

## 2016年度開催の研究委員会報告にみる地震防災の研究動向

東濃地震科学研究所 太田 裕

### *Characterization of Recent Research Topics on Earthquake Disaster Prevention Issues Based on Presentations at the 2016 Meeting promoted by the Institute*

Yutaka Ohta

Tono Research Institute of Earthquake Science

#### 概要

2016年7月に実施された地震防災研究委員会では、従来とは違って、伝統の理工学を大きく越え、多方面にわたる研究者20名強による研究成果が報告された。開催が2016年熊本地震の直後であったことから、必然この地震に注目した課題が多いが、一方、これらの発表に関わる課題群を質的資料とした上で—近年発展の著しい—Text Miningの手法を多用することで、発表の全体像把握に加え、さらに地震防災分野の現況とか近未来における拡がりや深さについて理解を進めるべく可能性があり、その予備的考察を行った。このようなアプローチは地震関連分野では未だ活用は未だしであるが、本研究は同様の手順・手法が今後さらに多用されて然るべきことを主張した研究でもある。

#### Synopsis

*This short paper aims at elucidating current trend on earthquake disaster prevention-related studies in positive use of recently developed Text Mining technology, since the basic information prepared by the 20 speakers, under the specific situation executed in a short time later than the occurrence of the 2016 Kumamoto earthquake, are given in terms of non-metric form at the special research meeting held in the Institute in the late July last year. What found are summarized as follows; first as that the earthquake disaster prevention discipline traditionally and so commonly accepted has gradually been renewed over the well-known scope and so variety of professionals should be welcome from the areas far over than the existing research field. Secondly, we ascertained the significance of use of the Text Mining technology for better understanding of non-numerical data and the necessity of positive use for plausible solution findings toward the future.*

#### 1. はじめに

本論は2016年7月に実施された当地震研究所防災分野の20編に及ぶ発表話題について、報告内容を話題毎にKeywords化した自然言語からなる「定性情報群」を用意することで、従来の量的分析限定とは違った分析を実施する手法-Text miningと総称される—の適用事例であり、通例の数値情報のみから想定される領域を越えて地震防災分野の近年の変容(=拡がりとか深さ等)について理解を深めようとする場合等、従来のアプローチとは一線を画す分析法として有用となるとの思いから考究を続けており、

このためのツールとなる *Text mining* 法の活用に向けて、*Data* の取得に始まる、一連の手順、そして成果の一端を記述しておく。

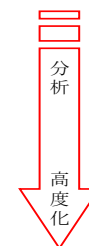
## 2. *Text mining* の手法：概観と本論における活用手順

まず、第 1 表を見ていただきたい。これは *Data* 処理の一般法について纏められたものであり、*Text mining* 技術の種類を難易度によって段階別に整理したものである。これらのうち、レベル 1 とか 2 については、対象がもっぱら数値 *Data* であり、よく知られている手法である。一方、レベル 3 の諸技術はいずれも「非数値情報を含む」*Data* 処理を可能とする手法であり、近年発展したものが多く、

表 1 *Text mining* 技術の段階別区分 (那須川, 2006)

処理のレベル	処理の概要	技術的要素
レベル 1	検 索	情報検索
レベル 2	分 類 整 理	基礎統計 系列分析
レベル 3	分 析 (新知見獲得)	多変量解析 自然言語処理 視覚化表現

那須川: テキストマイニングを使う技術・作る技術(2006)を簡略化



本論ではレベル 3 の手法のうち、特に「自然言語処理」を可能とし、さらに視覚化表現も可能な *Text mining* 手法の 1 つ「Polaris と総称」を活用させていただくこととしている (大澤, 2006 等)。ちなみに、筆者はこの手法を使って今までに地震災害分野のうち、とくに地震に伴う死者発生の実状を世界的視点で捉えるべく、医学分野で周知の PubMed (Public Medicine, National Inst of Health, USA) 活用による分析実施の経験を持っている (太田他, 2008)。また、我が国の地震工学研究の先達である金井清先生の研究活動を総括するための手法として活用した経験をもっている (太田, 2009)。本論はこういった経験を背景として研究展開されている。

