2L3-2in

複数人対話における視線交差のタイミング構造に基づく次話者と発話開始タイミングの予測

Prediction of next speaker and its start time based on timing structure of mutual gaze in multi-party meetings

石井亮[*1] 大塚和弘[*1] 熊野史朗[*1] 大和淳司[*1]
Ryo Ishii Kazuhiro Otsuka Shiro Kumano Junji Yamato

[*1] NTTコミュニケーション科学基礎研究所, 日本電信電話株式会社
NTT Communication Science laboratories, NTT Corporation

To build a model for predicting the next speaker and the start time of the next utterance in multi-party meetings, we focused on the timing structure of eye contact, that is who looks at first when the eye contact between a speaker and a listener happens, for constructing the prediction model. A result of analysis shows that a listener that becomes a next speaker tends to look at a speaker after the speaker looks at the listener in turn-taking. And, the start time of the next utterance is different depending on the timing structure of eye contact in turn-taking. As the result of evaluation of the prediction model using the timing structure of eye contact is useful to predict a next speaker and its stat time of next speaking.

1. はじめに

対面会話は, 社会生活において他者との情報交換, 意思決定を行う上で, 最も基本的なコミュニケーション形態である. 対面会話のような円滑なコミュニケーションを, 遠隔コミュニケーションや, 人と会話エージェントのコミュニケーションにおいても実現することが望まれている. そのため, 近年, 特に 3 人以上の対話(以後, 複数人対話と呼ぶ)を対象に, 円滑なコミュニケーションがそもそもどのように行われているかの機序の自動分析やモデル化に関する研究が注目を集めている [Otsuka 11].

複数人対話において, 話者が入れ替わる話者交替は特に重要な局面である. 本研究では, 以降, 着目する発話を行っている人物を“現話者”, 発話を行っていない人物を“非話者”, 次の発話を行う人物を“次話者”と呼ぶ. 複数人対話では, 複数の次話者の候補がいるため, 話者交替の機序が複雑になる. そのため参加者は, 現話者の発話の終了のタイミングや, 複数人の非話者のうち次に誰がいつ発話を開始するかを予測して, 参加者自身が発話をいつするべきか, または他の参加者に発話を促すかといった調整を行っている. もし, 複数人対話において次話者や発話開始のタイミングを予測可能になれば, その予測技術は, 適切なタイミングで発話を開始, 終了できる会話エージェントや, 遠隔コミュニケーションにおいて参加者に誰が次話者となるかを通知するなどして, 通信遅延による発話衝突を回避するシステムの実現に向けた基盤的な技術となることが期待される.

社会言語学の分野を中心に, 言語情報に加えて視線などの非言語情報が複数人対話における次話者の規定に重要な役割を担っていることが示されている [榎本 11, Jokinen 13]. 我々はこれまで, 視線, 頭部動作, 呼吸動作の情報を用いて, 複数人対話において誰が次話者になるか, さらに, 前の発話に対して, いつ発話が開始されるかのタイミングを予測可能なモデルの構築に取り組んできた [石井 14, Ishii 14, Ishii 15]. このとき, 次話者と発話開始タイミングを予測する上で, 特に発話末付近の現話者と非話者の視線の遷移パターンの情報が単体では有用であった. 今後さらに, 視線行動の遷移パターン以外の情報と次話者・発話開始タイミングとの関連性を明らかにして, より高精度な予測モデルを構築することが望まれている.

本研究では, より高精度な次話者・発話開始タイミングの予測モデルの構築を目指して, これまで着目がされていなかった視線行動のタイミング構造に着目し, 予測モデルの構築に取り組む. タイミング構造とは, 2 者の視線行動の時間的な順序であり, 本研究では, 特に話者交替に重要とされている現話者と非話者の視線交差に着目し, 視線交差が開始する際に, 現話者と非話者のどちらが先に視線を相手に向けるかといった, 視線交差時の 2 者の視線の順序に着目する. 先行研究 [石井 14] では, 発話末における各参加者の注視対象の遷移パターンに着目をしており, 2 者の視線行動の時間的な順序の情報を扱っていなかった. また, 話者交替時に現話者と視線交差を行った非話者が次話者になる傾向があることが示されている [榎本 11, 石井 14] が, 必ずしも, 視線交差した非話者が次話者になるわけではない. 視線交差のタイミング構造と, 次話者・発話開始タイミングの関連性が明らかになれば, タイミング構造の情報を用いて, より高精度な次話者・発話開始タイミングの予測モデルが構築できると考えられる.

