

ネットワーク分析によるプレスリリースデータを用いた企業行動の分析

知識科学研究科サービス経営コース

林 隆臣



1. 目的と背景

企業が顧客に提供するサービスの評価は、顧客の期待に大きく影響を受ける。顧客の期待を理解するためには、顧客側のサービス利用経験、うわさ、顧客のニーズなどに加え、企業が顧客に対して行う対外的なコミュニケーションも重要になる。

(A. ParasuramanらのDeterminants of Perceived Service Qualityによる。)

ここで、企業の対外的なコミュニケーションが、顧客の期待に影響を与えるプロセスを理解することは、サービスを企画する際に有効ではないかと考えた。対外的なコミュニケーションの一例としては、プレスリリースによる社外向けの発表があげられる。プレスリリースから、顧客に対する働きかけの行動を抽出することができれば、前述の有効性を確認するための要素を見いだすことができるのではないかと考えた。

本調査においては、プレスリリースを用いた企業行動の分析を行うために、ネットワーク分析が有効ではないかと考え、実際に企業行動の抽出が可能であるか検証を行った。

具体的には、モデルとして取り上げた企業がWeb上に発表したプレスリリースデータに出現するキーワードの共起ネットワークを作成し、出現頻度などから、その企業が示している行動の評価を行った。

2. 調査の概要

1. に挙げたプレスリリースデータからの企業行動を把握するための準備として、サンプルとなる企業を選定し、当該企業のプレスリリースデータを収集した。

プレスリリースデータの収集にあたっては、分析をしやすいために今回は、プレスリリースのエントリデータを用いることとした。プレスリリースのエントリデータとは、プレスリリース本体につけられた1～2行で作成される見出しのようなものである。

(1) 収集したデータの概要

- ・対象企業：大手ICT企業。ICTに関する研究開発から、ICT製品の製造販売、サービス提供まで幅広く行っている。
- ・対象データ：2001年1月から、2010年12月までに発表されたプレスリリースを対象とした。
ただし、データ収集最終日が、2010年12月14日までに発表されたものとなる。
10年間のエントリ数は以下の通りとなる。

表：サンプル企業の10年間の発表件数

年	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	合計
エントリ数	434	450	506	518	510	469	405	399	403	388	4482

(2) プレスリリースエントリデータのイメージ

多くの企業のプレスリリースについては、下記の様な形でアーカイブされている。

(以下は東芝社の事例: http://www.toshiba.co.jp/about/press/date_j.htm)

アーカイブの一覧

発表月別
2011年
1月 2月 3月 4月 5月 6月 7月 8月 9月 10月 11月 12月
2010年
1月 2月 3月 4月 5月 6月 7月 8月 9月 10月 11月 12月
2009年
1月 2月 3月 4月 5月 6月 7月 8月 9月 10月 11月 12月
2008年
1月 2月 3月 4月 5月 6月 7月 8月 9月 10月 11月 12月
2007年
1月 2月 3月 4月 5月 6月 7月 8月 9月 10月 11月 12月
2006年
1月 2月 3月 4月 5月 6月 7月 8月 9月 10月 11月 12月
2005年
1月 2月 3月 4月 5月 6月 7月 8月 9月 10月 11月 12月
2004年
1月 2月 3月 4月 5月 6月 7月 8月 9月 10月 11月 12月
2003年
1月 2月 3月 4月 5月 6月 7月 8月 9月 10月 11月 12月
2002年
1月 2月 3月 4月 5月 6月 7月 8月 9月 10月 11月 12月

月別のエントリー一覧

発表月別
2010年11月のニュースリリース
2010年11月29日
○ 米国大手エンジニアリング会社・グループとの原子力発電所建設に関する協力関係の構築について ○ アセアン市場向けのテレビ事業強化について ○ カザフスタンにおけるGISプロジェクトの実証試験の実施について
2010年11月26日
○ モンゴル MNFCC 社との資源開発・社会インフラ分野での協力に関する覚書締結について ○ 海外に展開したモノの流通の効率化とロジスティクスサービスの提供について (東芝 エイティック株式会社)
2010年11月24日
○ アイビエグループ・ソフトウェア・フォーラムが協賛第一回大会
2010年11月18日
○ 地産地消推進とワンセグチューナーを搭載したポータブルDVDプレーヤーなどの発売について ○ 東芝科学館で東芝エレクトロニクス25周年記念企画展を開催
2010年11月17日
○ 海外向けホームエアコン「あったか太陽熱」の発売について (東芝ホームアプライアンス株式会社)
2010年11月16日
○ 環境に配慮したモジュール型データセンターの受注開始について ○ 北オーストラリアの工業管理システムなどの産業システム向け 2.5 倍速 HDD の商品化について
2010年11月15日
○ 東芝経済システムが第7回エコプロダクツ大会(エコプロダクツ展)で、エコプロダクツ大賞を獲得
2010年11月11日
○ 東芝のスマートコミュニティ事業を推進するIT事業を東芝ソリューションが「スマートコミュニティ総合ソリューション」として発表 (東芝ソリューション株式会社)
2010年11月09日
○ 2010年度第2回半期「16か月累計」決算 ○ 米国向け事業開発を新たなブランド「スマートメント」として発表 (東芝ソリューション株式会社)
2010年11月08日
○ ステックタイプのSSD「Blade X-gate」の発売について
2010年11月05日
○ 新製品ソリューションズと金融機関向けソリューション事業で協業 (東芝ソリューション株式会社)
2010年11月04日
○ 米国における A B W R 設計技術の更新申請について ○ 東芝ソリューショングループが「環境フォーラム」を開催 (東芝ソリューション株式会社) ○ スマートコミュニティ事業推進の状況について

プレスリリース本体

発表月別
2010年11月26日
モンゴル MNFCC 社との資源開発・社会インフラ分野での協力に関する覚書締結について 当社は、モンゴル国 (以下、モンゴル) の MNFCC 社と、モンゴルにおけるウラン、レアアース、レアメタル等の鉱物資源開発、社会インフラ整備での協力検討を開始する覚書を締結しました。 本年 11 月 19 日、菅総理大臣がエルベグドルジ大統領と会談し、戦略的パートナーシップ構築に向けた共同声明が発表され、モンゴルの鉱物資源開発における足並の揃った協力を戦略的に推進していくべきことが確認されるなど、日・モンゴルの経済関係の更なる促進が期待されます。 当社は、モンゴルの官民協力プログラムでモンゴル政府に対し資源開発やエネルギー、社会インフラ整備を支援する MNFCC 社と互恵的協力関係構築に向けた協力を推進します。今後のモンゴルの経済発展に重要な火力、原子力、太陽光発電、電力系統を含む社会インフラ整備の可能性検討を行うとともに、鉱物資源開発の安定供給確保に向けた取り組みを進めます。 MNFCC 社の概要 1. 会社名: MNFCC LLC 2. 設立: 2010 年 9 月 3. 代表者: Dugersuren Dashdorj (ダグウルスレン ダシドルジ) 4. 事業内容: 資源開発、エネルギー、社会インフラ整備に関する官民協力プログラム実施

本調査では、この部分のデータを用いた

(3) 調査にあたっての仮説

プレスリリースのエントリデータから、

①企業が発表を通して行う一連の行動を読み取ることが可能である
(例えば、発表した製品やサービスとその取り扱い)

②個々の発表内容間に何らかの関係性を見いだすことが可能である

- ・時系列(バリューチェーンに基づく製品やサービスの展開、発表済み製品の強化、新技術や製品を活用した新しいサービスの開発内容)
- ・製品やサービスに関係ない活動