### 2.1 定性的 *Data* の取得：方法一般と今回の場合

当該研究委員会では「地域への公開講演を含めて」20 人近い研究者が 20 分/人程度の発表を実施しており、このときの発表内容を分析の原資源として話を進めていく。しかし、各人の発表内容をそのままに入力情報とするのは冗長に過ぎるし、内容の確保 (全記録の収録) は容易ではない。それ故、何らかの代替え案導入が必須となる。この有力な手段として自然言語処理 - *Text mining* - の援用法がある。一般に、(非数値からなる) 文書情報を入力とする自然言語処理の手順は大きく 2 段階に区分される。前段は原文書から単一語句の抽出に向けて構文分析を行うことであり、単一語句群に分解する作業がこれに続く。そして、いずれもが *Text mining* 手法適用に向けて不可欠の前段準備である。しかし、今回は、話題提供者全員に対して話題毎に『名詞系の語句集合 (Keywords)、つまり、各 30 語句の提出』をお願いすることで前段作業の代替えとさせていただいた。特に、能島氏 (岐阜大) には通常発表者の 3 倍程度 (~100 語) を用意していただくよう、特段の依頼をした。これは公開説明会で 1 時間弱にわたる特別講演を依頼していたことから、必然関係する Keywords 群も多彩となることを考慮したためである。そんな次第で、第 1 段の語句抽出作業はほぼパス出来ることとなった。なお、話題提供者全員の氏名・所属は末尾 (=謝辞の部) に記載させていただいた。

この結果をそのまま活用することで、主題となる語句間関係 (重要度、結合関係) の分析、そして背後に潜む「重要情報」を浮き彫りにしてゆく作業に直ぐにも移行できる筈であるが、実際には送られてきた Keywords 群が話題提供者毎に独自の判断で選別されていることから、本分析に先立って、若干の吟味が必要となる。

余談ながら、関係資料準備の最も手軽な手段として、各人が提案の「発表課題名をストレートに名詞化したものの組」を Keyword 群集合とする方法があり、これを簡便法として位置付けることも可能では

ある。しかし、この略式では多くの場合—情報量不足のため—、相応の成果しか期待できず、他に手段がない場合の「代替措置」として位置付けられるべきものであろう。

## 2.2 届けられた Keyword 群とその微細調整・度数分布等

上記の手順を経て、19人の方々から「30 Keywords/人」を平均値とする 600 語近い Keywords が用意された。これに能島氏からの 100 語を加えることで、合計で 700 個を越える Keywords 群を、研究委員会の後に集めることが出来た。第 2 表に各氏から Excel 形式で送付された原 Keywords 群の一部を載せておく。

こういった表では 1 行が 1 人分に相当し、30 語程度の Keywords が並ぶ。従って、こういった Keywords 群を原 Data として直ちに Text mining ソフトへの入力情報とすることも可能ではあるが、個別 Data に立ち入ってみると—各自が適宜(自由に) Keywords 群を与えているために—内容的には全く同一の対象(事象)が別の言語表示となっている場合が散見される。例えば、“人的被災度”と”人的被害度”とか、“避難所”と”地区避難所”等々、少なからざる語句が少しずつ違った形で届けられている。こういった問題は”話題提供者から送付された原 Keywords 群を個別に調べることで事前の”修正(ないしは同一化)”を済ましておく必要があり、当面、筆者自身が手動で改訂を試みた。そういった作業を終えた後の、Keywords 群に対する度数分布を第 1 図 (a), (b) に掲げておく。

第 1 図 (a) の高頻度 Keywords 群をみると第 1 位が「避難問題」となっているが、実際には講演者から届けられた Keywords は前述のような類似関係にあるものを筆者が集約した結果の全体数として扱ったものであることを特に注意しておきたい。他の Keywords 群についても適宜同様の少修正を行っている。2位の京都大は報告者数が多かったことを反映した結果である。次いで、3位が人的被害、4位が能島氏、5位が熊本地震等々となっている等、研究委員会直前の熊本地震発生を反映した状況が伺われる。参考までに、第 1 図 (b) には、出現項目数 2~1 の度数分布を掲げておくが、本論ではこれ等に対しては立ち入った議論はしない。ともあれ、上記の度数分布図はそれ自体が各 Keyword の重要度の 1 表現ではあるが、これらの図のみから読み取れる情報量はかなり貧弱なままであり、今一段の工夫が望まれるところであり、これに応えるべく(レベル 3 に属する) Text mining の手法が開発されているわけである。以下、大澤(2006)の説明を参考にしながら、Keyword 毎の重要度の判断・表現法について簡単

表 2 送付された原 Keywords(の部分).