以降, 分析のために構築したコーパスデータ, 視線交差のタイミング構造と次話者・発話開始タイミングの分析結果, タイミング構造の情報を用いた, 次話者・発話開始タイミングの予測モデルの構築と評価結果について報告する.

2. 複数人対話コーパス

コーパスデータ収集のために参加者 4 人対話の収録を行った. 参加者は, 初対面の同世代(20 代～30 代)の女性であった. 対話内容は, “男性と女性はどちらが得であるか”といった意見の分かれやすい議題を与え, 8 分間以内で議論を行った後に, 4 人で 1 つの答えを出させるように教示した. 発話の収録のために各被験者の胸につけられた指向性ピンマイクを用いた. また, 対話状況の全体的な外観や参加者のバストアップの様子の映像(30Hz)をビデオカメラで撮影した. 撮影された映像の一例を図 1 に示す. 参加者は全部で 20 名であり, 20 名を 4 名ずつに分けて 5 グループを構成し, それぞれのグループで 2 対話を

連絡先:石井亮, NTTコミュニケーション科学基礎研究所,
日本電信電話株式会社, 神奈川県厚木市森の里若宮3－1,
ishii.ryo@lab.ntt.co.jp

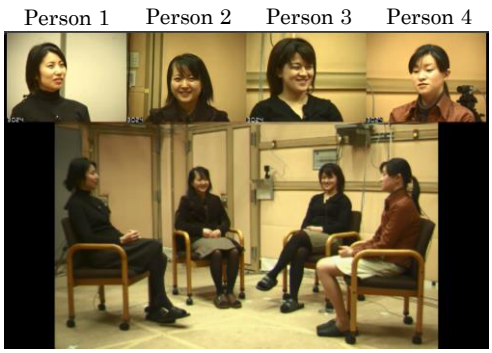

図 1　撮影した対話シーンの例

収録した. 収録した対話の音声・映像データから以下の言語・非言語情報のコーパスデータを作成した.

- 発話: 音声情報から発声言語の書き越しを行った後, 一定の音声の無音区間によって, 発話を区切る IPU(Inter-Pausal Unit)を利用して, 発話区間を生成した. IPU の具体的な生成方法として, 200ms 以上の無音区間で囲まれる発話部分を一つの IPU とした. 作成された IPU から, あいづちを除外し, 同一の人物による継続した IPU を 1 つの発話ターンとした. そして, 時間的に連続した IPU のペアを作成した. 作成された IPU の組は, 話者継続時で 1485, 話者交替時で 418 であった. 我々はこれまで, IPU が作成される時刻, すなわち IPU 区間の終端から 200ms 後の時点で, 次の発話が誰であるかを予測することに取り組んでおり [石井 14, Ishii 14], 本研究においても同様の時点で予測することに取り組む. そのため, 話者交替時に, 前 IPU の終端から 200ms よりも前に, 次の IPU が開始されていたデータを除外した. 残されたデータは, 全体の 74.2%であり, 310 ペアであった.
- 視線: 撮影された映像を観察して, 注視対象のラベリングを行った. ラベリングされた注視対象は, “4 人の各参加者(Person 1, 2, 3, 4)”および“Others”の 5 種類である.

3. 視線交差の予備分析とタイミング構造の定義

3.1 視線交差の予備分析

まず, 構築したコーパスにおいて, 発話末でどれくらいの視線交差が起きたか, また視線交差の開始や終了がいつ起きるかを予備分析し, どのような情報が予測に利用できそうかを議論する.

