



**HAL**  
open science

## Strength of Relationship Between Multi-labeled Data and Labels

Masahiro Kuzunishi, Tetsuya Furukawa

► **To cite this version:**

Masahiro Kuzunishi, Tetsuya Furukawa. Strength of Relationship Between Multi-labeled Data and Labels. 3rd International Conference on Information and Communication Technology-EurAsia (ICT-EURASIA) and 9th International Conference on Research and Practical Issues of Enterprise Information Systems (CONFENIS), Oct 2015, Daejeon, South Korea. pp.99-108, 10.1007/978-3-319-24315-3\_10 . hal-01466209

**HAL Id: hal-01466209**

**<https://inria.hal.science/hal-01466209>**

Submitted on 13 Feb 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Strength of Relationship between Multi-labeled Data and Labels

Masahiro Kuzunishi<sup>1</sup> and Tetsuya Furukawa<sup>2</sup>

<sup>1</sup> Faculty of Economics, Aichi Gakuin University  
Meijyo 3-1-1, Kita, Nagoya 462-8739, Japan  
kuzunisi@dpc.agu.ac.jp

<sup>2</sup> Dept. Economic Engineering, Kyushu University  
Hakozaki 6-19-1, Higashi, Fukuoka 812-8581, Japan  
furukawa@econ.kyushu-u.ac.jp

**Abstract.** Collected data must be organized properly to utilize well and classification of data is one of the efficient methods. Individual data or an object is classified to categories and annotated with labels of those categories. Giving ranks to labels of objects in order to express how close objects are to the categories enables us to use objects more precisely. When target objects are identified by a set of labels  $\mathcal{L}$ , there are various strength of relationship between objects and  $\mathcal{L}$ . This paper proposes criteria for objects with two rank labels, primary and secondary labels, such as a label relates to  $\mathcal{L}$ , a primary label relates to  $\mathcal{L}$ , every primary label relates to  $\mathcal{L}$ , and every label relates to  $\mathcal{L}$ . The strongest criterion which an object satisfies is the level of the object to express the degree of the strength of relationship between the object and  $\mathcal{L}$ . The results for two rank objects are extended to  $k$  rank objects.

**Keywords:** Multi-labeled Data, Ranks of Labels, Levels of Data, Criteria for the Strength

## 1 Introduction

With increasing various kinds of data such as numerical data, texts, images, and movies, utilization of collected data is becoming more important [4] [11] [12]. Such data must be organized properly to utilize well and classification of data is one of the efficient methods [3] [8]. Individual data is classified to a category by a certain attribute, for example, region or industry, and the data is annotated with the label of the category [1].

When data relates to multiple categories, the data is classified to those categories and annotated with the set of labels [9]. However, the strength of relationship between data and categories is different in general. Ranks of the labels which express how close the data is to categories enable us to use the data more precisely. Suppose that data are classified by business categories and annotated with sets of labels of the categories. If ranks are given to the labels of a data according to the financial figures (net revenues, net income, etc.) of the relating

business field, users can compare the data of automobile and finance as primary categories with the data of automobile as a primary category and finance as a secondary category.

Although ranks of labels give us richer information, target data identified by a set of labels are not clear because there are various strength of relationship between the data and the set of labels. Even if users need the data which relate to a set of labels more closely, there are cases that the strength can not be decided.

*Example 1.* For the strength of relationship between data and a set of labels  $\mathcal{L}_1 = \{Transportation, Electronics\}$ , it is not decided clearly whether the data on the company where automobile, mobile phone, and finance are primary relates to  $\mathcal{L}_1$  more closely than the data on the company where only automobile is primary and mobile phone and finance are secondary. While the former data relates to  $\mathcal{L}_1$  more closely if the primary categories related to  $\mathcal{L}_1$  are evaluated before the unrelated category to  $\mathcal{L}_1$ , the latter data is closer to  $\mathcal{L}_1$  if the primary categories after the unrelated category.

Criteria for the strength of relationship between data and a set of labels enable to evaluate the degree of the strength. The purpose of this paper is to propose such criteria by discussing the strength of relationship between data and a set of labels. This paper also refers to the strength of the criteria and introduces levels of data, which is decided by the strongest criterion that the data satisfies.

Ranks in classification of data have been studied in the field of information retrieval mainly. Those researches focus on the accuracy and the efficiency of automatic ranking of data [5] [6]. After ranking, personal preferences of queries are measured and the ranks of data based on those numerical values are applied to the advanced information retrieval [10]. Many researches focus on ranking data known as *Top-k*, where significant data are selected by measuring the degree of relationship between data and a set of labels [2] [7]. While their purpose is to evaluate data on a specific domain quantitatively to develop applications, the purpose of this paper is to build fundamental theories of multi-labeled data by evaluating data qualitatively.

This paper is organized as follows. Section 2 refers to the strength of relationship between data without ranks and a set of labels, and introduces criteria to decide the strength. In Section 3, criteria for data with ranks are discussed. Section 4 proposes the strength of the criteria and the levels of data to express the degree of the strength. The discussions of Sections 3 and 4 are for two rank data, which are extended to  $k$  rank data in Section 5. Section 6 concludes the paper.

