) も用意し，任意の部分木を抽出対象とする提案手法 (**任意部分木抽出**) との比較を行った．

本実験に用いたすべての文間依存木は，RST-DTB で人手付与された RST 構造を利用しており談話構造解析器は利用していない．考察において，談話構造解析器を用いた追加実験を行い，精度の変化について考察する．また，単語依存木は Suzuki らの提案した依存構造解析手法 [108] を用いて構築した．

なお，本実験では，単語の重要度 w_{ij} として以下を用いた:

$$w_{ij} = \frac{log(1 + tf_{ij})}{depth(i)^2}. \tag{3.18}$$

tf_{ij} は文書における単語 w_{ij} の単語頻度であり，$depth(i)$ は，文書の文間依存木における文 x_i の根からの深さである．また，制約 (3.10) における θ は 8 とした．

本研究では，ILP のソルバーとして ILOG CPLEX version 12.6.0 を用い，分枝限定法により解を求める．提案手法のモデルは他の多くの要約手法で用いられているモデルと同様に NP 困難問題であり多項式時間で厳密解を求める方法は知られていないが，規模の小さい問題であれば分枝限定法を用いるなどで厳密解が求めることができることがある．今回の規模のデータでは現実的な時間で解を得ることが出来る．[*4]．具体的には，計算機環境

[*3] 評価スクリプト実行時のオプションは，ROUGE-1 では “-a -m -s -n 1 -x”，ROUGE-2 では “-a -m -s -n 2 -x” である．また，stopword を含めた評価では “-s” を削除する．

[*4] コーパス中に含まれる原文書の平均単語数は 930.23 単語であり，参照要約は **short** セットで 39.43，**long** セットで 186.10 である．

として OS は Ubuntu 14.04 64bit server，Intel CPU は Xeon E5-2650(2.00GHz) を用いた場合の平均して 1 秒程度の計算時間であった．

3.2.2　ROUGE: Recall-Oriented Understudy for Gisting Evaluation

ここで，現在の文書要約研究において最もよく使われている自動評価尺度である ROUGE [74] について説明する．ROUGE は，機械翻訳における自動評価手法である BLEU を元に提案されたもので，原文書に対して用意された人手による参照要約との表層的な類似度 (n グラムの再現率) により評価を行う．ROUGE の計算式は，以下にのように表される:

$$\text{ROUGE-n}(C,R) = \frac{\sum_{S\in R}\sum_{t_n\in f(S,n)} min(cnt(t_n,C),cnt(t_n,S))}{\sum_{S\in R}\sum_{t_n\in f(S,n)} cnt(t_n,S)}. \tag{3.19}$$

ここで，n は評価に用いる n グラムの単位であり，t_n は実際の n グラムである．また，C が評価されるシステム要約であり，R は参照要約の集合 (一つの原文書に対して複数の参照要約が付与されることがあるため)．S は一つの参照要約であり，$f(X,n)$ は文書 X を n グラムの集合に変換する関数である．また，$cnt(t_n,X)$ は，文書 X 内で n グラム t_n が出現した回数である．

n グラムの単位として最も用いられるのは，ユニグラム ($n=1$) あるいはバイグラム ($n=2$) である．これは，Lin らの報告によれば人手評価との相関を調べた結果，その二つの単位が最も優れていることが確認されたためである[*5]．また，複数文書要約においては $n=2$，単一文書要約においては $n=1$ がより高い相関を示していた．

また，オプションとして事前に規定されたストップワードの除去や語幹抽出を行った上でカウントをする場合もある．これらのオプションを伴って評価したか否かにより結果の数値が大きく異なるため，論文にはこれらのオプションをどう指定したかを明記するべきである．

参照要約に含まれる表現は，原文書の重要な箇所であるとみなすことができるため，その再現率を用いる本指標は，システム要約による原文書の重要箇所の網羅性を評価するものであると考えることができる．

表 3.1　各手法による short 要約セットおよび long 要約セットにおける ROUGE-1,2 値．上 2 行が提案手法であり下 3 行は．[115] からの引用である．

	$\text{short}_{\text{without stopword}}$		$\text{long}_{\text{witout stopword}}$		$\text{short}_{\text{with stopword}}$		$\text{long}_{\text{with stopword}}$	
	ROUGE-1	ROUGE-2	ROUGE-1	ROUGE-2	ROUGE-1	ROUGE-2	ROUGE-1	ROUGE-2
任意部分木	**0.353**	**0.117**	**0.412**	**0.161**	**0.356**	**0.109**	0.455	0.160
根付き部分木	**0.353**	0.114	0.409	0.160	0.355	0.106	0.453	0.158
文選択	0.254	0.086	0.360	0.142	0.282	0.085	0.407	0.144
EDU 選択 [45]	0.321	0.112	0.405	0.157	0.344	0.106	**0.458**	**0.161**
$\text{Marcu}_{\text{ours}}$	0.272	0.093	0.362	0.143	0.315	0.094	0.432	0.155
LEAD_{EDU}	0.240	0.086	0.323	0.126	0.296	0.082	0.402	0.128
LEAD_{snt}	0.218	0.074	0.310	0.121	0.261	0.070	0.387	0.124
Marcu et al.	-	-	-	-	-	-	0.442	-
Ono et al.	-	-	-	-	-	-	0.456	-
Uzeda et al.	-	-	-	-	-	-	0.452	-

3.2.3　結果と考察

ROUGE による比較

表 3.1 に，各手法による ROUGE-1,2 値を示す．まず，short 要約，long 要約セット双方において，提案手法である任意部分木抽出と根付き部分木抽出の間に ROUGE 値の顕著な差はみられなかった．これについては両手法が抽出した実際の部分木を例に定性的な考察を行う．以下では，任意部分木を提案手法とし，他の手法との比較を行う．

また，表 3.1 の下 3 行は Uzeda らによる比較実験の結果から一部の数値を引用している[*6]．ここで，$\text{Marcu}_{\text{ours}}$(0.432) と Marcu et al.(0.440) は，どちらも [84] の手法による結果を示している．前者は我々による再実装の数値であり，後者は [115] において報告されていた数値である．数値が異なるのはトークナイゼーションなどの前処理の違いによるものであると考えられるが，Uzeda らの文献に前処理の詳細がないため，完全な比較とはならないことに注意されたい．とはいえ，両者の数値に大きな差異はないことから，ほぼ同じ実験条件での数値であると判断した．

まず，short 要約セットの stopword を除去した条件 (最も左のカラム) において，提案手法の評価値はホルム法による多重比較の結果，他の全ての手法を有意に上回っていることを確認した．個別の手法と比較すると，文選択手法すなわち文圧縮を一切行わない手法と比較すると，提案手法が大幅に上回っている事がわかる．これは，文をそのまま抽出する場合は，今回の要約設定（平均圧縮率が約 10%）では十分に情報を網羅できないことを示している．次に，EDU 選択手法と比較しても提案手法が上回っている．EDU 選択は文選択を有意に上回っていることから，文よりも細かい EDU を抽出粒度とすることで，要

[*5] 彼らは $n = 1$ から 9 まで比較を行っている

[*6] 具体的には [115] で報告されている結果のうち，本実験に最も条件が似ているもの（Table II の最左のカラム）の数値を引用している．

約文書の情報の網羅性を高めることができている．しかし，EDU という予め決められた長さのテキストスパンを抽出する手法よりも，部分木という可変長のテキストスパンを抽出できる提案手法の方が ROUGE 値は上回っており，その有効性がわかる．LEAD 法は，報道記事の単一文書要約問題において非常に強力なベースラインである．これは，報道記事ではしばしば記事の冒頭でその記事全体の小さなまとめが書かれる傾向にあるためである．今回の実験では抽出単位の異なる二種類の LEAD 法を用いたが，いずれも低い数値となった．これは要約対象となっている文書が，単純な報道記事ではなく，エッセイや社説によって構成されているためであり，冒頭に重要なまとめが記載されているわけではないことが原因である．

一方，long 要約セットでは，提案手法と EDU 選択手法との間に顕著な差は見られなかった．これは，25% という圧縮率が比較的緩く，いずれの手法，抽出単位でもある程度の情報が網羅できるために大きな差が生まれなかったためである．ただし，文を抽出単位とした手法 (文選択および LEAD$_{snt}$) の ROUGE スコアは低いことから，情報網羅性の向上のためには，文よりも小さいテキストスパンを抽出することが重要であるとわかる．

以上の結果から，提案手法のような要約長に柔軟な要約手法は，short 要約セットのように比較的圧縮率の高い設定において有効であることがわかる．

修辞構造の自動解析による精度の変化

表 3.1 の実験結果は，人手で与えられた RST に基づく修辞構造を用いていた．提案手法を任意の文書に適用する場合，文書の修辞構造を自動で解析する解析器が必須である．しかし，修辞構造の自動解析は難しいタスクの一つであり，人手で付与された談話構造を使用したときと比較して精度が劣化してしまうおそれがある．そこで本節では，既存の修辞構造解析器を用いて自動で解析した修辞構造を利用した場合の精度の変化を調べる．今回の実験では自動解析器として，サポートベクターマシンに基づく高い精度を持った解析器である HILDA[29, 43] を用いた．表 3.2 に実験結果を示す．$_{\text{HILDA}}$ と付いている行が，自動解析に基づく依存木を使用した場合の結果である．結果から，いずれの手法も人手で作成された修辞構造を用いたものより ROUGE 値が劣化していることがわかる．short 要約セットの場合は，提案手法の方が劣化が大きい．これは，提案手法が EDU 単位の依存構造を文単位に変更しているためである．HILDA を始めとする自動解析器は，ボトムアップに修辞構造木を組み上げていくため，それを用いて得た修辞構造木を談話依存構造へと変換すると，距離が近い EDU 間の依存関係は比較的高い精度で予測できるが，遠い依存関係の予測精度は低い．このため，遠距離の依存関係である文間の依存関係の同定に

表 3.2　修辞構造を自動解析した場合の精度の変化．上 2 行は表 3.1 より再掲している．

	$\text{short}_{\text{without stopword}}$		$\text{long}_{\text{without stopword}}$	
	ROUGE-1	ROUGE-2	ROUGE-1	ROUGE-2
提案手法	0.353	0.117	0.412	0.161
EDU 選択	0.321	0.112	0.405	0.157
提案手法 $_{\text{HILDA}}$	0.302	0.096	0.385	0.144
EDU 選択 $_{\text{HILDA}}$	0.284	0.093	0.404	0.157

失敗し，提案手法の ROUGE が大きく劣化したと考える．実際，依存先の正解率 [*7]を計算すると，EDU 単位で 0.590，文単位で 0.324 となった．しかしながら，short 要約セットにおいては，減少幅は大きいものの依然として提案手法の精度が，今回比較したどの手法の数値よりも高いことから，提案手法の有効性がわかる．

long 要約セットにおいては，EDU 選択手法の ROUGE 値の減少はほとんど見られなかった．前述の通り，long 要約セットは低い圧縮率であるため比較的多くの情報を要約に含めることができる．文単位の依存関係においても，正解率自体は低くとも選択できる文数が増えるため，short 要約セットよりも ROUGE 値の劣化が抑えられている．すなわち，short 要約セットのように圧縮率の厳しい設定では，より高い精度で抽出単位の依存先を推定する必要がある．

単一文書要約における重要箇所同定

前節で，提案手法の有効性を ROUGE 値によって確認した．本節では，談話構造，すなわち文間依存木の情報が文書中の重要箇所同定に有効かという点について考察を行う．現在，文書要約において主流な問題設定は，同じトピックについて書かれた文書の集合からひとつの要約文書を生成する，複数文書要約である．冒頭で説明した文抽出と文圧縮を組み合わせる手法も全て複数文書要約に取り組んでいるのに対して，今回我々が行なった実験は，一つの文書に対し一つの要約文書を生成する単一文書要約問題である．単一文書要約は複数文書要約と比較して，要約文書に含めるべき文書の重要部分の同定が難しい．なぜならば複数文書要約では文書集合全体として重要な話題は文書横断的に出現するため，その性質を利用できる[*8]が，単一の文書においてそのような情報は利用できないためである．

対象とする文書が報道記事である場合は，冒頭部分に記事全体の要約が書かれやすいと

[*7] ここで正解率とは，自動解析による談話構造木と gold standard の談話構造木を 3.1.2 節の手順で依存木に変換した場合の両者の一致率である．

[*8] その分，複数文書要約においてはいかに冗長な情報を要約に含めないかという点が重要となる．

表 3.3　RST に基づく文間依存木を利用しない場合の結果の変化．上 2 行は表 3.2 より再掲している．

	$\text{short}_{\text{without stopword}}$		$\text{long}_{\text{without stopword}}$	
	ROUGE-1	ROUGE-2	ROUGE-1	ROUGE-2
提案手法	0.353	0.117	0.412	0.161
提案手法 $_{\text{HILDA}}$	0.302	0.096	0.385	0.144
提案手法 ($w_{ij} = \log(1 + tf_{ij})$)	0.308	0.088	0.387	0.130
提案手法 $_{\text{HILDA}}$($w_{ij} = \log(1 + tf_{ij})$)	0.273	0.083	0.380	0.125
同時モデル	0.269	0.072	0.378	0.128

いう強力な基準があるが，そうでない場合に重要な部分を同定することは困難である[*9]．今回我々が用いた単一文書要約の評価セットは，報道記事ではなく社説やエッセイのような文書で構成されているため重要部分の同定が難しい．これは表 3.1 における LEAD 手法の ROUGE 値からも確認できる．

文間依存木の情報が文書の重要箇所同定に与える影響について検証するため，提案手法 (自動解析含む) の単語重要度から $depth^2$ を取り除いたもの，すなわち単語の重要度が単にその文書における出現頻度で決まる場合と，文間依存木の情報を一切用いない従来の同時モデルについても同様に実験を行い比較した．表 3.3 に，それぞれの結果を示す．提案手法の単語重要度から文間依存木の情報（依存木の根からの深さ）を除いた場合に十分な ROUGE スコアが得られないことから，文書の談話構造が単一文書要約における重要箇所の同定に寄与していることがわかる．なお，同時モデルと異なり，木構造の制約という形で文間依存木の情報は用いていることに注意されたい．同時モデルの結果を見ると，文抽出と文圧縮の同時最適化のみでは，本評価セットで有効に機能しないことがわかる．重要文の同定・抽出が困難であるならば，複数文書要約において盛んに取り組まれている文抽出と文圧縮の同時最適化を適用することも困難となり，要約長に柔軟な要約文書の生成も困難となる．本研究の結果は単一文書における重要部分の同定に対するひとつの手がかりとして，文書の談話構造が有効である可能性を示唆しているといえる．

異なる部分木抽出手法の定性評価

ここまで，ROUGE の観点から評価実験の結果についての考察を進めてきた．ここでは，単語間依存木からその部分木を抽出する方法として，任意の部分木を抽出することの有用性を，例を示して考察する．図 3.4 に，任意部分木抽出手法と根付き部分木抽出手法

[*9] ただし，報道記事であれば必ず記事冒頭に記事全体の要約が存在しているとは限らないことも分かっており，単純に記事冒頭を機械的に抽出しても必ずしも重要箇所が得られるとは限らない [128].

