1A2-04

文字列カーネルと動的計画法を用いたテキスト・音声のトピック分割アルゴリズム

A Topic Segmentation Algorithm for Spoken Documents Using String Kernels and DP

佐土原 健*1
Ken Sadohara

*1(独) 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

The goal of the present paper is to explore the feasibility of a topic segmentation algorithm for spoken documents without assuming topical and linguistic knowledge about the documents, such as vocabulary, the language model and the topic model. For erroneous sequences of phonemes obtained by using a continuous phoneme recognizer, the proposed method analyzes lexical cohesion of any sub-sequences by using string kernels employing soft-matching, and partitions the sequence into topically homogeneous blocks by using a dynamic programming. As an empirical study on the ICSI Meeting Corpus shows, by virtue of the error robust and exhaustive but efficient analysis of topically salient sub-sequences of phonemes, it performs on erroneous recognized sequences comparably with the algorithms on error free transcriptions utilizing the linguistic knowledge.

1. はじめに

トピック分割は，テキストや音声言語を意味的に等質な部分に分割する技術であり，トピック同定・分類・追跡，情報検索，要約，ブラウジングなどのより高次の情報処理のための基礎技術としてこれまでに多くの研究がなされてきた．

音声言語に対するトピック分割としては，ニュース音声や講義音声が分析の対象とされてきた．近年では，ビジネスの現場で多くの時間を占める会議の生産性向上を目的とした知的支援の要素技術として，会議音声の意味的構造化を行うトピック分割が研究されており，RWCP 会議音声コーパス,ICSI 会議コーパス,AMI コーパス等のコーパスの整備も進んでいる．

これら会議音声言語は，4〜7 人程度の参加者によるカジュアルな会議の音声であり，複数話者がいる，自由発話である，ノイズや発話の重なりが多い，文体や語彙がくだけている，トピックが多岐にわたる上に予測が困難である等，音声言語の処理・分析にとって様々な技術的課題を提示しているが，トピック分割の文脈では，トピックのモデルを事前に用意できないこと，大語彙連続音声認識 (LVCSR) の認識精度が低いことが特に大きな問題となる．音声言語に対するトピック分割手法のほとんどは，音声認識により得られたテキストに含まれる単語の出現頻度などの語彙的素性を利用しているので，認識精度の低下はトピック分割性能に大きな影響を与える．特にトピックを代表するような重要なキーワードが正しく認識できるかどうかが重要な意味を持つが，そのために必要な言語的知識を事前に用意することは一般に困難である．

本研究は，このような問題に対処するために，トピックモデルや，トピックに適応した語彙や言語モデルといった，分析する会議音声言語の内容に依存する知識を用いないトピック分割法について検討する．具体的には，会議音声を連続音素認識により音素の系列に変換し，単語の出現頻度を分析する代わりに，単語を構成する部分音素列の出現頻度を分析する．その際，認識音素列は，脱落・挿入・置換誤りを高頻度に含んでいるので，任意の不連続な部分列を網羅的に分析する必要がある．ただひ，分析音素列の長さを p, 音素の数を N とすれば，部分列の種類は N^p 個，長さ ℓ の音素列の中に ${}_{\ell}C_p$ 個の部分列が出現するので，分析を実行可能にするために，文字列カーネル [Lodhi 02] を用いたカーネル法さらに，このようなアルゴリズムを ICSI 会議コーパスに適応し，言語知識の欠落が与える影響，音素の認識誤りが与える影響についても考察する．

連絡先: 佐土原健，ken.sadohara [at] aist.go.jp

2. 関連研究

音声言語に対するトピック分割においては，イントネーション, 無音区間の長さや話者の交代など音声言語特有の素性を用いた研究がある一方，ほとんどの研究において，人間あるいは音声認識技術によって書き起こされたテキストに含まれる語彙的素性を利用している．

トピック分割で用いられている語彙素性としては，トピックが遷移する場所に現れる “Cue word”[Beeferman 99] なども用いられるが，トピックの分割に最も寄与する素性として，単語の出現の偏りに関連する “lexical cohesion” が多くの研究で用いられている [Hearst 94, Choi 00, Utiyama 01, Galley 03, Hsueh 06]. 本研究でも lexical cohesion を用いるが，単語の代わりに，音素の部分列の出現頻度を素性とする．同様の先行研究として [Sadohara 06] があるが，本研究ではアルゴリズムの本質的改善を行っている．