## 2 Strength of Relationship between Objects and Labels

Individual data or an object is classified by a certain type of characteristic, which is called an attribute. For example, an object is classified to manufacturing, transportation, automobile, etc., where the attribute is industry. This paper

assumes that an object is classified to the lowest category to which the object is related in a given classification hierarchy.

Let  $o$  be an object,  $L$  be a label, and  $\mathbf{L} = \{L_1, L_2, \dots, L_n\}$  be a set of labels. When  $o$  is classified to multiple categories,  $o$  is annotated with the set of labels of those categories, which is denoted by  $L(o)$  and called a set label of  $o$ .

For labels  $L_1$  and  $L_2$ ,  $L_1$  is lower than  $L_2$  ( $L_2$  is higher than  $L_1$ ) if the category of  $L_1$  is a lower concept of the category of  $L_2$ , and  $L_1 \preceq L_2$  denotes that  $L_1$  is lower than or equal to  $L_2$ .  $L_1$  relates to  $L_2$  if  $L_1 \preceq L_2$  and a label  $L$  relates to a set of labels  $\mathbf{L}$  if  $L$  relates to a label of  $\mathbf{L}$ . Label *Transportation* relates to label *Manufacturing* because *Transportation*  $\preceq$  *Manufacturing*.

For sets of labels  $\mathbf{L}$  and  $\mathbf{L}'$ , the set of the labels of  $\mathbf{L}$  which relate to  $\mathbf{L}'$  is described by  $Rel_{\mathbf{L}'}(\mathbf{L}) = \{L \mid L \in \mathbf{L}, \exists L' \in \mathbf{L}', L \preceq L'\}$ . Let  $\mathcal{L}$  be a set of labels to be used to identify target objects. If there is such a label in  $L(o)$  that relates to  $\mathcal{L}$  ( $Rel_{\mathcal{L}}(L(o)) \neq \phi$ ),  $o$  relates to  $\mathcal{L}$ . Let  $\overline{\mathcal{L}}$  be the set of objects relating to  $\mathcal{L}$ .

There are several kinds of objects in  $\overline{\mathcal{L}}$  such as objects which relate to multiple labels of  $\mathcal{L}$ , objects which relate to other labels than  $\mathcal{L}$ , and so on.

*Example 2.* For a set of labels  $\mathcal{L}_2 = \{Transportation\}$ ,  $\overline{\mathcal{L}_2}$  includes object  $o_1$  such that  $L(o_1) = \{Automobile\}$ , object  $o_2$  such that  $L(o_2) = \{Automobile, Motorcycle\}$  where multiple labels of  $L(o_2)$  relate to  $\mathcal{L}_2$ , object  $o_3$  such that  $L(o_3) = \{Automobile, Finance\}$  where *Finance* does not relate to  $\mathcal{L}_2$ , etc.

There are various ways to decide the degree of how strong an object  $o$  relate to  $\mathcal{L}$ , and the number of labels relate to  $\mathcal{L}$  may not be affected on the strength of relationship between  $o$  and  $\mathcal{L}$ . This paper focuses on the qualitiveness of relationship and does not mention the number of the labels which relate to  $\mathcal{L}$ .

There are several kinds of objects identified by  $\mathcal{L}$ , and the strength of relationship between objects and  $\mathcal{L}$  is different in general. Criteria for the strength enable us to select proper objects according to their purposes.

For objects  $o_1$  and  $o_2$ ,  $o_1 <_{\mathcal{L}} o_2$ ,  $o_1 < o_2$  if  $\mathcal{L}$  is obvious, denotes that  $o_2$  relates to  $\mathcal{L}$  more closely than  $o_1$ . A condition *cmd* is a criterion for strength of relationship between a set of labels and an object if  $o_2$  satisfies *cmd* and  $o_1$  does not for any objects  $o_1$  and  $o_2$  such that  $o_1 <_{\mathcal{L}} o_2$ .

### 3 Criteria for the Strength of Relationship

This section introduces ranks to express how close objects are to the categories. The more number of ranks give us the richer information and allow more precise analysis. This section discusses the cases where the number of ranks is two for simplicity. The discussions for two rank objects are extended to  $k$  rank objects in Section 5.

Let the ranks of labels of set labels be primary and secondary.

- Primary labels : If an object is classified to a category and relates to the category mainly, the label of the object is a primary label.

- Secondary labels : If an object is classified to a category and relates to the category but not mainly, the label of the object is a secondary label.

For an object  $o$ , let  $P(o)$  and  $S(o)$  be the set of the primary labels and the set of the secondary labels of  $L(o)$ , respectively.

*Property 1.*  $L(o) = P(o) \cup S(o)$ .

*Property 2.*  $P(o) \cap S(o) = \phi$ .

*Property 3.*  $P(o) \neq \phi$ .