原文	:	John Kriz, a Moody's vice president, {said} Boston Safe Deposit's performance has been hurt this year by a mismatch in the maturities of its assets and liabilities.
根付き部分木抽出	:	John Kriz a Moody's vice president [{said}] Boston Safe Deposit's performance has been hurt this year
任意部分木抽出	:	Boston Safe Deposit's performance has [been] hurt this year
原文	:	Recent surveys by Leo J. Shapiro & Associates, a market research firm in Chicago, {suggest} that Sears is having a tough time attracting shoppers because it hasn't yet done enough to improve service or its selection of merchandise.
根付き部分木抽出	:	surveys [{suggest}] that Sears is having a time
任意部分木抽出	:	Sears [is] having a tough time attracting shoppers

図 3.4　二つの手法が共通して要約に選択した文と，それぞれが抽出した部分木の例.

が共通して要約文書に含めた文と，そこから抽出した部分木に対応する二つの文を示す．なお，これは short 要約セットにおける例である．

ここで，{·} は依存構造解析器が単語間依存木の根であると出力した語であり，[·] は要約システムが部分木の根として選んだ語である．任意部分木抽出においては，例に示したいずれの文も解析器の根以外の単語を根として部分木を抽出している．例に見るように，目的節や that 節の内容の方が重要な情報を持つことが多いため，その部分のみを抽出することは，限られた長さで重要な情報のみ要約に含める上で有用であり，今回の実験ではこうした事例が少なかったこともあり ROUGE スコアで大きな差がでなかったが，特に圧縮率の高い設定[*10]では有効であろう．

ただし，単語間依存木からの単語の枝刈りに基づく文圧縮手法では，常に非文が生成されてしまうリスクが存在する．例えば，以下の文のように長く構造が複雑な文の場合は文法的に正しくない文が生成されてしまう:

> Ben earns any fees sent directly to charity and **[is] taxable on them, the IRS says**; of course, he also may take a charitable deduction for them

ここで，文中の太字になっている単語が，要約器が選択した単語，**is** が要約器が根として選択した単語である．

このような誤りは，文法性を担保するための制約が不十分な場合と，単語間依存機を構

[*10] すなわち，原文書そのものの長さに対し非常に短い要約を生成する必要があるような設定.

築するための構文解析に失敗している場合が存在する．このように長い文に対しても文法的に正しい圧縮を行うことは，今後実用を考える上でも重要な課題である．このような例を減らすための方法はいくつか考えられる．まずは，単語間依存木を構築するための構文解析器の精度向上が挙げられる．また，既存の汎用的な解析器ではなく，要約に特化して訓練された解析器を用いるという方法も存在する．たとえば，提案手法において文間依存木を構築するのに用いた修辞構造の解析器について，Wang らは文書要約タスクに特化した解析器を訓練することの有効性を報告している [122]．また，構文解析器を用いずに文圧縮を行う手法も登場しており，そのような手法を内部で用いることも考えられる [129, 32]．

抽出粒度と要約文書の文数の関係

EDU は RST における談話構造の基本単位であるが，抽出型要約の抽出単位として適切とは限らない．図 3.5 に，ある文[*11]とその文を構成する EDU の例を示す．図のように，EDU はおおよそ節に対応する文よりも細かな単位である．

抽出単位が文よりも細かい EDU であることは，EDU 抽出は，文圧縮を逐次適用した要約手法として考えることができる．つまり，EDU 抽出による要約は多くの文を事前に圧縮しつつ抽出していることに相当する．このように文よりも小さな断片を組み合わせて要約を生成すると，文を組み合わせる場合よりも長さ制約をちょうど満たすように要約を生成することができる可能性が高い．よって，ROUGE 値も上昇する傾向にある．しかし，EDU は文よりも短いため，たとえ一貫性があろうともそれを読んだ読者が違和感を覚えてしまうだろう．

抽出型要約において，文書中の多くの文から細かな断片を集めることで情報の断片化された要約の生成につながっているかどうかは，要約文書に含まれる抽出単位集合の元となる文の数が，その一つの指標となる．言い換えると，生成された要約を構成する文の数が，参照要約すなわち人間によって生成された要約に近い方が，自然で読みやすい要約に

- The scare over Alar, a growth regulator
- that makes apples redder and crunchier
- but may be carcinogenic,
- made consumers shy away from the Delicious,
- though they were less affected than the McIntosh.

図 3.5　本データセットにおける文と EDU の関係の一例．5 行全体で一つの文であり，各行が一つの EDU に対応している．

[*11] この文は wsj_1128 より選択した．

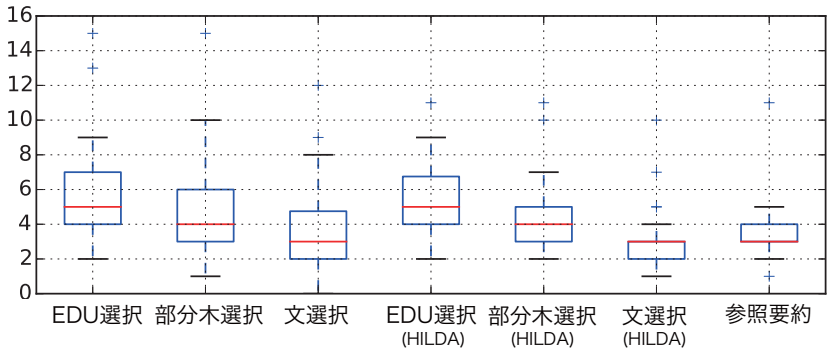

図 3.6　short 要約セットにおいて各手法が使用した原文書の文の数．(HILDA) と付いているものは自動解析による修辞関係を利用している．

なっていると考えられる．そこで本節では各手法が生成した要約文書に含まれる原文書の文数を比較した．比較に用いた手法は提案手法である任意部分木抽出の他に，文選択と EDU 選択である．文選択は原文書中の文の数，EDU 選択は原文書中の EDU に対応する文の数，部分木選択は，部分木に対応する文の数である．なお，参照要約は人間が自由に生成した要約であるため，必ずしも原文書の文とは対応していないことに注意されたい．

short 要約セットにおいて各選択手法が選択した文について箱ひげ図を図 3.6 に示す．各々の箱の上辺と下辺は，それぞれその手法が選択した文数の第一四分位点，第三四分位点を表しており，箱の中の線は中央値を表している．箱の上下に伸びる線（ひげ）の先は，それぞれ最大値，最小値を表し，ひげよりも外側に見られる + 印は，外れ値である．

図を見ると，EDU 選択手法が最も多くの文を用いて要約を生成していることがわかる．一方で文選択手法は，比較手法の中では最も参照要約の文に近いが，3.2.3 節で示した通り情報の網羅性という点で十分な要約を作成できない．部分木抽出は文選択と EDU 選択の間で，両者のように実際に抽出されるテキストスパンの長さを固定せずに要約システムが柔軟に各文から抽出する部分木を選択することができる．それにより情報の網羅性と要約としての自然さを両立出来ている．なお，部分木抽出手法の平均文数は 4.73 であり，中央値は 4 文であった．これに対し，EDU 選択の平均文数は 5.77 で中央値は 5 文であった．これは提案手法の方が有意[*12] に少ない文を用いて要約を生成していることを示している．自動解析を利用した場合も同様の傾向であるため詳細は割愛する．

同様に，long 要約セットについて図 3.7 に示す．圧縮率が低くなった場合も全体の傾向としては short 要約セットとの大きな差異はない．全体的にばらつき（箱の縦の長さ）が大きくなっているが，これは参照要約自体の長さのばらつきが short 要約セットよりも大きいことが原因である．

[*12] ウィルコクソンの符号順位検定 ($p < 0.05$)

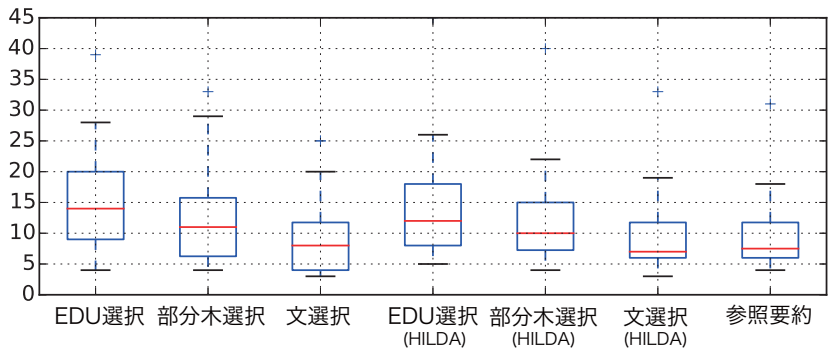

図 3.7　long 要約セットにおいて各手法が使用した原文書の文の数.

3.3　本章のまとめ

本研究では，単語間の依存構造解析に基づく単語間依存木と，RST に基づく文間依存木から入れ子依存木を構築し，そこから要約文書に含める単語を選択する要約生成問題を ILP として定式化した．提案手法は EDU の依存木の刈り込み手法に比べ，過剰に文を区切ることなく ROUGE を向上させることが確認できた．また，単語の依存木からその部分木を抽出する方法として，構文解析器が出力した根にこだわらない任意部分木抽出手法について，その有用性を定性的に分析した．さらに，人手で作成された修辞構造以外に，修辞構造解析器で推定された修辞構造も用いて，その精度への影響を確かめた．

提案手法の抱える課題は，任意の文書に対して適用する際に修辞構造の自動解析を利用しており，その精度の与える影響が大きいという点である．今回は EDU 単位の修辞構造解析結果を文単位の依存関係に変換して利用したが，はじめから EDU 単位の依存関係を獲得する研究 [131] も存在する．これらを踏まえ，今後はより良い文間依存木の獲得方法を検討していく．

今回は RST から得られる情報のうち文間の依存関係のみに着目したが，各文内における EDU 間の関係や修辞構造のラベルを考慮して文圧縮や文抽出を行うことが可能であり，今後取り組むべき課題として興味深い．

今後は，他のコーパスの文書や，複数文書要約においても提案手法を適用することを考えている．

第 4 章

要約器の効果的な訓練のための大規模要約資源の活用手法

4.1 New York Times Annotated Corpus

New York Times Annotated Corpus (NYTAC) は，約 180 万記事の New York Times 紙の新聞記事によって構成されており，そのうちの約 65 万記事には人手で生成された要約が付与されている．表 4.1 に主な統計量を，今回の実験で使用するテストセットと共に示す．また，本コーパスには人手による様々な種類の要約が混在している．たとえば，文抽出を行うことで作成されている要約もあれば，文抽出に加えて文圧縮や文融合などを行うことにより作成されている要約も存在する．加えて，言い換えなどの非抽出的な操作を行うことで作成されている要約や，メタな視点に立って書かれた説明口調な要約も存在する．このように異なる種類の要約が存在することは，文圧縮，文融合や非抽出的なアプローチに基づく要約など，異なる要約技術の発展に有益である [91, 49].

