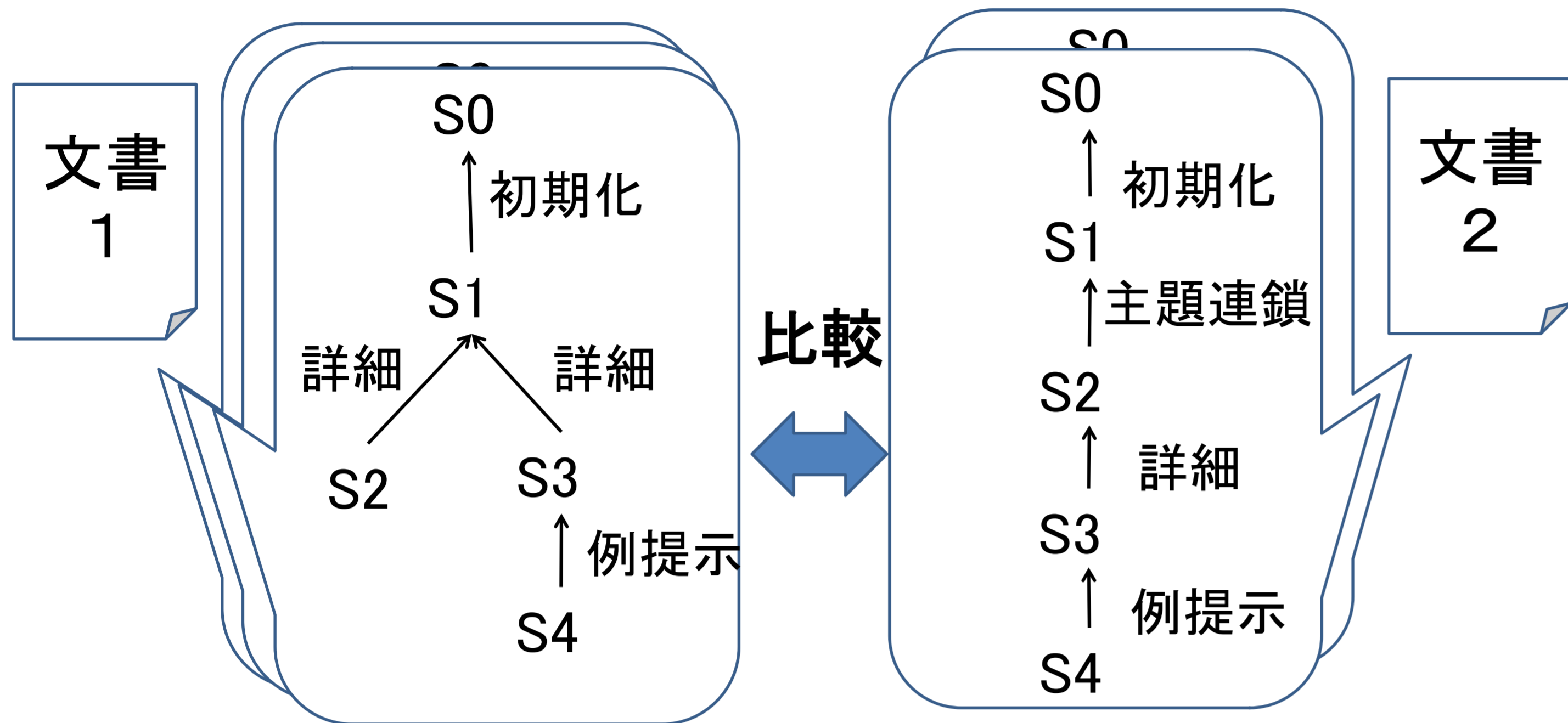


談話構造による文書の特徴付け

関西学院大学大学院 理工学研究科 佐藤瞳 金川絵利子 岡留剛

概要

文書を談話構造木で表し、文書同士の類似性を比較する



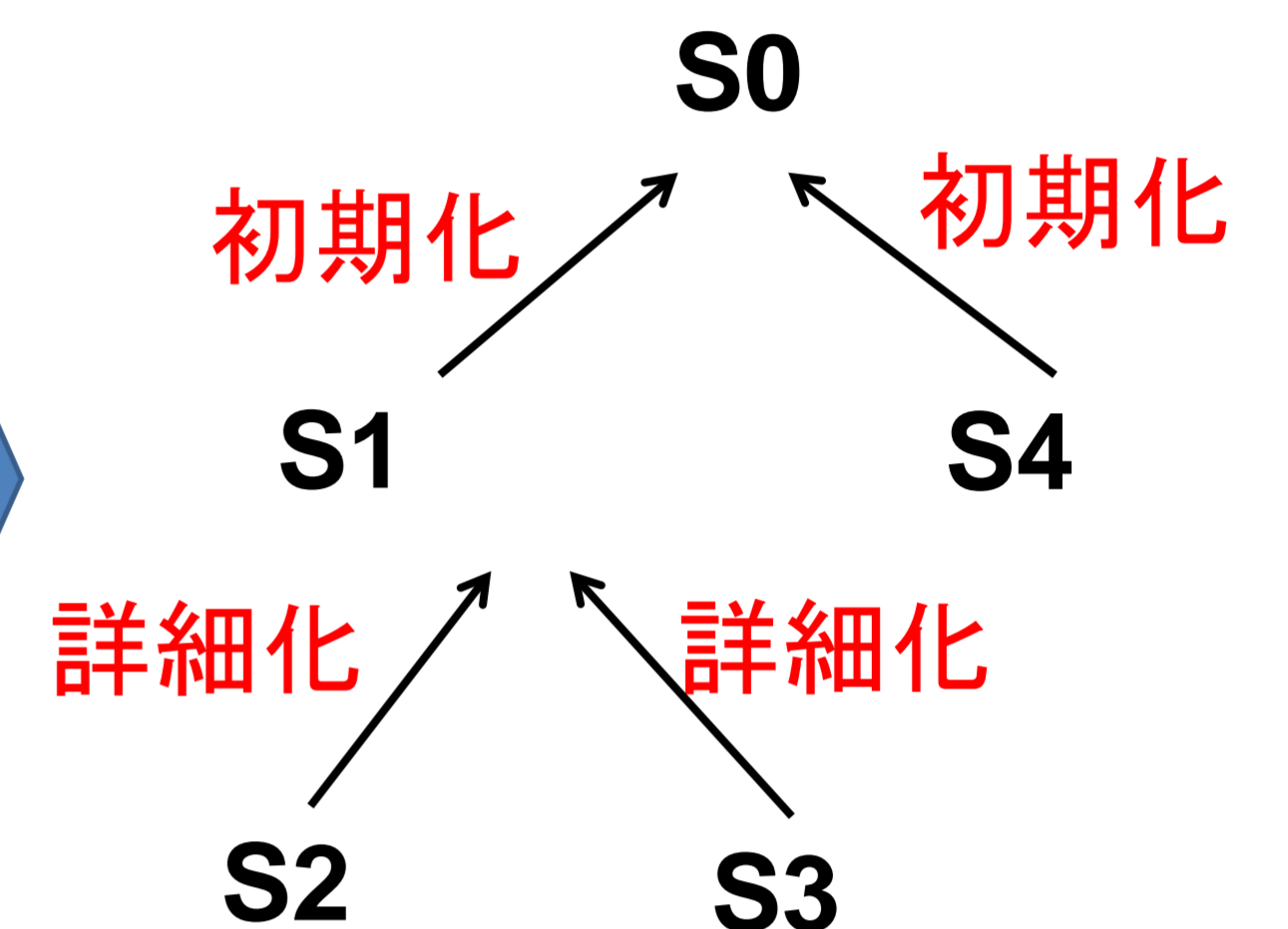
文書の生成・評価への利用

談話構造木

黒橋らの文章構造木(1994)をもとに設計構造

- ・基本単位は文
- ・1段落につき、1構造木
- ・初期ノードS0
- ・並列構造は木の形で表す

S1: 私は猫を飼っている。
S2: チャタという名前だ。
S3: 茶トラの猫である。
S4: 動物はかわいい。



文間関係ラベル

・10種(初期化を除く)