The rest of this section discusses criteria for two rank objects. It is reasonable to think that an object  $o_1$  relating to a set of labels  $\mathcal{L}$  ( $Rel_{\mathcal{L}}(L(o_1)) \neq \phi$ ) relates to  $\mathcal{L}$  more closely than an object  $o_2$  not relating to  $\mathcal{L}$  ( $Rel_{\mathcal{L}}(L(o_2)) = \phi$ ). Thus the condition whether a label of  $L(o)$  relates to  $\mathcal{L}$  is a criterion for the strength of relationship between  $o$  and  $\mathcal{L}$ .

$cmd_{LE}$  : For a set of labels  $\mathcal{L}$  and an object  $o$ , there exists a label of  $L(o)$  which relates to  $\mathcal{L}$ , that is,  $Rel_{\mathcal{L}}(L(o)) \neq \phi$ .

Since it is acceptable to think that an object  $o_1$  relating to a set of labels  $\mathcal{L}$  mainly ( $Rel_{\mathcal{L}}(P(o_1)) \neq \phi$ ) relates to  $\mathcal{L}$  more closely than an object  $o_2$  not relating to  $\mathcal{L}$  mainly ( $Rel_{\mathcal{L}}(P(o_2)) = \phi$ ), the condition whether a primary label of  $L(o)$  of an object  $o$  relates to  $\mathcal{L}$  is a criterion for the strength of relationship between  $o$  and  $\mathcal{L}$ .

$cmd_{PE}$  : For a set of labels  $\mathcal{L}$  and an object  $o$ , there exists a primary label of  $L(o)$  which relates to  $\mathcal{L}$ , that is,  $Rel_{\mathcal{L}}(P(o)) \neq \phi$ .

There are cases that secondary labels of  $L(o)$  affect the strength of relationship between  $o$  and  $\mathcal{L}$ . If  $Rel_{\mathcal{L}}(P(o)) = \phi$ , the strength by  $S(o)$  is equivalent to  $cmd_{LE}$ . Suppose  $Rel_{\mathcal{L}}(P(o)) \neq \phi$ . For an object  $o_1$  such that  $Rel_{\mathcal{L}}(P(o_1)) \neq \phi$  and  $Rel_{\mathcal{L}}(S(o_1)) \neq \phi$  and an object  $o_2$  such that  $Rel_{\mathcal{L}}(P(o_2)) \neq \phi$  and  $Rel_{\mathcal{L}}(S(o_2)) = \phi$ , both  $o_1$  and  $o_2$  do not relate to  $\mathcal{L}$  more closely than the other.

*Example 3.* Suppose the labels of object  $o_1$  and object  $o_4$  are  $P(o_1) = \{Automobile\}$ ,  $S(o_1) = \phi$ ,  $P(o_4) = \{Automobile\}$ , and  $S(o_4) = \{Mobile Phone\}$  for set of labels  $\mathcal{L}_3 = \{Manufacturing\}$ . Although *Mobile Phone* of  $L(o_4)$  relating to  $\mathcal{L}_3$  is a secondary label, it is not acceptable to think that  $o_1$  relates to  $\mathcal{L}_3$  more closely than  $o_4$  because  $o_1$  and  $o_4$  have primary label *Automobile* for  $\mathcal{L}_3$ , namely, both of them relate to *Manufacturing* mainly.

For object  $o'_4$  such that  $P(o'_4) = \{Automobile, Mobile Phone\}$  and  $S(o'_4) = \phi$ , it is not reasonable to think that  $o_4$  relates to  $\mathcal{L}_3$  more closely than  $o'_4$  because *Mobile Phone* of  $L(o'_4)$  is a primary label and *Mobile Phone* of  $L(o_4)$  is a secondary label. Since the number of the labels which relate to  $\mathcal{L}_3$  does not affect the strength, the strength of relationship between  $o'_4$  and  $\mathcal{L}_3$  is the same as the strength of relationship between  $o_1$  and  $\mathcal{L}_3$ .

For a set of labels  $\mathcal{L}$  and objects  $o_1$  and  $o_2$ , it can be regarded that  $o_1$  relates to  $\mathcal{L}$  more closely than  $o_2$  if  $L(o_1)$  does not include other labels than  $\mathcal{L}$  and  $L(o_2)$  includes such labels.

$cmd_{LN}$  : For a set of labels  $\mathcal{L}$  and an object  $o$ , there does not exist a label of  $L(o)$  which does not relate to  $\mathcal{L}$ , that is,  $Rel_{\mathcal{L}}(L(o)) = L(o)$ .

It is also acceptable to think that an object  $o_1$  relates to  $\mathcal{L}$  more closely than an object  $o_2$  if  $P(o_1)$  does not include other labels than  $\mathcal{L}$  while  $P(o_2)$  includes other labels than  $\mathcal{L}$ .

$cmd_{PN}$  : For a set of labels  $\mathcal{L}$  and an object  $o$ , there does not exist a label of  $P(o)$  which does not relate to  $\mathcal{L}$ , that is,  $Rel_{\mathcal{L}}(P(o)) = P(o)$ .