との仮説を持ち、プレスリリースエントリデータを分析することとした。

3. 作業内容詳細

(1) 作業の進め方

作業を進めるにあたり、以下の準備を行った。

1) プレスリリースエントリデータの準備

プレスリリースエントリデータについては、サンプル企業のホームページに掲載された、発表日別一覧データを入手することとした。

データの入手にあたっては、`wget`を用いて必要なデータを一括ダウンロードした。

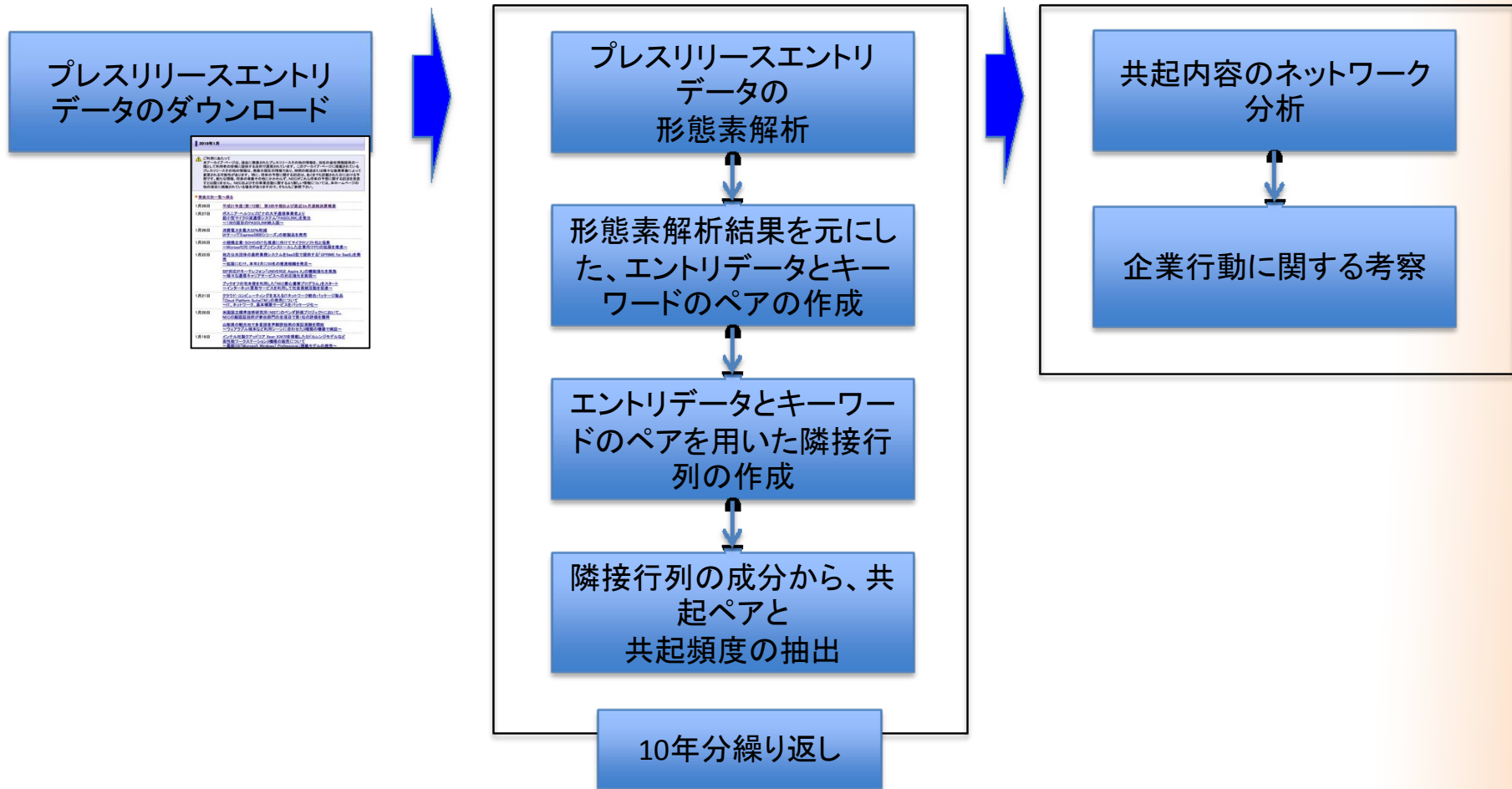
2) データ解析環境の準備

また、入手したデータについては、形態素解析を行い品詞分解し、必要となるキーワードを抽出することとした。形態素解析については、奈良先端科学技術大学院大学が提供している、`chasen`を用いた。

3) ネットワーク分析環境の準備

抽出したキーワード間の関係や、エントリデータとの関係について、ネットワーク分析を行うこととした。ネットワーク分析については、フリーのツールであるPajekを用いた。

(2) 作業フロー



(3) プレスリリースエントリデータの形態素解析について

今回解析の対象とした、プレスリリースエントリデータは、以下の様な特徴を持っている。

- ・すべてのエントリデータが、1～3行で構成されている。
- ・異なるエントリデータは、1行の行送りされている。
- ・同一エントリデータ中で改行が行われる場合は、行送りはしない。
- ・同一月内は、月末から月初の順に並んでいる(発表順)

(4) 形態素解析とエントリデータからのキーワードの抽出について

- ・形態素解析用の辞書については、chasenに付属している、ipadic-EUC-JP-2.7.0をそのまま用いた。
- ・キーワードの抽出にあたっては、名詞(非自立、代名詞除く)、記号(括弧、句読点除く)、未知語(辞書登録外)のみを抽出した。

抽出の単位は、年単位で行った。また、同一年内で1回しか出現しないキーワードは今回は集計対象から除外した。

(5) キーワードの共起情報の作成

共起とは、同一エントリ内に同時に出現するキーワードのペアを抽出したものである。キーワードのペア抽出にあたっては、次の手順を用いて行った。

1) 共起の抽出

抽出したキーワードを、「エントリ」×「キーワード」の対応行列を作成し、本行列を用いて、「キーワード」×「キーワード」行列を求めた。

- ・エントリ・キーワード行列A:「エントリ」×「キーワード」行列
- ・行列 tA : 行列Aの転置行列
- ・キーワード共起行列B: 行列 ${}^tA \times A$ 、「キーワード」×「キーワード」行列

ここで求めた、行列Bの右上の成分から、キーワードの共起の対応を抽出した。
計算例は下記の通り。(行列 ${}^tA \times A$ のグレー部分が抽出対象となる。)

例: エントリ3本と、キーワード10個

行列A:

エントリ \ KW	1	2	3	4	5	6	7	8	9	10
1	1	1	1	0	0	0	1	1	1	0
2	0	0	0	1	1	1	0	1	0	1
3	0	0	1	1	1	1	1	1	0	0

行列 tA :

	1	2	3
1	1	0	0
2	1	0	0
3	1	0	1
4	0	1	1
5	0	1	1
6	0	1	1
7	1	0	1
8	1	1	1
9	1	0	0
10	0	1	0

行列: $B = {}^tA \times A$

	1	2	3	4	5	6	7	8	9	10
1	1	1	1	0	0	0	1	1	1	0
2	1	1	1	0	0	0	1	1	1	0
3	1	1	2	1	1	1	2	2	1	0
4	0	0	1	2	2	2	1	2	0	1
5	0	0	1	2	2	2	1	2	0	1
6	0	0	1	2	2	2	1	2	0	1
7	1	1	2	1	1	1	2	2	1	0
8	1	1	2	2	2	2	2	3	1	1
9	1	1	1	0	0	0	1	1	1	0
10	0	0	0	1	1	1	0	1	0	1

行列Bにおいて、対角線上に現れる要素は、各キーワード(以下KW)がいくつのエントリに入っていたかを表している。また、行列Bの各要素は共起頻度(共起ペアの出現回数)を示している。

各行と列の対応は以下の通りとなる。

例:

- ・1行2列: KW1とKW2のペアが1つのエントリ中に現れる。
- ・1行4列: KW1とKW4のペアが1つのエントリ中に現れる。
- ・4行5列: KW4とKW5のペアが2つのエントリ中に現れる。

2)共起情報からネットワーク分析用のスクリプトの作成

1)の抽出した結果から、(KWa,KWb,共起頻度)のデータを作成し、共起頻度の4以上のデータを抽出し、ネットワーク分析用のデータを作成した。

(例)ある年のキーワードの共起の状況

(キーワード、キーワード、共起の出現頻度)

(発売,対応,26)	(開始,企業名,11)	(サーバ,企業名,10)
(発売,実現,19)	(開発,企業名,11)	(発売,R,10)
(提供,サービス,18)	(電力,サーバ,11)	(発売,サーバ,10)
(技術,開発,17)	(開始,サービス,11)	(支援,サービス,10)
(発売,製品,17)	(強化,サービス,11)	(強化,システム,10)
(システム,企業名,16)	(構築,システム,11)	(液晶,ディスプレイ,10)
(発売,搭載,15)	(通信,システム,11)	(製品,強化,10)
(サービス,企業名,14)	(発売,システム,11)	(発売,強化,10)
(販売,開始,14)	(製品,シリーズ,11)	(発売,支援,10)
(協業,企業名,12)	(実現,開発,11)	(電力,ECO,9)
(発売,モデル,12)	(強化,機能,11)	(強化,企業名,9)
(提供,開始,12)	(実現,技術,11)	(液晶,カラー,9)
(発売,環境,12)	(事業,強化,11)	(マイクロ,システム,9)
(発売,機能,12)	(実証,実験,11)	(小型,システム,9)
(発売,電力,12)	(電力,消費,11)	(発売,シリーズ,9)

...

※キーワード中に「企業名」とあるものは、サンプル企業の名称を置き換えたものである。

(例)ある年のキーワードの共起グラフの例

<Pajek用のスクリプト>

*Vertices 188

...

36 "お知らせ" ic Blue
37 "カメラ" ic Blue
38 "クラ" ic Blue
39 "グループ" ic Blue
40 "グローバル" ic Blue
41 "ケータイ" ic Blue
42 "ケーブル" ic Blue
43 "コスト" ic Blue
44 "コミュニケーション" ic Blue
45 "サーバ" ic Blue
46 "サービス" ic Blue
47 "システム" ic Blue
48 "シリーズ" ic Blue
49 "スイッチ" ic Blue
50 "セキュリティ" ic Blue
51 "センター" ic Blue

...

187 "利用" ic Blue
188 "連携" ic Blue

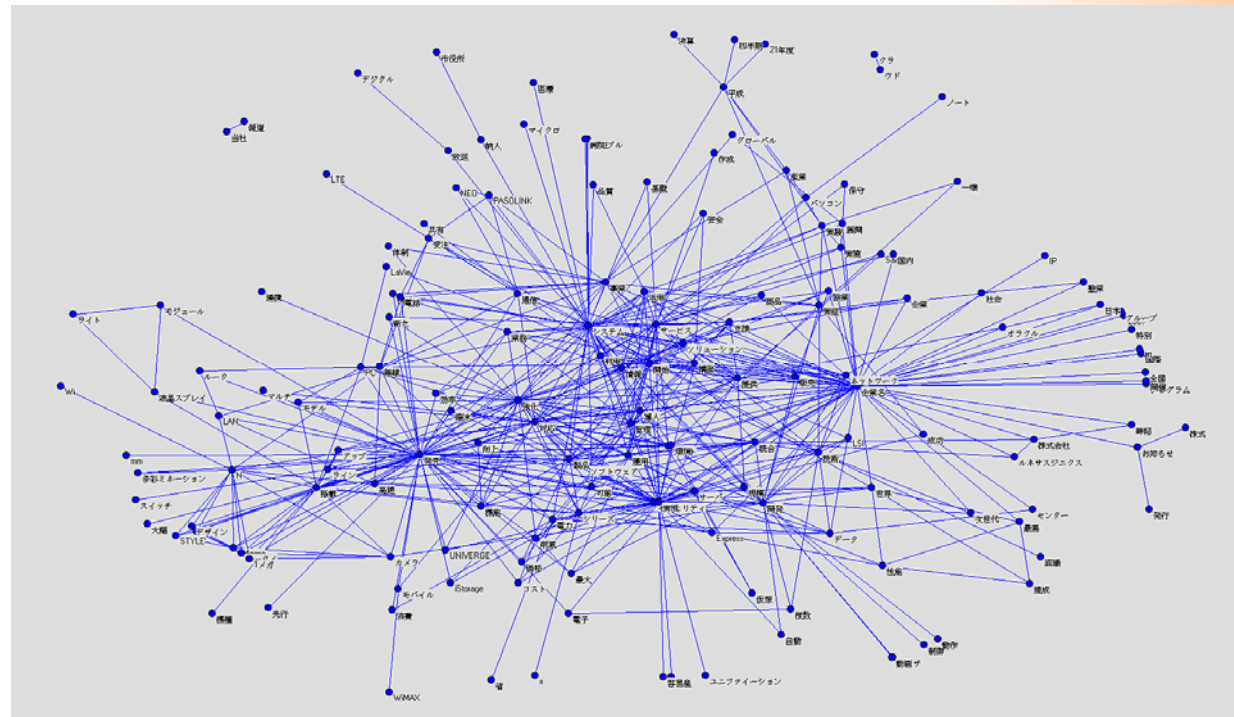
*Edges

16 1

...

