

2011年度 学生研究奨励賞

ニュースサイトにおける大学生の 嗜好要因の特定とサイト制作への応用

大木 基至

大阪大学大学院 基礎工学研究科 システム創成専攻

01. 背景と目的
02. 既存のニュースサイトの分類
03. 大学生ユーザー層の分類
04. クラスタ別嗜好要因の抽出
05. コンセプト決定とWebサイトの提案
06. まとめ

大学生に対する社会問題として大学生の政治離れ・選挙率の低下（図1）や学力低下[1]が危惧される。

学生の新聞または政治に対する情報への関心度が低くなる一方。

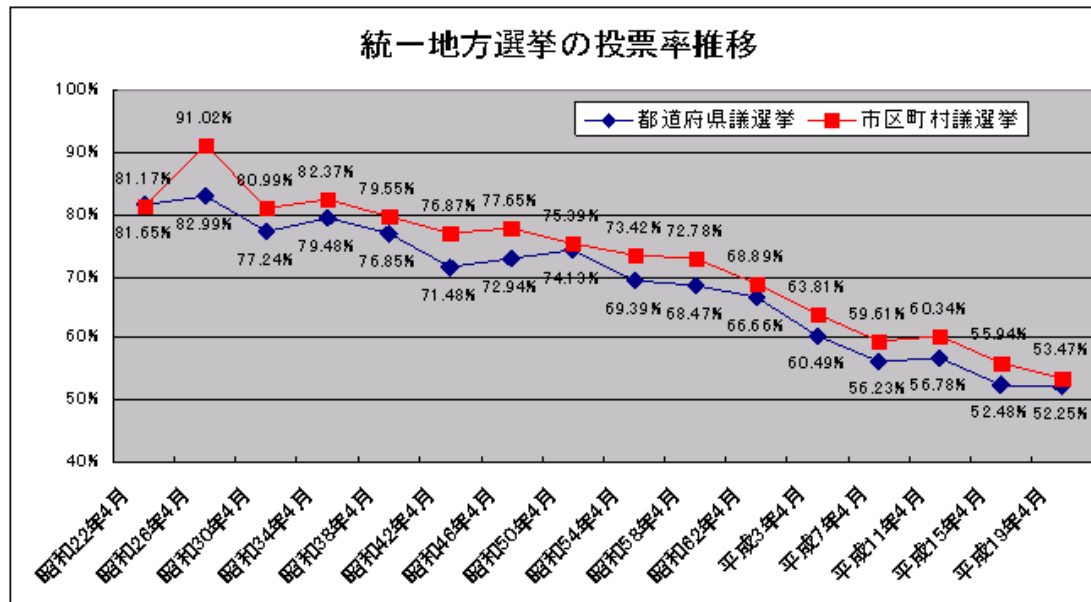
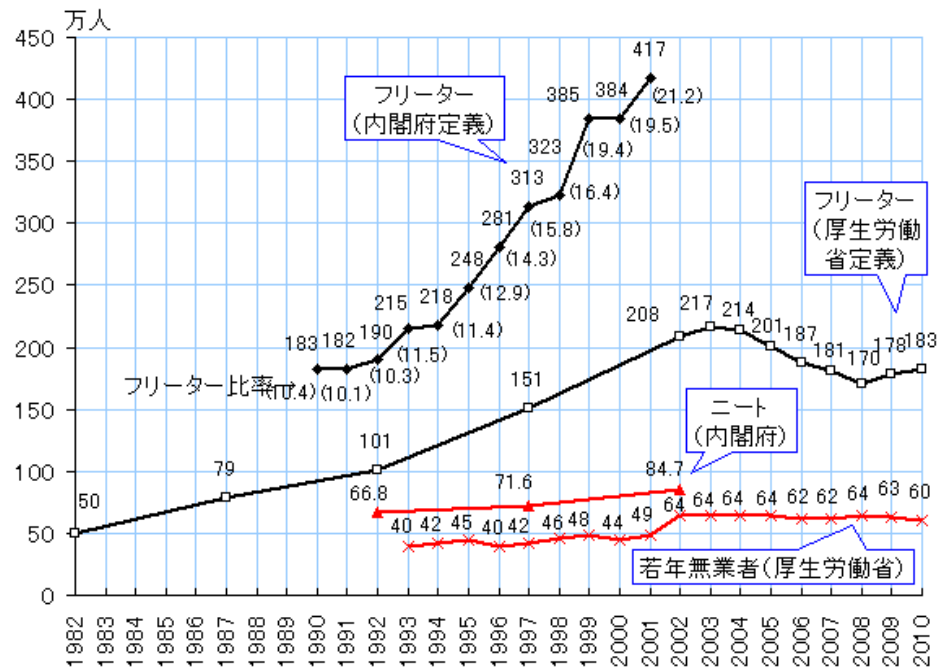


図1：選挙率の低下（「議会内閣制」の提案にあたって：
http://www.pref.osaka.jp/koho/chiji/220825_gikainaikaku.html

[1]深刻な大学生の学力低下 教員の6割問題視 Bennesse教育情報サイト：<http://benesse.jp/blog/20051201/p1.html>

01. 背景

また，大学生の就職内定率の低下が進み[1]，それに伴いフリーター数が増加している[2]。大学生たちの社会進出を促進するためにも，社会の変化や社会を知る事に敏感になることが必要である。



[1]時事ドットコム 大学生の就職率と就職内定率
: http://www.jiji.com/jc/v?p=ve_soc_tyosa-koyou-college-naitei

[2]フリーター数・ニート数の推移:
<http://www2.ttcn.ne.jp/honkawa/3450.html>

大学生の政治離れ・学力低下問題への改善が必要がある。また、大学生により多くの情報を効率よく知ってもらう必要がある。そこで本研究では、ニュースサイトを通じて大学生への政治への関心・就職活動の情報発信を行い、大学生のサポートを行うことを目的とする。

大学生のためのニュースサイトの構築

調査課題

「大学生が好むサイトとは？」を明確にする必要がある。そこで本研究では、アンケート調査によるデータからいくつかの多変量解析を行い、サイトのグループ化、大学生の分類、嗜好に寄与する要因の抽出を行う。

02. 既存のニュースサイトの分析

目的

既存のニュースサイトを収集し、既存のサイトの特徴、全体像を把握する。

-調査対象：48種類のニュースサイト（補遺1）

-収集方法：Google検索エンジンで「ニュース」という関連ワードから検索を行い、選出。

選出した48種類のニュースサイトの例



aolニュース
(トップ画面)



izaニュース
(トップ画面)

収集したニュースサイトをサンプルとして**数量化理論第Ⅲ類**による分析を行う。結果としてカテゴリマップ・プロダクトマップが作成され、既存のニュースサイトの特徴・全体像を把握することができる。

数量化理論第Ⅲ類とは？

- 林知己夫によって提唱された数量化理論の1つ。
- 対象を表すサンプルと対象の属性を表すアイテム・カテゴリで構成された属性表から分析を行う。アイテム・カテゴリは質的データを用いる。
- 属性表から固有値問題に定式化し、算出された第1, 2固有ベクトルを用いてサンプルとカテゴリを2次元平面上にプロットする。

ニュースサイトを構成するアイテム・カテゴリを抽出し、属性表を作成する。その後、ソフトウェア「エクセル統計2010 [1]」を用いて、数量化理論第Ⅲ類による分析を行う。