2016熊本地震	避難問題	避難問題	災害復旧	益城町	家具固定	ベッド同伴	小学校	防災教育
2016年台風	構造物被害	地盤変動	建物被害	耐力不足	地震動予測	微動計測	文化遺産	台南台地
2016熊本地震	被害想定	経済損失	直接被害	間接被害	推定式	少子化	高齢者化	南海トラフ
津波災害	2011年東日本	死亡率	年令特性	算定法	吟味	改善	東北3県	沿海域
微動計測	地中レーザ	鹿野断層	地盤変動		位相速度	地震動予測	震源断層	地質
避難問題	障害者	石巻市	民生委員	アンケート	インタビュー	地震動予測	要援護者	協力委員
ネパール	パタン	組構造	ゴルカ地震	地震動予測	世界遺産	微動計測	フリー変	固有振動
熊本震災	医療機関	DMAT	医師会	自衛隊	知事会	医師	看護師	薬剤師
2016熊本地震	木造	損傷度	暴露	人体損傷	避難問題	軽減率	死亡率	負傷
2016熊本地震	高齢者居住	人的被害	関連死	居住環境	室内被害	家具固定	高齢者	キッチン
2016年201	益城町	人的被害	要因	関連死	負傷者	全壊率	半壊	現地調査
個人世帯	身体的被害	資産的被害	精神的苦痛	円	ISS	K6	BGP	DCP
事業継続	地域特性	地区防災	多様性	機能継続	連携	戦略	ボトムアップ	提案制度
2016熊本地震	人口衛星	合成開口	画像	地盤変動	被害想定	干渉SAR	変動量	益城町カラ
2016年201	新聞報道	現地紙	全国紙	地元紙	新聞記事	記事内容	地震動予測	斜面崩壊
内陸地震	潜在断層	地震動予測	地表断層	建物被害	岩手宮城	福島県浜	長野神城	潜在断層
津波災害	避難問題	地域特性	石巻市	シミュレー	犠牲者	2011年東	逃げるこ	気象庁
リスク評価	津波災害	地震動予測	災害復旧	人的被害	グローバル	構築	災害復旧	2011年東
神戸市	防災コム	寄与	地域特性	活動特性	危険認識	課題点	要援護者	支援地震
避難問題	メカニズム	断層タイプ	布田川断層	日奈久断層	30年発生	被害想定	前震	本震
継続時間	兵庫県南	東北地方	KK-net	緊急地震	余裕時間	耐震化率	新耐震設	宅地被害
無鉄筋	地壊	建築基準	耐震改修	耐震補強	供給系	ライ	復旧・復興	電気
送電線	応急復旧	応急送電	一般道路	国道	県道	交通途絶	災害復旧	カーナビ情
老朽化	避難問題	医療機関	軒先避難	車中泊中	テント	体育館	支援物資	避難問題
防災士	防災リーダー							
沿道耐震	公助	自助	共助					

に列挙しておく。以下の通りである。

- 出現頻度 (*Term Frequency*) が高く、かつ共起性 (*Co-occurrence*) の高い語句を重要語句とみなす。つまり、出現度数に加えて、共起性を加味した形で「Keyword がもつ重要度」を評価するところに大きな特徴がある。それ故、従来法を越えた内容をもつ出力が期待できることとなる。
- 重要語句を核とするクラスター群を 2 次元表示することで「語句群間」の関係を可視化できる。
- 結果は語句をノードとして重要度を「黒、赤等」に色わけする。また語句間共起度の度合いに応じて結線 (リンク) することとし、語句間関係の全体を 2 次元グラフで視覚化表示し、これを *KeyGraph* と総称する。なお、(高頻度)、赤丸 (低頻度) の個数とかリンクの本数は適宜の値 (正の数) を指定できる。

### 3. 当該課題の KeyGraph 表示

上記の定義に従うことで、と赤丸の個数を適宜変えることによって種々の *KeyGraph* を描くことが出来る。ここでは易より難への常套法に沿う形で、ごく少数の重要 *Keyword* に注目した場合から話を始めるのが妥当であろう。