For the criteria for secondary labels of  $L(o)$  which relate to other labels than  $\mathcal{L}$ , if there are labels of  $P(o)$  which relate to other labels than  $\mathcal{L}$ , such labels of  $S(o)$  do not affect the strength of relationship.

*Example 4.* Suppose object  $o_3$  such that  $P(o_3) = \{Automobile, Finance\}$  and  $S(o_3) = \phi$  and object  $o'_3$  such that  $P(o'_3) = \{Automobile, Finance\}$  and  $S(o'_3) = \{Education\}$  are evaluated for  $\mathcal{L}_3 = \{Manufacturing\}$ . Although  $o'_3$  has secondary label *Education* which relates to other labels than *Manufacturing*, it is not acceptable to think that  $o'_3$  relates to  $\mathcal{L}_3$  more closely  $o_3$ .

For object  $o''_3$  such that  $P(o''_3) = \{Automobile, Finance, Education\}$ , it is not acceptable to think that  $o''_3$  relates to  $\mathcal{L}_3$  more closely than  $o'_3$  because  $o''_3$  has primary label *Education*. Thus  $o'_3 < o''_3$  is not satisfied. The strength of relationship between  $o''_3$  and  $\mathcal{L}_3$  is the same as the strength of relationship between  $o_3$  and  $\mathcal{L}_3$  because both of them have primary labels relating to  $\mathcal{L}_3$ .

If there are no labels of  $P(o)$  which relate to other labels than  $\mathcal{L}$ , the strength of relationship is decided by whether there exist labels of  $S(o)$  which relate to other labels than  $\mathcal{L}$ . If there exist such labels in  $S(o)$ , it means that there exist labels of  $L(o)$  which relate to other labels than  $\mathcal{L}$ , and this condition is equivalent to  $cmd_{LN}$ .

This section gave the criteria for the strength of relationship between  $o$  and  $\mathcal{L}$ .  $cmd_{LE}$  and  $cmd_{PE}$  are based on relationship between  $L(o)$  and the labels of  $\mathcal{L}$ , and  $cmd_{LN}$  and  $cmd_{PN}$  are based on relationship between  $L(o)$  and other labels than  $\mathcal{L}$ .

## 4 The Strength of the Criteria

This section discusses the strength of the criteria. When the strength of relationship between an object  $o_1$  and a set of labels  $\mathcal{L}$  is compared with the strength of relationship between an object  $o_2$  and  $\mathcal{L}$ , a criterion  $cmd_2$  is stronger than a criterion  $cmd_1$  for  $\mathcal{L}$  if the strength can be decided by whether an object satisfies  $cmd_2$  rather than  $cmd_1$ . In other words, if  $o_1 < o_2$  for  $o_1$  and  $o_2$  such that  $o_1$  does not satisfy  $cmd_2$  and  $o_2$  satisfies  $cmd_2$  regardless of whether  $o_1$  or  $o_2$  satisfies  $cmd_1$ ,

$cmd_2$  is stronger than  $cmd_1$  for  $\mathcal{L}$ . That's because the strength can be decided if  $o_1$  does not satisfy  $cmd_2$  while  $o_1$  may be relate to  $\mathcal{L}$  more closely than  $o_2$  by  $cmd_1$ .

**Definition 1.** For a set of labels  $\mathcal{L}$  and criteria  $cmd_1$  and  $cmd_2$ ,  $cmd_2$  is stronger than  $cmd_1$  for  $\mathcal{L}$  if  $o_1 <_{\mathcal{L}} o_2$  for any objects  $o_1$  and  $o_2$  such that  $o_1$  satisfies  $cmd_1$  but not  $cmd_2$  and  $o_2$  satisfies  $cmd_2$ , denoted by  $cmd_1 <_{\mathcal{L}} cmd_2$ , or  $cmd_1 < cmd_2$  if  $\mathcal{L}$  is obvious.

The set of the objects of  $\bar{\mathcal{L}}$  which satisfy a criterion  $cmd$  is denoted by  $\bar{\mathcal{L}}^{cmd}$  ( $=\{o \mid o \in \bar{\mathcal{L}}, o \text{ satisfies } cmd\}$ ). The strength of criteria can be decided by the inclusion relation among the set of the objects which satisfy each of the criteria.

**Lemma 1.** For a set of labels  $\mathcal{L}$  and criteria  $cmd_1$  and  $cmd_2$ ,  $cmd_1 < cmd_2$  if and only if  $\bar{\mathcal{L}}^{cmd_2} \subseteq \bar{\mathcal{L}}^{cmd_1}$ .

*Proof.* For objects  $o_1$  in  $\bar{\mathcal{L}}^{cmd_1}$  and  $o_2$  in  $\bar{\mathcal{L}}^{cmd_2}$ , if  $o_1 \notin \bar{\mathcal{L}}^{cmd_2}$ ,  $o_1 < o_2$  because  $cmd_2$  is a criterion. Since  $o_1$  satisfies  $cmd_1$  but not  $cmd_2$  and  $o_2$  satisfies  $cmd_2$ ,  $cmd_1 < cmd_2$ .