抽出したアイテムとカテゴリ（補遺 2）

- 広告の掲載がある or なし

- 画像の量

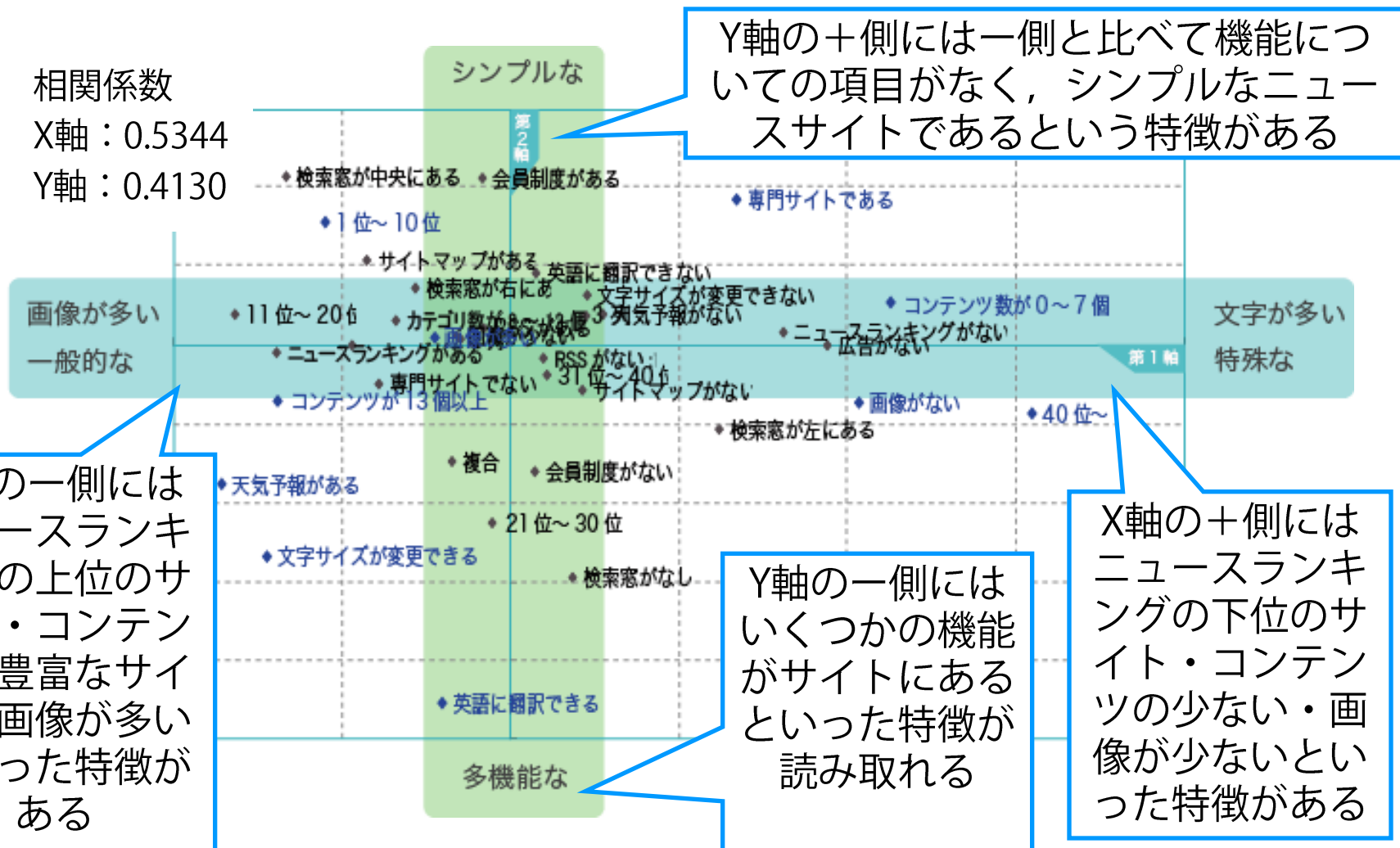
- レイアウトの構成

- 文字サイズ変更機能のある or なし

など。特に機能・TOPページの外見から把握できるカテゴリを抽出した。

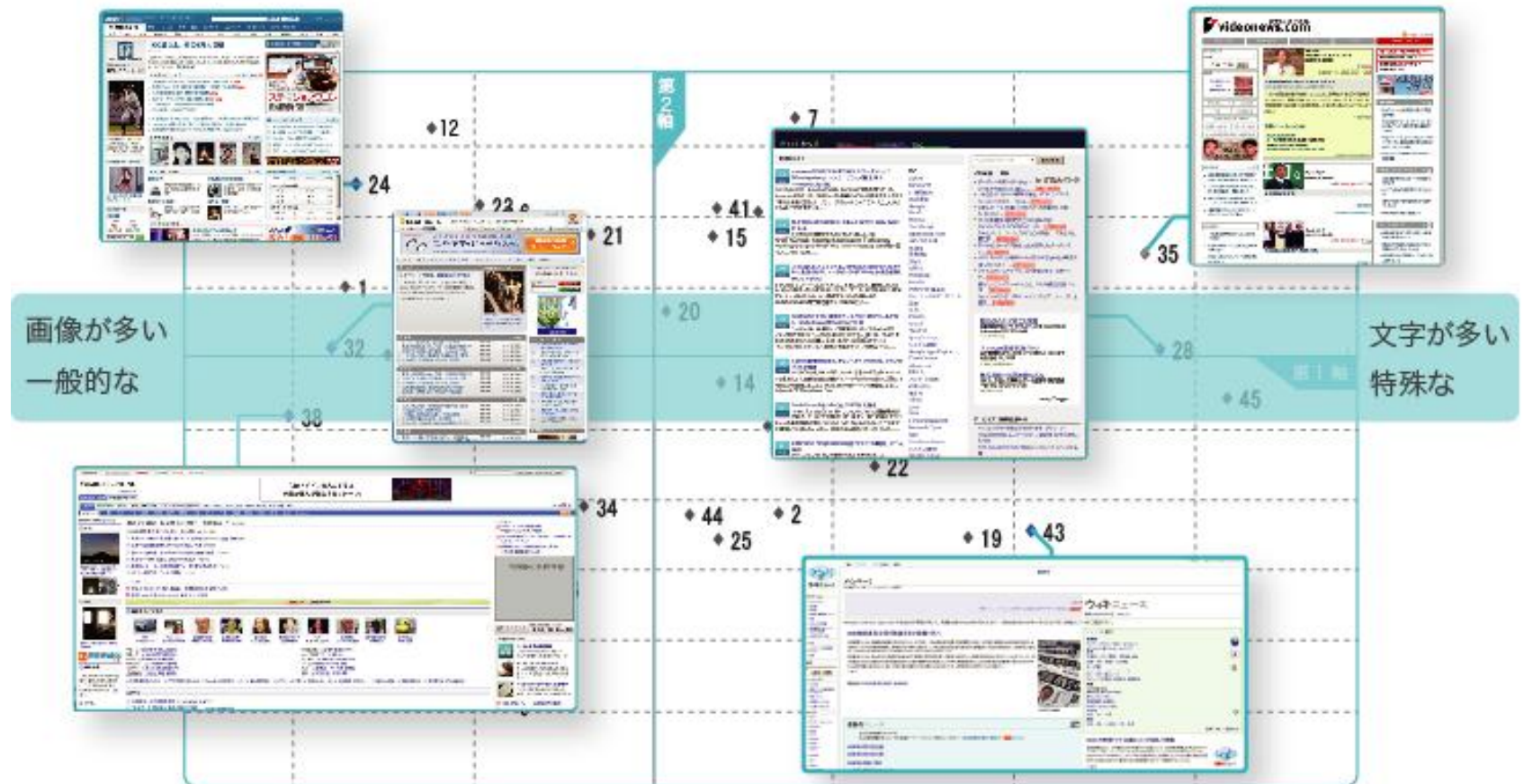
ニュースサイトのカテゴリマップ

相関係数
X軸：0.5344
Y軸：0.4130



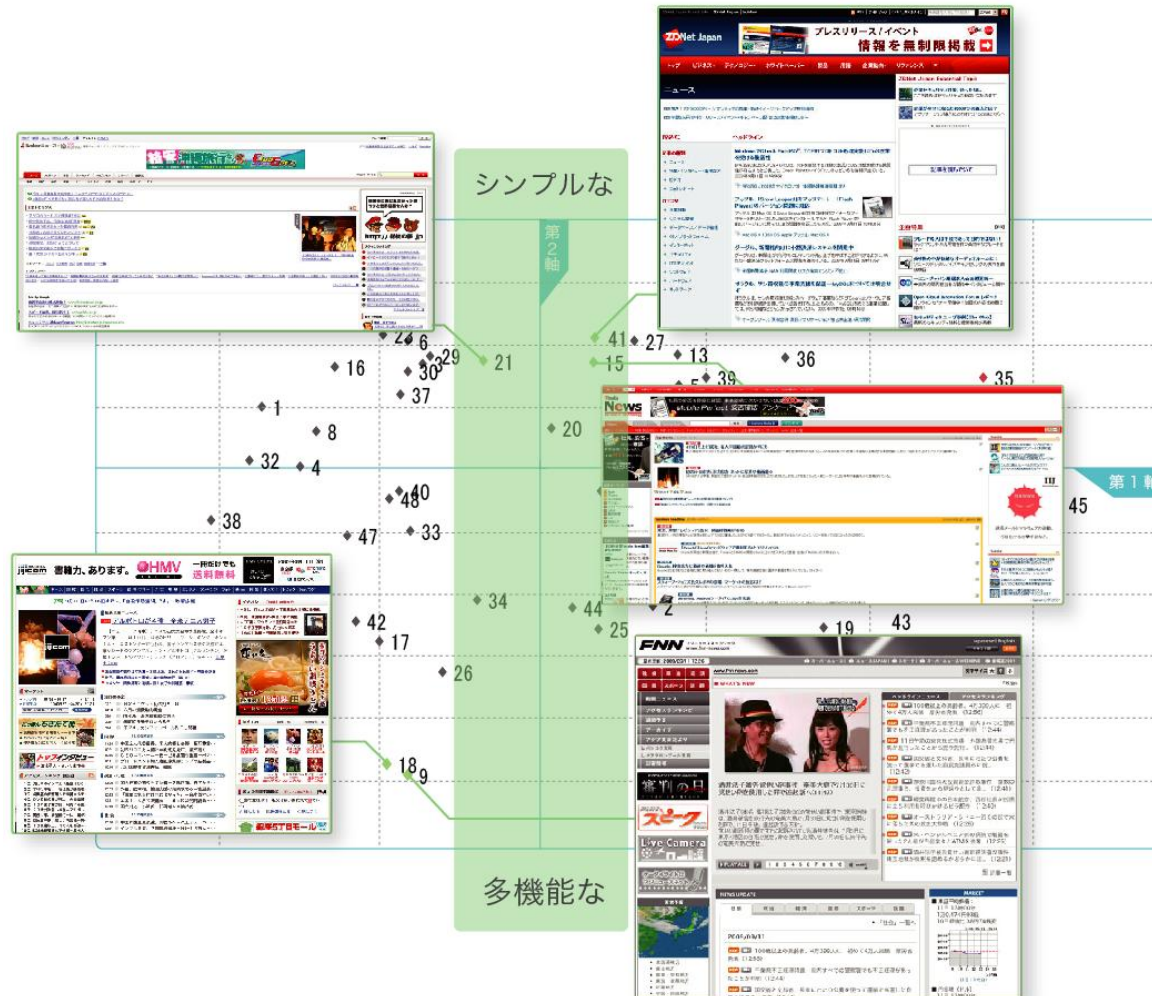
02. 既存のニュースサイトの分析

ニュースサイトのプロダクトマップ①



02. 既存のニュースサイトの分析

ニュースサイトのプロダクトマップ②



プロダクトマップの作成後，ニュースサイトのグルーピングを行うため，**クラスター分析**を行う。

クラスター分析とは？

階層的手法（ワード法や群平均法など）

与えられたデータセットの各データが1つのクラスタとなっている状態から，クラスタ間の距離や類似度に基づいて2つのクラスタを逐次的に併合してゆく手法

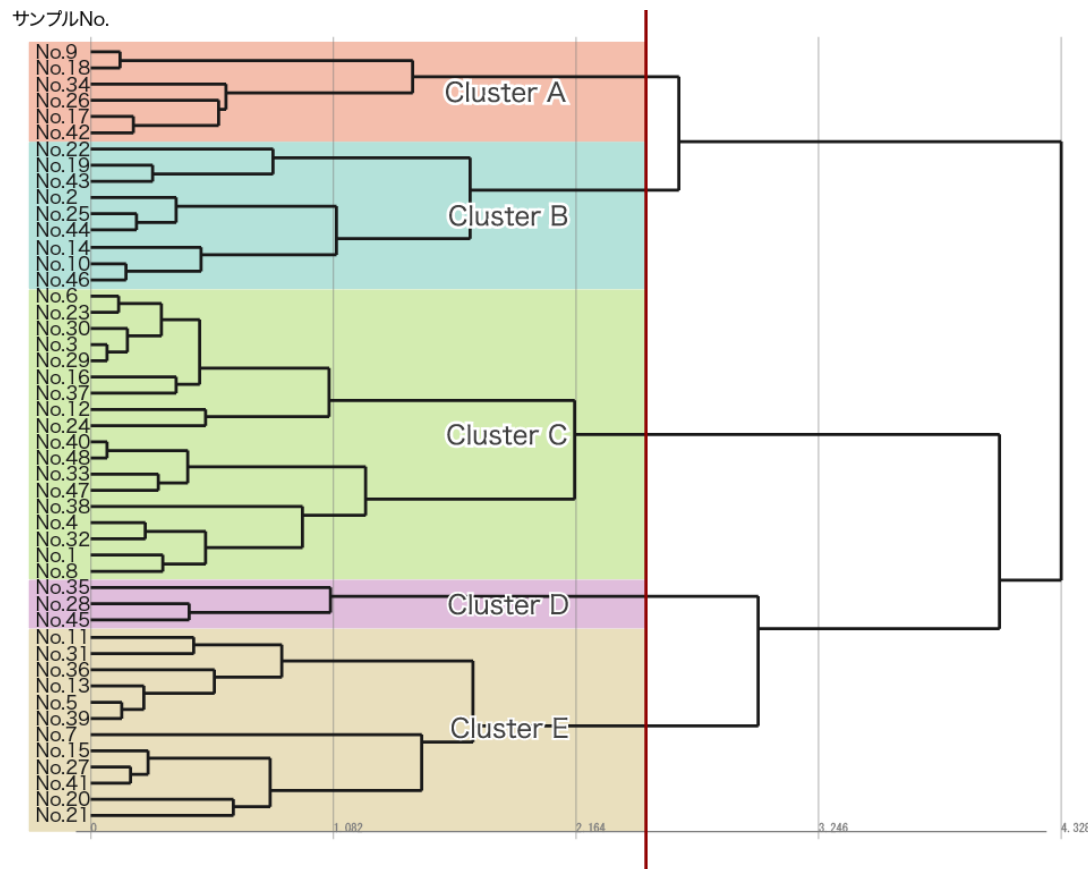
非階層的手法（K-means法など）

データの分割の良さを表すある評価関数を設定し，その評価関数に対する最適解を探索することでクラスタリングを行う手法

