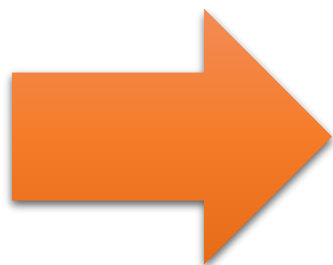


発話自動採点システム の研究と開発

近藤 悠介

発話評価の導入

- テストの実施・採点に時間がかかる。
- 評定者の訓練に時間がかかる。
- 訓練を施しても評価のブレは生じる。
- 信頼性・妥当性の高い複数の評定者の確保が困難。



自動採点

発話自動採点のイメージ



実用化されているもの

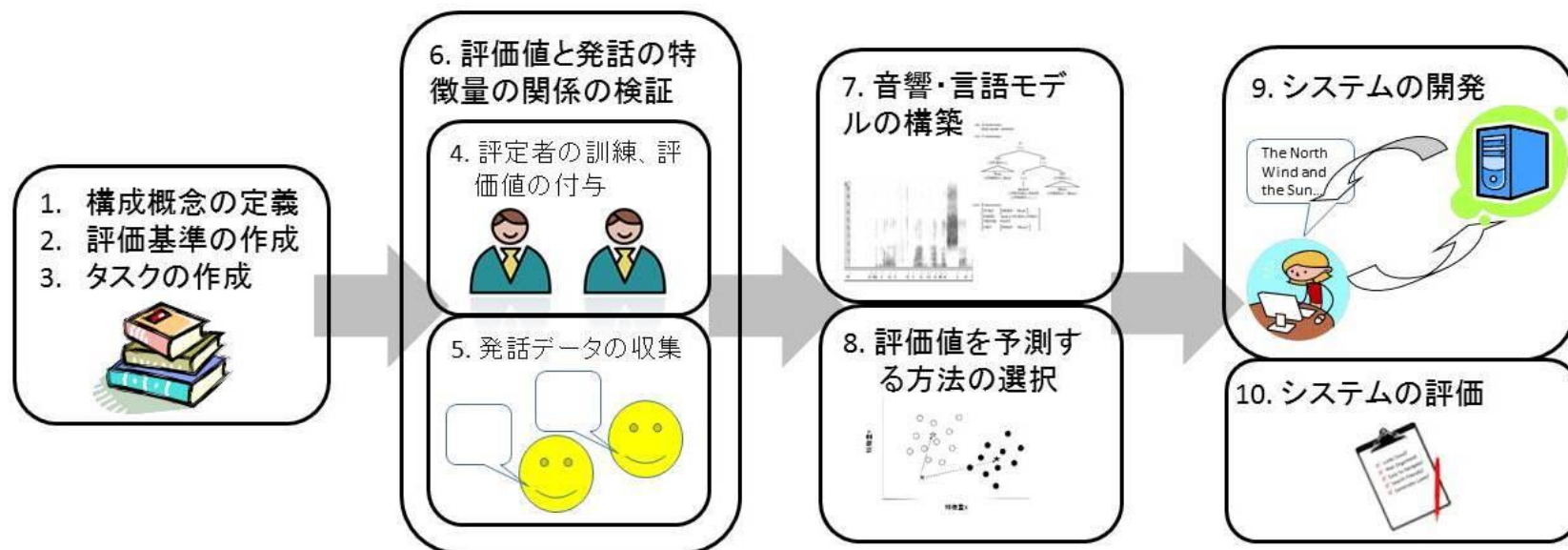
- SpeechRater

- 特定の話題に関する自由な発話を自動採点。
- 評定者との相関は.57-.70。
- 単語認識率は約50%。

- Versant

- 限定されたタスクを使用。
- 単語認識率は相当高い。
- 他のテストとの相関が高い。

システムの開発過程

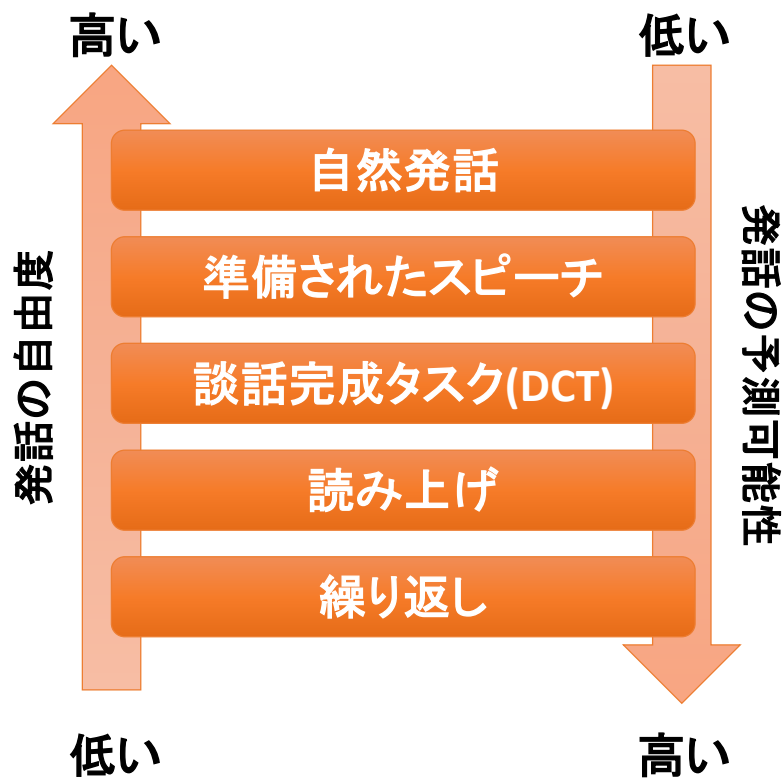


構成概念の定義、評価基準、タスクの作成

- 目的
- 言語使用の領域
- 発話誘出方法(タスク)
- 対象者
- 評価基準

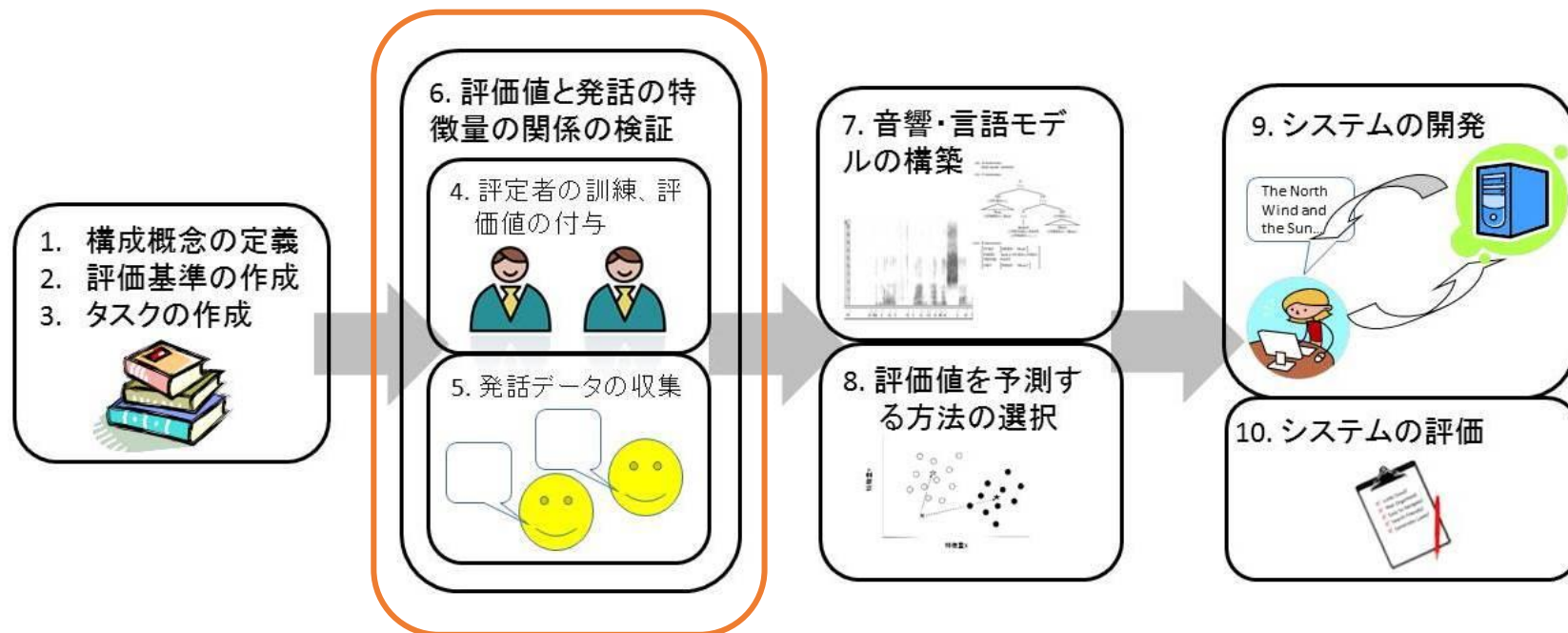
などが音声認識技術の精度によって制限されている。

タスクと音声認識



- 基本的に出現する単語の予想が難しくなると、認識の精度が落ちる。

システムの開発過程



評定者の信頼性

•一般化可能性理論による分析

	Estimated Variance (Before)	Estimated Variance (After)