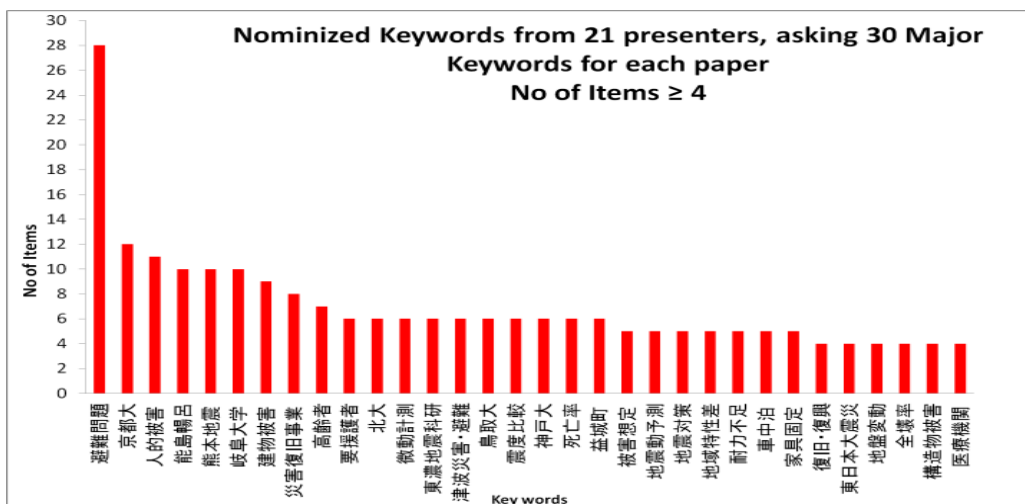


図 1a 改訂済み Keywords 群の度数分布 (項目数 4)

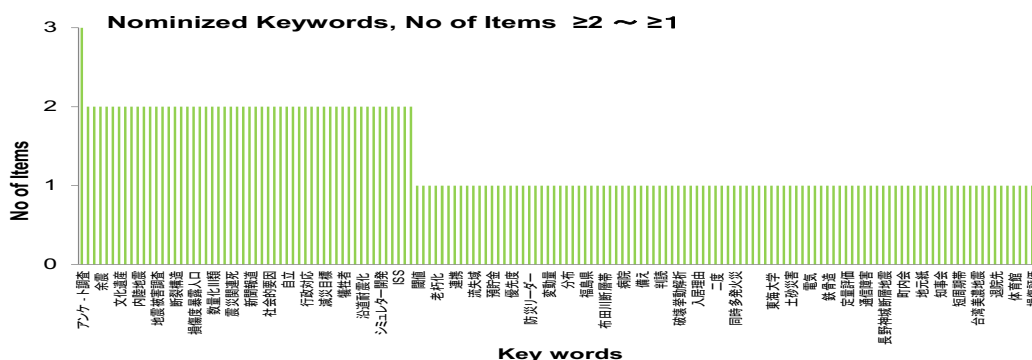


図 1b Keywords 群の度数分布 (項目数 2 ~ 1).

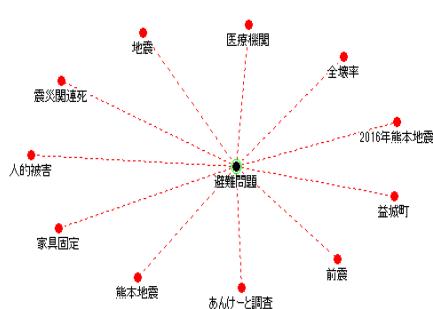


図2 ● 1点、● 1点とした場合.

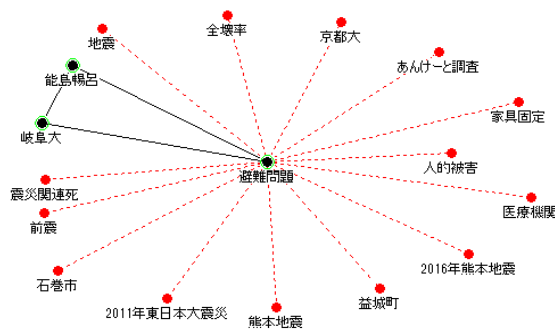


図3 ● 3点、● 14点とした場合.