If  $\bar{\mathcal{L}}^{cmd_2} \subseteq \bar{\mathcal{L}}^{cmd_1}$  is not satisfied, there exists such an object  $o_2$  that  $o_2 \in \bar{\mathcal{L}}^{cmd_2}$  and  $o_2 \notin \bar{\mathcal{L}}^{cmd_1}$ .  $o_2$  satisfies  $cmd_2$  but not  $cmd_1$ . On the other hand,  $o_2 < o_1$  for such an object  $o_1$  that  $o_1 \in \bar{\mathcal{L}}^{cmd_1}$  and  $o_1 \notin \bar{\mathcal{L}}^{cmd_2}$  because  $cmd_1$  is a criterion. Although  $o_2$  satisfies  $cmd_2$  and  $o_1$  satisfies  $cmd_1$  but not  $cmd_2$ ,  $cmd_1 < cmd_2$  is not satisfied because  $o_1 < o_2$  is not satisfied. Thus  $\bar{\mathcal{L}}^{cmd_2} \subseteq \bar{\mathcal{L}}^{cmd_1}$  if  $cmd_1 < cmd_2$ .

If a criterion  $cmd_1$  implies a criterion  $cmd_2$  ( $cmd_1 \Rightarrow cmd_2$ ),  $\bar{\mathcal{L}}^{cmd_1} \subseteq \bar{\mathcal{L}}^{cmd_2}$  and vice versa. Thus the strength of  $cmd_1$  and  $cmd_2$  is decided by implication of  $cmd_1$  and  $cmd_2$  because Lemma 1 shows that the strength of  $cmd_1$  and  $cmd_2$  is decided by the inclusion relation of the objects satisfying  $cmd_1$  and  $cmd_2$ .

**Theorem 1.** For criteria  $cmd_1$  and  $cmd_2$ ,  $cmd_1 < cmd_2$  is equivalent to  $cmd_2 \Rightarrow cmd_1$ .

*Proof.* If  $cmd_1 < cmd_2$ ,  $\bar{\mathcal{L}}^{cmd_2} \subseteq \bar{\mathcal{L}}^{cmd_1}$  by Lemma 1. If  $\bar{\mathcal{L}}^{cmd_2} \subseteq \bar{\mathcal{L}}^{cmd_1}$ ,  $cmd_2 \Rightarrow cmd_1$  because objects satisfying  $cmd_2$  also satisfy  $cmd_1$ . On the other hand, if  $cmd_2 \Rightarrow cmd_1$ ,  $\bar{\mathcal{L}}^{cmd_2} \subseteq \bar{\mathcal{L}}^{cmd_1}$ . Lemma 1 shows that  $cmd_1 < cmd_2$  if  $\bar{\mathcal{L}}^{cmd_2} \subseteq \bar{\mathcal{L}}^{cmd_1}$ .

If one criterion implies another criterion, those criteria are based on the same point of view, which means that the criteria can be used to evaluate the strength of relationship of objects and sets of labels consistently by Theorem 1. The rest of this section discusses implication of the criteria to decide the strength of the criteria.

If an object  $o$  satisfies  $cmd_{PE}$ , there exists a label of  $P(o)$  relating to  $\mathcal{L}$ , which is a label of  $L(o)$  and  $o$  satisfies  $cmd_{LE}$  too.

**Lemma 2.** For  $cmd_{PE}$  and  $cmd_{LE}$ ,  $cmd_{PE} \Rightarrow cmd_{LE}$ .

*Proof.* For an object  $o$  which satisfies  $cmd_{PE}$ , since  $Rel_{\mathcal{L}}(P(o)) \neq \phi$  by satisfying  $cmd_{PE}$  and  $L(o) = P(o) \cup S(o)$  by Property 1,  $P(o) \subseteq L(o)$ . Thus  $Rel_{\mathcal{L}}(L(o)) \neq \phi$ , and  $o$  satisfies  $cmd_{PE}$ .

If  $o$  satisfies  $cmd_{LN}$ , there is no label of  $L(o)$  which relates to other labels than  $\mathcal{L}$ . It means that there is no label of  $P(o)$  which relates to other labels than  $\mathcal{L}$  and  $o$  satisfies  $cmd_{PN}$  too.

**Lemma 3.** For  $cmd_{LN}$  and  $cmd_{PN}$ ,  $cmd_{LN} \Rightarrow cmd_{PN}$ .

*Proof.* For an object  $o$  which satisfies  $cmd_{LN}$ ,  $Rel_{\mathcal{L}}(L(o)) = L(o)$ , in other words,  $L(o)$  does not include a label relating to other labels than  $\mathcal{L}$ . Since  $L(o) = P(o) \cup S(o)$  by Property 1,  $L(o)$  does not include labels which relate to other labels than  $\mathcal{L}$ , and  $Rel_{\mathcal{L}}(P(o)) = P(o)$ . Thus  $o$  satisfies  $cmd_{PN}$ .