トピック分割アルゴリズムが用いる事前知識という観点では，分析対象の音声言語のみから分割を行うアルゴリズム [Galley 03, Utiyama 01] と，事前に訓練データからトピックやトピック境界のモデルを構築するアルゴリズムがあり，後者においては，決定木, 指数モデル, HMM, PLSA などが用いられる．本研究は，カジュアルな会議においては，話題の予測が難しいことを考慮し，分析対象の音声言語のみからトピック分割を行うアルゴリズムについて考察する．

3. トピック分割アルゴリズム

単語の出現頻度の代わりに音素部分列の出現頻度を分析するためには，以下のような問題に新たに対処せねばならない．

- 辞書を利用しないので，音素列の任意の部分列を分析の対象とせねばならない．単語の切れ目に関する情報が音素

列上では欠落しているため，音素列の任意の部分列を，謂わば一つの単語と考えて，頻度を数え上げる必要がある．

- 音素の脱落・挿入・置換誤りが高頻度に発生する場合においても，単語の頻度情報をある程度代替しうるような音素部分列の分析が必要になる．単語の正しい音素列がそのまま入力されることは期待できないので，脱落や挿入を考慮してギャップを許した任意の部分列の頻度を数える必要がある．さらに，置換誤りを考慮して，字面上異なる部分列である場合でも，音素の類似性を反映した連続量で部分列の頻度を評価する必要がある．

- ステミングやストップワード除去に相当する前処理を，語彙や品詞などの言語知識を用いずに行う必要がある．任意の音素部分列を対象にするので，除去すべき部分列のバリエーションは非常に大きく，静的リストを用いたストップワード除去を行うことは現実的ではない．したがって，音素部分列の分析過程の中で，ある部分列がトピック分割において有用であるか否かという観点で動的に分析・除去されねばならない．

3.1 文字列カーネル

連続音素認識の結果得られる音素列は，発話毎に分割されおり，時間的に重なりのある発話をひとまとめにし，これを原子セグメントと呼び分析の最小単位とする．すなわち，音声言語は，n 個の原子セグメント (音素列) の系列として表現される．そして，原子セグメントどうしの類似性を，ギャップを含む任意の部分文字列の頻度に基づいて以下のように文字列カーネル [Lodhi 02] を用いて計算する．

任意の文字列 $s = s_1 \cdots s_m$ に対して, $|s| = m$, $1 \le i \le |s|$ に対して $s[i] = s_1 \cdots s_i$ とする．さらに, $\boldsymbol{i} = i_1 \cdots i_\ell$ $(1 \le i_1 < \cdots < i_\ell \le |s|)$ に対して, $s(\boldsymbol{i})$ は不連続な部分文字列 $s_{i_1} \cdots s_{i_\ell}$ を表わし, $\ell(\boldsymbol{i}) = i_\ell - i_1 + 1$ とする．今, 文字列 s を長さ p の任意の文字列が張る空間に写像する次のような写像を考える．

$$\phi^p : s \mapsto (\phi^p_u(s))_{u \in \Theta^p},$$

ここで, Θ はアルファベット (e.g. 音素) の集合で,

$$\phi^p_u(s) = \sum_{\boldsymbol{i}: u = s(\boldsymbol{i})} \lambda^{\ell(\boldsymbol{i}) - p} \quad (0 \le \lambda \le 1).$$

このような写像により, 文字列 s は, 長さが p の部分文字列が張る空間のベクトルとして表現され, 各成分は, 部分文字列 u の s における全ての出現数になっている．ただし，各出現は, 含まれるギャップの数 g に応じた重み λ^g だけ減衰されており，通常の文字列カーネルのバリエーションの一つとなっている．本稿では，このような写像 ϕ^k $(1 \le k \le p)$ を用いて，各原子セグメント s を, 長さ p 以下の部分文字列が張る空間上のベクトル $\phi(s)$ として表現する．

原子セグメントの類似性は，表現ベクトルの内積として定義する．$k = p$ の場合だけを考えれば,

$$\kappa_p(s,t) = \langle \phi^p(s) \cdot \phi^p(t) \rangle$$

と定義され，以下のような動的計画法を用いて $O(p|s||t|)$ で計算可能であることが知られている [Lodhi 02].