183 166
176 167
176 170
185 176

<キーワードの共起状態のネットワーク>



4. 調査結果

(1) 共起ネットワークの結果から各年のエントリの中心をなすキーワードの抽出

各年に発表された内容を特徴づけるキーワードを抽出するために、各年における次数中心性の高いキーワードを、Pajekの機能を用いて抽出した。

次数中心性の高いキーワードは、当該年の多くのエントリに存在することから、その年に企業が発表した内容を代表するキーワードと見ることができる。

また、次数中心性の高いキーワードの出現内容に関して、経年変化を見ることで、企業が発表する内容のトレンドがどのように変化しているのかを見いだすこともできると考えられる。

実際に、出現したキーワードは以下の通りであった。

- ・10年間のプレスリリースデータに登場する次数中心性の高い語（各年の上位10以内）は全体で23語ある
- ・上位に10以内の出現パターンは次ページのグラフの通り

ちなみに、2010年に登場した、「クラ」、「ウド」については、形態素解析の辞書の関係で、「クラウド」というキーワードが分割して出現したと思われる。

<2001年から2010年の回数中心性の高いキーワード上位10>

2001	
発売	0.480446927
システム	0.312849162
実現	0.290502793
サービス	0.217877095
強化	0.217877095
搭載	0.217877095
対応	0.156424581
企業名	0.150837989
インターネット	0.139664804
携帯	0.111731844

2002	
発売	0.400943396
システム	0.363207547
実現	0.311320755
企業名	0.287735849
強化	0.268867925
開始	0.245283019
提供	0.216981132
サービス	0.202830189
製品	0.179245283
企業	0.146226415
構築	0.146226415

2003	
システム	0.37826087
発売	0.369565217
実現	0.365217391
サービス	0.273913043
強化	0.273913043
企業名	0.265217391
提供	0.234782609
開始	0.22173913
製品	0.217391304
サーバ	0.182608696

2004	
実現	0.495833333
システム	0.404166667
発売	0.3875
強化	0.3375
提供	0.2875
対応	0.279166667
サービス	0.266666667
可能	0.2625
企業名	0.254166667
開始	0.254166667

2005	
発売	0.451492537
実現	0.380597015
企業名	0.294776119
システム	0.27238806
強化	0.246268657
サービス	0.21641791
対応	0.197761194
提供	0.197761194
開発	0.182835821
サーバ	0.179104478

2006	
発売	0.468619247
実現	0.326359833
企業名	0.29707113
システム	0.288702929
強化	0.246861925
開発	0.225941423
電話	0.154811715
対応	0.146443515
サービス	0.142259414
製品	0.142259414

2007	
発売	0.365384615
実現	0.293269231
開発	0.230769231
強化	0.201923077
企業名	0.197115385
システム	0.1875
製品	0.182692308
製品	0.182692308
サービス	0.177884615
対応	0.144230769

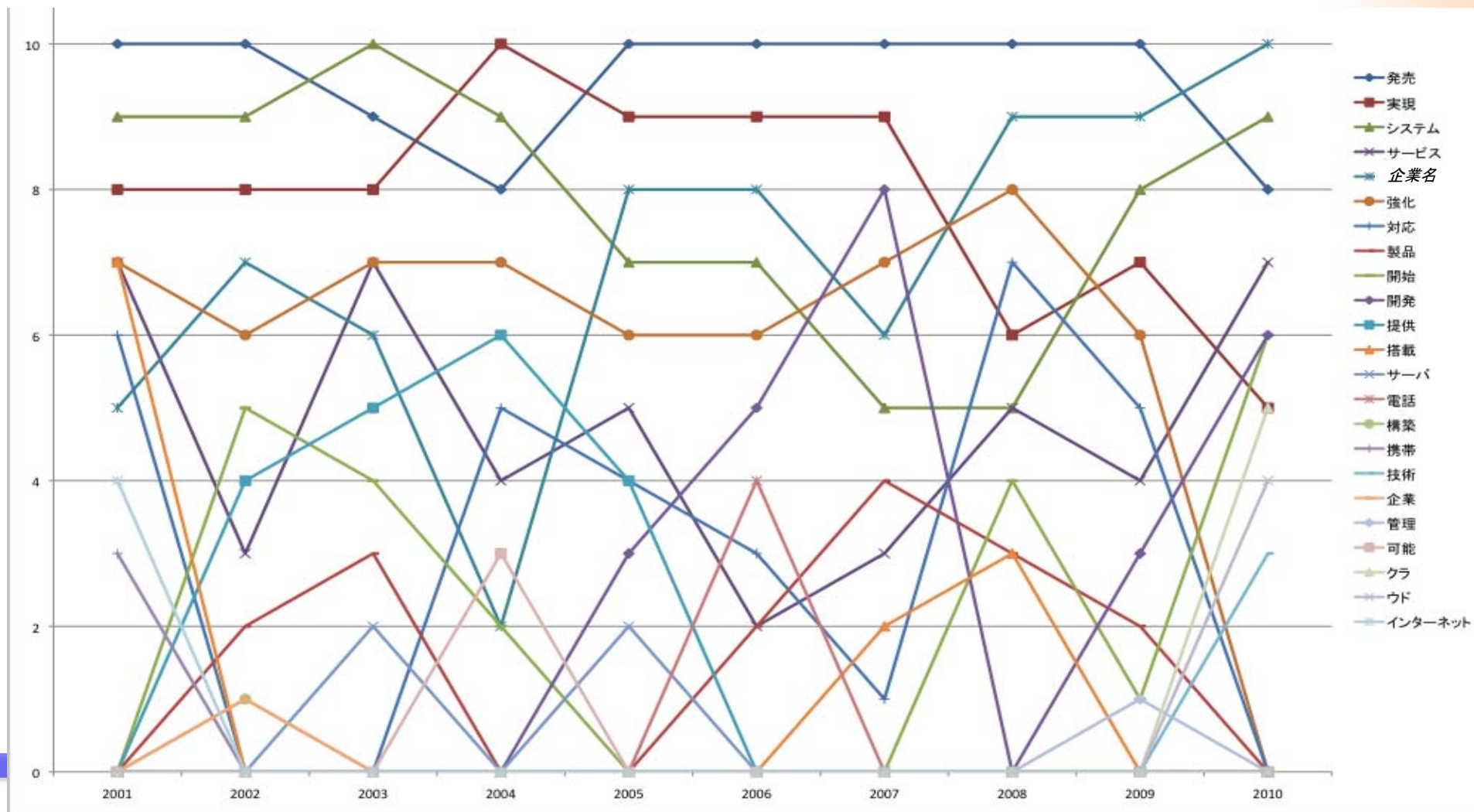
2008	
発売	0.417525773
企業名	0.201030928
強化	0.190721649
対応	0.18556701
実現	0.170103093
サービス	0.164948454
システム	0.164948454
開始	0.113402062
製品	0.108247423
搭載	0.108247423

2009	
発売	0.390374332
企業名	0.331550802
システム	0.294117647
実現	0.240641711
強化	0.203208556
対応	0.181818182
サービス	0.160427807
開発	0.155080214
製品	0.128342246
開始	0.122994652
管理	0.122994652

2010	
企業名	0.354651163
システム	0.319767442
発売	0.302325581
サービス	0.284883721
開始	0.197674419
開発	0.197674419
実現	0.180232558
クラ	0.180232558
ウド	0.174418605
技術	0.156976744

＜次数中心性の高いキーワードの年度別傾向＞

以下のグラフのY軸の数字は、10が次数中心性が一番高く、1が一番低いことを意味する。また、0の場合は、その年は、10位以内には出現していないことを意味する。



回数中心性の高いキーワードの状況をまとめると以下の通りとなる。

- ①毎年上位に出現するキーワード：
発売、実現、強化(2010だけなし)、システム、サービス、企業名
- ②最近中心的になり始めた語(初めの4年は全く出現しない)：
技術、管理、クラウド
- ③ばらつきがあるが10年間で複数回中心的になっている：
搭載、提供、対応、製品、開発、開始、サーバ
- ④途中一回のみ：
電話、技術、可能
- ⑤最初の3年のみ：
構築、携帯、企業、インターネット

このキーワードは、プレスリリースに出現するキーワードとしては、妥当なものであると考えられる。また、これらのキーワードは、大きく以下のように分類することができる。

- ・発表の対象自身を示すキーワード
(例)システム、サービス、企業名、サーバ、インターネット 等
- ・発表の対象となる行動を示すキーワード
(例)発売、実現、強化 等

また、発表の対象自身を示すキーワードを見ると、この企業がアピールしたいと思っているものの変化を見ることができる。

- ①毎年上位に出現するキーワード：
システム、サービス、企業名
- ②最近中心的になり始めた語(初めの4年は全く出現しない)：
技術、管理、クラウド
- ③ばらつきがあるが10年間で複数回中心的になっている：
製品、サーバ
- ④途中一回のみ：
電話、技術
- ⑤最初の3年のみ：
携帯、企業、インターネット