⇒本研究では，クラスタ間の結合の過程を把握し5つのグループに分類するため，階層的手法によりニュースサイトのグルーピングを行う。

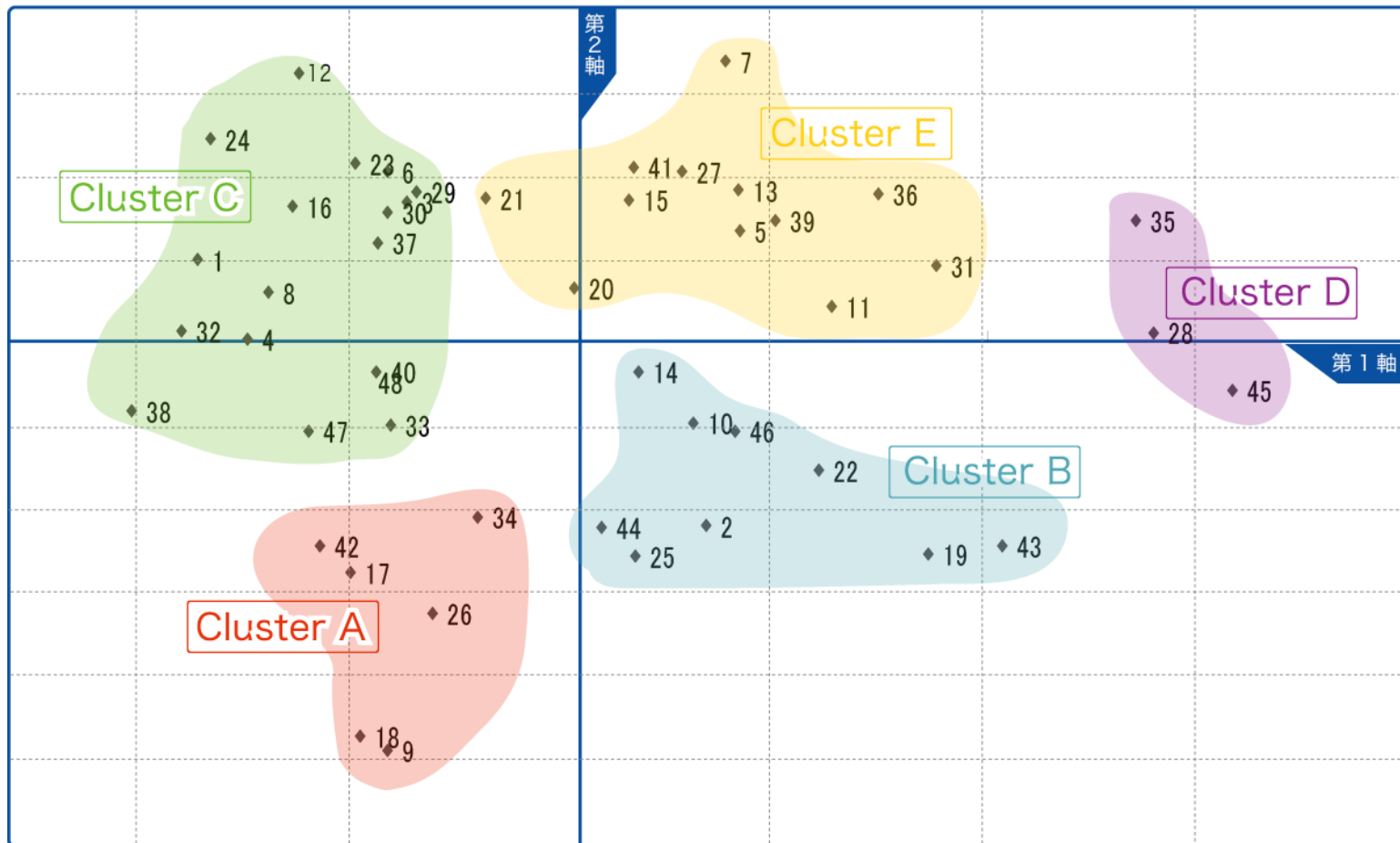
クラスタリング結果：デンドログラムの作成

数量化理論第Ⅲ類により算出された第1, 第2固有ベクトルを用いて, 階層的的手法(群平均法)を適用し, ニュースサイトのクラスタリングを行った。



02. 既存のニュースサイトの分析

クラスタリング結果：数量化理論Ⅲ類との統合



02. 既存のニュースサイトの分析

クラスタA：機能充実型クラスタ

- 他にない機能が豊富，やや情報が多く窮屈な印象を受ける



TBS News-i
ニュース



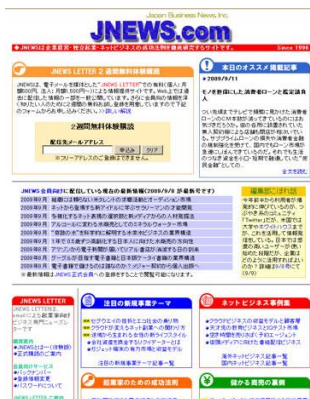
Jiji.com
ニュース



FNN
ニュース

クラスタB：バランス型(文字中心)クラスタ

- 文字多く寂しい印象，色によって情報を構造化している



JNEWS
.com



MBS
ニュース



朝日ニュー
スター

02. 既存のニュースサイトの分析

クラスタc: バランス型(画像中心)クラスタ

- アクセスが多い一般的なサイト, バランスの良い情報を集めている



MSN産経
ニュース



MBS
ニュース



mixi
ニュース

クラスタD: 特殊クラスタ

- 文字が多く, シンプルで飾り気の無い印象, 他サイトに無い特殊なコンテンツ