先行研究において, IPU 末付近(IPU 末の前 1000ms から IPU 末の後 200ms の間の 1200ms 区間)に起きた注視対象の情報は次話者・発話開始タイミングの予測に有用であった [石井 14]. 本研究においても, IPU 末付近 1200ms の間に起きた視線交差に着目する. IPU 末付近 1200ms の間に起きた現話者と非話者の視線交差の回数を集計した結果を表 1 に示す. 視線交差が起きたデータは, 話者継続の 1485 データの内の 718, 話者交替の 310 データの内の 182 であった. また, 話者交替時に視線交差が起きた 182 データの内, 視線交差していない非話者が次話者になったのが 90, 視線交差した非話者が次話者になったのが 92 であった. よって, 話者継続時の約 48%で現話者と非話者の視線交差が発生し, 話者交替時の約 59%で視線交差が発生している. また, 話者交替時では, 現話者と非話者の視線交差が起きた際に, 現話者と視線交差をした非話者が次話者になるのは, 約 50%である. これらの結果から, 話者継続, 話者交替時のどちらにおいても, 現話者と非話者の視線交差は多く発生し, また, 話者交替において, 必ずしも現話者と視線交差をした非話者が次話者になるわけではない. 従来研究では, 話者継続時には, 現話者と次話者になる非話者の視線交差が起こることが報告されている [榎本 11, 石井 14] が, 本コーパスにおいて, 視線交差の有無だけで, 次話者を予測するのは難しいと考えられる.

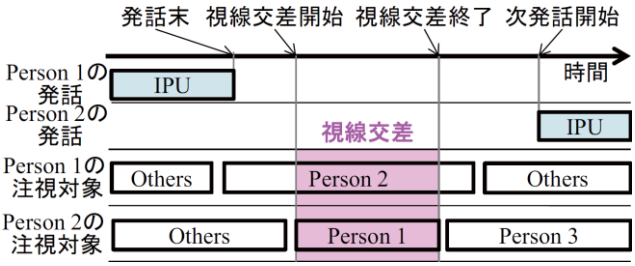

図 2　視線交差におけるタイミング構造

次に, 現話者と非話者の視線交差が起きた時に, 次話者が, “現話者であったとき(話者継続時)”, “視線交差していない非話者であったとき(話者交替時)”, “視線交差した非話者であったとき(話者交替時)”の 3 つの状況下で, 視線交差の開始・終了時刻と, 発話末・次発話開始時刻との関係を分析した(各時刻については図 2 を参照). その結果を, 表 1 に示す. まず, 発話末から視線交差開始までの平均時間は, -813.7ms から -979.7ms であり, 視線交差の開始は発話末よりも前から起きることが確認された. また, 視線交差開始から次発話開始までの平均時間は, 1589.1ms から 1849.7ms であった. よって, 視線交差は, 発話末の前から起こるため, 視線交差開始に伴うタイミング構造の情報は, 次発話や発話のタイミングを予測するために利用できる可能性があると考えられる.

次に, 発話末から視線交差終了までの平均時間は, 206.7ms から 587.0ms であった. また, 視線交差終了から次発話開始までの平均時間は, -574.4ms から 72.0ms であった. よって, 視線交差は, おおよそ次発話開始の直前から, 次発話開始の少し後の間に起こるため, 視線交差終了に伴うタイミング構造の情報は, 予測にはあまり利用できない可能性が示唆された.

3.2 視線交差時のタイミング構造

予備分析をふまえて, 次話者・発話開始タイミングの予測のための情報の候補として, 本研究では視線交差の開始時のタイ

表 1　次話者になった人物ごとに視線交差の継続長と発話開始タイミングの相関係数を算出した結果

	データ数	次話者	視線交差が起きたデータ数	発話末から視線交差開始までの平均時間	視線交差開始から次発話開始までの平均時間	発話末から視線交差終了までの平均時間	視線交差終了から次発話開始までの平均時間
話者継続	1485	現話者	718	-813.7 ms	1589.1 ms	587.0 ms	3.3 ms
話者交替	310	視線交差していない非話者	90	-789.7 ms	1849.7 ms	206.7 ms	-574.4 ms
		視線交差した非話者	92	-979.7 ms	1739.0 ms	354.7 ms	72.0 ms