表 4.2 に，NYTAC[*2]の 3 つの例を示す．例は，人手要約 (*abs*)，とその原文書 (*doc*)，両者で対応づいた部分テキストを共通の番号付きの下線で示す．文書 ID(*id*) の隣に示す括弧内の数字は，4.2.1 節で述べた文抽出オラクル要約が獲得した ROUGE-2 値である．各事例の *abs* を観察すると，要約作成者が文の抽出とみなせるような操作を行う際であっても，完全にコピーを行うわけではなく，様々な小さな変更を行っているがわかる．たとえば，表 4.2 であれば，冠詞を含む機能語の削除や時間を表す副詞 (表中の [e]) の削除，クォーテーション (表中の [b],[c]). ピリオド (表中の [d]) やハイフンなどの記号の削除，あるいは句レベルの言い換え (表中の [a]) などである．さらに，NYTAC 中の要約の中に

[*1] NYTAC に含まれる要約の総数は 658,874 であるが，原文書や要約文書が空であるなどのノイズを除去すると，最終的には 524,216 事例となった．

[*2] ©2016 The New York Times Annotated Corpus, used with permission

表 4.1　本研究で扱うコーパスとその統計量．事例数を除く全ての数値は平均値である．また，下から 3 つのコーパスは実験で用いるターゲットデータである．

	事例数	参照要約		原文書		圧縮率
		平均文数	平均単語数	平均文数	平均単語数	
NYTAC	524,216[*1]	2.42	38.79	29.36	607.62	0.128
DUC2002	533	5.62	101.09	27.40	549.61	0.363
$\text{RSTDTB}_{\text{long}}$	30	9.57	186.10	116.77	930.23	0.246
$\text{RSTDTB}_{\text{short}}$	30	3.5	39.43	116.77	930.23	0.093

は，原文書中の表現はほとんど用いず「この文書が何について述べられたものか」という，メタな視点で作成されたものも存在する．

抽出や圧縮など，各要約操作を計算機に行わせるためにはそれぞれ異なった技術が必要であり，それぞれにとって有効な訓練事例は異なることが予想できる．たとえば，文抽出による要約手法を用いる場合は，原文書からの表層的な抽出にこだわらない方法で作成された要約は訓練事例として適切ではないと考えられる．従って，本研究では NYTAC を訓練事例として利用するにあたり，訓練する要約器に合わせた事例の選択が必要であると考える．本研究では，現在の文書要約において最も広く用いられている文抽出に基づくアプローチを対象とする．そこで，適切な文抽出により人手の正解要約（参照要約）を再現できる度合いを基準とし，訓練事例の選別を行い，その有効性を確かめる．

4.2　NYTAC を利用した要約器の訓練手法

本節では，今回用いる要約器とそのパラメータ推定方法，そして訓練に NYTAC を利用する手法を説明する．NYTAC の利用法として，訓練方法そのものを規定する 5 つの手法と，それらの前処理として要約器の特性に合わせた事例選択の方法を提案する．

4.2.1　ナップサック問題による文抽出型要約器

本稿では，ナップサック問題に基づく要約器を用いる．すなわち，原文書中の各文に要約として選択した場合の利得を付与し，与えられた最大単語数以下で利得の総和を最大化する文の組み合わせを選択する．ナップサック問題による定式化は，単一文書要約を組合せ最適化問題として定式化する上で最も自然な方法の一つである [86, 140]．近年提案されている単一文書要約を対象とした多くの手法はナップサック問題の拡張であり，そのためベースラインとして頻繁に登場している [93, 45]．具体的には，以下のように定式化で

id	1996/07/28/0868107 (0.111)
abs	Neil Strauss recommends hearing reunion of German progressive-rock group Cluster at the Knitting Factory
doc	Cluster, the Knitting Factory, 74 Leonard Street, TriBeCa, (212) 219-3006. What can make reunions by obscure European groups like Cluster more exciting than classic-rock reunions is that the former often haven't performed in the United States before. [...] (7 sentences)
id	1988/03/23/0129961 (0.366)
abs	Gov James E McGreevey, whose insistence on staying in office until Nov 15 set off political, legal and public relations challenges, needs only to remain in office until midnight[1] Sept 3 to outlast state deadline that would automatically force special election to choose his successor[2]; his announcement Aug 12 that he was stepping down because of extra- marital affair with[3] **unidentified man**[a] led to coup attempt from within his own party, which he has managed to fend off; Richard J Codey, Senate president and fellow Democrat, will complete final 14 months of McGreevey's term[4] under provisions of New Jersey's Constitution; photos
doc	After revealing on national television that he is gay, seeing his name become the punch line of countless late-night comedy gags, and fending off a coup attempt from within his own party, Gov. James E. McGreevey needs only to remain in office until midnight[1] Friday to outlast the state deadline that would automatically force a special election to choose his successor. [2] Mr. McGreevey announced on Aug. 12 that he was stepping down because he had had an extramarital affair with[3] **a man he did not identify**[a], but his insistence on remaining in office until Nov. 15 set off political, legal and public relations challenges. [...] Mr. McGreevey will pass that deadline as of Saturday, allowing the Senate president, Richard J. Codey, a fellow Democrat, to complete the final 14 months of his term[4]. [...] (27 sentencces)
id	1996/08/30/0874273 (0.692)
abs	CBS announces that it has signed Diane English, executive producer of **Murphy Brown**[b], to take over as executive producer of its troubled new comedy, **Ink**[c], which stars Ted Danson and Mary Steenburgen[5]; Ink will have delayed debut of **Oct 21**[d]
doc	CBS announced **yesterday**[e] that it had signed Diane English, the longtime executive producer of **"Murphy Brown"**[b] to take over as executive producer of its troubled new comedy, **"Ink,"**[c] which stars Ted Danson and Mary Steenburgen. CBS also said the show, whose first four episodes fell so short of expectations that they were shelved, would come back with new episodes created by Ms. English starting on **Oct. 21**[d]. [...] (6 sentences)

表 4.2　NYTAC に収録されている要約の例．3 つの事例のそれぞれについて，*id*，*abs*，*doc* はそれぞれ文書 id，専門家による要約，原文書である．余白の都合上，*doc* に関しては途中で表示を打ち切り ([...]) と表示するとともに，最後に *doc* 全体での文数を示す．*id* の横の括弧内に書かれた数値はその事例において文抽出型オラクル要約の獲得する ROUGE-2 値である．文抽出型オラクル要約については 4.2.1 節にて述べる．各事例において，*abs* と *doc* で対応づいたテキストセグメントについては同じ番号と色の下線で対応を示している．太字と上付きアルファベット ([a]-[f]) で示された単語や句については 4.1 節で言及する．

きる:

$$\text{max.} \quad \sum_{i}^{n} b_i x_i$$

$$\text{s.t.} \sum_{i}^{n} c_i x_i \leq L; \tag{4.1}$$

$$x_i \in \{0,1\}; \forall i. \tag{4.2}$$

x_i は文 i を要約として選択した場合に 1 となる決定変数であり，b_i は文 i を要約として選択した場合に得られる利得である．また，c_i は文 i を要約として選択した場合のコスト（文 i の単語数）であり，L が要約に含めることのできる最大の単語数である．ここで，文 i の利得 b_i を素性関数 ϕ とその重みベクトル $\mathbf{w}$ として定義し，$\mathbf{w}$ を訓練により推定する: $b_i = \mathbf{w} \cdot \phi(s)$.

重みベクトル $\mathbf{w}$ を推定する方法として，本稿では教師あり構造学習の枠組みを採用する．具体的には Passive Aggressive 法 [22] の一種である PA-II 法に基づき，重みベクトルの更新を以下の最適化問題の解として定義する:

$$\mathbf{w}_{t+1} = \underset{\mathbf{w}}{\text{argmin}} \frac{1}{2}||\mathbf{w} - \mathbf{w}_t||^2 + C\xi^2 \tag{4.3}$$

$$\text{s.t.} loss(\mathbf{w}) \leq \xi. \tag{4.4}$$

$\mathbf{w}_t$ は t ステップ目の重みベクトルである．また，ξ は制約を破った場合のペナルティを意味するスラック変数であり，C はペナルティの影響を決めるパラメータである．ここで，損失関数について $loss(\mathbf{w}) = 1 - ROUGE(o, s)$ と定義する．s は重みベクトル $\mathbf{w}$ に基づき要約器が生成した要約である．o は，参照要約により求めた文抽出オラクル要約であり，次の項目で詳しく説明する．$ROUGE(o, s)$ はオラクル要約 o を正解としたときのシステム要約 s の ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [74] 値である．この最適化問題は，ラグランジュの未定乗数法により以下のような閉形式で表すことができる:

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \tau_t(\phi(o_t) - \phi(s_t)), \tag{4.5}$$

$$\tau_t = \frac{loss(\mathbf{w}_t)}{||\phi(o_t) - \phi(s_t)|| + \frac{1}{2C}}. \tag{4.6}$$

$$\tag{4.7}$$

$\phi(\cdot)$ は文書を素性ベクトルへ変換する関数であり，与えられた文書中に含まれる各文の素性ベクトルの和を取ったベクトルである．

$\phi(\cdot)$ の構築に用いる文単位の素性として，以下のものを用いる:

(a) 文に含まれる内容語の割合,

(b) 頻度上位 1 万語の bag-of-words,
(c) 文に含まれる単語数の対数,
(d) 文に含まれる単語数の，文書全体の単語数に対する相対的な長さ,
(e) 文の，文書における絶対位置の対数,
(f) 文の，文書における相対的な位置,
(g) 文が文書の前半 20% 以内に位置している場合に 1，それ以外に 0 となる二値素性,
(h) 文による文書中のユニグラム被覆率,
(i) 訓練済みの $NytOnly$ モデルが出力した文の利得（本素性は 4.2.2 節における $Featurize$ においてのみ利用される）.

また，訓練の際は Zhao らの報告に基づき並列化を行った [135].

要約のデコーディングは，Algorithm 1 で示す動的計画ナップサックアルゴリズムにより，擬多項式時間で厳密解を得ることができる．ここで，s,c はそれぞれ文書中の各文のスコア b_i，各文の長さが格納された配列である．また，L は制限として与えられる要約長，N は文の数である．最終的に出来上がる R は，要約として選択された文の番号を格納した配列である．

文抽出オラクル要約

オラクル要約とは，ある評価指標において，人手による参照要約に対して要約器が獲得できる評価値の上限となる要約である．本稿では評価指標として ROUGE-2 [74] を用いた．すなわち，文抽出型要約においては参照要約に対する ROUGE-2 が最も高くなる文の組み合わせを選択することでオラクル要約を生成する．

オラクル要約の評価値は要約器が獲得できる評価値の上限値であるため，その値が高い事例は，適切な文を選択することで参照要約の大部分を再現できることを意味する．

本研究では，NYTAC の参照要約-原文書対から計算されたオラクル要約を，二つの異なる目的で利用する．

- 一つ目の目的は，要約器の訓練における教師信号，すなわち式 (4.5)-(4.6) における o_t に，参照要約に代わって利用することである．構造化パーセプトロンに基づく学習手法においては，正解となる素性ベクトルを正解ラベルから構築する必要があるが，参照要約をそのまま変換するだけでは，文の位置など一部の素性が適切に学習されない.
- もう一つの目的は訓練時の事例選択への利用である．オラクル要約の獲得する ROUGE-2 値が低い場合，要約器がどのような文の組み合わせを選んだとしても高い評価値が得られない．そのため，文抽出要約器の訓練事例としてのふさわしさの指標としてオラクル要約の評価値を利用する．より詳細な説明は 4.2.3 節にて述

Algorithm 1 動的計画ナップサックアルゴリズム

```
INPUT: s, c, L, N
for l=0 to L do
    V[0][l] = 0
    Z[0][l] = 1
end for
for n=1 to N do
    for l=0 to L do
        V[n][l] = V[n-1][l]
        Z[n][l] = 0
    end for
    for l = c[n] to L do
        if s[n] + V[n-1][l - c[n]] ≥ V[n][l] then
            V[n][l] = s[n] + V[n-1][l-c[n]]
            Z[n][l] = 1
        end if
    end for
end for
l=L
R = {}
for i=N to 1 do
    if Z[i][l] = 1 then
        R = R ∪ i
        l = l - c[i]
    end if
end for
return R
```

べる．

本稿では，参照要約に含まれるバイグラムを被覆対象とする最大被覆問題を解くことで

オラクル要約を作成する:

$$\text{max.} \quad \sum_{j}^{m} z_j$$

$$\text{s.t.} \sum_{i}^{n} c_i x_i \leq L; \tag{4.8}$$

$$\sum_{i}^{n} a_{ij} x_i \geq z_j; \forall j \tag{4.9}$$

$$x_i \in \{0,1\}; \quad \forall i \tag{4.10}$$

$$z_j \in \{0,1\}; \quad \forall j. \tag{4.11}$$

ここで z_j は，要約として選択した文が参照要約中に出現する j 番目のバイグラムを含む場合に 1 となる決定変数である．また，式 (4.9) は文と被覆対象の整合性を取るための制約式であり，a_{ij} は文 i がバイグラム j を含む場合に 1 で，そうではない場合に 0 となる．なお，オラクル要約の計算時には全ての文書に対して，英数字以外の全ての文字を削除した上で語幹抽出を行った．これは，ROUGE の公式評価スクリプトをオプションは “-m -r 2” で実行した時と同じ前処理である．このように，人手により自由に作成された参照要約から関連する原文書の抜粋を構築するという試みは構造学習に基づく要約器の学習 [112, 105] の際に行われているほか，より一般的な問題として Marcu らによる試みがある [85].

4.2.2　NYTAC を利用した要約器の訓練

本節では NYTAC を訓練に利用する 5 つの手法について述べる．ターゲットデータを訓練に用いる際は，五分割交差検定を利用している．なお，ターゲットデータによる五分割交差検定のみで構築されたモデルを $TrgtOnly$ とし，以下に述べる全ての手法に対するベースラインとする．

> $NytOnly$ は NYTAC のみで訓練を行い，ターゲットデータの情報を一切利用しない手法である．そのため，この手法で訓練されたモデルのスコアは純粋に大量の訓練事例による効果を表しているといえる．
>
> $Mixture$ は，NYTAC の事例とターゲットデータの交差検定における訓練セットを混合した事例集合を用いて訓練を行う手法である．本手法は単純に訓練事例のサイズを増加させる目的で NYTAC を利用しているが，以下の 3 手法は訓練済みの $NytOnly$ モデルを用いる．

$LinInter$ は，訓練済みの $TrgtOnly$ モデルと $NytOnly$ モデルの重みベクトルの線形補間により作成した重みベクトルを用いて予測を行う手法である．重みパラメータは交差検定における検証セットを利用して決定する．

$Featurize$ は，訓練済みの $NytOnly$ モデルが出力する文の利得を追加の素性として，ターゲットデータで訓練を行う手法である．

$FineTune$ は，訓練した $NytOnly$ モデルの重みベクトルを初期値として利用し，そのモデルをターゲットデータで学習する手法である．NYTAC で学習した重みベクトルを初期値とすることは，PA 法をはじめとするオンライン学習手法で標準的なゼロベクトルによる初期化と比較してより良いパラメータの探索が可能となると期待できる．

4.2.3　オラクル要約による事例選択

4.1 節で述べた通り，NYTAC には多くの種類の参照要約が含まれる．ある要約は抽出型である，すなわち原文書から適切なテキストスパンを抽出することで生成されているが，ある要約は生成型，すなわち言い換えや一般化などにより原文書に存在しない表現を含んだ形で作成されている．本稿では文抽出に基づく要約器を用いるため，前者のような抽出に基づく要約事例を訓練事例として用いることは有用であると考えられる一方で，後者の要約を訓練事例として用いるのは適切ではないと考える．より一般的には，もし我々が特に統制のない状態で作成された大量の事例を手に入れた場合，その時の利用目的，今回は文抽出に基づく要約器の訓練をする上で有用であると思われる部分集合を自動で選択するという処理が必要になる．このような考えはドメイン適応の分野においても知られている [100, 11].