$$\begin{aligned}
\kappa_p(s,t) &= \lambda^{-2p} \sum_{i=1}^{|s|} \sum_{j=1}^{|t|} \kappa^S_p(s[i], t[j]) \\
\kappa^S_p(\epsilon, s) &= 0 \\
\kappa^S_p(s, \epsilon) &= 0 \\
\kappa^S_1(sa, tb) &= I_{a,b}\ \lambda^2 \\
\kappa^S_p(sa, tb) &= I_{a,b}\ \lambda^2 \sum_{i=1}^{|s|} \sum_{j=1}^{|t|} \lambda^{|s|-i+|t|-j+2} \kappa^S_{p-1}(s[i], t[j]),
\end{aligned}$$

ただし, $I_{a,b}$ は a と b が同じ文字のとき 1, そうでなければ 0 であり，これは字面上同じ部分文字列の頻度を数え上げるハートマッチングを実現する．

一方，文献 [Saunders 02] では, $I_{a,b}$ の代わりにアルファベットの類似度 $\boldsymbol{A}_{a,b}$ を用いて, ソフトマッチングを行うことができることを示している．このとき，$\boldsymbol{A}$ は対称かつ半正定値であることが求められるが，本研究では，音素の混同行列に基づいた以下のような $\boldsymbol{A}$ を用いる．

$$\boldsymbol{A}_{a,b} = \sum_{c \in \Theta} P(O=c) P(R=a|O=c) P(R=b|O=c)$$

ここで，$P(O=c)$ は音素 c の生起確率であり，$P(R=a|O=c)$ は，音素 c を音素 a と誤認識する確率であり，いずれも混同行列から推定される．このような行列が対称かつ半正定値であることは容易に確認することができる．

以上述べたような方法で，原子セグメントの表現ベクトル $\phi(s_i)$ どうしの内積を格納した行列

$$\boldsymbol{K}_{i,j} = \sum_{k=1}^{p} \gamma_k\ \kappa_k(s_i, s_j),\ (1 \le i, j \le n)$$

が計算される．本稿では，以降，$\gamma_k = 1$ と仮定する．

3.2 対角成分の縮減

上述した，$\phi(s_i)$ には，s_i にしか出現しない部分文字列の成分が非常に多く含まれるが，このような成分は原子セグメントどうしの類似性を議論する際には冗長であるので削減することが望ましい．しかし，$\phi(s_i)$ の制限を直接計算することは計算量の観点から困難であるので，$\boldsymbol{K}$ の対角成分を縮減することで実現する． s_i にしか出現しない成分を取り除くことは, $\boldsymbol{K}$ の対角成分にしか影響を与えないからである．新たな対角成分は，以下のように計算される．

$$\tilde{\boldsymbol{K}}_{i,i} = \|P_{U_i}(\phi(s_i))\|^2.$$

ここで，P_{U_i} は，$\phi(s_1), \ldots, \phi(s_{i-1}), \phi(s_{i+1}), \ldots, \phi(s_n)$ が張る部分空間 U_i への射影を表している．この射影は，$\boldsymbol{K}$ の i 行と i 列を取り除いた行列の固有値 $\lambda_1 > \cdots > \lambda_m > 0$, 固有ベクトル $\boldsymbol{v}^1, \ldots, \boldsymbol{v}^m$ を用いて以下のように計算できる．

$$\begin{aligned}
\|P_{U_i}(\phi(s_i))\|^2 &= \sum_{j=1}^{m} \left(\sum_{k=1}^{n} \beta^j_k \boldsymbol{K}_{i,k} \right)^2 \\
\beta^j_k &= \begin{cases} \dfrac{1}{\sqrt{\lambda_j}} \boldsymbol{v}^j_k & k < i \\ 0 & k = i \\ \dfrac{1}{\sqrt{\lambda_j}} \boldsymbol{v}^j_{k-1} & k > i \end{cases}
\end{aligned}$$

3.3 冗長成分の除去

機能語や，トピックに無関係に出現する内容語などを構成する音素列を除去する目的で，平均ベクトルの直交補空間へ,

各原子セグメントベクトルを射影する．このような冗長な成分は，どのような部分区間であっても平均ベクトルと同じ頻度パターンで出現すると期待されるからである．原子セグメントの表現ベクトル $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$，平均ベクトルを $\boldsymbol{g} = (1/n)\sum_{i=1}^{n} \boldsymbol{z}_i$ とするとき，$\boldsymbol{g}$ の直交補空間への $\boldsymbol{z}_i$ の射影は，

$$P_{\perp}(\boldsymbol{z}_i) = \boldsymbol{z}_i - \frac{\langle \boldsymbol{g} \cdot \boldsymbol{z}_i \rangle}{\|\boldsymbol{g}\|}\boldsymbol{g}$$