Publickey
ニュース



Videonews
.com

クラスタE：専門クラスタ

- 1つのコンテンツを扱っているサイトが多い，専門的な情報を多く扱っている



WebBCニュース



Gigazineニュース



ZDNetニュース

数量化理論第Ⅲ類とクラスター分析により，48種のサイトを分類し，5つのクラスタの特徴を把握する事ができた。

目的

ターゲット層である大学生が好むニュースサイトを分析するため、思考・普段の生活習慣に回答するアンケート調査から大学生を3つのグループに分類する。

- 調査対象：大学生(18歳～24歳, 50名)

アンケート9項目+ 自由回答設問2項目(P.~)

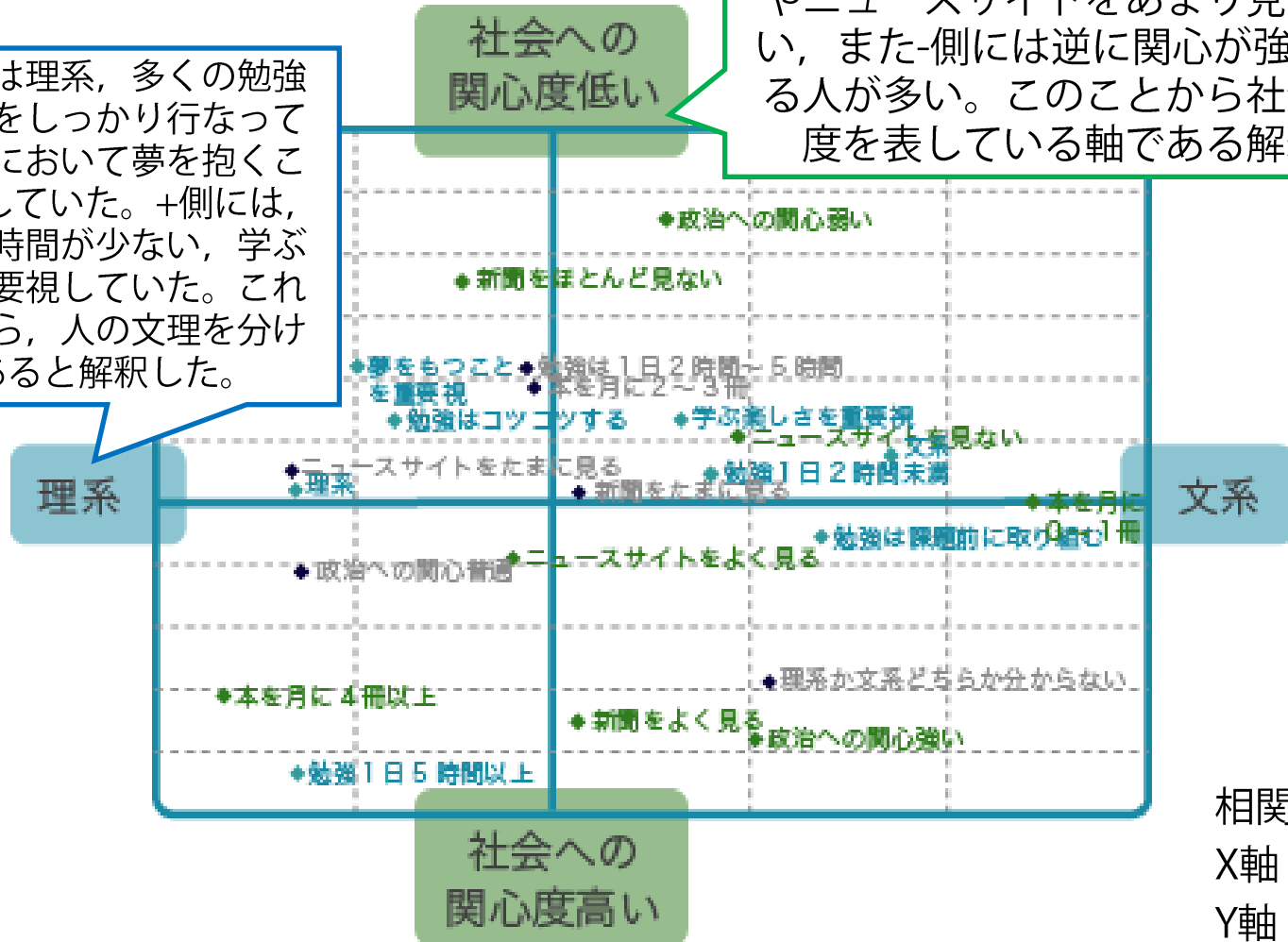
- ①性別, ②文理, ③一月に読む本の数, ④新聞を読む頻度
- ⑤政治への関心度, ⑥ニュースサイトを見る頻度
- ⑦1日の勉強時間, ⑧勉強への取り組み方(毎日行う or 提出前にまとめて行う)
- ⑨勉強をするために重要視すること(夢を持つ事を重要視 or 楽しさを重要視)

03. 大学生ユーザ層の分類

作成されたカテゴリマップ

X軸の-側には理系，多くの勉強時間の確保をしっかりと行っており，勉強において夢を抱くことを重要視していた。+側には，文系，勉強時間が少ない，学ぶ楽しさを重要視していた。これらのことから，人の文理を分ける軸であると解釈した。