e (examinee)	0.43	0.63
t (item)	0.51	0.08
r (rater)	0.12	0.11
et	0.28	0.08
er	0.08	0.25
tr	0.18	0.09
etr (residual)	0.49	0.41
Sum	2.08	1.66

大きいほど受検者の能力を
よく識別できている

大きいほど評価項目の難易
度にばらつきがある

大きいほど評定者間のばら
つきがある

大きいほど項目によって受
検者の順序が入れ替わる

大きいほど評定者によって
受検者の順序が替わる

Kondo (2010). *The development of automatic speech evaluation system for learners of English.*

評定者と自動採点の一致

Group	Correlation	% Exact Agreement	% Adjacent Agreement	Kappa	Weighted Kappa
Naïve	.77	41	89	.27	.75
Untrained	.61	31	73	.16	.59
Certificated (Average)	.92	70	99	.62	.91
Certificated (Exemplary)	.95	80	100	.76	.94

Powers, Escoffery, and Duchnowski (2015) *Applied Measurement in Education*

Untrained < Naïve < Certificated (Average) < Certificated (Exemplary)

評定者の信頼性

- クロンバックの α
- 一致の割合
- コーエンの κ
- スピアマンの順位相関
- ピアソンの相関係数
- 項目反応理論のInfitとOutfit
- 平均平方根偏差
- ポリコリック相関係数
- マクドナルドの ω
- 級内相関係数
- 相関係数の標準偏差
- 相関係数の平均

さまざまな指標があり、明確な基準はない。

評定者の信頼性

- 信頼性の程度
- 信頼性の指標

などを検討しなければならない

- これらをどのように受検者に知らせるか

発話の特徴量

特徴量	定義
フィラーの数	mm, ehhなどに代表される非語彙の挿入数
無音ポーズの長さ	100ms以上の無音の時間的長さ
Mean length of run	2つのフィラーあるいは無音ポーズの間の音節の数の平均値
言い直し	言い直しで挿入された音節の数
Pruned syllable per second	言い直し、フィラーを除いた音節の数を総発話時間で割ったもの
強勢音節と非強勢音節の比	非強勢音節の長さの平均値を強勢音節の長さの平均で割ったもの
文間のポーズの数	文の境界で置かれたフィラーおよび無音ポーズの数
フレーズ間のポーズの数	フレーズの境界で置かれたフィラーおよび無音ポーズの数
/i/と/ɪ/の識別率	F1およびF2による/i/と/ɪ/の識別率(判別分析による)
/u:/と/ʊ/の識別率	F1およびF2による/u:/と/ʊ/の識別率(判別分析による)
/æ/と/ʌ/の識別率	F1およびF2による/æ/と/ʌ/の識別率(判別分析による)