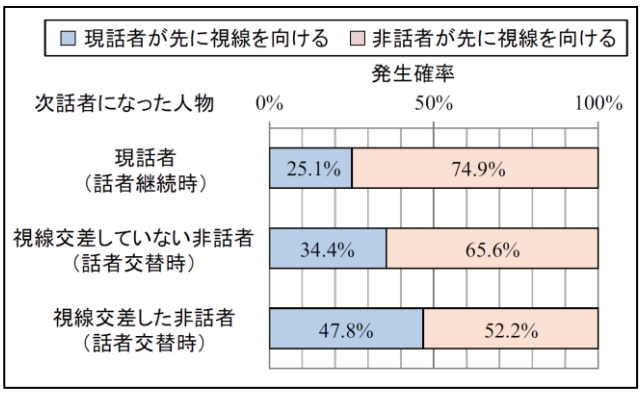

図 3　視線交差開始時におけるタイミング構造と次話者の分析結果

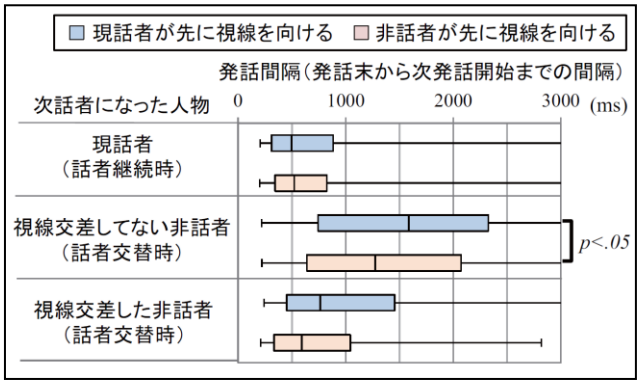

図 4　視線交差開始時のタイミング構造と発話開始タイミングの分析結果

ミング構造に着目する. 図 1 は, Person 1 と person 2 の発話（IPU）と注視対象を時系列で示したものであり, 現話者である Person 1 が発話を終えた後, 話者交替が起こり, 非話者であった Person 2 が次話者になるシーンの一例である. このとき, Person 1 の発話末付近の視線に着目すると, Person 1 と Person 2 の間で視線交差が起きている. このときの視線交差のタイミング構造は, 現話者（Person 1）が先に視線を向け, 非話者（Person 2）が後に視線を向けている. このような視線交差開始のタイミング構造に応じて, 次話者と発話開始タイミングが変化するのかを次に分析する.

4. 視線交差のタイミング構造の分析

4.1 タイミング構造と次話者の関係

次話者が, 現話者であるとき（話者継続時）, 話者交替時に, 視線交差した非話者以外の非話者であるとき, 視線交差した非話者であるときの 3 つの状況下で, IPU 末付近で発生した, 現話者と非話者の視線交差が開始される時に, 現話者と非話者のどちらが先に視線を相手に向けるのかを分析した. その結果を, 図 3 に示す. 次話者になった人物の条件が, 視線交差開始時のタイミング構造に影響を与えるのかを検証するために, カイ二乗検定を行った. その結果, 有意傾向が確認された ($\chi^2(2)=10.60, p<.10$). 次に, 残差分析を行い, どの条件に特徴がみられるかを分析した. その結果, 現話者が次話者になるとき（話者継続時）には, 視線交差開始時に非話者が先に現話者に視線を向ける頻度が有意に高く ($p<.10$), 話者交替時に視線交差した非話者が次話者になるときは, 現話者が先に視線を向ける頻度が有意に高いことが示唆された ($p<.05$).

よって, 次話者が現話者であるとき, すなわち話者継続時のときに, 非話者が先に現話者に視線を向けている確率が 74.9% と高い. これに対して, 次話者が, 話者交替時に視線交差した非話者であるときは, 非話者が先に現話者に視線を向けている確率が 52.2%と低いことが明らかになった.

4.2 タイミング構造と発話開始タイミングの関係

4.1 節と同様に, 3 つの次話者の条件下で, 現話者と非話者の視線交差開始時のタイミング構造によって, 発話末から次発話開始までの発話間隔が異なるかを分析した. 発話間隔を集計した結果を, 箱ひげ図で示したものを図 4 に示す. 箱ひげ図の箱の左端は第 1 四分点, 右端は第 3 四分点, 中央線は中央値, ひげの左端は最小値, 右端は最大値を示す.

まず, 分散分析をおこなった結果, 次話者になった人物と, 視線交差のタイミング構造の 2 要因の交互作用が有意であった ($F(2,884)=11.11, p<.01$). さらに, 次話者になった人物の各条件下で, 視線交差のタイミング構造の単純主効果を評価した結果, 次話者が視線交差をしていない非話者であったという条件下において, 非話者が先に視線を向けたときに比べて, 現話者が先に視線を向けたときの発話間隔の平均が有意に長かった ($t(88)=-2.20, p<.05$).