順序	理由
対比	原因-結果
主題連鎖	例提示
焦点-主題連鎖	質問
詳細化	定義

実験

対象

- ・2015年度人工知能学会全国大会発表資料
 - ・Web上で公開されている論文
- 計12文書

実験方法

- ・文書から人手で談話構造木を作成
- ・木カーネル値の総当たり平均
 1. 順序を考慮しない総当たり平均
 2. 文書を3分割し、それぞれにおいて総当たり平均

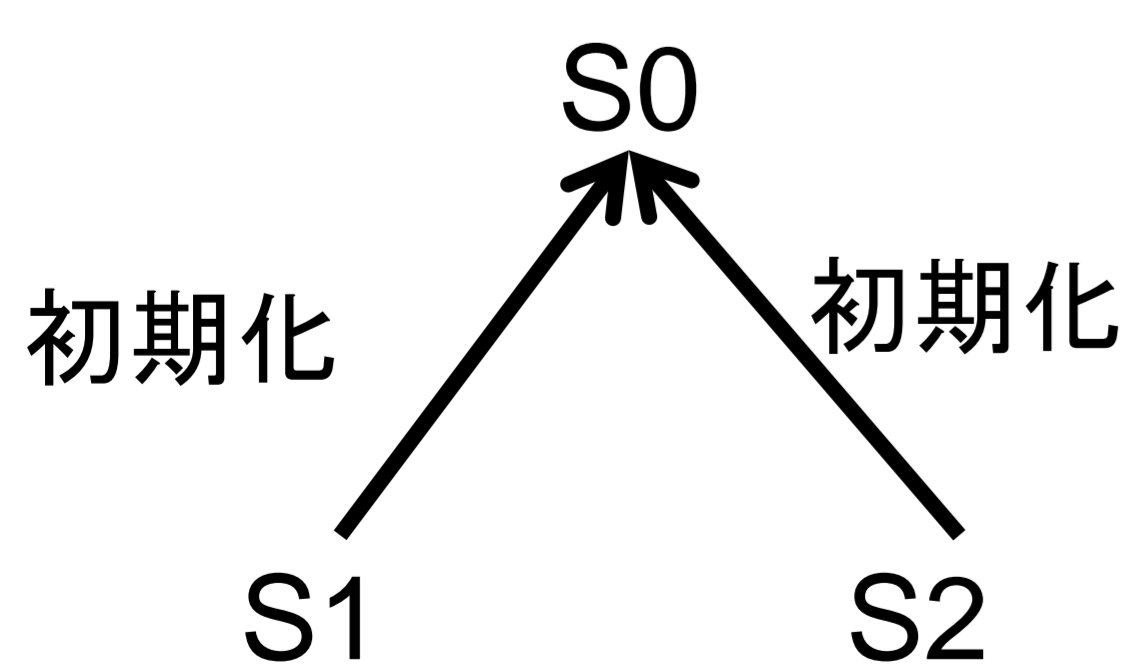
実験1

他文書との総当たり平均値の平均値

	1	2	3	4	5	6	7	8	9	10	11	12
平均	0.28	0.31	0.34	0.42	0.34	0.42	0.33	0.30	0.39	0.38	0.37	0.36