評価値と特徴量の関係

評定者による評価

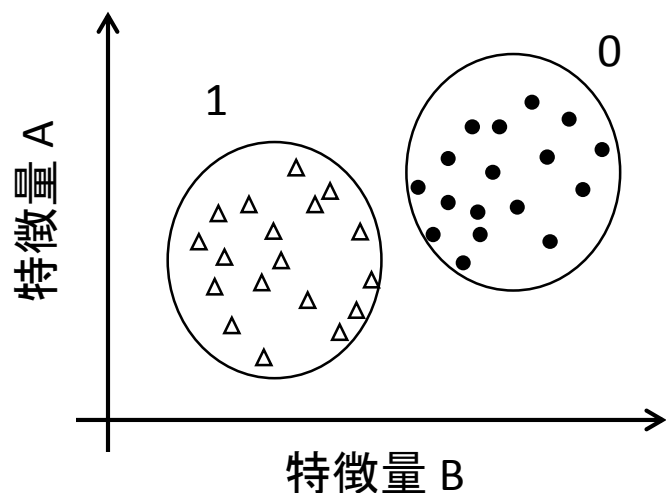
$$\hat{y} = x_1 + x_2 + \cdots + x_n$$

発話の特徴量

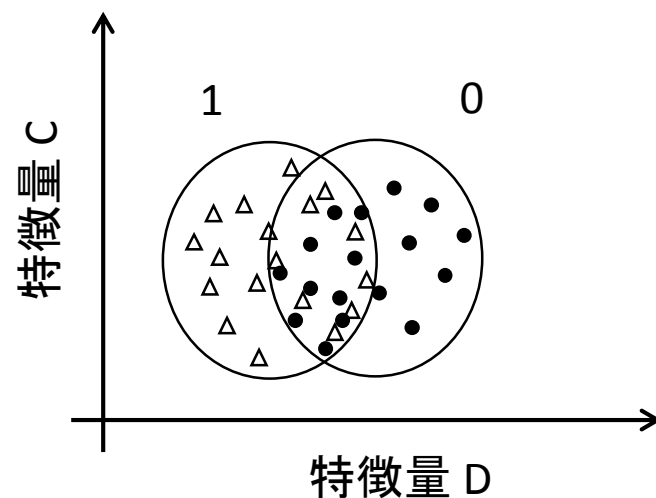
総語数、ピッチ幅、ポーズの頻度

評価値と特徴量の関係

良い特徴量

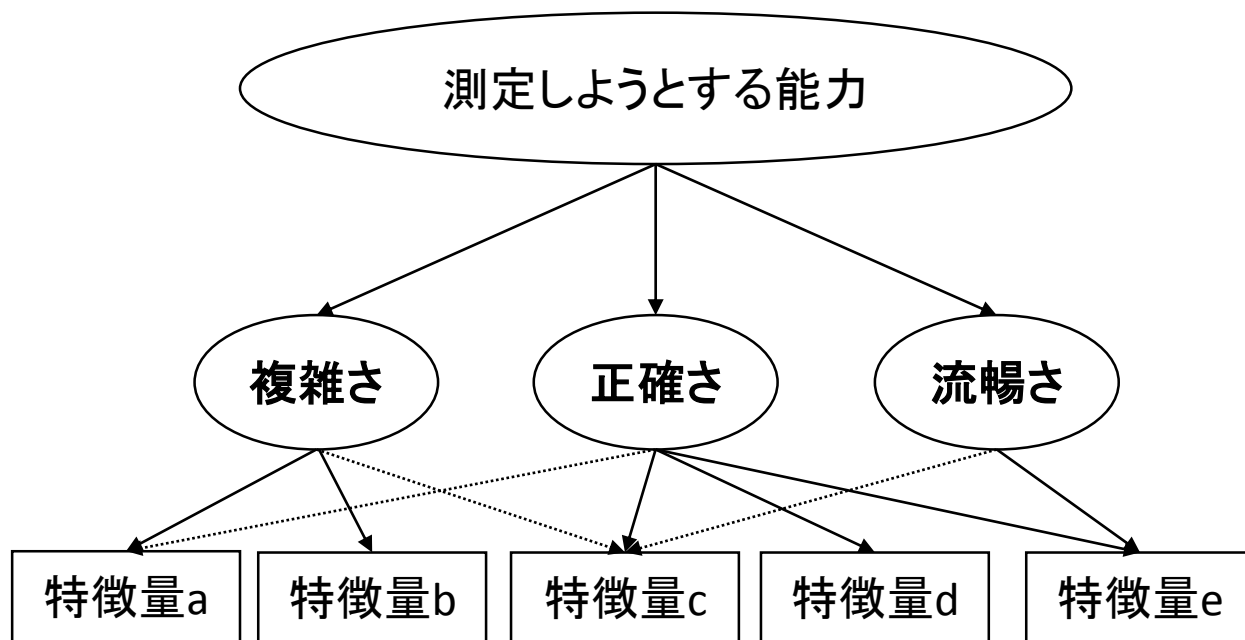


悪い特徴量



CAF研究

「ある指標がある構成概念の測定指標である」と分かれば、特徴量と評価の関係を検証することなく、特徴量のみでスコアを算出することが可能。