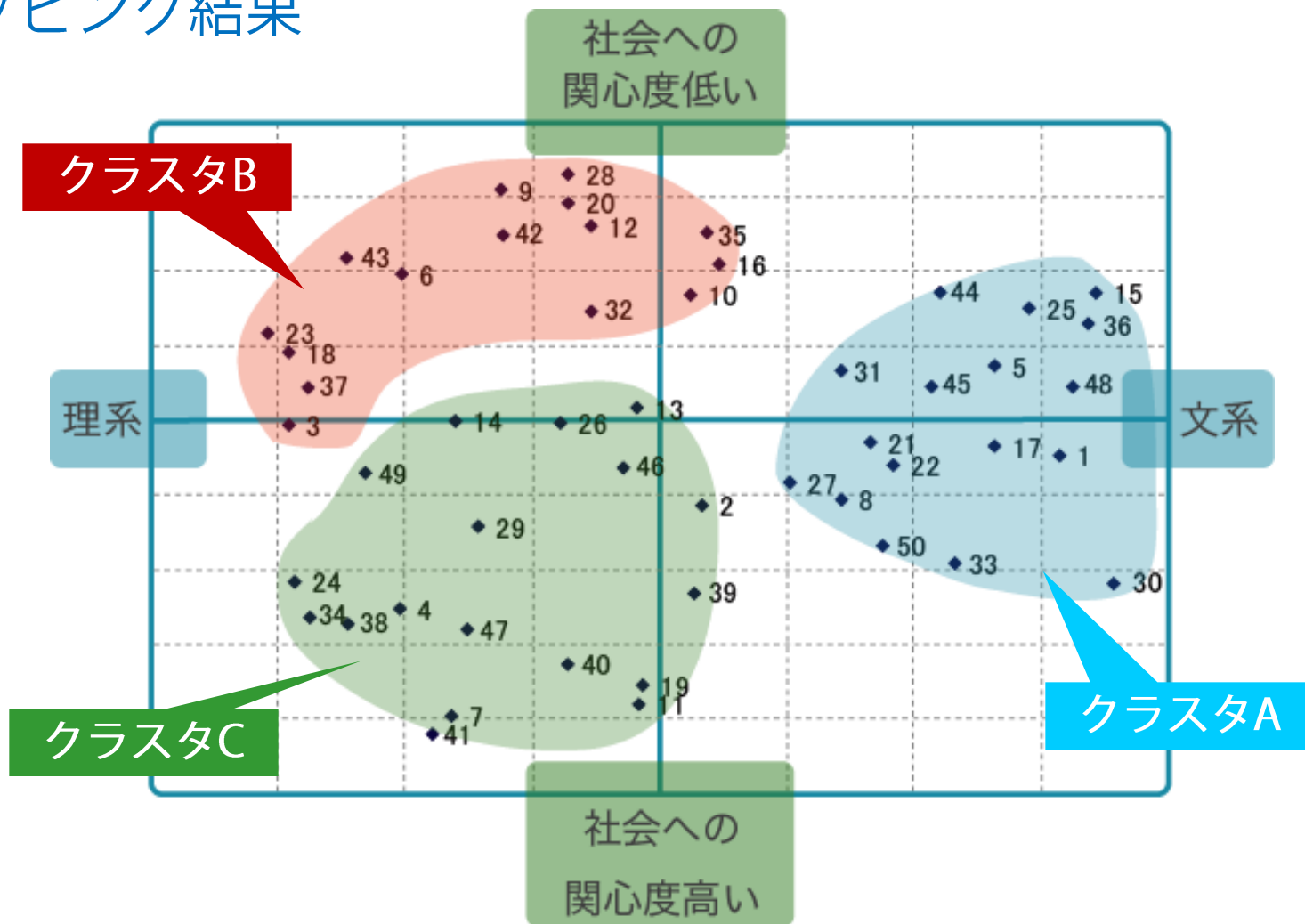
Y軸の+側には政治への関心が弱い，新聞やニュースサイトをあまり見ない人が多い，また-側には逆に関心が強く，よく見る人が多い。このことから社会への関心度を表している軸であると解釈した。



相関係数
X軸：0.5720
Y軸：0.5007

03. 大学生ユーザ層の分類

数量化理論第Ⅲ類とクラスター分析による大学生の
マッピング結果



クラスタA：文系型クラスタ

文系の大学生が多い，社会への関心度は様々，勉強はやる時に集中して取り組む，学ぶ楽しさを重要視

クラスタB：理系型クラスタ

理系の大学生が多く，社会への関心度は低い，勉強はコツコツする，夢を持つことを重要視

クラスタc：社会への関心度が高いクラスタ

理系の大学生が多く，社会への関心度は高い，多くの勉強時間の確保・政治への関心度も十分高い。

目的

3つのクラスタに分類した各被験者クラスタそれぞれが好むWebサイトのデザイン・特徴を明らかにする。

調査方法

- ①自由記述回答形式の文章(計50個)から各クラスタの特徴語をテキストマイニングにより把握する。
⇒ Text Mining Studioによる**ことばネットワーク**
- ②5つのクラスタに分類した各Webサイトについて6つの評価尺度を用いて評価を行ってもらい、嗜好に寄与する要因を特定する。
⇒ **重回帰分析**

①ことばネットワーク

大学生ユーザ層の分類のためのアンケート調査によって収集した以下の自由回答欄項目のアンケートデータを用いてことばネットワークを行い、3つのクラスタが理想とするWebサイトの特徴の全体像を把握する。本研究では、質問項目2について分析を行った。

- ことばネットワーク動作：共起関係(1回以上)
- 抽出することば：話題一般

質問項目

1. あなたがよく見るニュースサイトの特徴を
できる限り記入してください。
2. あなたの理想のニュースサイトとは何ですか？
具体的なコンテンツ・デザインなどをできる限り記入してください。

クラスタA（文系型クラスタ）の回答例

質問1：「いろんな情報が掲載している。」「このサイトでしか扱っていないニュースがある。」...

質問2：「楽しく経済や政治のことを知れる」「ゲームをしながら政治を学べる」...

クラスタB（理系型クラスタ）の回答例

質問1：「コンピュータの記事が中心のサイト」「レイアウトがシンプル」

質問2. 「カスタマイズ性のあるサイト」「最新技術の特集がある」...

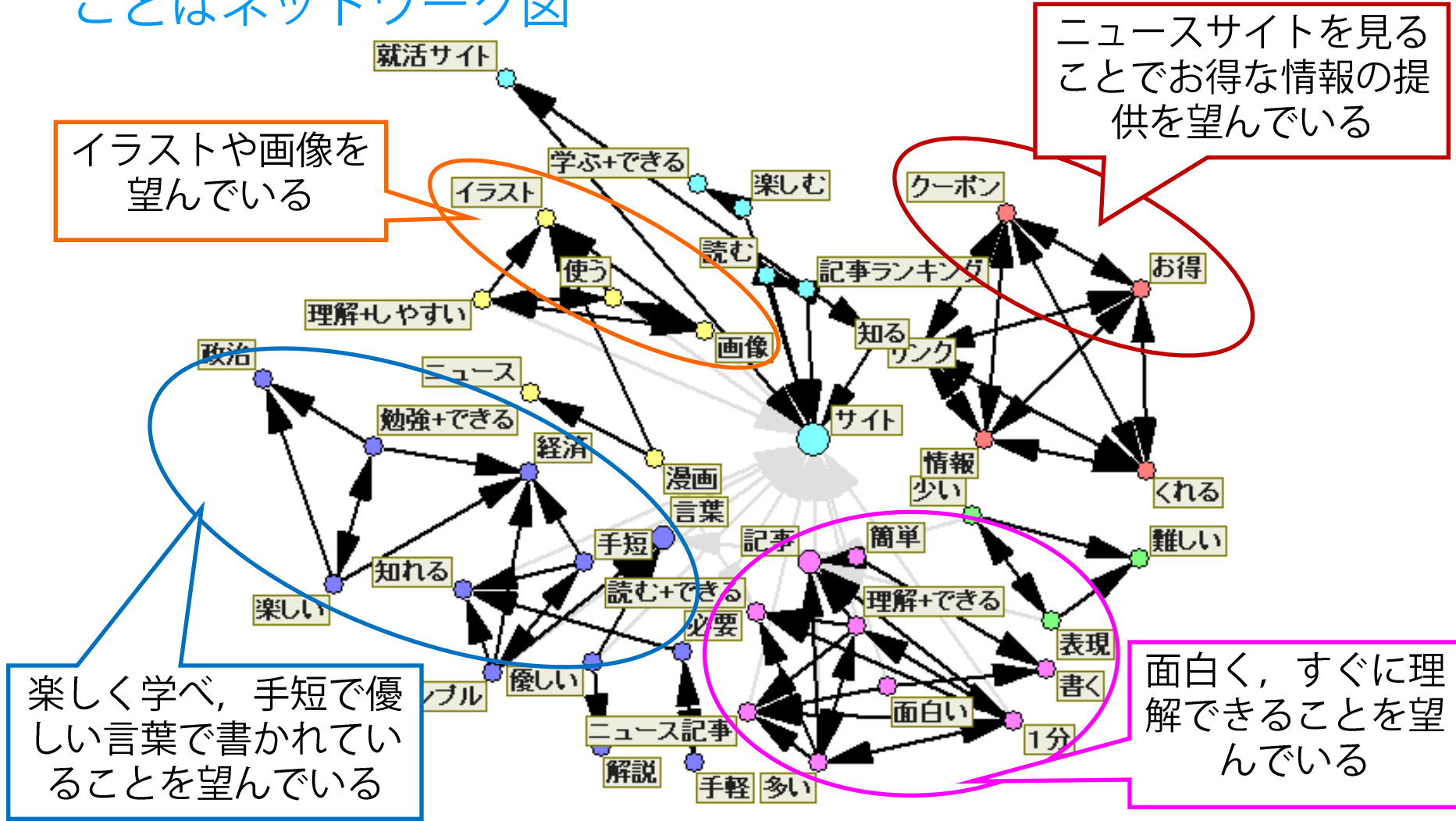
クラスタC（社会への関心度が高いクラスタ）の回答例

質問1. 「レポーターの深い考察がある」「ニュース記事がビジネス向け」

質問2. 「自分の好きな記事だけを推薦してくれるサイト」
「まとめて必要な情報を知れるサイト」...