よって, 話者交替時に現話者と非話者の視線交差がおきたとき, 視線交差していない非話者が次話者になるときに, 視線交差開始時に現話者が先に（次話者にならない）非話者に視線を向けたときは, 次話者の発話開始タイミングが長くなる. 逆に, 非話者が先に非話者に視線を向けたときは, 発話開始タイミングが短くなることが示唆された.

5. 次話者と発話開始タイミングの予測モデル構築

4 章の分析結果から, 現話者と非話者の視線交差開始時のタイミング構造が, 次話者・発話開始タイミングと関連性があることが明らかとなった. 次に, 視線交差開始時のタイミング構造の情報が, 次話者と発話開始タイミングの予測に有用であるかを, 予測モデルを構築し, 性能を評価することで検証を行った. これまで, IPU 末付近 1200ms の視線データを用いて, 下記の 3 つのステップで次話者と発話開始タイミングの予測を行うモデルを構築してきた [石井 14].

① 話者継続/話者交替のどちらが起こるかの予測
② 話者交替時に誰が次話者になるかの予測
③ 発話開始タイミングの予測

本研究においても, 発話末から 200ms 後の時点までで得られる視線交差開始時のタイミング構造の情報を用いて, 先行研究と同様に, 3 つのステップで次話者と発話開始のタイミングを予測するモデルを構築する. 予測モデルの評価にあたっては, 先行研究で用いた視線遷移パターンの情報を特徴量として用いた予測モデルとの比較を行う.

5.1 話者継続/話者交替の予測モデルと評価

4.1 節の分析結果から, 次話者が現話者であるとき, すなわち話者継続時のときに, 非話者が先に現話者に視線を向けている確率が高く, 次話者が, 話者交替時に視線交差した非話者であるときは, 現話者が先に非話者に視線を向けている確率が高いことが明らかになった. よって視線交差開始時におけるタイミング構造の情報は話者交替, 話者継続の予測に利用できると考えられる. そこで, 実際に, 視線交差のタイミング構造の情報が話者継続/交替の予測に有用であるかを検証するために, SVM を用いた予測モデルを構築し, モデルの性能を評価した. 利用したアルゴリズムは, Weka [Bouckaert 10] に実装された

表 2 話者継続/交替の予測モデルの評価結果（F 値）

	話者継続	話者交替
従来モデル	0.680	0.652
提案モデル	0.696	0.690

表 3 話者交替時の次話者の予測モデルの評価結果

	予測精度
従来モデル	61.0%
提案モデル	62.6%

SMO [Keerthi 01] を用いた. SVM のカーネルには RBF を使用した. クラスとして, 話者交替, 話者継続の 2 クラスを設定した. 構築した, 従来モデルと, 視線交差のタイミング構造を用いた提案モデルで用いた特徴量を下記に示す.

- 従来モデル:現話者の人物, IPU 末 1000ms から IPU 末 200ms の間の 1200ms における参加者 4 人の視線行動の遷移パターンを特徴量として用いる. 詳細は, [石井 14] を参照されたい.
- 提案モデル:従来モデルで用いた特徴量に加えて, 視線交差が起きた際に, 現話者と非話者どちらが先に視線を向けたかの情報を特徴量として用いる.

学習・テストに用いたデータは, 話者継続と交替時のデータ数を均等にするために, 4 章の分析で使用した話者交替の 310 のデータと, 話者継続の 1485 のデータの中からランダムに抽出された 310 のデータの計 620 の データを用いた.

10 分割交差検定を行ってモデルの性能を評価した結果を, 表 2 に示す. 従来モデルに比べて, 提案モデルで話者継続, 交替の F 値が 0.696, 0.690 と向上した. よって, 視線交差開始時のタイミング構造の情報は話者継続/交替の予測に有用であることが示唆された.

5.2 話者交替時の次話者の予測と評価

5.1 節と同様にして, 視線交差のタイミング構造の情報が話者交替時の次話者の予測に有用であるかを検証するために, SVM を用いた予測モデルを構築し, モデルの性能を評価した. 利用したアルゴリズム, 構築したモデルの特徴量は, 5.1 節と同様である. クラスとして, 話者交替時に 3 人の非話者の中から次話者になる人物の 3 クラスを設定した. 使用したデータは, 4 章の分析で使用した話者交替の 310 のデータである.

10 分割交差検定を行ってモデルの性能を評価した結果を, 表 3 に示す. 従来モデルに比べて, 提案モデルで話者交替時の次話者の予測精度が 62.6%と向上した. よって, 視線交差開始時のタイミング構造の情報は, 話者交替時の次話者の予測に有用であることが示唆された.