本稿では，事例選択の基準として 4.2.1 節で述べた文抽出オラクル要約の ROUGE-2 値を用いる．オラクル要約の ROUGE-2 値が十分に高い場合は，原文書から適切な文の組み合わせを選択することで参照要約の大部分を再現できる．反対に，オラクル要約の ROUGE-2 が低い場合はどのような文の組み合わせを選択したとしても参照要約とは表層的に異なる要約しか生成することが出来ない．このようなケースでは，言い換えや一般化など，より複雑な操作が必要となる．

具体的な事例選択の方法としては，まず NYTAC 中の全ての要約-文書対に対しオラクル要約を計算し，その ROUGE-2 値が閾値 thr を超えた事例のみで構成された部分集合を選択する．実験を行う際は，thr 毎に $NytOnly$ モデルを訓練しておき，交差検定の検証セットにおいて，どのモデルを利用するかを決定する．

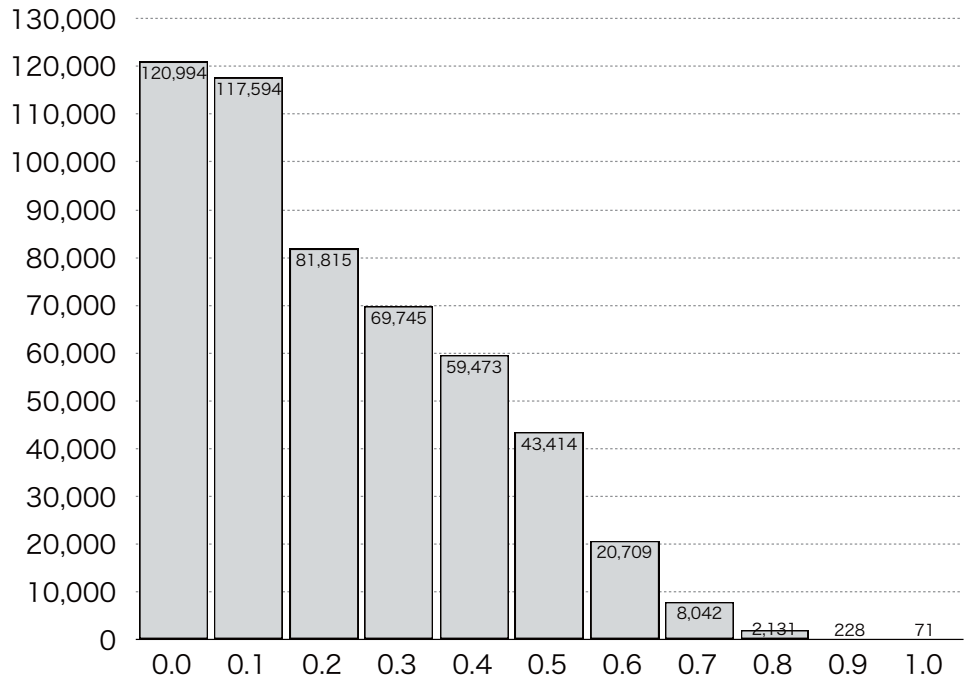

図 4.1　thr ごとに選択される訓練事例の数.

図**??**に thr ごとの事例数を示す．表により，NYTAC の中の 2,430 事例 ($thr = 0.8$) は，原文書中の文を適切に選択することで人手により作成された要約の 8 割以上を被覆した要約文書が作成可能であることが分かる.

4.3　評価実験

4.3.1　実験設定

本稿では，ターゲットデータとして，単一文書要約におけるテストデータとして用いられることの多い三種類のコーパスを用いる．以下，各データを DUC2002，RSTDTB$_{long}$，RSTDTB$_{short}$ と呼び，それぞれの統計量を表 4.1 に示す．DUC2002[27] は評価型ワークショップである Document Understanding Conference (DUC) において単一文書要約が共通課題として設定された際のテストデータである[*3]．567 個の原文書-参照要約対で構成されており，各参照要約の長さは 100 単語以下[*4]となるように生成されている RSTDTB$_{long}$ および RSTDTB$_{short}$ は，Rhetorical Structure Theory Discourse Treebank (RSTDTB，LDC2002T07) と呼ばれる，修辞構造理論 (RST) に関するコーパスに含まれ

[*3] 2002 年は単一文書要約が共通課題として設定された最後の年度である.

[*4] 評価の際には，全ての要約を 100 単語に切り詰める.

ている要約データである．RSTDTB は Penn Treebank 中の 385 記事の Wall Street Journal 記事によって構成されており，各記事には RST に基づく談話構造が人手でアノテーションされている．さらに，そのうち 30 記事には人手で生成された参照要約が付与されている．RSTDTB$_{long}$ と RSTDTB$_{short}$ では参照要約の長さが異なる．具体的には，RSTDTB$_{long}$ では要約者は原文書の 25% の長さとなるよう要約を作成するよう指示されており，RSTDTB$_{short}$ は 2,3 文と指示されている．RST では，文書を文よりも細かいおおよそ節に相当する elementary discourse unit (EDU) に分割し，談話構造を考える上での最小単位としている．そのため，RSTDTB を利用した要約研究はこの EDU を抽出単位とすることが多く，本研究においても同様に文ではなく EDU の抽出を行う．

要約器に与える制限要約長 L は，DUC2002 を用いた実験においては 100 単語とし，RSTDTB$_{long}$ および RSTDTB$_{short}$ を用いた実験ではそれぞれの参照要約の単語数と等しくなるように設定した．

評価尺度としては，単一文書要約の評価によく用いられている ROUGE-1 を用いる．この際，DUC2002 において人手との相関が最も高かった設定 [74] である，ストップワードを削除した上で語幹の抽出を行った上での数値[*5]を報告する．また，参考のために同様に人手評価との高い相関を示した ROUGE-2 による評価結果も付与する．ROUGE-2 においては，ストップワードの有無による差は報告されていないため，語幹抽出のみを行った上で評価を行う[*6]．

ここで，三種類のテストセットの参照要約の使途の違いについて言及しておく．DUC2002 および RSTDTB$_{long}$ の参照要約は報知的要約と呼ばれる要約であり，対して RSTDTB$_{short}$ は指示的要約である[*7]．本稿で用いる要約器を含め，先行研究における多くの要約手法は報知的要約を対象としている．そのため，まず DUC2002 および RSTDTB$_{long}$ に対して提案手法の評価を行い，次に RSTDTB$_{short}$ おける実験を行うことにより NYTAC が指示的要約の生成についての学習にも寄与するかを確認する．

4.3.2　DUC2002 における評価結果

DUC2002 による評価結果を表 4.3 に示す．まず，$Featurize$ を除くすべての提案手法が $TrgtOnly$ を有意に上回る結果となった[*8]．これは，NYTAC を追加の訓練データとし

[*5] 具体的には，ROUGE の公式評価スクリプト（バージョン 1.5.5）においてオプション “-a -x -n 1 -m -s” で実行した時のスコアである．

[*6] 具体的には，ROUGE の公式評価スクリプト（バージョン 1.5.5）においてオプション “-a -x -n 2 -m” で実行した時のスコアである．

[*7] これら二つの要約についての具体的な説明は，例えば Nenkova による著作 [91] では以下のように記述されている: *A summary that enables the reader to determine about-ness has often been called an indicative summary, while one that can be read in place of the document has been called an informative summary*.

[*8] 検定にはウィルコクソンの符号順位検定 ($p \leq 0.05$) を用いた．

て用いることの有効性を示している．$NytOnly$ は訓練時に DUC2002 の情報を一切用いていないにもかかわらず，ROUGE 値が $TrgtOnly$ を有意に上回っていることは興味深い結果であり，訓練データの規模の重要性を示唆している．5 つの提案手法間で比較すると，$FineTune$ が NYTAC を用いる上で最も有効な手法であることが分かる．

名前の末尾に $_{slct}$ がついた結果は，4.2.3 節で提案した事例選択を事前に行った上で訓練した場合の結果である．その差は有意ではないものの，事例選択を行う前と比較して僅かな向上が見られた．また，$FineTune_{slct}$ は事例選択を行わない場合と同様に他の手法を有意に上回っている．なお，五分割交差検定によって選択された thr の値は，それぞれ 0.2, 0.3, 0.3, 0.3, 0.3 となった．

最後に，いくつかの既存手法と提案手法を比較する．LEAD は原文書の冒頭から 100 単語を機械的に抽出する手法である．LEAD は単純な手法であるにもかかわらず，単一文書要約において非常に強力なベースラインであることが知られている．TextRank [87] も単一文書要約において知られた手法であり，PageRank アルゴリズムに基づくグラフベースの重要文ランキングを行う．s21-31 は DUC2002 の単一文書要約タスクに参加したシステムのうち上位 5 件のシステムである．$FineTune_{slct}$ は，LEAD，s27，s29，s31 を有意に上回る性能を示している一方で，s28 を有意に下回る結果となっている．

4.3.3　RSTDTB$_{long}$ における評価結果

RSTDTB$_{long}$ による評価結果を表 4.4 に示す．概ねの結果は DUC2002 と同様である．提案手法の中では $FineTune$ が最も高い評価値を得ており，その評価値は $TrgtOnly$ をはじめ，他の提案手法を有意に上回っている．加えて，事例選択により更なる向上が見られる．

ここで，比較する先行研究について説明する．LEAD$_{EDU}$ は要約長 L に至るまで原文書の冒頭の K 個の EDU を抽出する手法である．Marcu は原文書の RST に基づいて EDU の重要度をランキングすることで要約を生成する手法である [84]．Hirao は，平尾らの提案した RST の依存構造木からの刈り込みに基づく要約手法であり，EDU 抽出による要約手法の中で state-of-the-art な手法である．

これらの手法と比較しても，$FineTune_{slct}$[*9]が高い評価値を得ていることが分かる．Marcu および Hirao は人手でアノテーションされた原文書の修辞構造を利用した手法であり，そのような情報を一切用いずに state-of-the-art な評価値を獲得したことは，大規模な訓練データの重要性を物語っているといえる．

[*9] 五分割交差検定によって選択された thr の値は，それぞれ 0.1, 0.1, 0.1, 0.1, 0.1 であった．

表 4.3　DUC2002 における各手法の ROUGE 値．表は便宜上横線により 4 つの区画に分けられており，それぞれの区画で最も高い ROUGE-1 値を太字で示している．

	ROUGE-1	ROUGE-2
$TrgtOnly$	**0.400**	**0.218**
$NytOnly$	0.409	0.217
$Mixture$	0.411	0.219
$Featurize$	0.404	0.220
$LinInter$	0.409	0.217
$FineTune$	**0.415**	**0.221**
$NytOnly_{slct}$	0.411	0.224
$Mixture_{slct}$	0.412	0.224
$Featurize_{slct}$	0.403	0.219
$LinInter_{slct}$	0.411	0.224
$FineTune_{slct}$	**0.418**	**0.225**
LEAD	0.413	0.224
TextRank	0.409	0.206
s21	0.416	0.223
s27	0.401	0.212
s28	**0.428**	**0.228**
s29	0.400	0.213
s31	0.393	0.203

4.3.4　RSTDTB$_{short}$ における評価結果

RSTDTB$_{short}$ による評価結果を表 4.5 に示す．$FineTune$ が他の手法に比較して高い評価値を獲得しているという点では DUC2002, RSTDTB$_{short}$ と同様だが，いずれの差も有意では無いという点で異なる．事例選択[*10]による影響を見ると，$FineTune_{slct}$ を除き有効な効果は得られなかった．

RSTDTB$_{short}$ における性能向上は DUC2002 や RSTDTB$_{long}$ と比較して大きなものではなかったが，指示的要約と報知的要約という両者の違いを考慮すれば妥当な結果といえる．

[*10] 五分割交差検定によって選択された thr の値は，それぞれ 0.3, 0.6, 0.3, 0.3, 0.6 であった．

表 4.4　RSTDTB$_{long}$ における各手法の ROUGE 値．表は便宜上横線により 4 つの区画に分けられており，それぞれの区画で最も高い ROUGE-1 値を太字で示している．

	ROUGE-1	ROUGE-2
$TrgtOnly$	**0.324**	**0.121**
$NytOnly$	0.313	0.128
$Mixture$	0.320	0.132
$Featurize$	0.359	0.150
$LinInter$	0.313	0.128
$FineTune$	**0.385**	**0.158**
$NytOnly_{slct}$	0.320	0.134
$Mixture_{slct}$	0.323	0.133
$Featurize_{slct}$	0.376	0.156
$LinInter_{slct}$	0.320	0.134
$FineTune_{slct}$	**0.408**	**0.173**
LEAD$_{EDU}$	0.323	0.128
Marcu	0.362	0.155
Hirao	**0.405**	**0.161**

4.4　本章のまとめ

本研究では，文抽出に基づく要約器の訓練に NYTAC を活用するため 5 つの手法を提案し，3 つのテストデータをターゲットとした実験により比較を行った．また，要約器の特性に併せた事例選択を行い，その効果を確かめた．