04. クラスター別嗜好要因の抽出

クラスターA（文系型クラスター）の質問2の回答についてのことばネットワーク図



イラストや画像を望んでいる

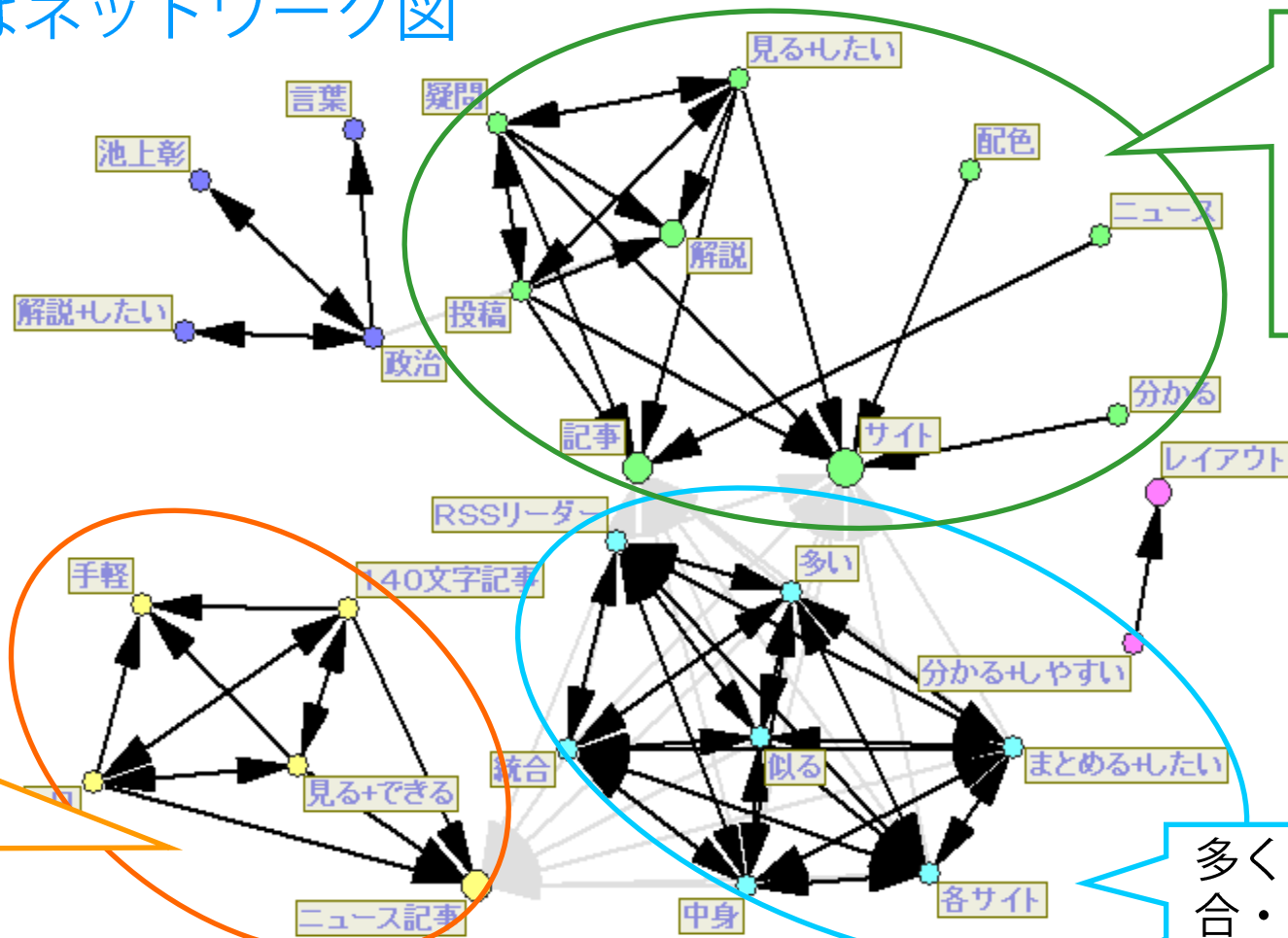
ニュースサイトを見ることでお得な情報の提供を望んでいる

楽しく学べ、手短で優しい言葉で書かれていることを望んでいる

面白く、すぐに理解できることを望んでいる

04. クラスタ別嗜好要因の抽出

クラスタB（理系型クラスタ）の質問2の回答についてのことばネットワーク図



疑問点を解説してくれるサイトや配色が好ましいサイトが望まれる。

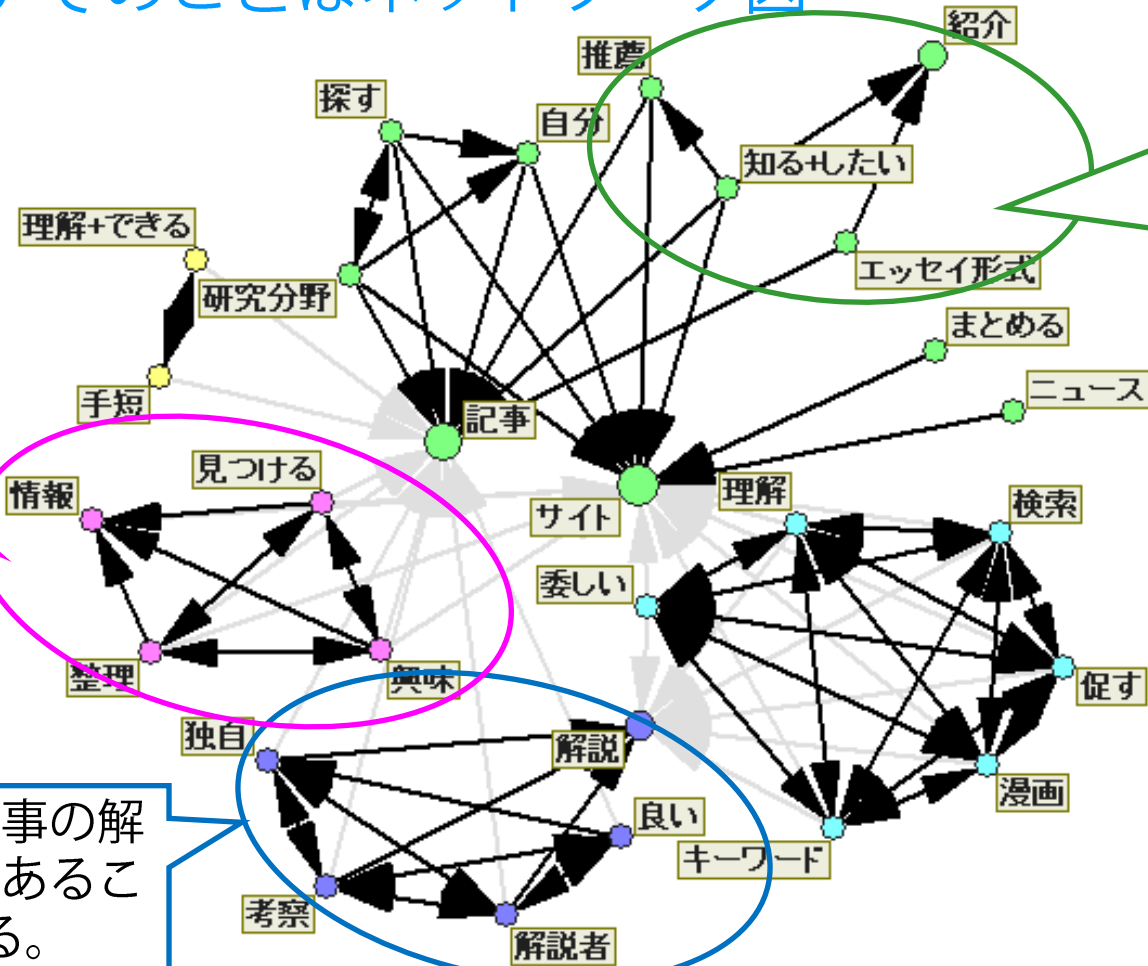
手軽に1日の情報を知る機能や工夫が望まれる。

多くの情報を統合・まとめる機能が望まれる

04. クラスタ別嗜好要因の抽出

クラスタc（社会への関心度が高いクラスタ）の質問2の回答についてのことばネットワーク図

整理された情報の提示・興味のある記事を見つけやすいことが望まれる



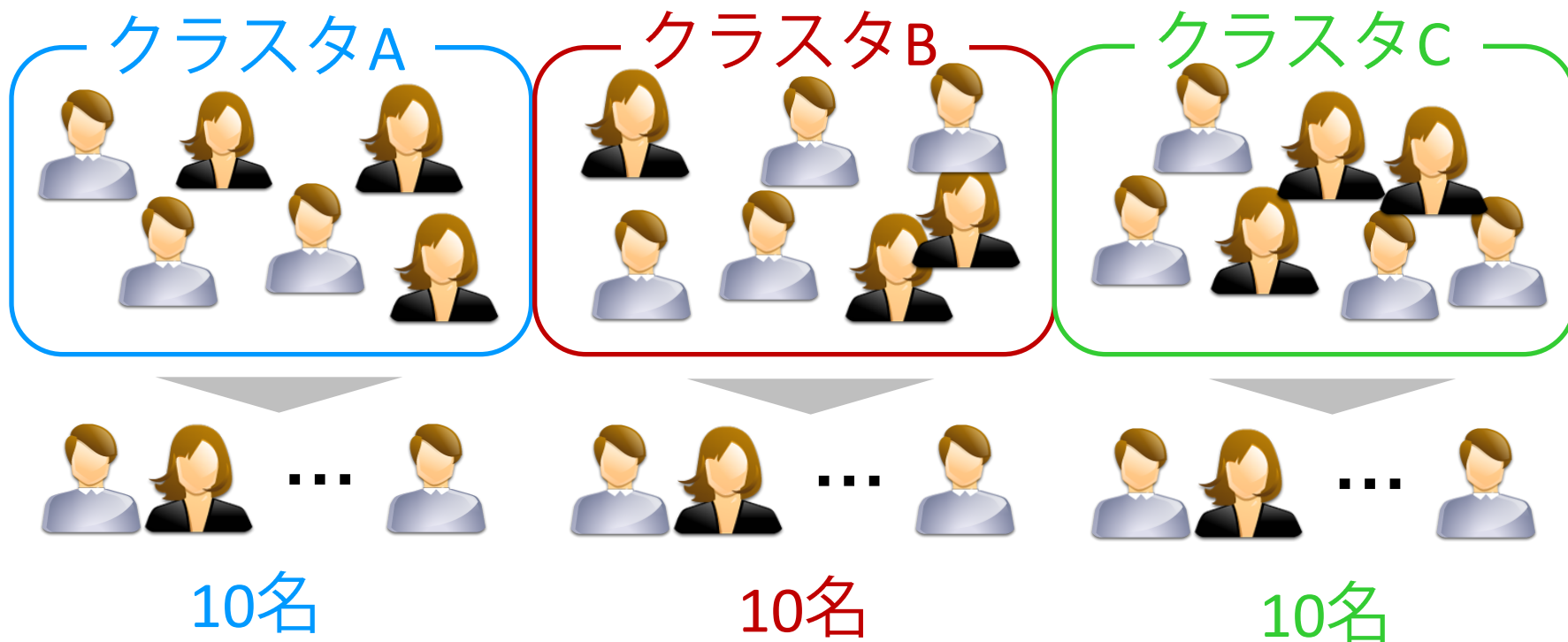
知りたい・読みたい記事を推薦・紹介してくれるサイトが望まれる

解説者による記事の解説や考察結果があることが望まれる。

04. クラスタ別嗜好要因の抽出

②重回帰分析による魅力度に影響を及ぼすイメージ評価項目抽出

STEP1： 3つの被験者クラスタから代表者10名ずつを無作為に選び、計30名の被験者を選ぶ。



04. クラスタ別嗜好要因の抽出

28 / 37

STEP2 : 被験者30名に対して, 5つのクラスタそれぞれから3つのサイトを無作為に抽出し, 1人15個のWebサイトを約1分間ブラウジングしてもらう。



STEP3 : その後, 各サイト毎に6つの評価項目に対して5段階評価(1 : あまりよくない~5 : とてもよい)をつけてもらう。

- ①楽しさ, ②配色, ③レイアウト,