であるから，以下のような行列の変換を行えばよい．

$$\begin{aligned}\tilde{\boldsymbol{K}}_{i,j} &= \langle P_{\perp}(\boldsymbol{z}_i) \cdot P_{\perp}(\boldsymbol{z}_j) \rangle \\ &= \boldsymbol{K}_{i,j} - \frac{\left(\sum_{k=1}^{n} \boldsymbol{K}_{i,k}\right)\left(\sum_{k=1}^{n} \boldsymbol{K}_{j,k}\right)}{\sum_{k=1}^{n}\sum_{\ell=1}^{n} \boldsymbol{K}_{k,\ell}}.\end{aligned}$$

直交補空間上に射影された原子セグメントベクトルの重心は原点となる．平均ベクトルを差し引く通常のセンタリングとの違いは，この変換が線型であり，原子セグメントベクトルの和を保存する点である．このことは，トピック分割が原子セグメントの粒度に依存しないという意味で望ましい性質である．

3.4 最適分割

原子セグメントの類似性行列 $\boldsymbol{K}$ に基づいて，原子セグメントの最適な分割 $T^* = \{T_1, \ldots, T_L\} \subseteq \{1, \ldots, n\}, (T_0 = 0 < T_1 < \cdots < T_L)$ を求める．セグメント T_k の表現ベクトルを $\boldsymbol{t}_k = \sum_{i=T_{k-1}+1}^{T_k} \boldsymbol{z}_i$ と考えると，以下の事実に着目すれば，

$$\cos\theta = \frac{\langle \boldsymbol{t}_i \cdot \boldsymbol{t}_j \rangle}{\|\boldsymbol{t}_i\|\|\boldsymbol{t}_j\|} \geq 0 \Leftrightarrow \|\boldsymbol{t}_i + \boldsymbol{t}_j\|^2 \geq \|\boldsymbol{t}_i\|^2 + \|\boldsymbol{t}_j\|^2$$

セグメントのノルムが増えるようなセグメントの併合を行うことで，同じ方向を向いたベクトルが併合されることが分かる．実際，原子セグメントが一様に分布すると仮定するとき，ノルムの最大化はセグメント内分散の最小化，セグメント間分散の最大化に相当する [Kobayashi 08]．そこで，最大のノルムを持つ分割を最適な分割とする．

$$\begin{aligned}T^* &= \operatorname*{argmax}_{T=\{T_1,\ldots,T_\ell\}} \sum_{k=1}^{\ell}\sum_{i=T_{k-1}+1}^{T_k}\sum_{j=T_{k-1}+1}^{T_k} \boldsymbol{K}_{i,j} \\ &= \operatorname*{argmin}_{T=\{T_1,\ldots,T_\ell\}} Cost(T) \\ Cost(T) &= \sum_{k=1}^{\ell}\sum_{i=T_{k-1}+1}^{T_k-1}\sum_{j=i+1}^{T_k} -\boldsymbol{K}_{i,j}\end{aligned}$$

ノルムの最大化は，以下のような再帰式に基づく動的計画法により $C(n) = Cost(T^*)$ を求めることで解くことができる．

$$\begin{aligned}C(0) &= 0 \\ C(j) &= \min_{0\leq i<j} (C(i) + c(i,j)), \qquad (1 \leq j \leq n) \\ p(j) &= \operatorname*{argmin}_{0\leq i<j} (C(i) + c(i,j)), \qquad (1 \leq j \leq n) \\ c(i,j) &= -\sum_{k=i+1}^{j-1}\sum_{\ell=k+1}^{j} \boldsymbol{K}_{k,\ell}, \qquad (1 \leq i,j \leq n)\end{aligned}$$

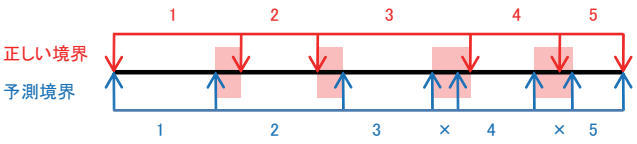

図 1: 分割誤り率: 最適な対応付け 1 – 5 の元で，正しく対応しない領域（色付けされた部分）の割合

このとき，最適分割の長さを ℓ　とすれば，

$$p^{\ell}(n) = 0, p^{\ell-1}(n), \ldots, p^1(n), p^0(n) = n,$$

が最適分割となる．ただし，$p^k(n) = \underbrace{p(\cdots p(}_{k} n)\cdots)$.