サンプル企業は大手ICT企業であり、システムやサービスを中心にビジネスを行っていると見ることができる。また、最近の流行同様、クラウドや管理といったキーワードをアピールしている。

(2) 次数中心性の高いキーワードと共起するキーワード

(1)にて、次数中心性の高いキーワードは大きく、2つに分類できるとした。

①発表の対象自身を示すキーワード

②発表の対象となる行動を示すキーワード

上記の分類を係り受けの関係を構成する要素と見立てて評価を行った。

①に相当するものを主語、目的語の役割になるもの。②に相当するものを述語の役割になるもの。その他に修飾語の役割になるものの3タイプとした。

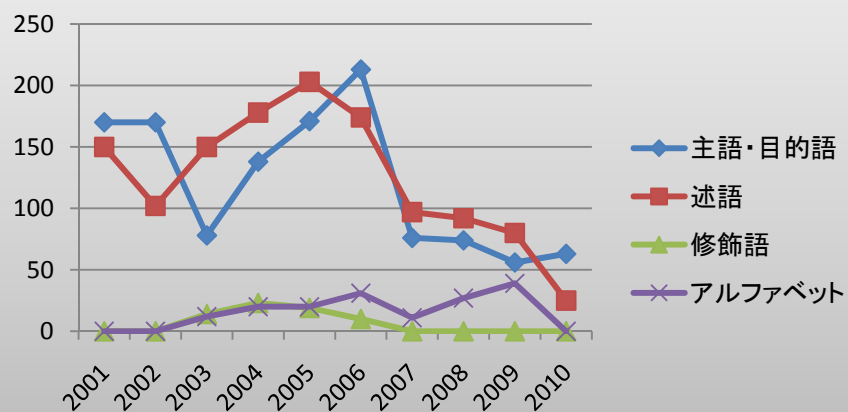
次数中心性の高いキーワードのうち、②に相当する「発売」、「実現」、「強化」については、主語・目的語と他の述語との共起が同様に出現している。また、①に相当する「企業名」、「システム」、「サービス」は述語との共起が多く出現する結果となった。

なお、「システム」、「サービス」と共起していたキーワード(10年間でトータル50回以上出現するもの)は以下の通りであった。

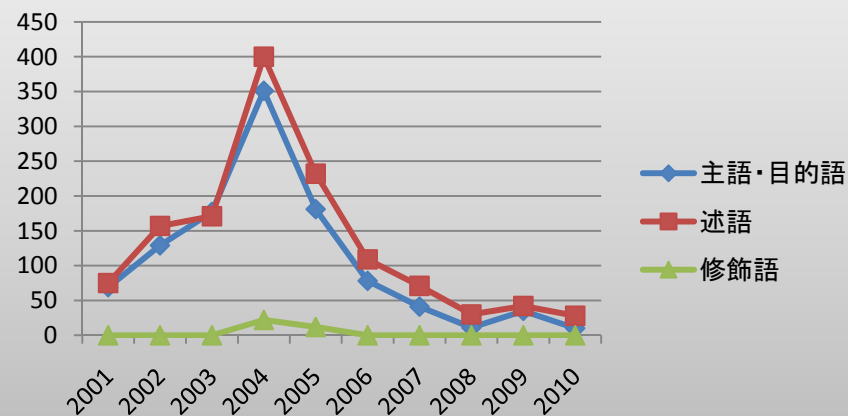
- ・システム: 実現、企業名、構築、開発、発売、業務、提供、管理、サービス、情報、強化、開始、活用、導入、受注
 - ・サービス: 提供、開始、企業名、実現、強化、発売、システム、サービス名、対応
- ※ サービス名: サンプル企業が有するサービスの名称を置き換えたもの

＜次数中心性の高いキーワードと共起するキーワードの種別＞

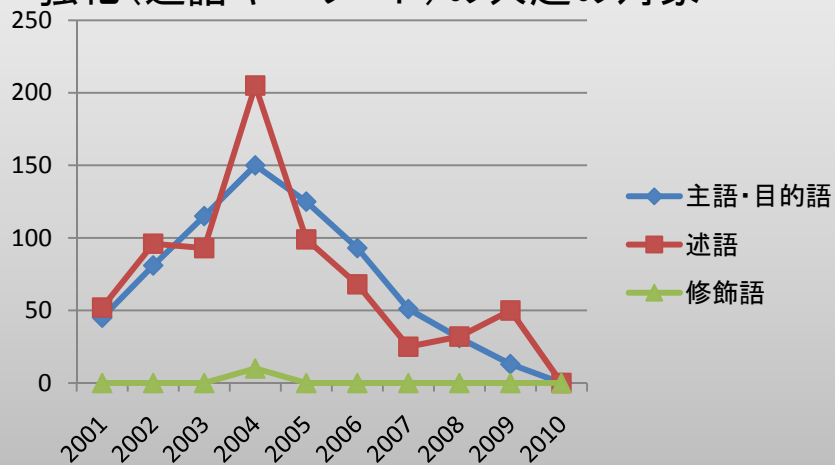
発売(述語キーワード)の共起の対象



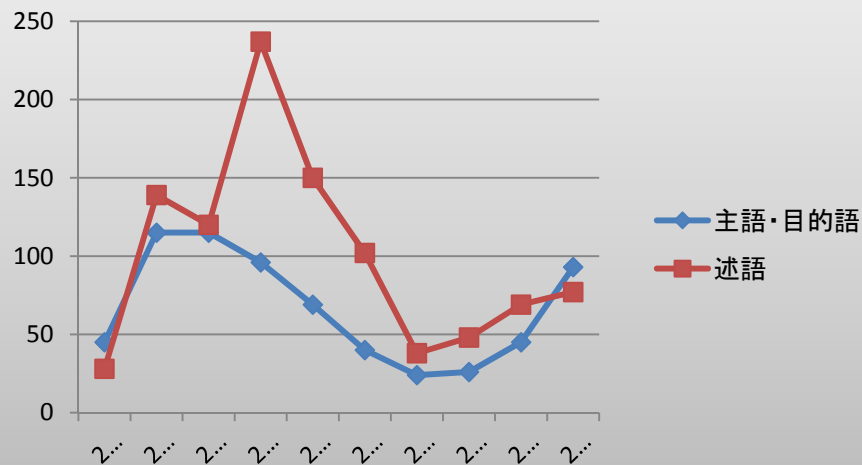
実現(述語キーワード)の共起の対象



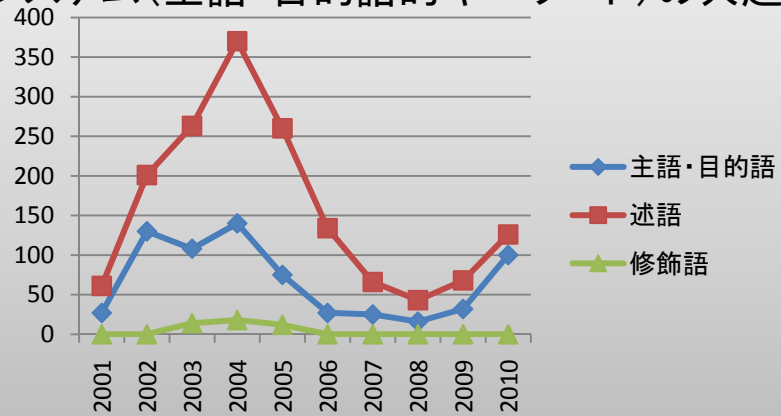
強化(述語キーワード)の共起の対象



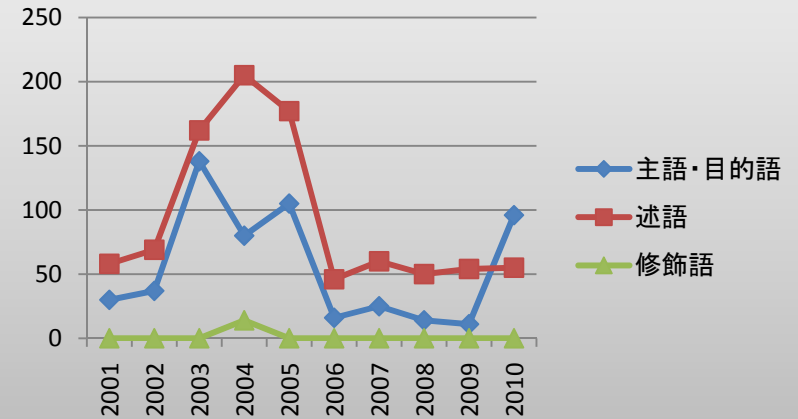
企業名(主語・目的語的キーワード)の共起の対象



システム(主語・目的語的キーワード)の共起の対象



サービス(主語・目的語的キーワード)の共起の対象



(3) 次数中心性の高いキーワード同士の共起状況

次数中心性の高いキーワード同士の共起については、その年の発表内容を代表する行動を抽出できると考え、それらの共起状況について抽出した。

まず、年度別に見た場合、以下の特徴を見いだすことができる。

① 2004年は実現が突出している。

2004年は、実現の出現頻度がNo1(518エントリで、132回登場)。

この年は、この表現をするのがサンプル企業のトレンドだったと思われる。

② サービスと発売の共起状況

2007年以降、サービスと共起する発売は以下の回数であった。

2007年:5、2008年:8、2009年:9、2010年:8

この年以降、サービスと10回以上共起するものは、以下の通り。

(全て10回以上共起している。)

- ・2007年:提供、開始、企業名、実現、強化、基盤、開発

- ・2008年:提供、開始、企業名、強化、支援

- ・2009年:提供、開始、企業名、実現、強化

- ・2010年:提供、開始、企業名、システム、活用、クラ、ウド

であり、サービスは発売するという表現を用いなくなってきたことが分かる。また、2010年にクラウドというキーワードが出てきてから、(サービス、クラウド、システム)という関係が見いだせる。