- ④充実度, ⑤分かりやすさ, ⑥魅力度

- 評価用語は, 以下の文献[1]を参考にした。

[1]原田利宣, 田中良介: ラフ集合によるWeb画面デザイン仕様の明確化, デザイン学研究, vol.52, no.5, pp.1-6, 2006

04. クラスタ別嗜好要因の抽出

STEP4：各被験者毎にSTEP3で評価した①～⑤の評価語を説明変数，⑥の”魅力度”を目的変数として重回帰分析を行い，偏回帰係数を算出する。算出された10名分の偏回帰係数の平均を取り，1つのクラスタの総合影響度とする。

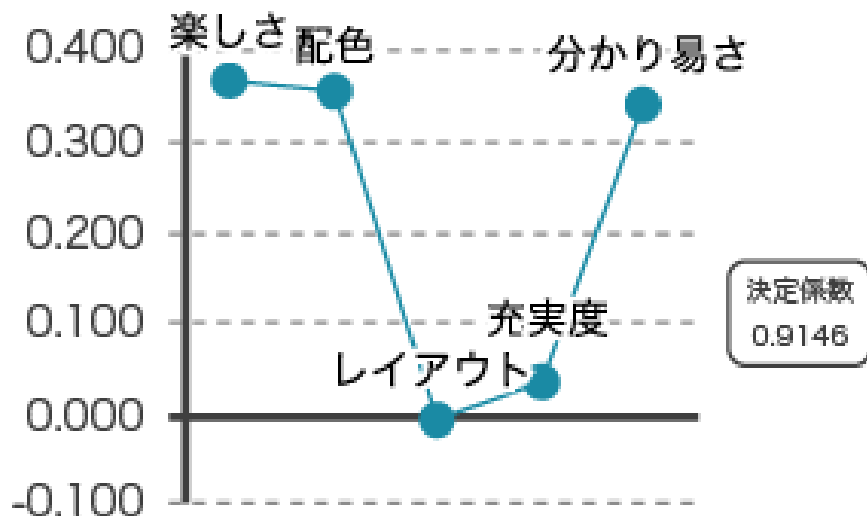
ある被験者(A_1)のデータ表



		楽しさ	配色	レイアウト	充実度	分かり易さ	魅力度
被験者A_1	A_9_FNN	3	4	3	3	3	4
	A_18_時事	2	3	2	3	2	4
	A_26_日経	3	2	4	2	4	3
	E_7_CNET	3	5	3	3	2	5
	E_27_ORI	5	5	4	4	4	5
	E_39_YUC	3	3	3	2	3	3

クラスタA：文系型クラスタについての分析結果

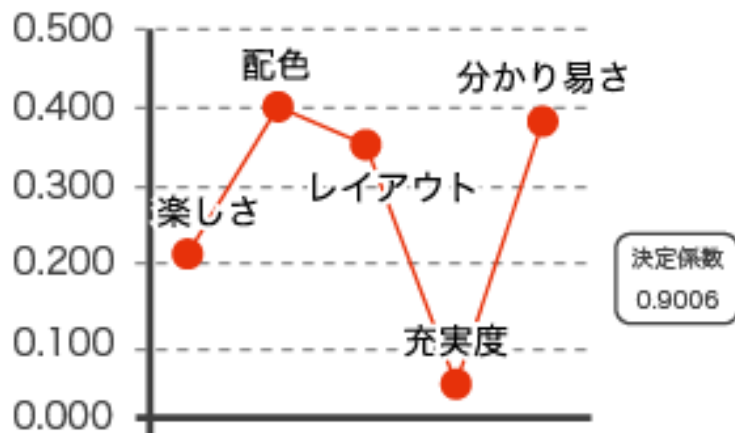
	A-1	A-2	A-3	A-4	...	A-10	平均値
楽しさ	0.5676	0.1863	0.4621	0.4478	...	0.0539	0.3755
配色	0.1635	0.1961	0.4183	0.3635	...	0.8476	0.3469
レイアウト	-0.0400	0.1480	0.0409	0.2420	...	-0.1388	0.0123
充実度	0.0709	0.1161	-0.0912	0.1233	...	0.0201	0.0478
分かりやすさ	0.3106	0.5417	0.2725	0.3064	...	0.2545	0.3372
決定係数	0.9271	0.9005	0.9505	0.9179	...	0.8769	0.9146