### 3.1 少数の重要 Keyword に注目した場合

内容とか課題間の関係等はいきなり複雑な KeyGraph を描いても容易に把握できるものではない。「易より難に向かって進む」のが常道であろう。以下の KeyGraph 群はそういった方針で、重要 Keyword と位置付けている一の個数を”1”に始まり、順次増やしながら状況変化をみていくのが順当であろう。

そこで、先ず●1点（避難問題）を囲んで●11個の場合から話を始めることとする。これが第2図である。これにみるように第1義的に重要な Keywords 群 [●印] は避難問題となっており、度数分布図（第1図）の Top となった語句である。これが●印として「熊本地震、益城町」他の Keyword 群と one set で繋がる形となっている。第3図は●印を3点とした場合を対比的に並列表示してみたものである。ここでは「避難問題、能島暢呂-岐阜大」が●群を形成する。一方、●印として東日本大震災が新たに浮上していること、つまり、複数の地震が対象となっていることなど、また京都大の活躍の程も見て取れる形へと変化している。その他、「人的被害」、「家具固定」、「前震」、「全壊率」等など、在来枠に属す周知の Keywords 群が先行する様子が顕著にみられる。

以下、同様にして●を漸増し、併行して●を増やした場合の KeyGraph (2次元図) を以下に示す。

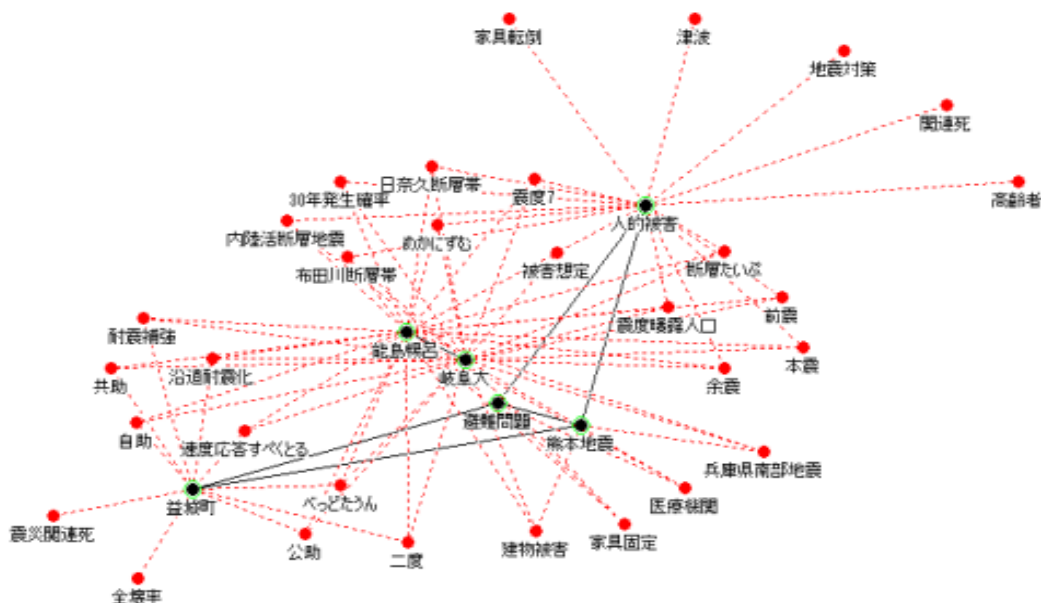


図4 ● 6点、● 30点とした場合

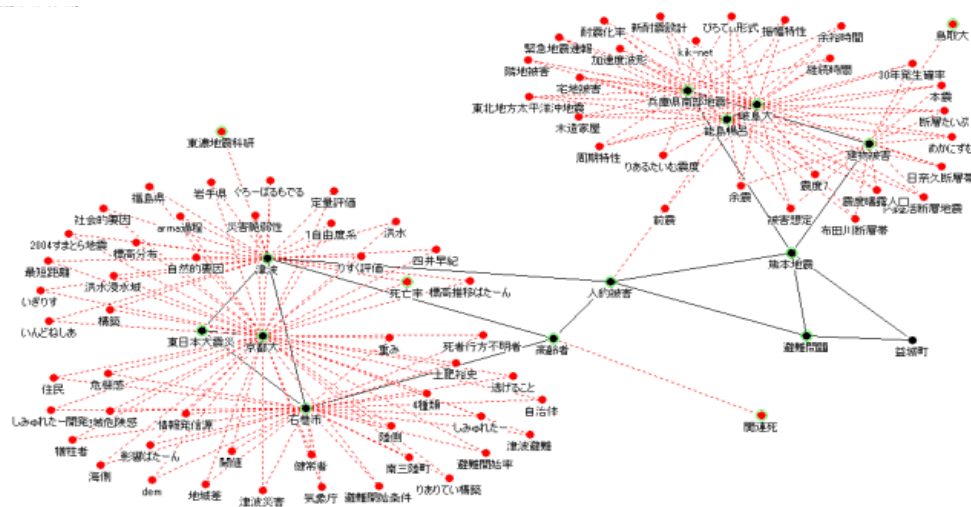


図5 ●15点, ●80点とした場合

第4図は●6点, ●30点とした場合の KeyGraph である。このとき重要語句となる●は「避難問題」を軸として「熊本地震・益城町」が続き、一方「人的被害」が他の地震を挟む形で出現している。能島(岐阜大)は当該研究委員会における長時間報告を反映したものであろう。一方, ●に注意を向けると兵庫県南部地震他の内陸起震断層系への関心とか、津波問題も議論の対象となっていることが判る。