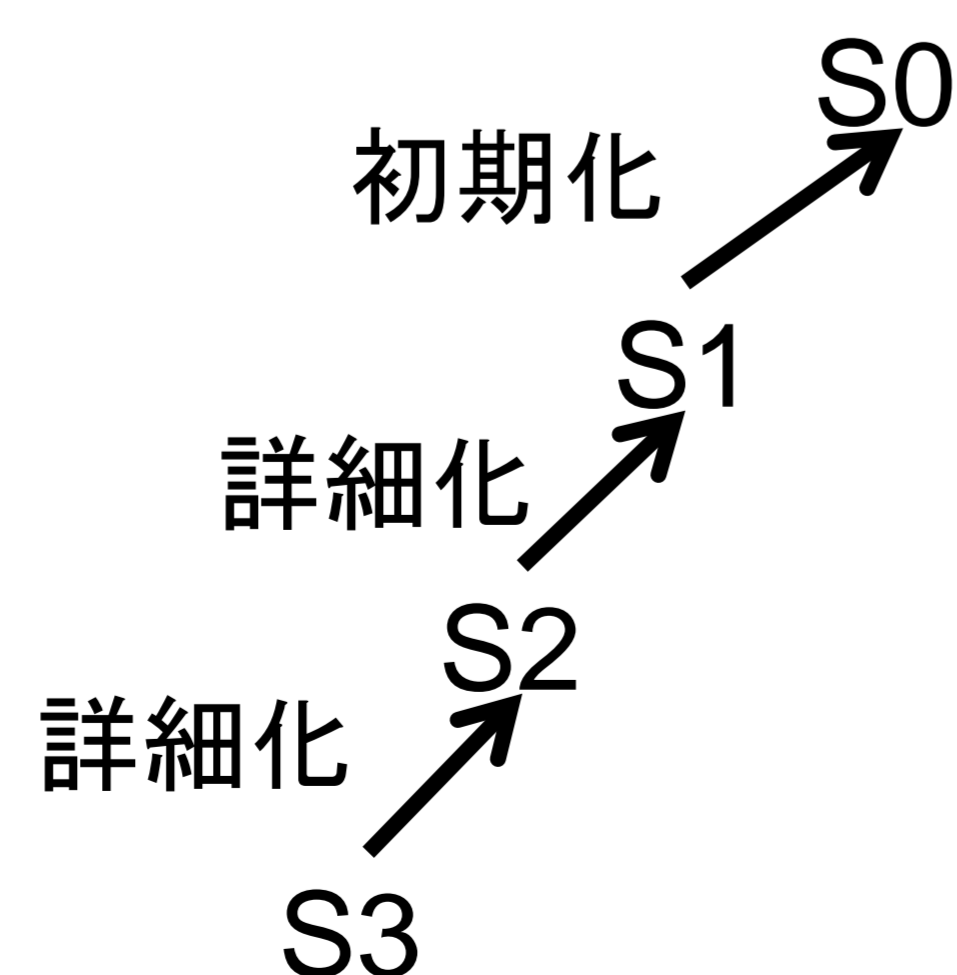
文書1

・初期化の並列構造が多い



文書4,6

・詳細化が多い



- ・詳細化関係により話題を展開させていく構造が一般的
- ・1文目が並列構造をとることは少ない

実験2

	1	2	3	4	5	6	7	8	9	10	11	12
はじめに	0.33	0.45	0.42	0.42	0.28	0.35	0.29	0.34	0.34	0.28	0.29	0.40
関連研究等	0.25	0.29	0.32	0.43	0.35	0.44	0.33	0.29	0.40	0.40	0.38	0.35
まとめ	0.32	0.27	0.32	0.12	0.39	0.39	0.42	0.33	0.24	0.20	0.42	0.33
平均	0.30	0.34	0.35	0.32	0.34	0.39	0.35	0.32	0.33	0.29	0.36	0.36
分割なし	0.28	0.31	0.34	0.42	0.34	0.42	0.33	0.30	0.39	0.38	0.37	0.36

・各部分の値に強く影響される

文書を談話構造木で表すことにより、文書の構造的な特徴を捉えられることを確認