また，分割数 L が所与の場合は，以下のような動的計画法を用いればよい．

$$\begin{aligned}C_1(0) &= \infty \\ C_1(j) &= c(0,j),\ p_1(j) = 0, \qquad (1 \leq j \leq n) \\ C_\ell(j) &= \infty, \qquad (0 \leq j < \ell,\ 2 \leq \ell \leq L) \\ C_\ell(j) &= \min_{0\leq i<j} (C_{\ell-1}(i) + c(i,j)), \ (\ell \leq j \leq n,\ 2 \leqq \ell \leq L)\end{aligned}$$

以上述べたように，文字列カーネルを用いて，原子セグメントの類似性を格納した行列 $\boldsymbol{K}$ を計算してしまえば，冗長な成分の除去と最適分割いずれの計算も，表現ベクトルを陽に取り扱うことなく，$\boldsymbol{K}$ を用いて計算することが出来る．

4. 実験

言語知識を用いず，単語の代わりに音素部分列の頻度を分析することによるトピック分割性能の劣化と，連続音素認識によって得られる音素列に高頻度に含まれる置換・脱落・挿入誤りによる性能劣化を評価するための実験行った．この際，比較のために，テキストに対するトピック分割アルゴリズムとして `Textseg`[Utiyama 01] と `LCSeg`[Galley 03] を用いたが，`LCSeg` は，良い分割性能が得られなかったので割愛する．

評価データとしては，ICSI Meeting Corpus[Janin 03] を用いる．このデータは，主として ICSI の研究グループで定期的に行われている音声と言語に関する少人数 (平均 6.5 人) の会議 75 回分を，話者毎のヘッドセットマイクと複数のバウンダリマイクを用いて収録したものである．各会議の長さはおよそ 60 分程度で，発話毎の開始時間，終了時間と書き起こしテキストの他，さまざまなアノテーションが付与されている．

文献 [Galley 03] では，トピック分割性能評価のためのテストセットとして 25 の会議を選び，少なくとも 3 人により行われたトピック分割を元にして，各会議あたり 7 個程度のトピックに分割している．本研究でも，この 25 個の会議をテストセットとし実験を行い，その際上記分割を正解分割とする．

分割の評価指標としては，予測分割, 正解分割それぞれのセグメントを対応付ける時に，正しく対応付けされない領域の割合を用いる（図 1 参照）．セグメントの対応付けは複数の可能性があるので，任意の対応付けの下で計算された値の最小値を分割誤り率と定義し評価指標とする．

表 1 は，書き起こしテキストに対して，`Textseg`(U00 と記述) と提案手法 (S10 と記述) を適用した結果得られた分割誤り率の平均値を示している．提案手法においては，単語列中の単語を 1 つのシンボルとみなして本手法を適用した結果を S10(単語) として，音素列の音素を 1 つのシンボルとみなして

データ	書き起こし			音声認識
アルゴリズム	U00	S10(単語)	S10(音素)	S10(音素 ASR)
分割数なし	0.322 (± 0.020)	0.328 (± 0.018)	0.318 (± 0.023)	0.346 (± 0.026)
分割数あり	0.303 (± 0.019)	0.312 (± 0.016)	0.306 (± 0.026)	0.342 (± 0.027)

表 1: 分割誤り率：上段は正解分割数を与えない場合．下段は与える場合．括弧内は標準誤差.

本手法を適用した結果を S10(音素) として示している．S10(音素) については，音声認識によって得られた音素列に適用して得られた結果 S10(音素 ASR) も同時に示している．

U00, S10(単語) の実験では，テキスト中の単語列に対して，ステミングとストップワード除去を適用した後，テキストの分割を行っている．U00 と S10(単語) では，文献 [Choi 00] で用いられた, ストップワードリストと Porter のステミングアルゴリズムを共通に用いている．S10(単語) は，$p = 1$, でハードマッチングを用いた．

S10(音素) においては，Force Alignment によりテキストを音素列に変換し, 提案手法を適用している．また，S10(音素 ASR) に置いては，連続音素認識の結果得られる音素列に提案手法を適用している．いずれの場合も，パラメータは，$p = 40$, $\lambda = 0.1$ を用い, ソフトマッチングを用いている．