### 3.2 ●を漸増させた場合

図5は●15点, ●80点とした場合の KeyGraph である。このとき重要語句となる●の分布は大きく左右に2分されており、この図では右側が「熊本地震、兵庫県南部地震」等に関わる語句群となり、一方、左側は主に「東日本大震災」関連の Keywords 群となっている。そして、これら両者の接合点に「高齢者・人的被害」項が位置し、古い/新しい地震群を繋ぐ共通項の役割を果たしている様子が明瞭である。このように一人一人の提供課題の多くは”深さ方向の考究を目指すものの”, そういった個別研究を2次元グラフに投影することで”地震防災課題がもつ広がり”を具体的に描き出すことが出来るといえる。

次いで、図6に●20点, ●80点とした場合を示しておくが、左側が東日本大震災関連の諸研究となるのに対して、右側が熊本地震・兵庫県南部地震関連の研究群となり、両者を結ぶ主要 Keyword として「人的被害、高齢者等々」となる一般性は●15点の辺りで既に確定しているように見える。従って、●, ●1をさらに増やしても本課題における KeyGraph の骨格構成はほとんど不変と思っていいていであろう。逆に Graph が赤・黒の個点数が増えることによって却って見かけ上複雑になり過ぎる恐れもあろう。

## 4. 結果と今後の課題

### 4.1 結果の概要

大澤らによる *Text mining Soft* を活用させていただくことで、昨年7月に当研究所が実施した「地震防災関連研究委員会」における20人に及ぶ報告内容を“地震防災分野の広がりと深さを探る”といった視点で分析を試み、かなり目的に近い成果を得ることができた。そして、その結果を KeyGraph という2次元座標上に—Keywords 群を重要度に応じて「●, ●」表示することで視覚的認識にも耐える形で—実現することが出来た。これを一口で言えば、「●となる Keywords を20個程度用意することで、話題提供者集団が主題とするところ、そして内容についても」相当程度まで理解できることである。