実験の結果，すべてのターゲットデータにおいて，NYTAC により事前に学習した要約器のパラメータを初期値として追加的に学習する手法（$FineTune$）が最も学習に寄与することが分かった．

また，事例選択を加える事で RSTDTB$_{long}$ において ROUGE 値のさらなる向上が確認できた．しかしながら，他の二つの評価セットにおいては有効性を確認することが出来なかったため，より良い事例選択の方法を開発する必要がある．

今後は，大規模な訓練事例の活用による，より柔軟で効果的な素性の設計を行う．また，標準的なナップサックモデル以外の文書要約モデルに対する NYTAC の有効性を検証する必要がある．

表 4.5　RSTDTB$_{short}$ における各手法の ROUGE 値．表は便宜上横線により 4 つの区画に分けられており，それぞれの区画で最も高い ROUGE-1 値を太字で示している．

	ROUGE-1	ROUGE-2
$TrgtOnly$	**0.258**	**0.086**
$NytOnly$	0.272	0.085
$Mixture$	0.272	0.085
$Featurize$	0.264	0.087
$LinInter$	0.272	0.085
$FineTune$	**0.283**	**0.094**
$NytOnly_{slct}$	0.265	0.087
$Mixture_{slct}$	0.264	0.088
$Featurize_{slct}$	0.246	0.075
$LinInter_{slct}$	0.265	0.087
$FineTune_{slct}$	**0.286**	**0.092**
LEAD$_{EDU}$	0.240	0.082
Marcu	0.272	0.094
Hirao	**0.321**	**0.106**

第 5 章

結論と今後の課題

5.1 まとめ

本論文では，単一文書要約の高度化のため，単一文書要約において重要な二つの問題に取り組んだ．一つ目の取り組みとして，古くから文書要約における手がかりとして有効に働くという知見のある文書の修辞構造の情報を，現在の文書要約において最も盛んに研究されている手法の一つである文抽出と文圧縮の同時最適化モデルに組み込む手法について述べた．そのため，文書の修辞構造を文間の依存関係で表現するための変換を行い，解の実行可能解に対する制約としてモデルに組み込んだ．実験の結果，提案手法は特に短い要約が要求される設定において，他の RST を要約手法に組み込む手法 [84, 45] を有意に上回ることを確認した．また，RST の情報を使わずに同時最適化モデルを単純に単一文書要約に適用したモデルでは十分な結果を得ることが出来なかった．これは，複数文書要約と比較したときに単一文書要約における重要箇所同定が難しい問題である [90] ことが理由であり，本研究における結果は，RST が単一文書要約の重要箇所同定における重要な手がかりの一つであることを示している．また，圧縮率に対するより高い柔軟性を獲得するため，文圧縮の方法についても拡張を加えた．具体的には，従来手法が文の構文木からその根付き部分木を抽出していたのに対し，提案手法では任意の動詞を根とする部分木の抽出を許容する．これにより，たとえば that 節の中身のみを要約に含める事が可能となり，より柔軟な要約生成が行えることを定性的に示した．

二つ目に，これまで要約の研究において十分に活用されてこなかった NYTAC を，要約器の訓練に効率的に活用するための取り組みについて述べた．具体的には，目標となる非常にサイズの小さい評価セット (ターゲットデータ) と，それとは独立に作成された大規模な要約資源 (NYTAC) が存在するときに，後者をどのように用いることで評価セットにおける最終的な精度を向上させられるかという問題に取り組んだ．特定の評価セットに依存した結果が出ることを避けるため，単一文書要約においてよく用いられる，原文書の長

さや要求される圧縮率が異なる 3 つの評価セットを用意し，それぞれをターゲットデータとして独立に実験を行った．単純に事例を混合する手法や訓練後にパラメータを線形補間する手法など，ドメイン適応の分野を参考に 5 つ手法を比較した．実験の結果，まず NYTAC で訓練を行い，その後得られたパラメータを初期値としてターゲットデータで再び訓練を行う手法 ($FineTune$) が全てのターゲットデータにおいて安定して最も高い ROUGE 値を獲得したことを確認した．他の 4 つの手法は $FineTune$ と比較して低い ROUGE 値であっただけでなく，ターゲットデータによってベースライン ($TrgtOnly$) よりも結果が悪化したもの ($NytOnly$,$Mixture$ および $LinInter$) や，ベースラインは常に上回るものの，ターゲットデータによっては上記の 3 手法を下回る ($Featurize$) など，評価セットの違いによってその結果が安定しないことが分かった．

加えて，要約器の特性に併せて訓練に用いる事例の選択を行うことによる結果について確認した．本研究では標準的な文抽出に基づく要約器を用いたため，参照要約の文抽出度合いが高い事例を文抽出オラクル要約によって求め，そのオラクルが獲得する ROUGE 値を閾値とする事例選択を行った．実験の結果，事例選択が有効に働いたターゲットデータは一部のみ (DUC2002) ではあるものの，その上昇は大きく適切な事例選択が非常に有効に働く場合があることが分かった．

5.2　今後の課題

本研究では，単一文書要約における二つの異なるレベルの課題に対し取り組み，それぞれの課題において有効である手法を提案した．今後の展望として，まず最初に挙げられるのは，3 章で取り組んだ新たな要約を，4 章で取り組んだ大規模な要約資源の活用を同時に行うことが挙げられる．前者に取り組んだ際は，抽出単位の重要度はヒューリスティックに基づくものであり，後者に取り組んだ際はシンプルな文抽出型の要約器のみを訓練するに留まっていた．近年，文圧縮を伴う抽出型要約に対するオラクル要約を求める手法が提案されており，4 章の枠組みを，文圧縮を含んだ複雑なモデルに適用することが可能となっている．近年の複数文書要約では，複雑な要約器の訓練に取り組む研究は増えているものの，有効な素性や最適な機械学習手法の適用方法，大規模なデータの効果的な活用方法などは明らかになっていない [28, 2, 93]．そのためそのような課題に取り組むことは今後の文書要約研究における重要な課題のひとつであるといえる．

最後に，本節では今回取り組んだ問題において残されている課題，また単一文書要約や文書要約一般において解決されるべき課題について述べる．

5.2.1　文間の依存関係を考慮した文抽出と文圧縮の同時最適化手法

まず，我々がはじめに取り組んだ，文間の依存関係を利用した文抽出と文圧縮の同時最適化手法について，今後検討すべき課題を述べる．

修辞構造の自動解析性能の向上

本研究で我々が用いた文間依存木は，文書の修辞構造に基づく句構造木を変換することで構築していた．実験においては，人手で付与された修辞構造と自動解析による修辞構造の二つを利用したが，両者の違いが最終的な ROUGE 値に与える影響は大きかった．単語間依存木の構築に用いた依存構造解析は，古くから取り組まれており現在においても最も重要で活発に研究が行われているタスクの一つであり，毎年その精度が向上している [3]．一方，修辞構造の自動解析に関してはその絶対数が少ない上，解析対象が文書全体であるため非常に難しいタスクであることが知られている．しかしながら，近年盛んに研究が行われているニューラルネットワークを用いたモデルが提案される [67, 66] など，また，従来の修辞構造解析器は句構造に基づくものが多かったが，近年は依存構造を直接予測する取り組みも提案 [131, 70, 42] されており，今後これらの精度の向上に伴い本研究の精度も向上していくことが期待できる．

修辞構造のラベルの利用

本研究では，文書の修辞関係のうち，その構造のみ入れ子依存木の構造的制約として利用した．RST には補足や例示など約 90 の修辞関係 [81] が定義されており，それらの情報も要約を作成する上で役に立つだろう．たとえば，単語や文の重要度を予測する手がかりとして利用する方法などが考えられる．もちろん，これには先に述べた修辞構造解析器の精度向上が不可欠である．

非文の生成を避ける文圧縮

文圧縮は人手で作成した文の一部を削除する操作であり，多くの要因で非文が生成されてしまう．本研究では非文を避けるために，構文解析器が出力した木構造の部分木に実行可能解を制限した上で，冠詞や固有名詞列など，最低限の文法的な制約を加えることで対処したが，非文の生成を完全に防ぐことは出来ない．そもそも構文解析器が解析を失敗した場合は，非文の生成を避けることは非常に難しい．

評価対象のデータおよびドメインの拡大

本研究は，主にエッセイで構築された評価セットで評価を行った．報道記事など他のドメインの文書においてその有効性がどのように変化するかは興味深い．

複数文書要約への適用

単一文書要約のための提案手法である入れ子依存木からの枝刈り手法を，複数文書要約のために拡張することも興味深い課題である．複数の文書集合における関係についての理論としては，Cross-document Structure Theory (CST) [98] などがあり，実際に一部の研究者が複数文書要約における手がかりとして利用している [15, 134, 126]．しかしながら，CST，すなわち複数文書における修辞構造の自動解析は単一文書以上に困難な課題である．そのため，それぞれの文書に独立に修辞関係を解析し入れ子依存木を構築した後にそれらを何らかの手段で統合する方法も考えられる．

5.2.2　要約器の効果的な訓練のための大規模要約資源の活用手法

次に，大規模要約資源の追加的な利用による単一文書要約器の訓練について，今後検討すべき課題は以下のとおりである．

より複雑な要約器における実験

本研究では，標準的な文抽出器にのみ焦点を絞って実験を行った．その結果，$FineTune$ が最も高い精度を得ることを確認したが，これが任意の要約器の訓練に有効かどうかは自明ではない．今後は，より複雑なモデルでも，その訓練に与える影響を調査する．

より良い事例選択手法

本研究では，文抽出オラクル要約が獲得する ROUGE 値を基準として事例の選択を行った．結果としては常に事例選択が有効に働くわけではないものの，有効に働いた場合にはその精度が有意に向上することが確認できた．そのため，今後より良い事例選択の手法について研究を行うことには非常に価値がある．考えられる候補として，原文書と参照要約の間で単語レベルのアライメントを取る手法 [23, 48] などが考えられる．

文圧縮を含む要約器のための事例選択　上に述べた 2 つの課題と関連するが，現在の主流な要約アプローチである文圧縮を含んだ要約手法のために有効な事例選択手法を検討することも重要である．本研究で文抽出型オラクルの有効性が一部で確認できたことから，圧縮型オラクル [142] の利用が有効に働くことが期待できる．

要約に有効な特徴量

従来の要約研究では，その訓練事例の不足の影響もあり，機械学習の特徴量として，古くから有効に働くことが分かっている手がかりが使われてきた．本研究により大量の訓練事例を効率的に活用することができるようになったことで，文書要約

に有効に働く特徴量の調査がより進むことが期待できる．

多彩なデータやドメインにおける有効性の検証

今回は三種類の評価セットによりその有効性を評価したが，医療や特許など，報道記事やエッセイに限らない他のドメインにおいても文書要約の必要性は確認されており，そのようなデータにおいても提案手法が有効に働くかを検証することは興味深い．その場合はドメインに特化した素性設計など，ドメイン毎の調整が重要になると考えられる．

5.2.3　文書要約における課題

本節では，本研究で直接取り組んだ問題にかぎらず，文書要約研究に残された課題について述べる．

要約の評価指標

現在最も用いられている評価指標は ROUGE である．これは ROUGE 以外の指標で評価する際の人手コストが大きいことに起因するが，ROUGE は表層的な再現率しか考慮出来ないという問題がある．これは以前から研究者の間で問題として挙がっていたが，有効な解決策はまだ発見されていない．特に非抽出型要約の評価では ROUGE を用いた評価が妥当であるとは考えにくいため，今後は評価の信頼性と人手コストのトレードオフを取ったより良い評価指標が必要である．ROUGE に拡張を加えることで，自動評価の質を高める取り組みも存在している．Ng らは ROUGE に単語の埋め込みベクトルを利用することで類義語も考慮した自動評価指標を提案した [92]．このように，新たな技術を取り込むことで人手評価との高い相関を持った自動評価尺度を研究する研究者も出ており，今後も重要なトピックであるといえる．

抽出型要約の精度向上

非抽出型の要約への期待が高まっている一方で，たとえ要約の問題を抽出に限ったとしてもまだ多くの課題が残されていることも定量的に示されている．具体的には，文圧縮操作を含むか否かに関わらず，抽出型要約は，その獲得できるであろう評価値の上限値（オラクル要約の評価値）と比較した時にまだ十分な精度が獲得出来ているとはいえない [142, 141] ということが近年報告されている．ここにどのような要因が絡んでいるかはこれまで不透明であったが，2015 年に行われた言語処理学会主催のエラー分析ワークショップである Project Next NLP [*1] において，日

[*1] https://sites.google.com/site/projectnextnlp/ws2015

本の要約研究者による文書要約手法の誤りの分析や整理が行われた[*2]．また，西川らはそこで得られた知見を利用して要約器に修正を加えることで実際に要約の精度が向上したことを報告した [138]．抽出型要約のさらなる向上のために，今後も抽出型要約のもつ現状の問題を体系的に整理することは非常に重要である．

文圧縮の適用方法

文抽出と文圧縮をどのように組み合わせるかという枠組みについて，同時最適化が現在の主流なアプローチの一つである一方で，文圧縮を要約へ利用する方法は他にも存在する．代表的な手法は，multi-candidate reduction (MCR)[132] というアプローチに基づくものである．MCR は，オリジナルの文から，文圧縮によるその文の亜種を複数用意しておくという，文圧縮を重要文抽出の前処理としてパイプライン的につなぐアプローチであり，近年においてもこの枠組に基づく要約手法は報告されている [121, 93, 137] ．同時最適化手法と MCR では，どちらも利点と欠点が存在する．同時最適化に基づく手法は，要約のデコード時に，選択した他の文や単語に併せて動的に圧縮の方法が変化するため柔軟な文圧縮が可能である反面，文法性や可読性が崩れやすいため，構造的な制約や重要度の計算方法によって対応する必要がある．一方，MCR に基づく手法は事前に文の亜種を用意するため，文法的な正しさを事前に判定することができる．田中らは，ルールにもとづき生成した文の亜種集合の各文に対し言語モデルの生成確率にもとづき計算される acceptability(容認度) スコアリング [61][*3]を行い重要度の手がかりにすることで最終的な ROUGE 値が向上することを報告している．しかしながら，各文を独立に文圧縮すると，文書全体でどの情報が必要であるかを判断することができないという課題も存在する．これらの手法を体系的に比較し，両者の欠点を補う方法を検討することは重要である．