If  $o$  satisfies  $cmd_{PN}$ , there is no label of  $P(o)$  which relates to other labels than  $\mathcal{L}$ . Since there always exist labels of  $P(o)$ , there are labels of  $P(o)$  which relate to  $\mathcal{L}$  and  $o$  satisfies  $cmd_{PE}$  too.

**Lemma 4.** For  $cmd_{PN}$  and  $cmd_{PE}$ ,  $cmd_{PN} \Rightarrow cmd_{PE}$ .

*Proof.* For an object  $o$  which satisfies  $cmd_{PN}$ ,  $Rel_{\mathcal{L}}(P(o)) = P(o)$ , in other words,  $P(o)$  does not include a label relating to other labels than  $\mathcal{L}$ . Since  $P(o) \neq \phi$  by Property 3, there always exists a label of  $P(o)$  which relates to  $\mathcal{L}$ , and  $Rel_{\mathcal{L}}(P(o)) \neq \phi$ . Thus  $o$  satisfies  $cmd_{PE}$ .

The strength of the criteria is decided by implication of  $cmd_{LE}$ ,  $cmd_{PE}$ ,  $cmd_{PN}$ , and  $cmd_{LN}$  as shown by Lemmas 2, 3, and 4.

**Theorem 2.** For  $cmd_{LE}$ ,  $cmd_{PE}$ ,  $cmd_{PN}$ , and  $cmd_{LN}$ ,  $cmd_{LE} < cmd_{PE} < cmd_{PN} < cmd_{LN}$ .

*Proof.*  $cmd_{LN} \Rightarrow cmd_{PN} \Rightarrow cmd_{PE} \Rightarrow cmd_{LE}$  because  $cmd_{LN} \Rightarrow cmd_{PN}$ ,  $cmd_{PN} \Rightarrow cmd_{PE}$ , and  $cmd_{PE} \Rightarrow cmd_{LE}$  by Lemmas 3, 4, and 2, respectively. Thus  $cmd_{LE} < cmd_{PE} < cmd_{PN} < cmd_{LN}$  by Theorem 1.

Since Theorem 2 shows the strength of the criteria, the strongest criterion satisfied by  $o$  is to express the strength of relationship between  $o$  and  $\mathcal{L}$ , which is defined as the level of  $o$  for  $\mathcal{L}$ . Levels of objects enable us to evaluate the strength of the objects for  $\mathcal{L}$ .

*Example 5.* For  $\mathcal{L}_1 = \{Transportation, Electronics\}$ , object  $o'_5$  such that  $P(o'_5) = \{Finance\}$  and  $S(o'_5) = \{Automobile, Mobile Phone\}$  satisfies only  $cmd_{LE}$  and  $o'_5$  is at the lowest or the weakest level on the strength. Since object  $o''_5$  such that  $P(o''_5) = \{Automobile, Finance\}$  and  $S(o''_5) = \{Mobile Phone\}$  includes primary label *Automobile* for  $\mathcal{L}_1$ ,  $o''_5$  satisfies  $cmd_{PE}$  too, namely,  $o''_5$  is at the second level on the strength. Moreover, object  $o'''_5$  such that  $P(o'''_5) = \{Automobile, Mobile Phone\}$  and  $S(o'''_5) = \{Finance\}$  satisfies  $cmd_{PN}$  because  $o'''_5$  does not have the primary labels which relate to other labels than  $\mathcal{L}_1$ , and  $o'''_5$  is at the third level



on the strength. Finally, object  $o'_4$  ( $P(o'_4) = \{Automobile, Mobile Phone\}$  and  $S(o'_4) = \phi$ ) satisfies  $cdn_{LN}$  because  $o'_4$  has no label which relates to other labels than  $\mathcal{L}_1$ , and  $o'_4$  is at the highest or the strongest level on the strength. Thus  $o'_5 < o''_5 < o'''_5 < o'_4$ .

By Lemma 1, the inclusion relation of objects according to the levels is shown in Figure 1 and users can select the range of the objects on their purposes.

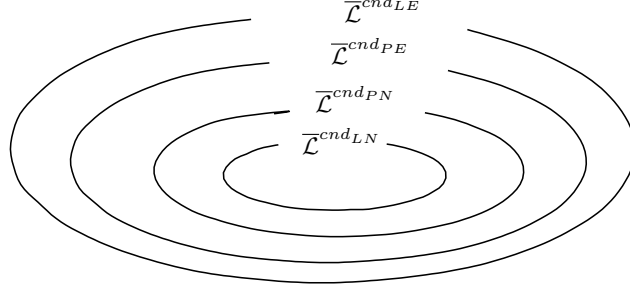


Fig. 1. Inclusion Relation of Objects Satisfying each of the Criteria

## 5 $k$ Ranks

The discussion for two rank objects is extended to  $k$  rank objects. Let  $k$  ranks be  $R_1, \dots, R_i, \dots, R_k$  ( $1 \leq i \leq k$ ) where  $R_i$  is stronger than  $R_j$  ( $i < j$ ). In the case of  $k = 2$ ,  $R_1 = P$  and  $R_2 = S$ . For an object  $o$ , let  $R_i(o)$  be the set of labels of  $L(o)$  whose rank is  $R_i$ , and  $R_i^*(o) = \bigcup_{j=1, i} R_j(o)$  be the set of labels of  $L(o)$  whose ranks are stronger than or equal to  $R_i$ .