③サービスと提供の共起状況

・2007年:14、2008年:18(No1)、2009年:15(No2)、2010年:23(No3)

サービスについては、②の発売よりも提供という表現を用いるようになってきたことが分かる。

④2009年以降のシステムと発売の共起状況

2009年以降、システムと共起する発売は以下の回数であった。

・2009年:9、2010年:9

ちなみに、2009年以降、システムと共起するものは、以下の通り。

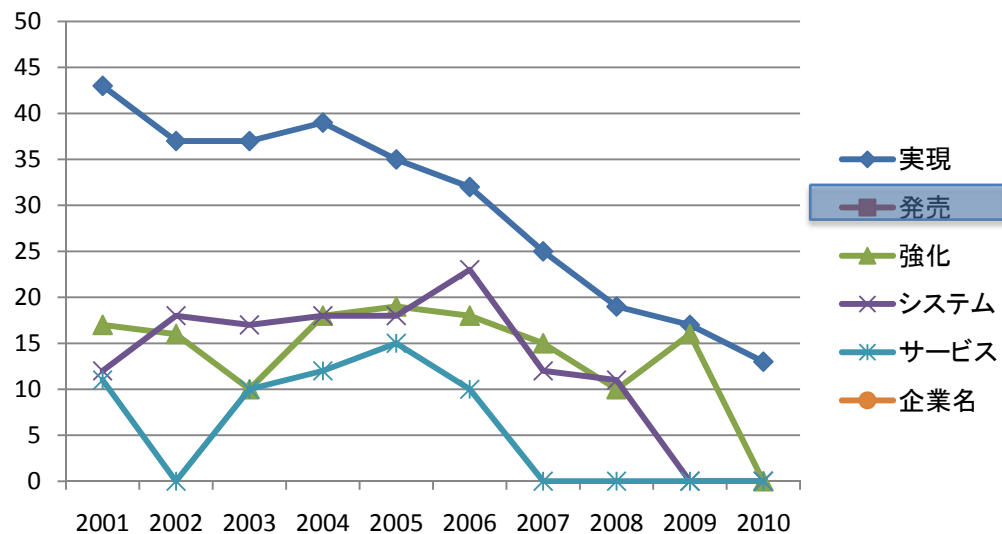
・2009年:企業名、実現、構築、開発、業務、受注、事業

・2010年:企業名、実現、構築、業務、受注、事業、管理、提供、開始、サービス、事業、基幹、
納入、通信、支援、クラ、ウド

この中で、構築、受注については、発売しているもの(システム)が実際に顧客へ提供され始めたことを示すものであり、提供開始の発表後の行動の成果と見ることができる。

< 次数中心性の高いキーワード同士の共起状況 >

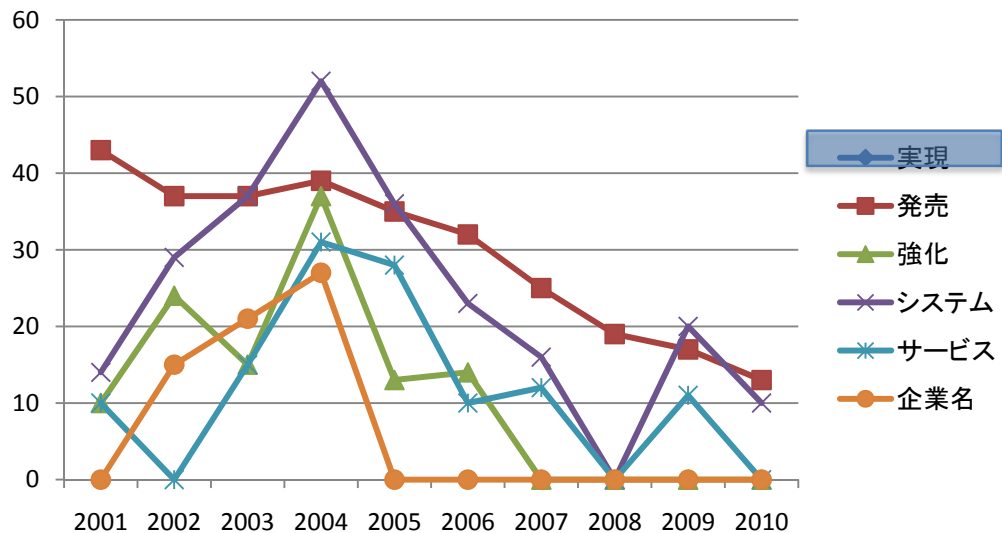
ここでは、1年間のエントリ中で共起する頻度をY軸に取っている



発売との共起

強化(する)、実現(する)行動と、発売の対象がシステムやサービスであるとみることができる。

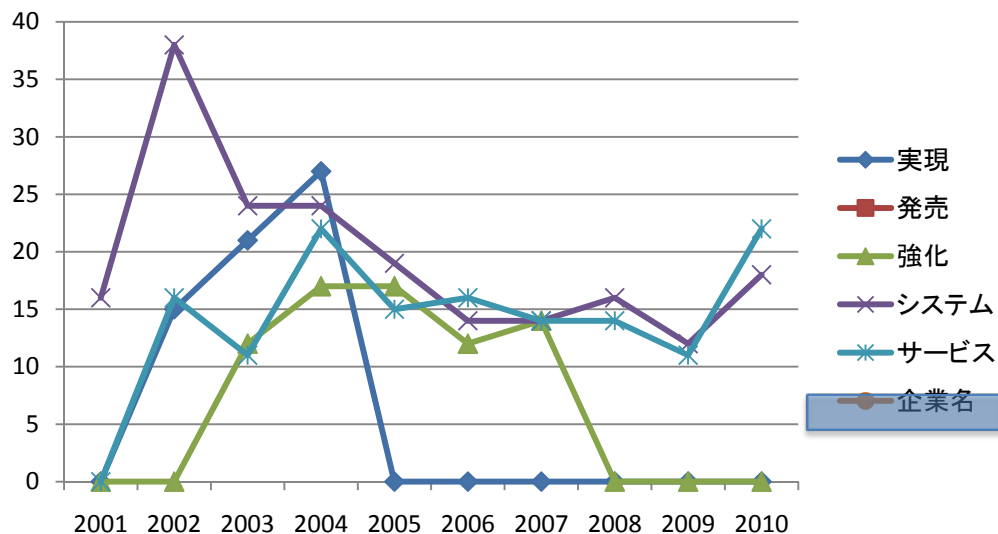
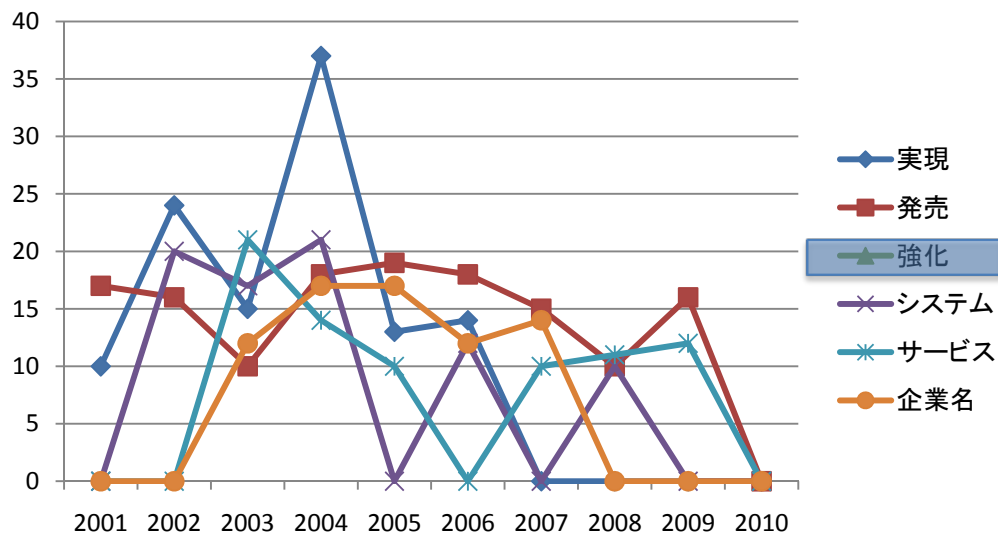
(ただし、2010年を除く)