抽出に基づく要約の限界

前述のとおり，抽出型要約にまだ向上の余地が多く残されていることは事実であるが，同時に抽出型要約では絶対に生成できない，抽出型よりも有用である要約が存在することは明白である．たとえば，言い換えにより簡潔な表現を生成することが考えられるが，より難しい例も存在する．Takamura らは前述のエラー分析ワークショップにおいて，原文書に含まれる数値について，参照要約ではその数値に対する解釈やそのような数値になった背景が述べられるという事例が存在することを報告した．このような現象は単純な言い換えで生成することは困難である．同様の

[*2] 本ワークショップにおいて分析に用いられた要約器はすべて抽出型要約に基づくものである．

[*3] ある文を母語話者が読んだ時に許容，容認できる度合いをスコア化したもの．Lau ら [61] によると，最も人手によるスコアとの相関が高かったものは recurrent neural network を使用したニューラル言語モデルである．

理由で，表や図など，構造化された情報が原文書に含まれおり，要約者がその情報も要約に含めようとした場合も，基本的には抽出型の要約アプローチでは再現できない．

ほかに，要約者が原文書そのものの情報やその内容を客観的に説明するような論調で, 原文書の著者の視点とは異なった第三者の視点 (しばしば要約者自身の視点) に立って書かれた要約も存在する．そのような要約は，抽出型要約ではもちろん，言い換えによって生成することも不可能であり，個別の問題として今後研究していく必要がある．このような問題は，非抽出型要約に属する問題であり，今後の発展が期待される．近年，ニューラルネットワークや深層学習と呼ばれる技術の発展に伴い，非抽出な要約生成を可能とする可能性が示されはじめている．これについて本節の最後の項目で説明する．一方で，深層学習を抽出型要約に用いる研究も近年増加しており，これについて次の項目で述べる．

ニューラルネットワークあるいは深層学習の抽出型要約への適用

近年，画像処理や音声処理など様々な分野においてニューラルネットワーク及び深層学習に基づいた手法が盛んに研究されている．自然言語処理においてもさまざまな研究が報告されており，抽出型要約に取り組んだ研究もいくつか存在する [12, 58, 57, 16, 92]．Cao ら [12] は，recursive neural network[106] を用いて文単位のスコア関数を学習し複数文書要約における文の重要度として利用した．Kobayashi ら [57] は，単語の埋め込みベクトルを用いて，要約文書に選んだ文に含まれる各単語のベクトル空間上の分布が，原文書に含まれる単語の形成する分布と類似するように要約をモデル化した．具体的には，原文書中の全ての単語に対し，その単語と要約における最近傍の単語の距離を求め，その総和を最小化するよう要約の探索を行う．レビュー文書の要約問題 [37] においてその有効性を確かめた．Cheng ら [16] は，encoder-decoder モデルによる文抽出モデルを提案した．Cheng らは，入力系列が文書に含まれる文の系列，出力系列が各文を抽出するか否かの系列を学習するネットワークを提案した．また，階層的な拡張を行い選択した文のうちさらにどの単語を抽出するかという系列を予測するモデルも提案し評価を行った．またモデルではなく評価指標に関して，Ng らは ROUGE に単語の埋め込みベクトルを利用することで類義語も考慮した自動評価指標を提案した [92]．このように，深層学習に基づく抽出型要約の研究はすでにいくつも報告されており，今後も注目されるトピックであると予想できる．

Encoder-decoder モデルによる文の要約

近年，ニューラルネットワークに基づいた encoder-decoder モデル [51, 107, 17] と呼ばれる枠組みが盛んに研究されている．Encoder-decoder は，機械翻訳 [51, 17, 107, 8, 78], に適用されて以降, 画像キャプション生成 [118, 125], 構文解析 [117], 対

話応答生成 [65, 104] などを含む多くの系列生成タスクにおいて encoder-decoder モデルを適用した研究が報告されている．2015 年に Rush ら [101] がニュース記事とタイトルから大規模な訓練データを整備して以降，文要約は encoder-decoder モデルの新たな適用タスクとして盛んに研究され始めている [99, 75, 7, 40, 41, 89, 53]．本タスクは，入出力が共に文であり，これまで本論文で扱ってきた，文書を対象とした要約とは異なるが，文書要約やヘッドライン生成[*4] と呼ばれるタスクの構成要素として活用できる可能性がある．文書要約という観点において，本タスクにおける興味深い点は，生成される要約が非抽出型要約に属するものであるという点である．文書要約で用いられるアプローチが長らく抽出型要約に限られていたことを考えると，今後どのように従来の文書要約の技術と融合していくか興味深い．今後，encoder-decoder に基づく要約生成モデルを文単位から文書単位への拡張，あるいは既存の文書要約の枠組みにひとつの構成要素として組み込むことができれば，非抽出的な文書要約研究の高度化に寄与することが期待される．

[*4] ヘッドライン生成は，報道記事（の一文目）からタイトルを生成するタスクで，2000 年台初頭に研究され始めた問題である [26, 133, 123, 36, 35, 33, 32, 55]

謝辞

本研究を遂行するにあたり，多くの方々のご支援，ご指導を頂きました．

指導教員である高村大也准教授には，修士課程から五年半に渡り丁寧かつ熱心なご指導を賜りました．研究を遂行していくにあたり必要な心構えや知識を教えていただき，時には一緒に悩みつつ，研究を進めていく楽しさを含め多くの事を学ばさせていただきました．心より感謝申し上げるとともに，これからも良き仲間の一人としてよろしくお願いします．また大也氏とともに，定期的に食事会を開いていただくなど激励してくださった高村綾氏，高村光氏にも深く感謝申し上げます．

奥村学教授には，奥村高村研究室の一員として多くのご指導，ご助言を頂いたこと心より感謝申し上げます．研究者としての大先輩として俯瞰的な視点から頂く示唆に富んだ助言の数々は，視野が狭まり方向性を見失ってしまう度に研究を適切な方向にを導いてくださいました．大学院で初めて自然言語処理の世界に入り不安もあったものの，二年間楽しく修士課程を過ごし，博士課程に進学することをそれほど躊躇せず決断できたことも，奥村高村研究室でお二人に指導を受けることができたからこそ思っています．

また，より学生に近い立場から多くのアドバイスを頂きました笹野遼平助教に感謝申し上げます．同じ居室で気軽に相談できる研究者の先輩として，時に夜遅くまで議論をして頂き多くの示唆を得ることができました．

研究室の秘書としてあらゆる事務処理をこなし，研究室運営において間違いなくかけがえのない存在であった飯山信子氏に感謝申し上げます．

学位論文審査の過程で審査をしてくださった，小野田功先生，寺野先生，新田克己先生に感謝いたします．先生方からは，異なる分野の研究者ならではの視点から多くの貴重なご助言，ご指導を頂きました．

修士課程からの同期である林正頼氏に深く感謝いたします．博士課程の在学中，年を追う毎に後輩の割合の方が多くなっていく中，常に同期として近い立場で議論や気分転換に付き合ってくださり，在学中に健全な精神を保つ上で非常に大きな存在でした．

修士課程，博士課程在学中には，いくつかのインターンシップやアルバイトなどを経験させていただきましたが，そのいずれにおいてもとても貴重な経験をすることができま

した．

NTT コミュニケーション基礎科学研究所の平尾努氏には，インターンシップのメンターとして熱い指導を賜っただけでなく，その後も国際会議や論文誌への投稿をはじめ，様々な機会に多くのご助言を頂きました．心より感謝申し上げます．平尾氏の下で行った研究の成果は本論文への影響も大きく，平尾氏のご指導なしでは本論文は存在していませんでした．けいはんなで過ごした 3 ヶ月は充実したかけがえのないものになっています．インターンシップの期間中，暖かく受け入れてくださった NTT コミュニケーション基礎科学研究所のみなさまにも心より感謝申し上げます．

また，初めてのインターンシップを経験させていただいた岩倉友哉氏をはじめとする富士通研究所の方々，金融という分野で新鮮な経験をさせていただいた平澤英司氏をはじめとする金融工学研究所の方々，学生最後のアルバイトでのメンターであり，現在の上司でもある海野裕也，福田昌昭をはじめとする株式会社 Preferred Networks および株式会社 Preferred Infrastructure の方々など，多くの貴重な経験の場を提供してくださった，全ての方々に感謝致します．

直接の指導教官やインターンシップ，アルバイト先の方々以外にも，多くの研究者としての先輩あるいは人生の先輩方に数えきれないご助言，ご支援を頂きました．自然言語処理内外で活躍されている多くの方々に貴重なご助言，ご支援を賜りました．特に，同じ文書要約の研究者として示唆に富んだ多くの議論を交わして下さった東京工業大学の西川仁氏をはじめ，甲南大学の永田亮氏特許庁の目黒光司氏，Carnegie Mellon University の Graham Neubig 氏，富士通研究所の横野光氏，茨城大学の古宮嘉那子氏，IBM 東京基礎研究所の吉川克正氏からは，公私ともに多くのご助言を頂きました．ありがとうございました．また，定期的に合同研究会という形で多くのご助言や刺激を下さった，お茶の水女子大学の小林一郎教授ならびに小林研究室の学生の方々に感謝申し上げます．

OB の松田耕史氏，森田一氏，牧野拓哉氏には普段の研究に関するご助言に加え，私が主催した SR 勉強会や牧野氏が主催したすずかけ論文読み会の場において多くの活発な議論やサポートを賜りましたこと感謝致します．この三名をはじめとする勉強会参加者の協力なければ今日の私は存在していません．また，研究室外部からの参加者としてすずかけ論文読み会で多くの議論を交わして頂いた富士通研究所の斎藤淳哉氏も感謝申し上げます．

また，折に触れて集まり食事などの気分転換に付き合ってくださった五十嵐沢馬氏，木曽鉄男氏，久保光証氏，久保田敦氏，鈴木雄登氏，富田紘平氏，中村直哉氏，に感謝を申し上げます．

研究室の皆様には，研究の議論や日常の気分転換まで，同じ研究室の仲間として多くの貴重な経験をさせていただきました．歳を重ねる毎に多くの後輩たちと過ごしましたが，どの世代も特徴が異なっており，多くの多様な思い出を残す事ができました．

木更津工業高等専門学校の米村恵一准教授は，私の研究歴における最初の指導教員であり，大学院へ進学したいというきっかけを与えて下さった人物でもあります．その意味で本論文を執筆する道のりの最初の一歩を後押しして下さった先生であり，深く感謝申し上げます．

最後に，これまでの道のりを常に温かく辛抱強く見守って下さった両親および兄弟に深い感謝の意を評して謝辞と致します．

参考文献

[1] *Deep Learning for Automatic Summary Scoring*. Canadian AI, May 2011.

[2] Miguel Almeida and Andre Martins. Fast and robust compressive summarization with dual decomposition and multi-task learning. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 196–206, August 2013.

[3] Daniel Andor, Chris Alberti, David Weiss, Aliaksei Severyn, Alessandro Presta, Kuzman Ganchev, Slav Petrov, and Michael Collins. Globally normalized transition-based neural networks. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2442–2452, Berlin, Germany, August 2016. Association for Computational Linguistics.

[4] Chinatsu Aone, Mary Ellen Okurowski, and James Gorlinsky. Trainable, scalable summarization using robust nlp and machine learning. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, ACL '98, pages 62–66, Stroudsburg, PA, USA, 1998. Association for Computational Linguistics.

[5] Chinatsu Aone, Mary Ellen Okurowski, James Gorlinsky, and Bjornar Larsen. A scalable summarization system using robust nlp. In *Intelligent Scalable Text Summarization*, pages 66–73, 1997.

[6] Amittai Axelrod, Xiaodong He, and Jianfeng Gao. Domain adaptation via pseudo in-domain data selection. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 355–362, 2011.

[7] Ayana, S. Shen, Z. Liu, and M. Sun. Neural Headline Generation with Minimum Risk Training. *CoRR*, abs/1604.01904, 2016.

[8] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *Proceedings of ICLR15*, 2015.

[9] Tal Baumel, Raphael Cohen, and Michael Elhadad. Query-chain focused summariza-

tion. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 913–922, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[10] Taylor Berg-Kirkpatrick, Dan Gillick, and Dan Klein. Jointly learning to extract and compress. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 481–490, Portland, Oregon, USA, June 2011.

[11] Ergun Biçici. Domain adaptation for machine translation with instance selection. *The Prague Bulletin of Mathematical Linguistics*, 103:5–20, 2015.

[12] Ziqiang Cao, Furu Wei, Li Dong, Sujian Li, and Ming Zhou. Ranking with recursive neural networks and its application to multi-document summarization. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, pages 2153–2159. AAAI Press, 2015.

[13] Jaime Carbonell and Jade Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '98, pages 335–336, New York, NY, USA, 1998. ACM.

[14] Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of the 2nd Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL)*, pages 1–10, 2001.

[15] Maria Lucia Castro Jorge and Thiago Pardo. Experiments with cst-based multidocument summarization. In *Proceedings of TextGraphs-5 - 2010 Workshop on Graph-based Methods for Natural Language Processing*, pages 74–82, Uppsala, Sweden, July 2010. Association for Computational Linguistics.

[16] Jianpeng Cheng and Mirella Lapata. Neural summarization by extracting sentences and words. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 484–494, Berlin, Germany, August 2016. Association for Computational Linguistics.