The properties of set labels of two rank objects in Section 3 are extended to  $k$  ranks.

*Property 4.*  $L(o) = \bigcup_{i=1, k} R_i(o)$ .

*Property 5.*  $R_i(o) \cap R_j(o) = \phi$  ( $i \neq j$ ).

*Property 6.*  $R_1(o) \neq \phi$ .

The rest of this section discusses the criteria for  $k$  rank objects. The criteria for the strength of relationship between a set of labels  $\mathcal{L}$  and the set labels of an object  $o$  can be divided into two types, which are the criteria for relationship between  $\mathcal{L}$  and the labels of  $L(o)$  and the criteria for relationship between  $\mathcal{L}$  and other labels than  $L(o)$ . Each of them is extended for  $k$  rank.

$cdn_{R_i E}$  ( $1 \leq i \leq k$ ): For a set of labels  $\mathcal{L}$  and an object  $o$ , there exists a label of  $L(o)$  which relates to  $\mathcal{L}$  and whose rank is stronger than or equal to  $R_i$ , that is,  $Rel_{\mathcal{L}}(R_i^*(o)) \neq \phi$ .

$cdn_{R_i N}$  ( $1 \leq i \leq k$ ): For a set of labels  $\mathcal{L}$  and an object  $o$ , there does not exist a label of  $L(o)$  which relates to other labels than  $\mathcal{L}$  and whose rank is stronger than or equal to  $R_i$ , that is,  $Rel_{\mathcal{L}}(R_i^*(o)) = R_i^*(o)$ .

Lemmas 2, 3, and 4 correspond to the following lemmas for  $k$  rank objects.

**Lemma 5.** For  $cdn_{R_i E}$  ( $2 \leq i \leq k$ ),  $cdn_{R_{i-1} E} \Rightarrow cnd_{R_i E}$ .

*Proof.* For an object  $o$  which satisfies  $cdn_{R_{i-1} E}$ ,  $Rel_{\mathcal{L}}(R_{i-1}^*(o)) \neq \phi$ . Since  $R_{i-1}^*(o) \subseteq R_i^*(o)$ ,  $Rel_{\mathcal{L}}(R_i^*(o)) \neq \phi$ , and  $o$  satisfies  $cdn_{R_i^* E}$ .

**Lemma 6.** For  $cdn_{R_i N}$  ( $2 \leq i \leq k$ ),  $cdn_{R_i N} \Rightarrow cnd_{R_{i-1} N}$ .

*Proof.* For  $Rel_{\mathcal{L}}(R_i^*(o)) = R_i^*(o)$ . Since  $R_{i-1}^*(o) \subseteq R_i^*(o)$ ,  $Rel_{\mathcal{L}}(R_{i-1}^*(o)) = R_{i-1}^*(o)$ , and that  $o$  satisfies  $cdn_{R_{i-1}^* N}$ .

**Lemma 7.** For  $cdn_{R_1 N}$  and  $cdn_{R_1 E}$ ,  $cdn_{R_1 N} \Rightarrow cnd_{R_1 E}$ .

*Proof.* For an object  $o$  which satisfies  $cdn_{R_1 N}$ ,  $Rel_{\mathcal{L}}(R_1(o)) = R_1(o)$ , in other words,  $R_1(o)$  does not include a label relating to other labels than  $\mathcal{L}$ . Since  $R_1(o) \neq \phi$  by Property 6, there always exists a label of  $R_1(o)$  which relates to  $\mathcal{L}$ , and  $Rel_{\mathcal{L}}(R_1(o)) \neq \phi$ . Thus  $o$  satisfies  $cdn_{R_1 E}$ .

Since implication of the criteria is shown by Lemmas 5, 6, and 7, the following theorem on the strength of the criteria is satisfied.

**Theorem 3.** For  $cdn_{R_i E}$  and  $cdn_{R_i N}$  ( $1 \leq i \leq k$ ),  $cdn_{R_k E} < cnd_{R_{k-1} E} < \dots < cnd_{R_1 E} < cnd_{R_1 N} < \dots < cnd_{R_{k-1} N} < cnd_{R_k N}$ .