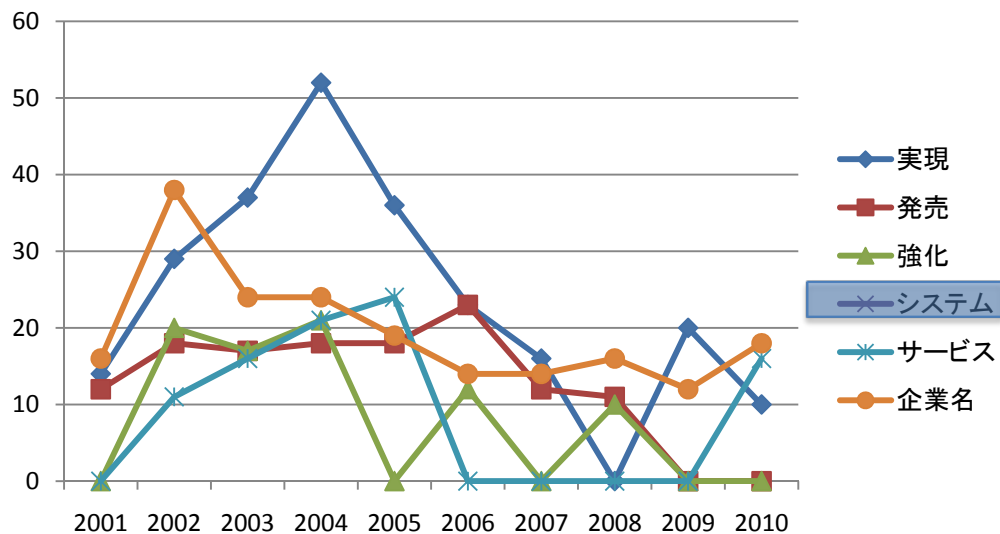


実現との共起

発売(する)行動と、実現の対象がシステムやサービスであるとみることができる。

また、実現の対象は、システムがサービスよりも多い。



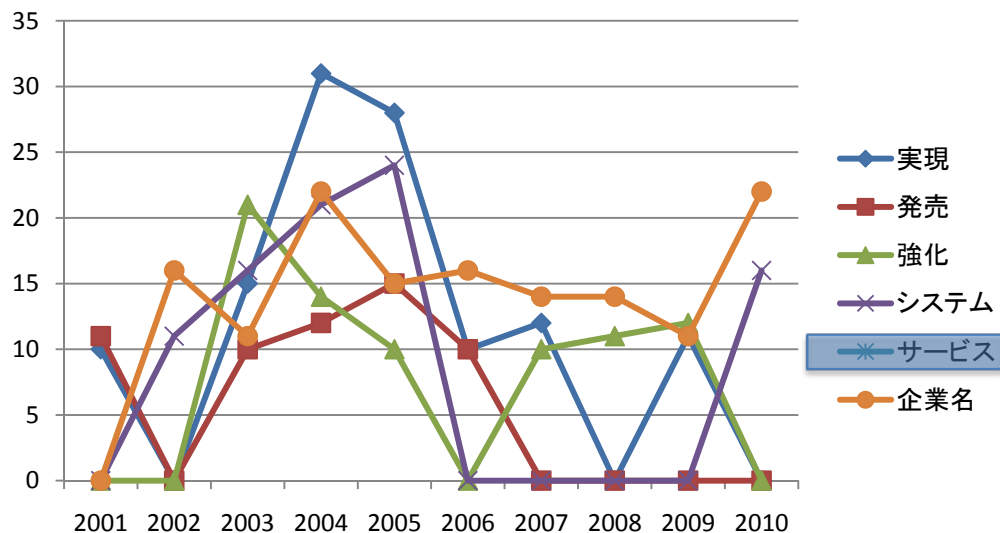


システムとの共起

実現(する)、強化(する)、発売(する)行動の対象になっている。強化は他の2つよりは弱い。システムは何かを実現するものであると主張していると見ることができる。

強化と発売は2009年以降対象にならなかった。

サービスとシステムの関係は2005年までは強かったが、それ以降はあまり強くない。ただし、2010年で復活した。



サービスとの共起

実現(する)行動の対象になっている。

しかし、対象となる行動についてはかなりばらつきが多い。(0個の年、と多い年に差が見られる。)

(4) バリューチェーンと「システム」、「サービス」と共起するキーワードの状況

サンプル企業のプレスリリースエントリデータの次数中心性の高いキーワードのうち、この企業が市場に提供しているものとして、システムやサービスが非常に重要なポジションを締めていることがこれまでの考察結果から分かる。ICT企業においては、システムやサービスはビジネスの中心となるアイテムであり、これらは、企画／開発し、営業などの行為を通して、顧客へ提供されていく。顧客に提供されるシステムやサービスとバリューチェーンの状況を、共起するキーワードを用いて表現した。（ここでは、便宜的にバリューチェーンを、「発売時：まだ顧客へ提供する前の段階」、「発売後：市場に投入を開始し、実際に顧客へ提供、納入されている段階」、「その他」のフェーズで分割した。）

①バリューチェーンの判断のポイント

システム、サービスと共起する頻度の高いキーワードのうち、バリューチェーンの状態を表すものとして 以下

のキーワードが抽出できた。なお、各キーワードがバリューチェーンのどのフェーズにあたる ものなのかは以下の通りの分類とした。

<システム>

- ・発売時：開発、発売、販売、提供
- ・発売後：構築、受注、運用、導入、納入、採用
- ・その他：強化、対応

<サービス>

- ・発売時：開発、提供、発売
- ・発売後：構築、運用、導入
- ・その他：強化、対応

②システムの状況

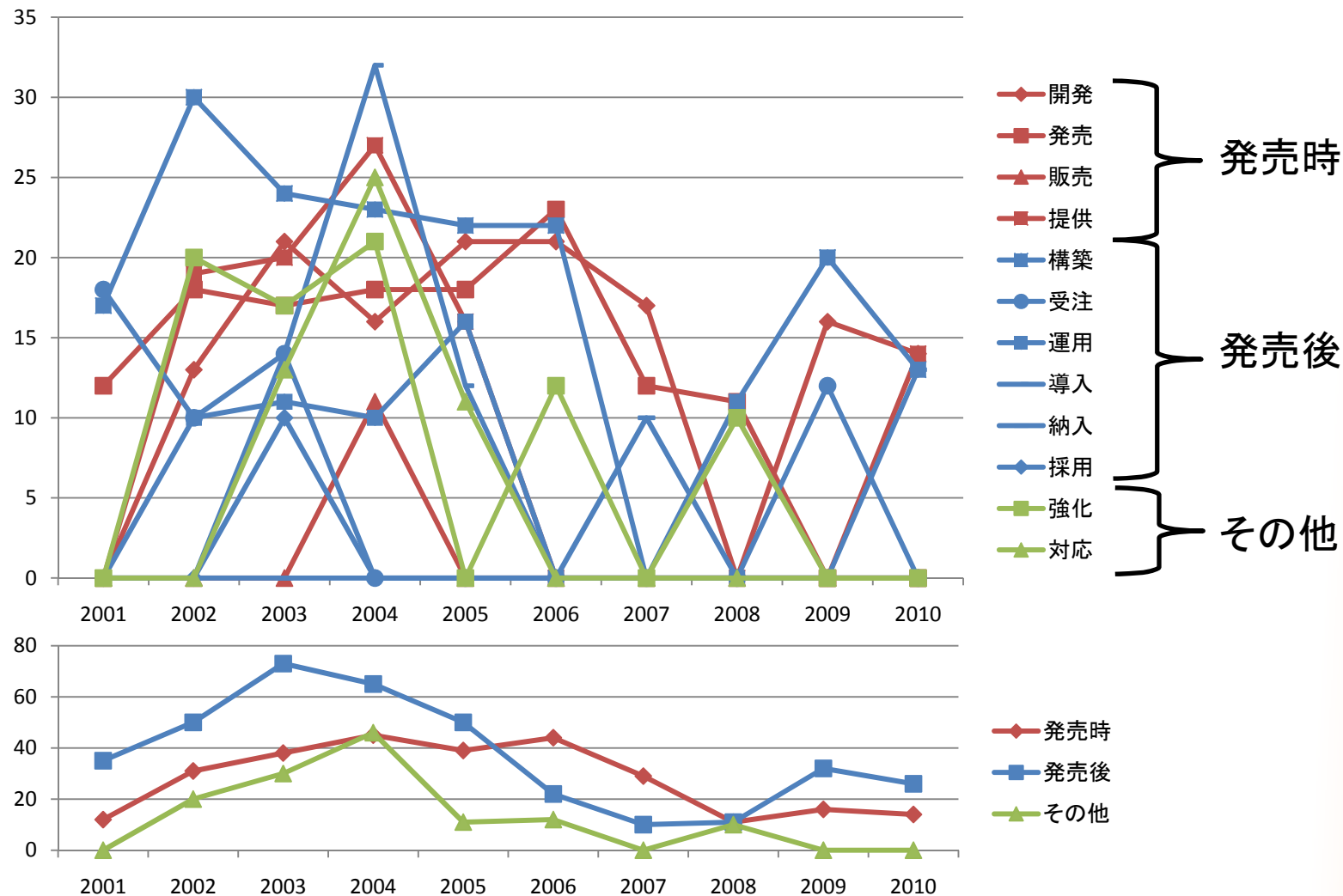
ここで抽出したキーワードの共起状況で発表内容を評価すると、新システムリリースの発表とそれらが売れたことはバランスよく行われてるといえる。