[17] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1724–1734, 2014.

[18] Janara Christensen, Mausam, Stephen Soderland, and Oren Etzioni. Towards coherent multi-document summarization. In *NAACL:HLT*, pages 1163–1173, 2013.

[19] Janara Christensen, Stephen Soderland, Gagan Bansal, and Mausam. Hierarchical

summarization: Scaling up multi-document summarization. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 902–912, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[20] James Clarke and Mirella Lapata. Global inference for sentence compression an integer linear programming approach. *Journal of Artificial Intelligence Research*, 31:399–429, 2008.

[21] Linguistic Data Consortium. Hansard corpus of parallel english and french. In *Linguistic Data Consortium*, http://www.ldc.upenn.edu/, 1997.

[22] Koby Crammer, Ofer Dekel, Joseph Keshet, Shai Shalev-Shwartz, and Yoram Singer. Online passive-aggressive algorithms. *The Journal of Machine Learning Research*, 7:551–585, 2006.

[23] Hal Daumé, III and Daniel Marcu. Induction of word and phrase alignments for automatic document summarization. *Computational Linguistics*, 31(4):505–530, 2005.

[24] Hal Daumé III. Frustratingly easy domain adaptation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL)*, pages 256–263, 2007.

[25] Hal Daumé III and Daniel Marcu. A noisy-channel model for document compression. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 449–456, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.

[26] Bonnie Dorr, David Zajic, and Richard Schwartz. Hedge trimmer: A parse-and-trim approach to headline generation. In *Proceedings of the HLT-NAACL 03 Text Summarization Workshop*, pages 1–8, 2003.

[27] DUC. Document understanding conference. In *ACL Workshop on Automatic Summarization*, 2002.

[28] Greg Durrett, Taylor Berg-Kirkpatrick, and Dan Klein. Learning-based single-document summarization with compression and anaphoricity constraints. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1998–2008, Berlin, Germany, August 2016. Association for Computational Linguistics.

[29] David duVerle and Helmut Prendinger. A novel discourse parser based on support vector machine classification. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 665–673, Suntec, Singapore, August 2009. Associa-

tion for Computational Linguistics.

[30] H. P. Edmundson. New methods in automatic extracting. *J. ACM*, 16(2):264–285, April 1969.

[31] Elena Filatova and Vasileios Hatzivassiloglou. A formal model for information selection in multi-sentence text extraction. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING)*, pages 397–403, 2004.

[32] Katja Filippova, Enrique Alfonseca, Carlos A. Colmenares, Lukasz Kaiser, and Oriol Vinyals. Sentence compression by deletion with lstms. In *Proceedings of EMNLP15*, pages 360–368, 2015.

[33] Katja Filippova and Yasemin Altun. Overcoming the lack of parallel data in sentence compression. In *Proceedings of EMNLP13*, pages 1481–1491, 2013.

[34] Katja Filippova and Michael Strube. Dependency tree based sentence compression. In *Proceedings of the 2nd International Natural Language Generation Conference (INLG)*, pages 25–32, 2008.

[35] Katja Filippova and Michael Strube. Dependency tree based sentence compression. In *Proceedings of INLG08*, pages 25–32, 2008.

[36] Dimitrios Galanis and Ion Androutsopoulos. An extractive supervised two-stage method for sentence compression. In *Proceedings of NAACL-HLT10*, pages 885–893, 2010.

[37] Kavita Ganesan, ChengXiang Zhai, and Jiawei Han. Opinosis: A graph based approach to abstractive summarization of highly redundant opinions. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 340–348, Beijing, China, August 2010. Coling 2010 Organizing Committee.

[38] Dan Gillick and Benoit Favre. A scalable global model for summarization. In *Proceedings of the NAACL HLT Workshop on Integer Linear Programming for Natural Language Processing (ILP)*, pages 10–18, 2009.

[39] Jade Goldstein, Vibhu Mittal, Jaime Carbonell, and Mark Kantrowitz. Multi-document summarization by sentence extraction. In *Proceedings of the 2000 NAACL-ANLPWorkshop on Automatic Summarization - Volume 4*, NAACL-ANLP-AutoSum '00, pages 40–48, Stroudsburg, PA, USA, 2000. Association for Computational Linguistics.

[40] Jiatao Gu, Zhengdong Lu, Hang Li, and Victor O.K. Li. Incorporating copying mechanism in sequence-to-sequence learning. In *Proceedings of ACL16*, pages 1631–1640, 2016.

[41] Caglar Gulcehre, Sungjin Ahn, Ramesh Nallapati, Bowen Zhou, and Yoshua Bengio.

Pointing the unknown words. In *Proceedings of ACL16*, pages 140–149, 2016.

[42] Katsuhiko Hayashi, Tsutomu Hirao, and Masaaki Nagata. Empirical comparison of dependency conversions for rst discourse trees. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 128–136, Los Angeles, September 2016. Association for Computational Linguistics.

[43] Hugo Hernault, Helmut Prendinger, David A. duVerle, and Mitsuru Ishizuka. HILDA: A Discourse Parser Using Support Vector Machine Classification. *Dialogue and Discourse*, 1(3):1–33, 2010.

[44] Tsutomu Hirao, Hideki Isozaki, Eisaku Maeda, and Yuji Matsumoto. Extracting important sentences with support vector machines. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, volume 1, pages 1–7, 2002.

[45] Tsutomu Hirao, Yasuhisa Yoshida, Masaaki Nishino, Norihito Yasuda, and Masaaki Nagata. Single-document summarization as a tree knapsack problem. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1515–1520, 2013.

[46] Jerry R Hobbs. *On the coherence and structure of discourse*. CSLI, 1985.

[47] Kai Hong and Ani Nenkova. Improving the estimation of word importance for news multi-document summarization. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 712–721, 2014.

[48] Hongyan Jing. Using hidden markov modeling to decompose human-written summaries. *Computational Linguistics*, 28:527–543, 2002.

[49] Hongyan Jing and Kathleen R. McKeown. Cut and paste based text summarization. In *Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference (NAACL)*, pages 178–185, 2000.

[50] D. S. Johnson and K. A. Niemi. On knapsacks, partitions, and a new dynamic programming technique for trees. *Mathmatics of Operations Research*, 8(1):1–14, 1983.

[51] Nal Kalchbrenner and Phil Blunsom. Recurrent continuous translation models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1700–1709, Seattle, Washington, USA, October 2013. Association for Computational Linguistics.

[52] Yuta Kikuchi, Tsutomu Hirao, Hiroya Takamura, Manabu Okumura, and Masaaki Nagata. Single document summarization based on nested tree structure. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume*

2: Short Papers), pages 315–320, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[53] Yuta Kikuchi, Graham Neubig, Ryohei Sasano, Hiroya Takamura, and Manabu Okumura. Controlling output length in neural encoder-decoders. In *Proceedings of EMNLP16*, 2016.

[54] Yuta Kikuchi, Akihiko Watanabe, Sasano Ryohei, Hiroya Takamura, and Manabu Okumura. *Learning from Numerous Untailored Summaries*, pages 206–219. Springer International Publishing, 2016.

[55] Sigrid Klerke, Yoav Goldberg, and Anders Søgaard. Improving sentence compression by learning to predict gaze. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1528–1533, San Diego, California, June 2016. Association for Computational Linguistics.

[56] Kevin Knight and Daniel Marcu. Statistics-based summarization - step one: Sentence compression. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 703–710, 2000.

[57] Hayato Kobayashi, Masaki Noguchi, and Taichi Yatsuka. Summarization based on embedding distributions. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1984–1989, Lisbon, Portugal, September 2015. Association for Computational Linguistics.

[58] Mikael Kågebäck, Olof Mogren, Nina Tahmasebi, and Devdatt Dubhashi. Extractive summarization using continuous vector space models. In *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality (CVSC)*, pages 31–39, Gothenburg, Sweden, April 2014. Association for Computational Linguistics.

[59] Mitsumasa Kubo, Ryohei Sasano, Hiroya Takamura, and Manabu Okumura. Generating live sports updates from twitter by finding good reporters. In *Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) - Volume 01*, WI-IAT '13, pages 527–534, Washington, DC, USA, 2013. IEEE Computer Society.

[60] Julian Kupiec, Jan Pedersen, and Francine Chen. A trainable document summarizer. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '95, pages 68–73, New York, NY, USA, 1995. ACM.

[61] Jey Han Lau, Alexander Clark, and Shalom Lappin. Unsupervised prediction of ac-

ceptability judgements. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1618–1628, Beijing, China, July 2015. Association for Computational Linguistics.

[62] Chen Li, Yang Liu, Fei Liu, Lin Zhao, and Fuliang Weng. Improving multi-documents summarization by sentence compression based on expanded constituent parse trees. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 691–701, Doha, Qatar, October 2014. Association for Computational Linguistics.

[63] Chen Li, Yang Liu, and Lin Zhao. Using external resources and joint learning for bigram weighting in ILP-based multi-document summarization. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, pages 778–787, 2015.

[64] Chen Li, Xian Qian, and Yang Liu. Using supervised bigram-based ILP for extractive summarization. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1004–1013, 2013.

[65] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. In *Proceedings of NAACL-HLT16*, pages 110–119, 2016.

[66] Jiwei Li, Dan Jurafsky, and Eudard Hovy. When are tree structures necessary for deep learning of representations? *arXiv preprint arXiv:1503.00185*, 2015.

[67] Jiwei Li, Rumeng Li, and Eduard H Hovy. Recursive deep models for discourse parsing. In *EMNLP*, pages 2061–2069, 2014.

[68] Junyi Jessy Li and Ani Nenkova. Fast and accurate prediction of sentence specificity. In *Proceedings of the Twenty-Ninth Conference on Artificial Intelligence (AAAI)*, pages 2281–2287, 2015.

[69] Qi Li. Literature survey: Domain adaptation algorithms for natural language processing. Technical report, Department of Computer Science. The Graduate Center, The City University of New York, 2012.

[70] Sujian Li, Liang Wang, Ziqiang Cao, and Wenjie Li. Text-level discourse dependency parsing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 25–35, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[71] Chin-Yew Lin. Training a selection function for extraction. In *Proceedings of the Eighth International Conference on Information and Knowledge Management*, CIKM

'99, pages 55–62, New York, NY, USA, 1999. ACM.

[72] Chin-Yew Lin. Improving summarization performance by sentence compression: A pilot study. In *Proceedings of the Sixth International Workshop on Information Retrieval with Asian Languages - Volume 11*, pages 1–8, 2003.

[73] Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. In *Proc. ACL workshop on Text Summarization Branches Out*, pages 74–81, 2004.

[74] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Proceedings of the Workshop on Text Summarization Branches Out*, pages 74–81, 2004.

[75] Konstantin Lopyrev. Generating news headlines with recurrent neural networks. *CoRR*, abs/1512.01712, 2015.

[76] Annie Louis, Aravind Joshi, and Ani Nenkova. Discourse indicators for content selection in summarization. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 147–156, Tokyo, Japan, September 2010. Association for Computational Linguistics.

[77] H. P. Luhn. The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2):159–165, April 1958.

[78] Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation. In *Proceedings of EMNLP15*, pages 1412–1421, 2015.

[79] Inderjeet Mani. *Automatic Summarization*. Natural language processing. John Benjamins Publishing Company, 2001.

[80] Inderjeet Mani and Eric Bloedorn. Machine learning of generic and user-focused summarization. In *Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence*, AAAI '98/IAAI '98, pages 820–826, Menlo Park, CA, USA, 1998. American Association for Artificial Intelligence.

[81] William C. Mann and Sandra A. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281, 1988.

[82] Daniel Marcu. From discourse structures to text summaries. In *Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization*, pages 82–88, 1997.

[83] Daniel Marcu. The rhetorical parsing of unrestricted natural language texts. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pages 96–103, Madrid, Spain, July 1997. Association for Computational Linguistics.

[84] Daniel Marcu. Improving summarization through rhetorical parsing tuning. In *Proceedings of the 6th Workshop on Very Large Corpora*, pages 206–215, 1998.

[85] Daniel Marcu. The automatic construction of large-scale corpora for summarization research. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 137–144, 1999.

[86] Ryan McDonald. A study of global inference algorithms in multi-document summarization. In *Proceedings of the 29th European Conference on Information Retrieval (ECIR)*, pages 557–564, 2007.

[87] Rada Mihalcea and Paul Tarau. Textrank: Bringing order into texts. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, pages 404–411, 2004.

[88] Hajime Morita, Ryohei Sasano, Hiroya Takamura, and Manabu Okumura. Subtree extractive summarization via submodular maximization. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1023–1032, 2013.

[89] Ramesh Nallapati, Bing Xiang, and Bowen Zhou. Sequence-to-sequence rnns for text summarization. *CoRR*, abs/1602.06023, 2016.

[90] Ani Nenkova. Automatic text summarization of newswire: Lessons learned from the document understanding conference. In *Proceedings of the 20th National Conference on Artificial Intelligence - Volume 3*, AAAI'05, pages 1436–1441. AAAI Press, 2005.

[91] Ani Nenkova and Kathleen McKeown. Automatic summarization. In *Foundations and Trends® in Information Retrieval*, volume 2-3, pages 103–233, 2011.

[92] Jun-Ping Ng and Viktoria Abrecht. Better summarization evaluation with word embeddings for rouge. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1925–1930, Lisbon, Portugal, September 2015. Association for Computational Linguistics.

[93] Hitoshi Nishikawa, Kazuho Arita, Katsumi Tanaka, Tsutomu Hirao, Toshiro Makino, and Yoshihiro Matsuo. Learning to generate coherent summary with discriminative hidden semi-markov model. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, pages 1648–1659, Dublin, Ireland, August 2014. Dublin City University and Association for Computational Linguistics.