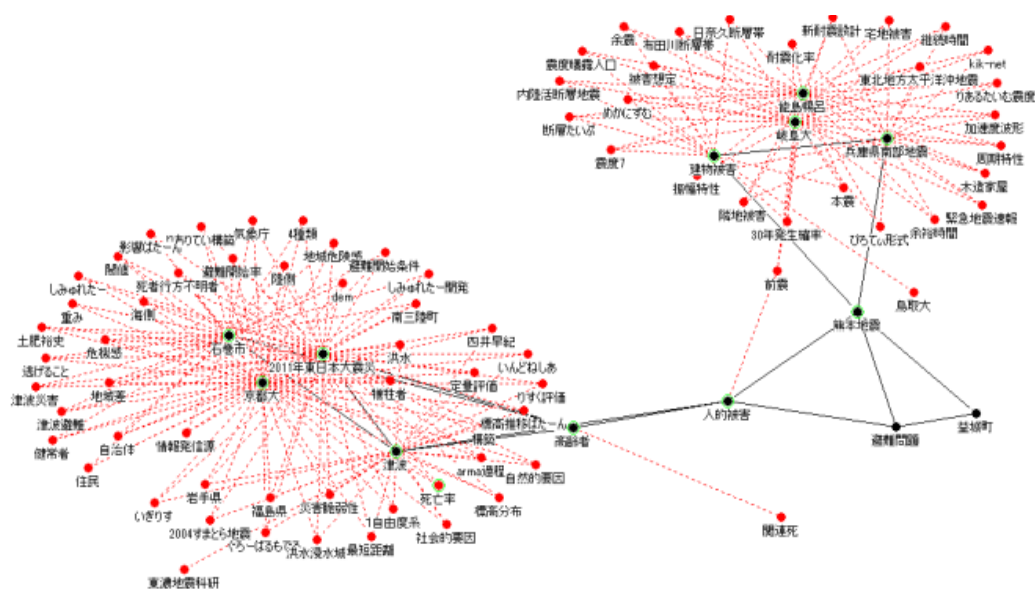


図6 ● 20点, ● 80点とした場合

## 4.2 今後の課題

今回の試行で一番の問題は話題提供者から届けられた *Keywords* 群の事前整理である。この問題は小さい。これは方針さえ決めてしまえば、後は力仕事で済むが故に時間労力の問題に還元されることから本質的な問題とはならない。

しかし、今回は取り上げなかったが、*Keyword* 群整理の段階で大きな問題が残っている。端的に言えば *Keywords* 間の階層構造を全く無視している点である。本題の事例でいえば、筆者が簡単のためということで避難関連の諸事項（実際には「避難誘導」に始まり「車中泊」等々に至る 10 数語がある）等がある。また断層関連でもやはり（地表断層、潜在断層、正・逆断層）等々のように多数の *Keywords* 群がある。つまり、各話題提供者による *Keywords* 群は実は”何段階かの階層構造を持っている”のが実情である。今後、検討すべき重要問題の 1 つである。

## 謝辞

この報告は東濃地震科研の地震防災研究委員会開催時（2016年7月）に発表いただいたメンバーの方々に話題提供毎に 30 語前後の *Keywords* 群を用意していただくことをお願いし、その集まりを原 *Data* として活用させていただいた。こういったお願いに協力いただいたのは次の方々である。以下、お名前と所属を列挙して謝意を表明しておきたい（敬称略、順不同）。・鈴木光（減災アトリエ）、清野純史（京都大）、野口竜也（鳥取大）、大西一嘉（神戸大）、古川愛子（京都大）、村坂建史（金沢医大）、中嶋唯貴（北海道大）、志垣智子（高齢者住宅研）、村上ひとみ（山口大）、岡田成幸（北海道大）、磯打千雅子（香川大）、松岡昌志（東工大）、北浦勝（金沢職人大）、大保直人（地震予知振興会）、吉田昌平（鳥取大）、土井裕史（京都大）、四井早紀（京都大）、岡田尚子（神戸大）、能島暢呂（岐阜大）の諸氏と筆者自身を加えた 20 名である。なお、当該研究委員会を主導された谷口仁士（東濃地震科研副首席主任研究員）は当該研究委員会直後に急逝されたため、上記メンバーの一員としてのお願いは出来なかった。

一方、こういった *Keywords* 群を原 *Data* とした文書分析には大沢幸生教授（東大工学部）主導の *Text Mining Soft* - 呼称” *Polaris* -”，を活用させていただいた。記して、厚くお礼申し上げます。

## 参考文献

1. 那須川哲哉, テキストマイニングを使う技術/作る技術, 東京電機大学出版局, 1-236, 2006.
2. 大澤幸生, チャンス発見のデータ分析, 1-273, 東京電機大学出版 1-273, 2006.
3. 太田裕・野添篤竹・榊原真奈美, 地震に起因する文献学研究 (1 ~ 3), 東濃地震科研報告, Seq. No 22, 271-3930, 2008.
4. 太田裕, 資料: 金井先生の *Academic Activities* を概観する - 文献書誌学的視点から - 東濃地震科研報告, Seq. No 26, 187-198, 2010.