③サービスの状況

ここで抽出したキーワードの共起状況で発表内容を評価すると、2006年以降はサービスが売れている行動の発表は非常に少ないということになる。

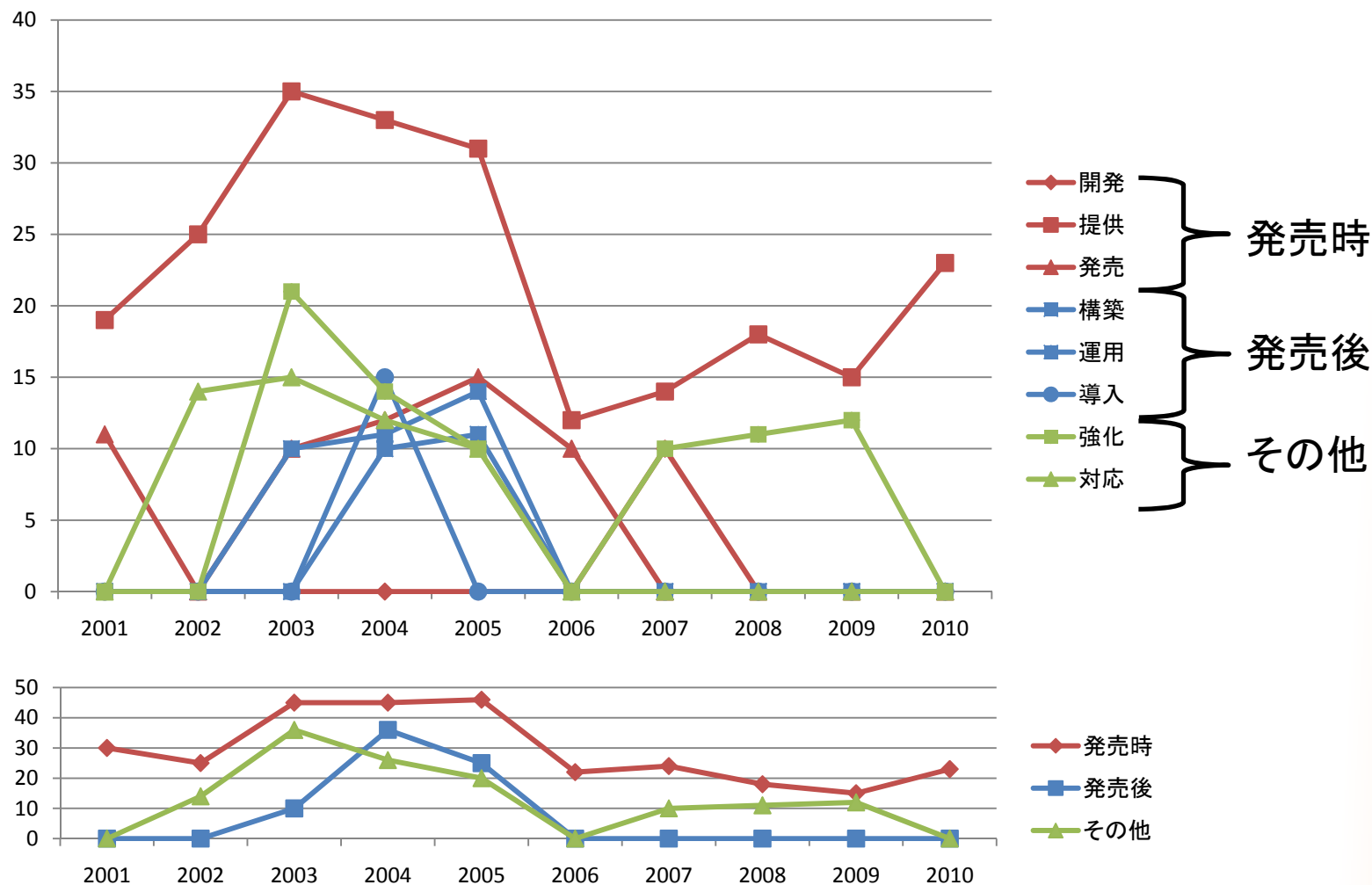
<システムとバリューチェーンの関係>

ここでは、システムと共起するキーワード全てではなく10回以上共起しているVCに関するキーワードのみの関係であるため、極端な結果になっている。

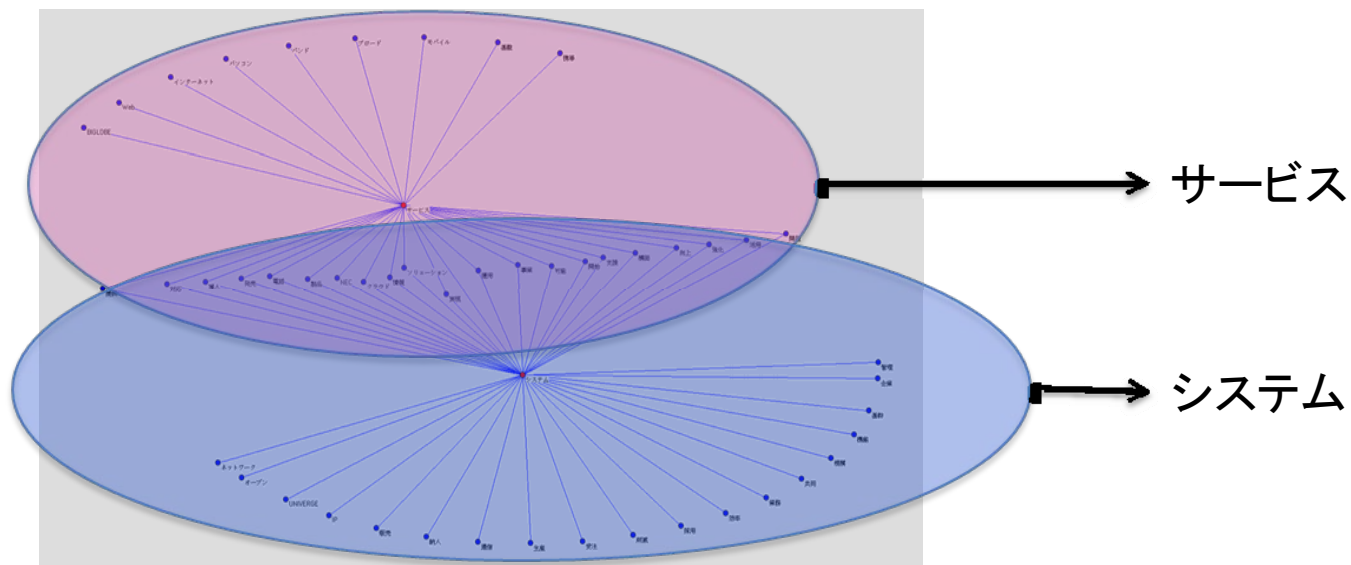


<サービスとバリューチェーンの関係>

ここでは、サービスと共起するキーワード全てではなく10回以上共起しているVCIに関するキーワードのみの関係であるため、極端な結果になっている。



④サンプル企業の方向性:システムか?サービスか?



システムに共起する行動を示すキーワードをサービスに共起する行動を示すキーワード(お互いに、共起頻度の高いもの)を比較した結果、

- ・双方に共通なもの:提供、対応、導入、発売、実現、運用、開始、支援、構築、向上、強化、活用、開発
- ・システムのみ:販売、納入、受注、採用
- ・サービスのみ:なし

との結果になった。

(5) 次数中心性は高くなくても特徴的なものについて

(4)までは、次数中心性の高いキーワードの共起を取り出し、評価を行ったが、キーワードの共起をネットワーク分析を行い可視化した際に次数中心性は低くても、特徴的なパターンを持つ共起のペアを発見することができた。

多くの共起パターンは、次数中心性の高いキーワードを介して様々キーワードにつながっていくが、一部のキーワードの中には特定のキーワードしか共起しないものが存在する。

発見したパターンは、次ページの様なものになるが、これらを分類すると以下の通りとなる。

①製品名等で辞書に載っていないもの

サンプル企業が提供している製品やサービス名、製品と製品の説明語の組み合わせ、新しい言葉

②特定の行動を示すもの

多くは、システムやサービス、製品の発表とは関係なく行われている行動を示すもの。

例. 役員一異動、XX年度一平成一決算一概要

①については、サンプル企業特有であり、形態素解析辞書への登録を行うべきものと見ることができる。前項で取り上げた、「クラ」、「ウド」等は、昨今のクラウドの登場により新しく用いられるようになったキーワードの典型例と見ることができる。

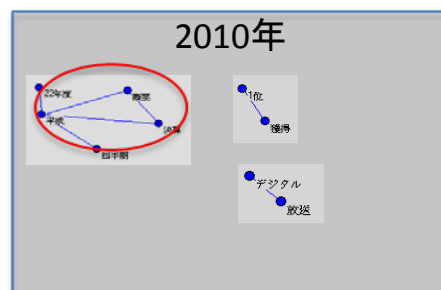
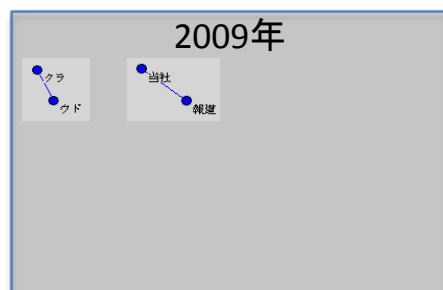
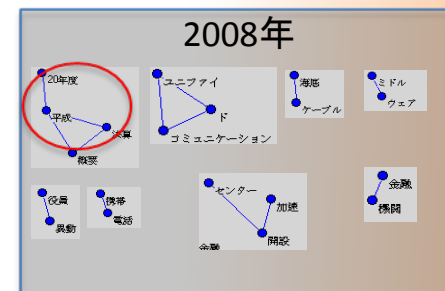
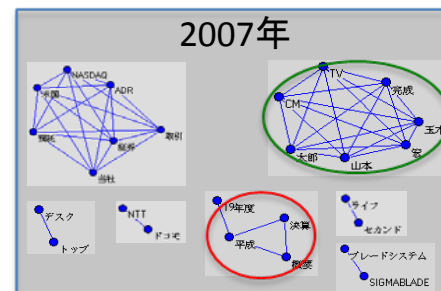
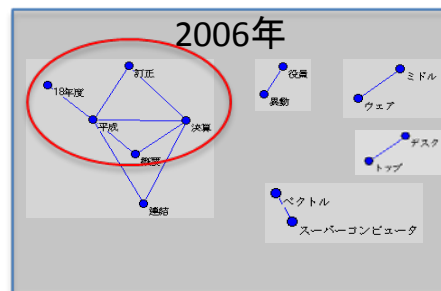
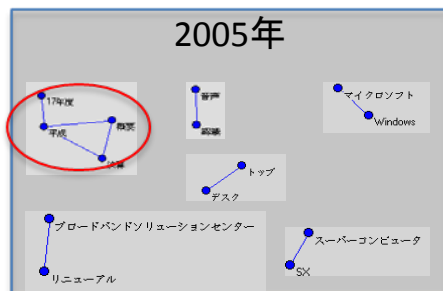
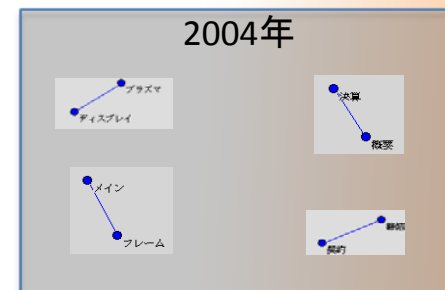
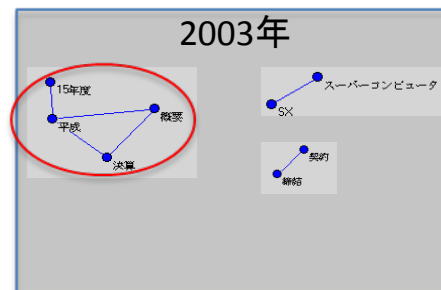
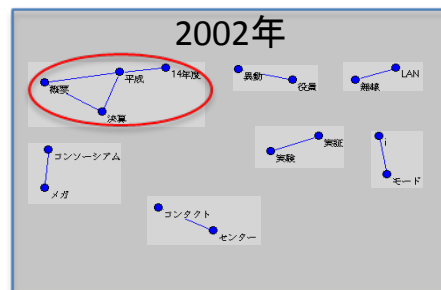
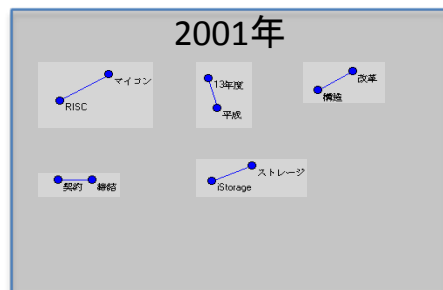
②については、企業のある側面の行動を意味するキーワードの組み合わせであり、ある行動の連想語の組み合わせになっていると見ることができる。