たとえば、Koizumi & In'nami. (2014). Modeling Complexity, Accuracy, and Fluency of Japanese Learners of English: A Structural Equation Modeling Approach

CAF研究

- 操作的定義が異なる。
- 独自の測定法。
- 指標の妥当性が不明。

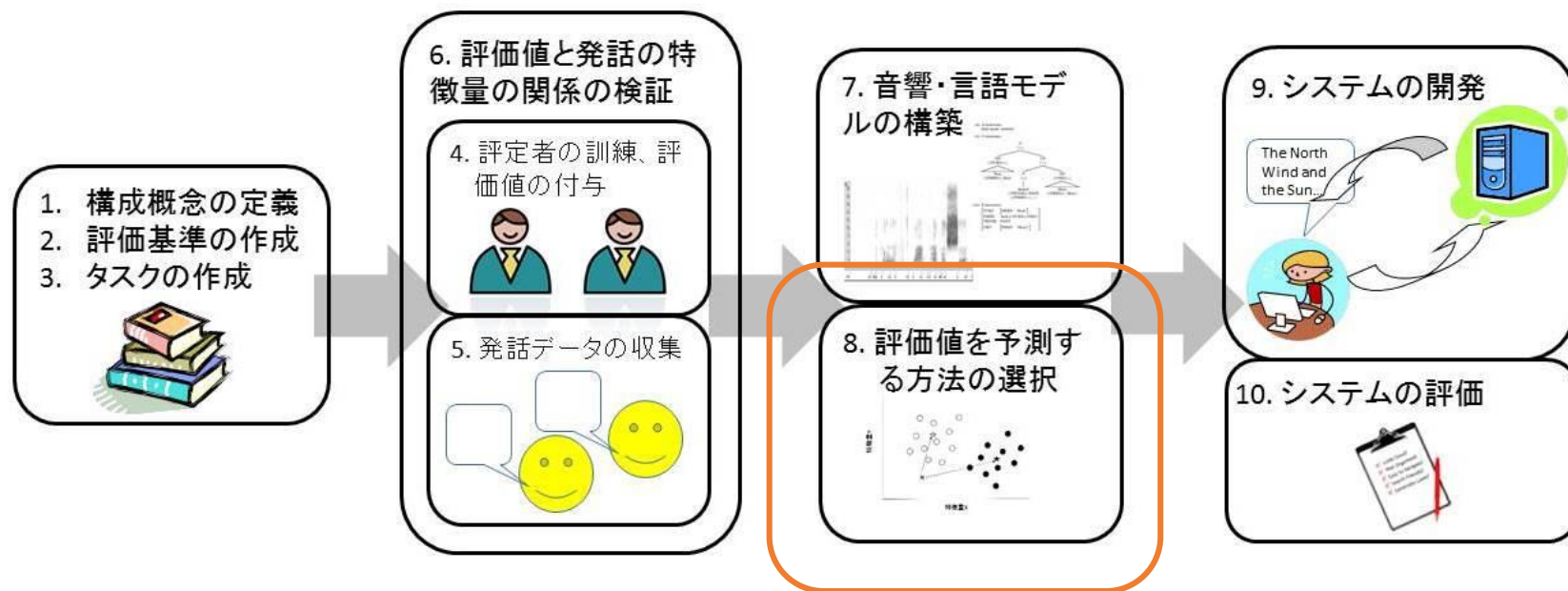
Ellis (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production.