連続音素認識に用いたデコーダーは，音声対話技術コンソーシアム [ISTC] が配布している Julius[Lee 01] の english dictation kit を用いた．言語モデルは，訓練データから学習させた音素トライグラムを用いている．音響モデルとしては，同梱されている WSJ から学習されたトライフォンモデルをベースに，訓練データを用いて MLLR で話者適応を行ったものを使用した．連続音素認識の認識率は，評価データに対する音素正解率が 63.8%, 音素認識精度が 50.1% であった．

4.1 考察

まず，S10(単語) は，U00 と比較して，paired t-test において有意な性能差はなかった．ここでは，ユニグラムの頻度のみを分析しているが，N グラムに拡張すると提案手法の性能が向上することは確認しているが，テキスト分割アルゴリズムとしての性能については稿を改めて考察したい．

次に，書き起こしテキスト得られた音素列を分割する S10(音素) を S10(単語) と比較すると，有意な性能劣化は確認出来ない．S10(音素) では，語彙，品詞，語形変化，ストップワードリスト等の言語知識を用いることが出来ないが，任意の部分音素列の頻度を，冗長成分の除去などの前処理行い分析する提案手法によれば，このような言語知識の不足を補うことができることを示している．

さらに，連続音素認識結果を用いた S10(音素 ASR) を S10(音素) と比較しても，有意な性能劣化は確認できなかった．これは，ソフトマッチングを用いた文字列カーネルによる分析が，音素の脱落・挿入・置換誤りを適切に補償しているためであると考えられる．

5. おわりに

本稿では，トピックモデル, 語彙や言語モデル等，音声言語の内容に依存する知識を仮定しない，トピック分割アルゴリズムを提案した．このアルゴリズムは，連続音素認識により得られた音素列を，カーネル法を用いた部分音素列の網羅的な分析を行うことによって，トピック分割を行う．実際に，このアルゴリズムを ICSI 会議音声コーパスに適用し，言語知識を用いなくても有意なトピック分割性能の劣化が認められないこと，音素認識誤り率がおよそ 50% の認識精度の低い認識音素列を対象としても有意な性能劣化が認められないことを確認した．

このアルゴリズムは，単語を音素と考えれば，ギャップを許した不連続な単語 N グラムに基づくテキストのトピック分割を行うことも可能である．さらに，データ間に距離が定義された任意の系列データの分割にも適用可能である．これら音声とは異なるデータに対する有効性を今後検証していきたい．

参考文献

[Janin 03] Janin, A., et al.: The ICSI Meeting Corpus, in *Proc. of ICASSP*, pp. 364–367 (2003)

[Lee 01] Lee, A., et al.: Julius — an open source real-time large vocabulary recognition engine, in *Proc. of EURO SPEECH*, pp. 1691–1694 (2001)

[Beeferman 99] Beeferman, D., et al.: Statistical models for text segmentation, *Machine Learning*, Vol. 34, No. 1–3, pp. 177–210 (1999)

[Choi 00] Choi, F.: Advances in doman independent linear text segmentation, in *Proc. of NAACL*, pp. 26–33 (2000)

[Lodhi 02] Lodhi, H., et al.: Text Classification using String Kernels, *Journal of Machine Learning Research*, Vol. 2, pp. 419–444 (2002)

[ISTC] Interactive Speech Technology Consortium (ISTC): `http://www.astem.or.jp/istc`

[Kobayashi 08] Kobayashi, T., et al.: Motion image segmentation using global criteria and DP, in *Proc. of FG*, pp. 1–6, (2008)

[Sadohara 06] Sadohara, K., et al.: Domain-independent topic segmentation using a string kernel on recognized sub-word sequences, in *Proc. of SLT* (2006)

[Hearst 94] Hearst, M.A.: Multi-paragraph segmentation of expository text, in *Proc. of ACL* (1994)

[Galley 03] Galley, M., et al.: Discourse segmentation of multi-party conversation, in *Proc. of ACL*, pp. 562–569,(2003)

[Utiyama 01] Utiyama, M. and Isahara, H.: A statistical model for domain-independent text segmentation, in *Proc. of ACL*, pp. 491–498 (2001)

[Hsueh 06] Hsueh, P., et al.: Automatic Segmentation of Multiparty Dialogue, in *Proc. of EACL*, (2006)

[Saunders 02] Saunders, C., et al. : Syllables and other string kernel extensions, in *Proc. of ICML*, pp. 530–537 (2002)