[94] Hitoshi Nishikawa, Takaaki Hasegawa, Yoshihiro Matsuo, and Genichiro Kikui. Opinion summarization with integer linear programming formulation for sentence extraction and ordering. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*, pages 910–918, 2010.

[95] Chikashi Nobata, Satoshi Sekine, Masaki Murata, Kiyotaka Uchimoto, Masao Utiyama, and Hitoshi Isahara. Sentence extraction system assembling multiple evidence. In *Proceedings of the Second NTCIR Workshop Meeting*, pages 5–213, 2001.

[96] Miles Osborne. Using maximum entropy for sentence extraction. In *Proceedings of the ACL-02 Workshop on Automatic Summarization*, pages 1–8, Phildadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.

[97] Xian Qian and Yang Liu. Fast joint compression and summarization via graph cuts. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1492–1502, 2013.

[98] Dragomir Radev. A common theory of information fusion from multiple text sources step one: Cross-document structure. In *1st SIGdial Workshop on Discourse and Dialogue*, pages 74–83, Hong Kong, China, October 2000. Association for Computational Linguistics.

[99] Marc'Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. Sequence level training with recurrent neural networks. *CoRR*, abs/1511.06732, 2015.

[100] Robert Remus. Domain adaptation using domain similarity- and domain complexity-based instance selection for cross-domain sentiment analysis. In *Proceedings of the 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW 2012) Workshop on Sentiment Elicitation from Natural Text for Information Retrieval and Extraction (SENTIRE)*, pages 717–723, 2012.

[101] Alexander M. Rush, Sumit Chopra, and Jason Weston. A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 379–389, 2015.

[102] Evan Sandhaus. The New York Times annotated corpus. In *Linguistic Data Consortium*, https://catalog.ldc.upenn.edu/LDC2008T19, 2008.

[103] Natalie Schluter and Anders Søgaard. Unsupervised extractive summarization via coverage maximization with syntactic and semantic concepts. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 840–844, Beijing, China, July 2015. Association for Computational Linguistics.

[104] Iulian Vlad Serban, Alessandro Sordoni, Yoshua Bengio, Aaron C. Courville, and Joelle Pineau. Building end-to-end dialogue systems using generative hierarchical neural network models. In *Proceedings of AAAI16*, pages 3776–3784, 2016.

[105] Ruben Sipos, Pannaga Shivaswamy, and Thorsten Joachims. Large-margin learning of submodular summarization models. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 224–233, 2012.

[106] Richard Socher, Cliff C. Lin, Andrew Y. Ng, and Christopher D. Manning. Parsing

Natural Scenes and Natural Language with Recursive Neural Networks. In *Proceedings of the 26th International Conference on Machine Learning (ICML)*, 2011.

[107] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Proceedings of NIPS14*, pages 3104–3112, 2014.

[108] Jun Suzuki, Hideki Isozaki, Xavier Carreras, and Michael Collins. An empirical study of semi-supervised structured conditional models for dependency parsing. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 551–560, Singapore, August 2009. Association for Computational Linguistics.

[109] Krysta Svore, Lucy Vanderwende, and Christopher Burges. Enhancing single-document summarization by combining RankNet and third-party sources. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 448–457, Prague, Czech Republic, June 2007. Association for Computational Linguistics.

[110] Hiroya Takamura and Manabu Okumura. Text summarization model based on maximum coverage problem and its variant. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 781–789, Athens, Greece, March 2009. Association for Computational Linguistics.

[111] Hiroya Takamura and Manabu Okumura. Text summarization model based on the budgeted median problem. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM)*, pages 1589–1592, November 2009.

[112] Hiroya Takamura and Manabu Okumura. Learning to generate summary as structured output. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1437–1440, 2010.

[113] Hiroya Takamura, Hikaru Yokono, and Manabu Okumura. *Summarizing a Document Stream*, pages 177–188. Springer Berlin Heidelberg, 2011.

[114] Simone Teufel. Sentence extraction as a classification task. In *Intelligent Scalable Text Summarization*, pages 58–65, 1997.

[115] Vinícius Rodrigues Uzêda, Thiago Alexandre Salgueiro Pardo, and Maria Das Graças Volpe Nunes. A comprehensive comparative evaluation of rst-based summarization methods. *ACM Trans. Speech Lang. Process.*, 6(4):4:1–4:20, May 2010.

[116] Lucy Vanderwende, Hisami Suzuki, Chris Brockett, and Ani Nenkova. Beyond sumbasic: Task-focused summarization with sentence simplification and lexical expansion. *Information Processing and Management: an International Journal archive*, 43(6):1606–1618, November 2007.

[117] Oriol Vinyals, Lukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey E. Hinton. Grammar as a foreign language. In *Proceedings of NIPS15*, pages 2773–2781, 2015.

[118] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164, 2015.

[119] Xiaojun Wan, Jianwu Yang, and Jianguo Xiao. Towards an iterative reinforcement approach for simultaneous document summarization and keyword extraction. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 552–559, Prague, Czech Republic, June 2007. Association for Computational Linguistics.

[120] Lu Wang, Hema Raghavan, Vittorio Castelli, Radu Florian, and Claire Cardie. A sentence compression based framework to query-focused multi-document summarization. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1384–1394. Association for Computational Linguistics, 2013.

[121] Lu Wang, Hema Raghavan, Vittorio Castelli, Radu Florian, and Claire Cardie. A sentence compression based framework to query-focused multi-document summarization. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1384–1394, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.

[122] Xun Wang, Yasuhisa Yoshida, Tsutomu Hirao, Katsuhito Sudoh, and Masaaki Nagata. Summarization based on task-oriented discourse parsing. *IEEE Transactions on Audio, Speech, and Language Processing*, 23(8):1358–1367, August 2015.

[123] Kristian Woodsend, Yansong Feng, and Mirella Lapata. Title generation with quasi-synchronous grammar. In *Proceedings of the EMNLP10*, pages 513–523, 2010.

[124] Rui Xia, Chengqing Zong, Xuelei Hu, and Erik Cambria. Feature ensemble plus sample selection: Domain adaptation for sentiment classification. *IEEE Intelligent Systems*, 28(3):10–18, 2013.

[125] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In David Blei and Francis Bach, editors, *Proceedings of ICML15*, pages 2048–2057. JMLR Workshop and Conference Proceedings, 2015.

[126] Yong-Dong Xu, Xiao-Dong Zhang, Guang-Ri Quan, and Ya-Dong Wang. Mrs for

multi-document summarization by sentence extraction. *Telecommunication Systems*, 53(1):91–98, 2013.

[127] Yinfei Yang and Ani Nenkova. Detecting information-dense texts in multiple news domains. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI)*, pages 1650–1656, 2014.

[128] Yinfei Yang and Ani Nenkova. Detecting information-dense texts in multiple news domains. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*, pages 1650–1656, 2014.

[129] Norihito Yasuda, Masaaki Nishino, Tsutomu Hirao, and Masaaki Nagata. Sub-sentence extraction based on combinatorial optimization. In *Proceedings of the 35th European Conference on Advances in Information Retrieval*, ECIR'13, pages 812–815, 2013.

[130] Wentau Yih, Joshua Goodman, Lucy Vanderwende, and Hisami Suzuki. Multi-document summarization by maximizing informative content-words. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence (IJCAI)*, pages 1776–1782, 2007.

[131] Yasuhisa Yoshida, Jun Suzuki, Tsutomu Hirao, and Masaaki Nagata. Dependency-based discourse parser for single-document summarization. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1834–1839, Doha, Qatar, October 2014. Association for Computational Linguistics.

[132] David Zajic, Bonnie J. Dorr, Jimmy Lin, and Richard Schwartz. Multi-candidate reduction: Sentence compression as a tool for document summarization tasks. *Inf. Process. Manage.*, 43(6):1549–1570, November 2007.

[133] David Zajic, Bonnie J Dorr, and R. Schwartz. Bbn/umd at duc-2004: Topiary. In *Proceedings of NAACL-HLT04 Document Understanding Workshop*, pages 112 – 119, 2004.

[134] Zhu Zhang, Sasha Blair-Goldensohn, and Dragomir R. Radev. Towards cst-enhanced summarization. In *Eighteenth National Conference on Artificial Intelligence*, pages 439–445, Menlo Park, CA, USA, 2002. American Association for Artificial Intelligence.

[135] Jiayi Zhao, Xipeng Qiu, Zhao Liu, and Xuanjing Huang. Online distributed passive-aggressive algorithm for structured learning. In *Proceedings of Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data - 12th China National Conference, CCL*, pages 120–130, 2013.

[136] 奥村学, 原口良胤, 望月源. 決定木学習を用いたテキスト自動要約手法に関するいく

つかの考察. 情報処理学会第 *59* 回全国大会講演論文集, 1999.

[137] 田中 駿, 笹野 遼平, 高村 大也, 奥村 学. 要約長，文長，文数制約付きニュース記事要約. 言語処理学会第 *22* 回年次大会 *(NLP2016)*, pages 342 – 345, 2016.

[138] 西川 仁. 自動要約における誤り分析の枠組み. 自然言語処理, 23(1):3–36, 2016.

[139] 平尾 努, 磯崎 秀樹, 前田 英作, 松本 裕治. Support vector machine を用いた重要文抽出法. 情報処理学会論文誌, 44(8):2230–2243, 2003.

[140] 平尾 努, 鈴木 潤, 磯崎 秀樹. 最適化問題としての文書要約. 人工知能学会論文誌, 24(2):223–231, 2009.

[141] 平尾 努, 西野 正彬, 鈴木 潤, 永田 昌明. オラクル要約の列挙. 言語処理学会第 *20* 回年次大会 *(NLP2014)*, pages 650 – 653, 2014.

[142] 平尾 努, 西野 正彬, 永田 昌明. 圧縮型要約のオラクルに関する考察. 言語処理学会第 *22* 回年次大会 *(NLP2016)*, pages 350 – 353, 2016.

[143] 富田紘平, 高村大也, 奥村学. 重要文抽出と文圧縮を組み合わせた新たな抽出的要約手法. *IPSJ SIG Technical Report 2009-NL-189*, pages 13–20, 2009.

[144] 野本忠司，松本裕治. 人間の重要文判定に基づいた自動要約の試み. 電子情報通信学会技術研究報告. *NLC,* 言語理解とコミュニケーション, 97(200):1–6, 1997.

研究業績

論文誌

1. 菊池悠太, 平尾努, 高村大也, 奥村学, 永田昌明. ”入れ子依存木の刈り込みによる単一文書要約手法”, 自然言語処理, Vol.22, No.3, pp.197-217, 2015.
2. 菊池悠太, 米村恵一, 北崎充晃. ”視野闘争時の知覚交替に伴う瞳孔反応の検討：注意指標の抽出を目指して”, 電子情報通信学会論文誌 A（レター）, Vol.J95-A, No.6, pp.535-538, 2012.

査読付き国際会議

1. Yuta Kikuchi, Graham Neubig, Ryohei Sasano, Hiroya Takamura, Manabu Okumura. “Controlling Output Length in Neural Encoder-decoder”, In proceedings of EMNLP16, 2016.
2. Yuta Kikuchi, Akihiko Watanabe, Ryohei Sasano, Hiroya Takamura, Manabu Okumura. “Learning from Numerious Untailored Summaries”, In proceedings of PRICAI16, 2016
3. Yuta Kikuchi, Tsutomu Hirao, Hiroya Takamura, Manabu Okumura, Masaaki Nagata. ”Single Document Summarization based on Nested Tree Structure”, In Proceedings of ACL14, 2014.
4. Yuta Kikuchi, Hiroya Takamura, Manabu Okumura and Satoshi Nakazawa. “Identifying a Demand towards a Company in CGM”, In proceedings of CICLing14, 2014.

国内学会発表

1. 菊池悠太, Graham Neubig, 笹野遼平, 高村大也, 奥村学, “Encoder-Decoder モデルにおける出力長制御”, 第 227 回自然言語処理研究会, 2016.

2. 菊池悠太, 渡邉亮彦, 高村大也, 奥村学, “大規模要約資源としての New York Times Annotated Corpus”, 第 224 回自然言語処理研究会, 2015.
3. 菊池悠太, 渡邉亮彦, 高村大也, 奥村学. “重要箇所同定用コーパスの構築 New York Times Annotated Corpus の文書要約資源化に向けて“, 言語処理学会第 21 回年次大会, 2015.
4. 目黒光司, 笹野遼平, 榊原隆文, 菊池悠太, 高村大也, 奥村学. “F ターム概念ベクトルを用いた特許検索システムの改良”, 言語処理学会第 21 回年次大会, 2015.
5. 菊池悠太, 平尾努, 高村大也, 奥村学, 永田昌明. ”修辞構造と係り受け構造を制約とした単一文書要約手法”, 言語処理学会第 20 回年次大会, 2014.
6. 菊池悠太, 高村大也, 奥村学. ”属性-評価ペアを単位とした評判情報の要約”, 人工知能学会第 27 回年次大会, 2013.
7. 菊池悠太, 高村大也, 奥村学, 中澤聡. “CGM 中からの企業への要望文の同定”, 言語処理学会第 19 回年次大会, 2013.
8. 菊池悠太, 高村大也, 奥村学. “属性-評価ペアを単位とした評判情報の要約”, 情報処理学会技術報告, 2012.
9. 菊池悠太, 米村恵一, 北崎充晃. “安全運転支援システムへの応用を目的とした瞳孔反応の基礎検討”, 電子情報通信学会技術報告, 2010.
10. 菊池悠太. “非注意性盲目と輻輳を考慮した瞳孔径変化の検討”, 豊橋技術科学大学 2008 年度高専連携教育研究プロジェクト学生成果報告会, 2009.

特許

1. 特開 2015-170224, 文書要約装置、方法、及びプログラム, 平尾努, 菊池悠太, 高村大也, 奥村学, 2015 年 9 月 28 日.