*Proof.*  $cdn_{R_k N} \Rightarrow cnd_{R_{k-1} N} \Rightarrow \dots \Rightarrow cnd_{R_1 N} \Rightarrow cnd_{R_1 E} \Rightarrow \dots \Rightarrow cnd_{R_{k-1} E} \Rightarrow cnd_{R_k E}$  because  $cdn_{R_i N} \Rightarrow cnd_{R_{i-1} N}$ ,  $cdn_{R_1 N} \Rightarrow cnd_{R_1 E}$ , and  $cdn_{R_{i-1} E} \Rightarrow cnd_{R_i E}$  by Lemmas 6, 7, and 5, respectively. Thus  $cdn_{R_k E} < cnd_{R_{k-1} E} < \dots < cnd_{R_1 E} < cnd_{R_1 N} < \dots < cnd_{R_{k-1} N} < cnd_{R_k N}$  by Theorem 1.

## 6 Conclusion

This paper introduced the criteria for the strength of relationship between an object  $o$  and a set of labels  $\mathcal{L}$ , which are  $cdn_{LE}$ ,  $cdn_{PE}$ ,  $cdn_{PN}$ , and  $cdn_{LN}$ .  $o$  satisfies  $cdn_{LE}$ ,  $cdn_{PE}$ ,  $cdn_{PN}$ , and  $cdn_{LN}$  if  $o$  relates to  $\mathcal{L}$ ,  $\mathcal{L}$  mainly, only  $\mathcal{L}$  mainly, and only  $\mathcal{L}$ , respectively. Referring to the strength of the criteria, it was shown that  $cdn_{LE} < cnd_{PE} < cnd_{PN} < cnd_{LN}$ . Thus the strongest criterion which  $o$  satisfies can be used as the levels to express the degree of the strength. Those results can be applied to  $k$  rank objects.

The strength of the criteria is decided by implication of the criteria. If one criterion implies another criterion, both of them can be used to evaluate the strength of relationship between objects and sets of labels because their criteria are based on the same point of view. The criteria proposed in this paper enable to evaluate data with ranks qualitatively and consistently. Thus users can select

the range of data at each of levels based on the criteria according to the purpose of the utilization of data. In general, the range of data is decided by changing the set of labels to identify the data. For example, if users want to utilize data widely, lower concept labels are replaced by higher concept labels or additional labels are inserted. In such utilization of data, originally intended purpose may not be realized because the set of labels itself has been changed. The levels based on the criteria solve this issue because the range of data can be changed according to the levels even if the set of labels is no change.

## References

1. Cardoso-Cachopo, A. and Oliveira, A.: Semi-supervised Single-label Text Categorization using Centroid-based Classifiers, *Proc. Symposium on Applied Computing (SAC '07)*, pp. 844–851 (2007).
2. Chakrabarti, K., Ganti, V., Han, J., and Xin, D.: Ranking Objects by Exploiting Relationships: Computing Top-K over Aggregation, *Proc. ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD '06)*, pp. 371–382 (2006).
3. Ding, B., Wang H., Jin, R., Han, J., and Wang, Z.: Optimizing Index for Taxonomy Keyword Search, *Proc. ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD '12)*, pp. 493–504 (2012).
4. Fazzinga, B., Flesca, S., and Furfaro, F.: RFID-Data Compression for Supporting Aggregate Queries, *ACM Transaction on Database Systems*, Vol. 38, Issue 2, pp. 11–45 (2013).
5. Kim, J. and Croft, B.: Ranking using Multiple Document Types in Desktop Search, *Proc. ACM Int'l Conf. on Research and Development in Information Retrieval (SIGIR '10)*, pp. 50–57 (2010).
6. Lopez, V., Prieta, F., Ogihara, M., and Wong, D.: A Model for Multi-label Classification and Ranking of Learning Objects, *Expert Systems with Applications*, Vol. 39, Issue 10, pp. 8878–8884 (2012).
7. Lu, J., Senellart, P., Lin, C., Du, X., Wang, S., and Chen, X.: Optimal Top-k Generation of Attribute Combinations based on Ranked Lists, *Proc. ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD '12)*, pp. 409–420 (2012).
8. Silla, C. and Freitas, A. : A Survey of Hierarchical Classification Across Different Application Domains, *Data Mining and Knowledge Discovery*, Vol. 22, Issue 1–2, pp. 31–72 (2011).
9. Tang, L., Rajan, S., and Narayanan, V.: Large Scale Multi-label Classification via MetaLabeler, *Proc. Int'l Conf. on World Wide Web (WWW '09)*, pp. 211–220 (2009).
10. Wang, H. and He, X., Chang, M., Song, Y., White, R., and Chu, W.: Personalized Ranking Model Adaptation for Web Search, *Proc. ACM Int'l Conf. on Research and Development in Information Retrieval (SIGIR '13)*, pp. 323–332 (2013).
11. Zhang, X., Yang, Y., Han, Z., Wang, H., and Gao, C.: Object Class Detection: A Survey, *ACM Computing Surveys*, Vol. 46, Issue 1, pp. 10–53 (2013).
12. Zhu, X., Song, S., Lian, X., Wang, J., and Zou, L.: Matching Heterogeneous Event Data, *Proc. ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD '14)*, pp. 1211–1222 (2014).