・IR: XX年度一平成一決算一概要

・機構改革: 役員一異動

＜孤立しているが常に同じ共起パターンの行動について＞

以下は、各年の共起ネットワークから、孤立しているパターンを抜き出したものである。



○:「XX年度ー平成ー決算ー概要」のパターンを含むもの

○: 2007年の新CM発表内容

5. まとめ

(1) プレスリリースエントリからネットワーク分析を用いた企業の行動分析

4. までの評価の結果、プレスリリースエントリデータに対して、ネットワーク分析を適用することで、様々な企業の行動を抽出することができた。特に、今回の評価においては、プレスリリースエントリに出現するキーワードの共起関係をネットワーク分析における次数中心性の高いキーワードの観点で、評価を行った結果、サンプル企業が「システム」や「サービス」を事業の中心としていることが見てとれた。実際に、サンプル企業は、IRにおいてもITサービスが売上げの大半を占めており、評価結果を裏付ける内容となっている。

また、次数中心性の高いキーワードに絞って評価を行った場合、キーワードの共起の頻度のボリュームから、当該年の企業行動の状況を把握することができると思われる。例えば、4-(4)で評価した、「システム」と「サービス」というキーワードについてみると。

①「システム」の場合：

- ・新システムの発表よりも、顧客へ提供した実績の発表が多い
- ・2006年、2007年だけは、逆転している

②「サービス」の場合：

- ・新サービスの発表数が、顧客へサービスを提供した実績の発表より多い
- ・2006年以降は、顧客へのサービス提供の発表はあまり行われていない

しかし、実際の行動では②の2項はあり得ないと考えられる。これは、直接的なキーワード(例えば、サービスの発売は、「サービス:発売」、「サービス:提供」のように、サービスに共起する行動で評価している)が出現するものしか、評価していないことが原因であると考えられる。

(2) 行動解析に合致したデータの抽出

今回のネットワーク分析に用いたデータは、プレスリリースエントリデータであった。プレスリリースエントリでは、新聞の見出しのように少ないキーワードで読み手に、イメージを伝えるためより直接的で分かりやすいキーワードを用いる傾向があるとも考えられるため、キーワードもプレスリリースエントリデータに即した形で抽出するなどの方法で、より精度を上げることが可能であると考えられる。

精度を上げるため1つの方法として、ネットワーク分析を行う前に、形態素解析を行っているが、形態素解析を行う際の、抽出キーワードの精度を上げるアプローチがある。

例えば、4-(5)で抽出したもののうち、常に同じ組み合わせで出てくるキーワード同士については、1つの単語として形態素解析の辞書へ登録してしまうことも考えられる。実際に、「クラ」、「ウド」について「クラウド」として辞書登録したところ、「クラウド」とキーワードが抽出されることから、この選択は正しかったと思われる。(相変わらず、「クラ」、「ウド」が共起する場合は、文中に連続して出現しないことであり、1つの単語と

としてみるものが正しくないと言える。)

また、別のアプローチとしては、複数の語が必ず共起する場合は、連想語辞書として管理し、ここに相当するキーワードの組み合わせで出現した場合は、ある意味を持つ組み合わせであると評価する方法が考えられる。

上記も含め、さらにより企業の行動を正確に読み取るためには、キーワード間の意味的な関係を抽出し、複数のキーワードの共起状況から判断する方式が考えられる。これは、今回は2キーワードの共起だが、3キーワード以上の共起などとするのが考えられる。また、直接的なキーワードを含まない(例えば、「サービス」、「システム」)場合の行動分析が可能なキーワードの組み合わせをつくりだすことも可能であればやってみたい。

(3) 他のパターンへの応用について

今回の評価では、キーワードの共起のみを扱ったが、3-(5)で説明した、キーワード間の関係行列を作成する作業を見直すことで、エントリ間の関係行列を作ることも可能である。

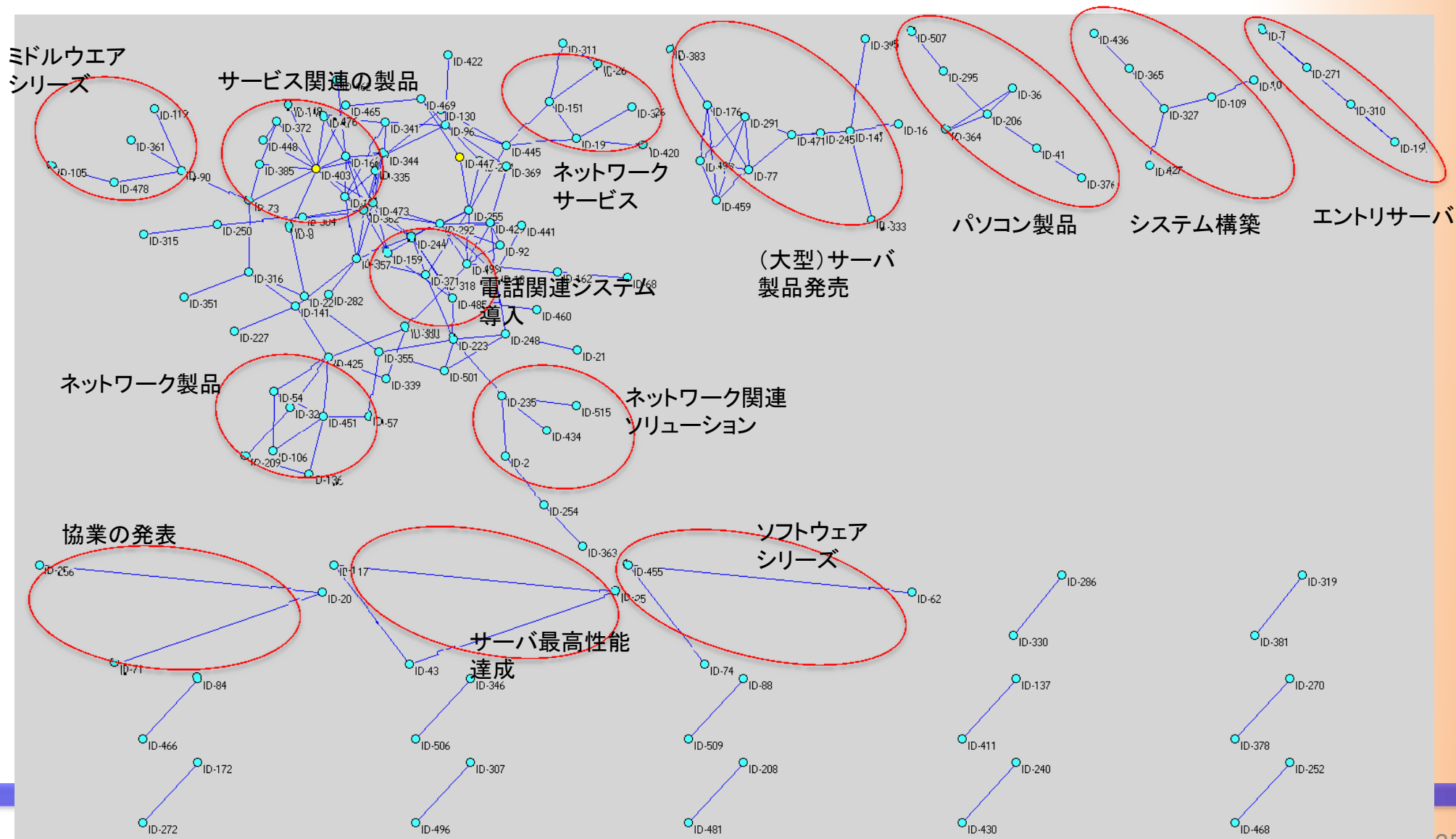
このアプローチを取ることで、その年の発表した内容同士の関係性を見ることが可能である。また、これらの経年変化を追うことで、あるシステムやサービスのバリューチェーンの変化をトレースすることも可能ではないかと思う。

ちなみに、ある年のエントリ間の共起状況をネットワーク分析したものが次ページである。

この年は、発表は518回行われており、1月から発表順にエントリIDを各エントリに付番している。また、ここに挙げたエントリの共起は、共起の際に共有するキーワードが7以上のペアを用いたネットワークを作成している。

赤丸でかこっている部分を見る限り、エントリの類似性について、評価が可能であると考えられる。

＜ある年のエントリー間のネットワーク＞



参考文献

- ・安田 雪(2001)「実践ネットワーク分析」、新曜社
- ・日本数理学会(2004)「社会をくモデル>でみる 数理社会学への招待」、頸草書房
- ・木下 栄蔵(1992)「わかりやすい 意思決定論入門」、啓学出版
- ・岡 宏樹(2009)「プレスリリースのキーワード検索に基づく企業戦略情報の獲得」、第7回日本データベース学会年次大会
- ・鮫島 正樹(2007)事業キーワードに関係する数値と時期記述の共起条件を用いたプレスリリースからの時系列数値データ抽出方法」、電気学会論文誌C Vol.127 No.6 2007
- ・中居 隆(2009)「テキストマイニング(データマイニング)技術紹介」、特技懇 252号