考察

「楽しさ」と「配色」と「分かりやすさ」を重視する。サイトのデザイン案として楽しさを与えられる配色や分かりやすいコンテンツが有効だと考えられる。

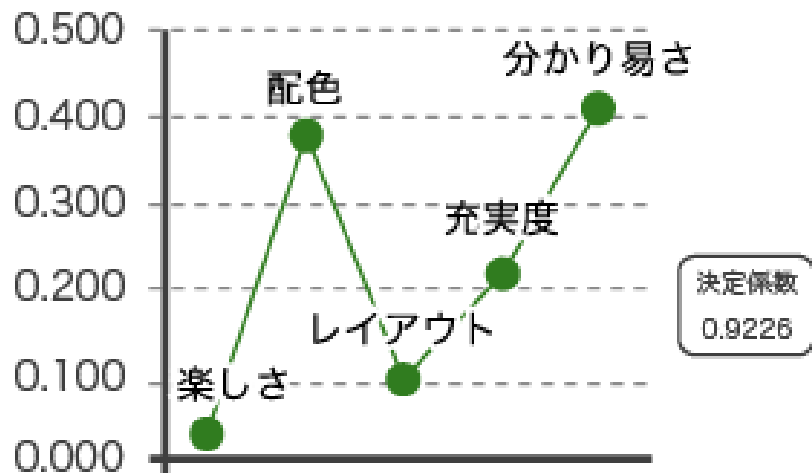
クラスタB：理系型クラスタについての分析結果



考察

「レイアウト」と「配色」と「分かりやすさ」を重視する。サイトのデザイン案として、楽しさよりも配色で構造化されたレイアウトや分かりやすい情報の見せ方が有効だと考えられる。

クラスタc：学習意欲が高いクラスタについての分析結果



考察

「配色」と「分かりやすさ」に加えて、他のクラスタとは異なる点として「充実度」を重視する。サイトのデザイン案として、専門的な情報を掲載するなど、サイトの充実度を上げる必要がある。

目的

本研究では、被験者クラスタB（理系型クラスタ）に焦点を当てる。クラスタBは社会への関心度が低い層であるため、今回のターゲットとした。これまでの分析からWebサイトデザインのコンセプトを決定し、Webサイト制作を行う。

Webサイトデザインのコンセプト

- ・ **ことばネットワーク分析**により、**配色**を重要視、多くの情報をまとめて表示、手軽に本日のニュース記事などを紹介できるページが望まれている。
- ・ **重回帰分析**の結果より、「**配色**」「**分かりやすさ**」「**レイアウト**」が選好に寄与している。

以上の点に注目し、Webサイト制作を行う。

制作したWebサイト (HTML + CSS : 開発期間1週間)



今日の情報を簡単に説明できる場所。重要度の高い情報だけを閲覧できる。

各テーマの記事をパネルインタフェースで構造的に表示する。

ユーザそれぞれにあった情報の提示を行う。

配色変更ボタン

機能. ニュース記事に付属したクーポン情報の追加

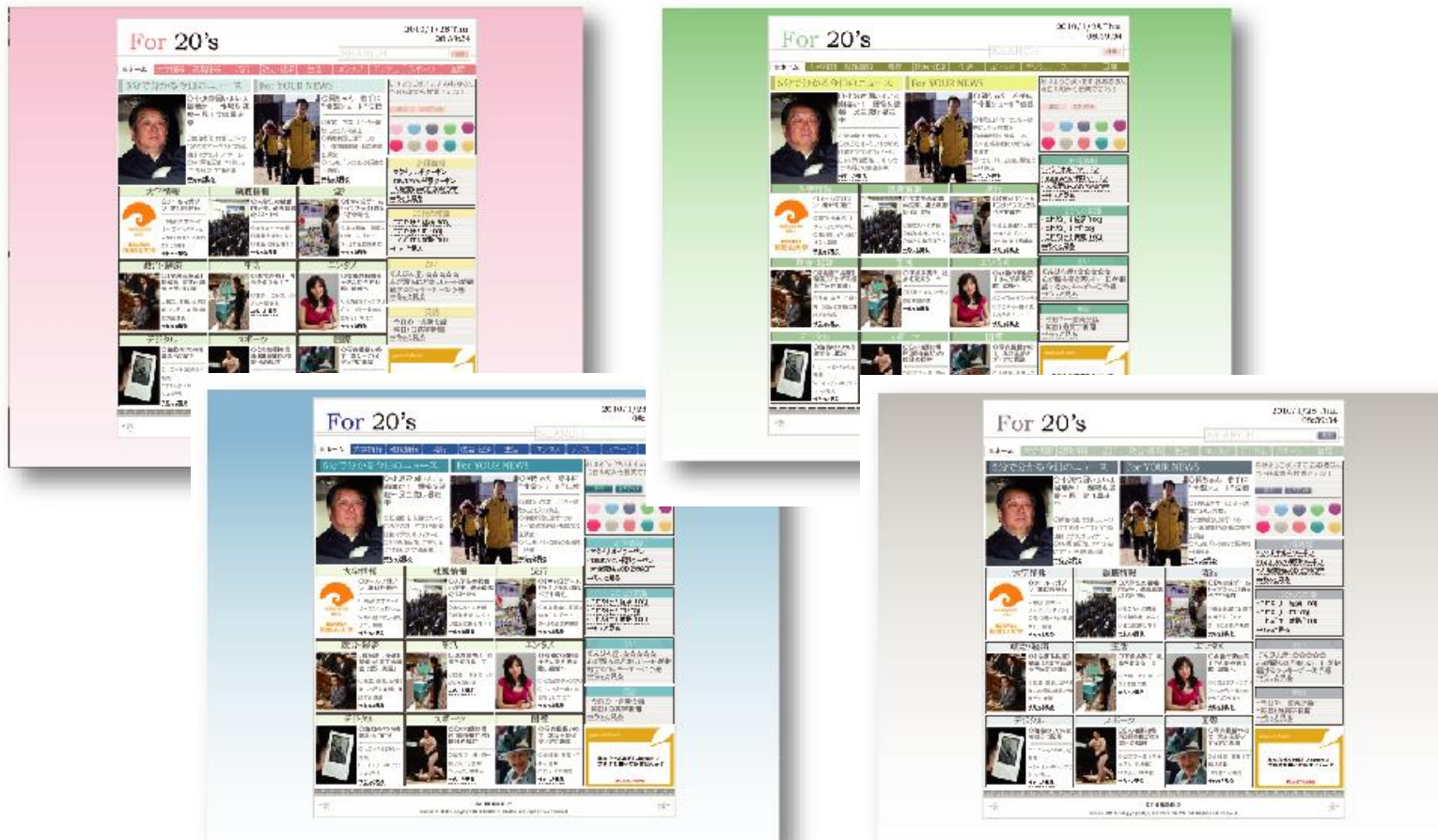
政治・経済などの用語集をまとめて把握できる箇所

05. Webサイトの提案

34 / 37

制作したWebサイト

- 配色変更ボタンで配色のカスタマイズが可能
- Java Scriptで開発



要約

- 既存のニュースサイト48種のグルーピングを行った。
- アンケート調査から大学生50名のプロファイリングを行い, 3つのクラスタに分類した。
- 各クラスタ毎に嗜好に寄与するカテゴリを特徴分析・重回帰分析を用いて抽出し, デザインコンセプトの立案を行った。
- Webサイト制作(TOPページ)を行った。

今後の展望

- 制作したサイトのリンクページの制作と機能実装後, 一部のユーザーを対象に一般公開化。
- 制作したサイトと既存サイトを比較し, 大学生クラスタ B に制作したサイトが好まれるかどうかを分析する。

補遺 1.48種類のニュースサイト

36 / 37

1. yahoo! ニュース	2. NIKKEI NET	3. asahi.com ai	4. msn 産経ニュース
5. goo ニュース	6. NHK ニュース	7. Infoseek 楽天 ニュース	8. google ニュース
9. 毎日 jp	10. livedoor ニュース	11. ITmedia news	12. excite ニュース
13. BIGLOBE ニュース	14. CNET JAPAN	15. YOMIURI ONLINE	16. TBS News i
17. @nifty ニュース	18. AOL ニュース	19. YUCASEE MEDIA	20. ORICON STYLE
21. jijj.com	22. ZDNet Japan	23.FNN 富士ニュース	24. mixi ニュース
25. JUGEM ニュース	26. ANN News	27. Slashdot	28. au one ニュース
29. アメーバニュース	30. Fresh eye ニュース	31. REUTERS ロイター	32. Gigazine
33. 日テレ news24	34. 朝日ニュースター	35. Publickey β	36. Jcast ニュース
37. Web BCN	38. zak zak 情報サイト	39. sponichi annex	40. iza
41. Seachina ニュース	42. 版ニュース	43. MBS ニュース	44. so-net ニュース
45. ウィキニュース	46. videonews.com	47. とれまがニュース	48. JNEWS.com

補遺 2. 選定したアイテム・カテゴリ 37 / 37

広告	ある	コンテンツ数	0～7 種類
	なし		8～12 種類
画像	多い		13 種類～
	少ない	簡易天気予報	ある
会員制度	ある		なし
	なし	専門サイト	である
文字サイズ	変更可能		でない
	変更不可	レイアウト	2 列
検索窓	左寄り		3 列
	中央寄り		複合
	右寄り	アクセスランキング	1 位～10 位
	なし		11 位～20 位
RSS 機能	ある		21 位～30 位
	なし		31 位～40 位
多言語へ翻訳	ある		41 位～
	なし	ニュースランキング	ある
サイトマップ	ある		なし
	なし		