5.3 発話開始タイミングの予測と評価

4.2 節の分析結果から, 現話者と非話者の視線交差開始時に, 視線交差をしていない非話者が次話者になるときに, 視線交差のタイミング構造によって, 発話開始タイミングが異なることが明らかとなった. よって視線交差のタイミング構造の情報は, 話者交替時の発話開始タイミングの予測に利用できると考えられる. そこで, 実際に, 視線交差のタイミング構造の情報が話者交替時の発話開始タイミングの予測に有用であるかを検証するために, SVR（Support Machine Regression）を用いて, 発話末から次発話開始までの発話間隔を予測するモデルを構築し, 性能を評価した. 利用したアルゴリズムは, Weka に実装された SMOreg [Keerthi 01] を用いた. 予測モデル構築に用いた特徴量は, 5.1, 5.2 節と同じである. 使用したデータは, 4 章の分析で使用した話者交替の 310 のデータである.

表 4 話者交替時の発話開始タイミング（発話間隔）の予測モデルの評価結果

	予測誤差の平均
従来モデル	965 ms
提案モデル	911 ms

10 分割交差検定を行ってモデルの性能を評価した結果を, 表 4 に示す. 話者交替時の発話間隔の予測誤差の平均は, 従来モデルが 965ms だったのに比べて, 提案モデルで 911ms と小さくなった. よって, 視線交差開始時のタイミング構造の情報は, 話者交替時の次話者の発話開始タイミングの予測に有用であることが示唆された.

6. まとめ

複数人対話における次話者と発話開始タイミングの予測モデルの構築に向けて, これまで扱われていなかった, 発話末付近に起こる現話者と非話者の視線交差のタイミング構造に着目した. 分析から, 話者継続時には, 非話者が先に現話者に視線を向けている確率高く, 話者交替時に視線交差した非話者が次話者になるときは, 現話者が先に非話者に視線を向けている確率が高いことが明らかになった. また, 話者交替時に, 視線交差してない非話者が次話者になる際に, 視線交差開始時に現話者が非話先に先に視線を向けたとき, 発話開始のタイミングが遅くなることが分かった. このような視線交差開始時のタイミング構造の情報を用いて予測モデルを構築した結果, 視線交差のタイミング構造の情報は, 次話者と発話開始タイミングの予測に有効であることが示唆された.

今後は, 現話者と非話者の視線交差以外の視線行動のタイミング構造に着目し, 次話者・発話開始タイミングとの関連性を明らかにする. また, 視線, 頭部動作, 呼吸動作といったマルチモーダル情報を利用した予測モデルの構築に取り組む予定である.

参考文献

[Otsuka 11] Otsuka, K.: Conversational scene analysis, *IEEE Signal Processing Magazine*, vol.28, pp.127–131 (2011).

[榎本 11] 榎本美香, 伝康晴, 話し手の視線の向け先は次話者になるか, *社会言語科学*, Vol.14, pp.97‐109 (2011).

[Jokinen 13] Jokinen, K., et al.: Gaze and turn-taking behavior in casual conversational interactions, *TiiS* 3(2): 12 (2013).

[石井 14] 石井 亮, 大塚 和弘, 熊野 史朗, 松田 昌史, 大和 淳司:複数人対話における注視遷移パターンに基づく次話者と発話開始タイミングの予測, *電子情報通信学会論文誌*, Vol.J97-A, No.6, pp.453-468 (2014).

[Ishii 14] Ishii, R., et al.: Analysis of Respiration for Prediction of “Who Will be Next Speaker and When” in Multi-Party Meetings, in *Proc. ICMI*, pp.18-25 (2014).

[Ishii 15] Ishii, R., et al.: Predicting Next Speaker using Head Movement in Multi-party Meetings, in *Proc. ICASSP* (2015).

[Bouckaert 10] Bouckaert, R. R., et al.: WEKA–Experiences with a Java Open-Source Project, *Journal of Machine Learning Research*, Vol.11, pp.2533–2541 (2010).

[Keerthi 01] Keerthi, S. S., et al.: Improvements to Platt’s SMO Algorithm for SVM Classifier Design, *Neural Computation*, Vol.13, No.3, pp.637–649 (2001).