発話の特徴量の選択

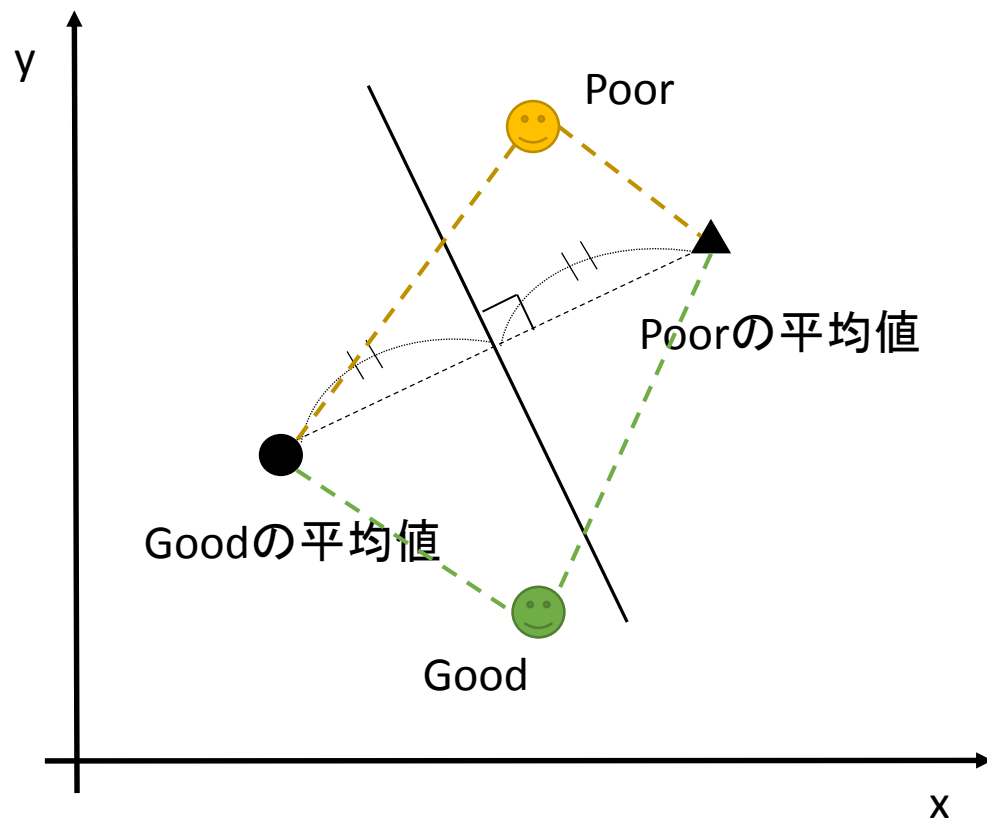
- 測定方法
- 特徴量の構成概念
- 評価との関係

などを検討しなければならない

システムの開発過程



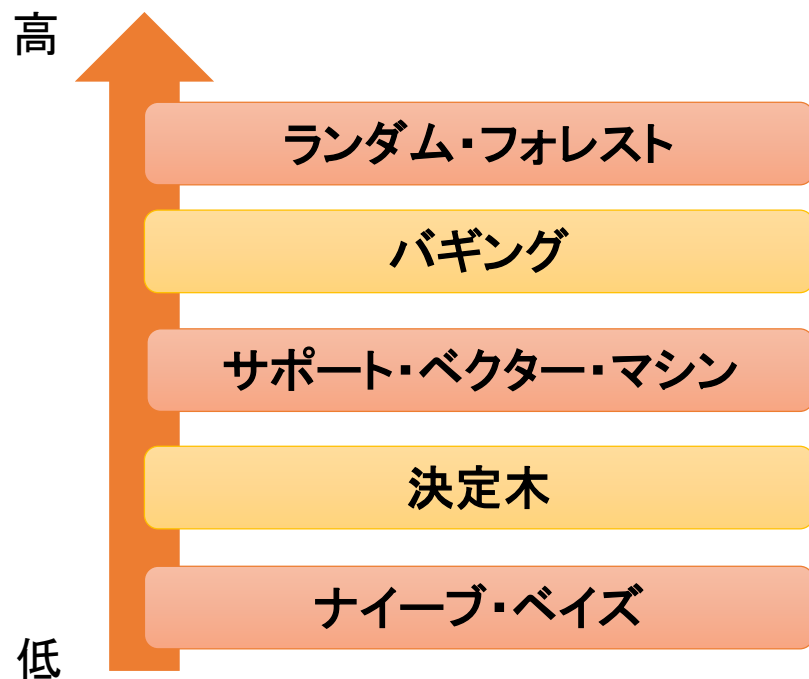
予測方法



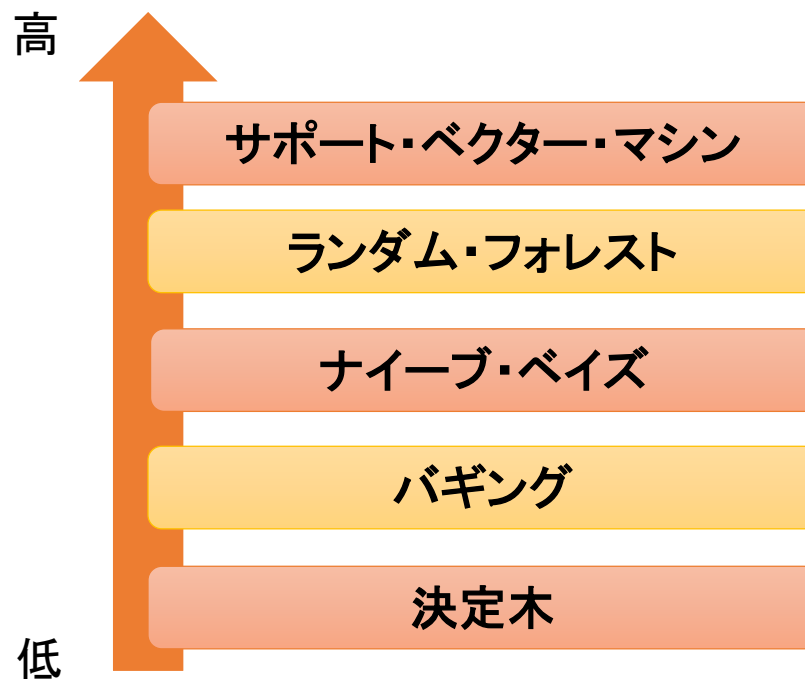
値が小さい方のカテゴリが評価となる。sdはその次元の標準偏差。

$$D(p, X) = \sqrt{\frac{(x_i - \bar{x})^2}{sd_x} + \frac{(y_i - \bar{y})^2}{sd_y}}$$

手法による予測精度の違い



小林, 田中, 富浦. (2011). メタ談話
標識を素性とするパターン認識を
用いた英語科学論文の質判定.



私の研究

「勿論、分類精度は用いたテキストや素性に依存するため絶対的なものではない」

回答 1/3

- 評価の信頼性をどの程度確保すべきなのか。
 - 検討する必要がある。
- 評定者訓練の効果はあるか。
 - 効果はあるが、解釈が難しい。

回答 2/3

- 音声認識技術によって得られる特徴量のみで測定したい能力が測定できるか。
 - テストの目的によっては可能。
- 音声認識によってタスクはどの程度制限されるか。
 - かなりの程度制限される。

回答 3/3

- 最適な予測方法は？

- 予測の手順も含め、検討する必要がある。

- 発話の特徴量をどのように選択するか。

- 個々の特徴量の測定方法など検討すべきことが多い。

参考文献

- 小林雄一郎・田中省作・富浦洋一 (2011年12月10日). 「メタ談話標識を素性とするパターン認識を用いた英語科学論文の質判定」 情報処理学会 人文科学とコンピュータ研究会「じんもんこん2011—「デジタル・アーカイブ」再考」(於 龍谷大学)
- Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, 30 (4), 474-509.
- Koizumi, R., & In'nami, Y. (2014). Modeling complexity, accuracy, and fluency of Japanese learners of English: A structural equation modeling approach. *JALT Journal*, 36, 25-46.
- Kondo, Y. (2010). *The development of automatic speech evaluation system for learners of English*. Unpublished doctoral dissertation, Waseda University, Tokyo, Japan.
- Zechner, K., Higgins, D., Xi, X., & Williamson, D. M. (2009). Automatic scoring of non-native spontaneous speech in tests of spoken English. *Speech Communication*, 51, 883-895.