

NTUSocialRec: An Evaluation Dataset Constructed from Microblogs for Recommendation Applications in Social Networks

Chieh-Jen Wang¹, Shuk-Man Cheng¹, Lung-Hao Lee¹, Hsin-Hsi Chen¹,
Wen-shen Liu², Pei-Wen Huang² and Shih-Peng Lin²

¹Department of Computer Science and Information Engineering, National Taiwan University

²Smart Network System Institute, Institute for Information Industry

¹No. 1, Sec. 4, Roosevelt Rd., Taipei 106, Taiwan

²11F. No. 106, Sec. 2, Heping E. Rd., Taipei 106, Taiwan

E-mail: {cjwang, smcheng, lhlee, hhchen}@nlg.csie.ntu.edu.tw, {wsliau, athena, shihpeng}@iii.org.tw

Abstract

This paper proposes a method to construct an evaluation dataset from microblogs for the development of recommendation systems. We extract the relationships among three main entities in a recommendation event, i.e., *who* recommends *what* to *whom*. User-to-user friend relationships and user-to-resource interesting relationships in social media and resource-to-metadata descriptions in an external ontology are employed. In the experiments, the resources are restricted to visual entertainment media, movies in particular. A sequence of ground truths varying with time is generated. That reflects the dynamic of real world.

Keywords: evaluation dataset construction, microblog analysis, resource recommendation

1. Introduction

Recommendation becomes a very important application nowadays due to the large scale e-Business on the web. For example, Amazon.com recommends books to buyers; Netflix recommends movies and television shows to subscribers, and so on. Many recommendation algorithms based on different cues such as semantic similarity of resources (Passant, 2010), users' preferences and behaviours (Yu, Zheng, Zhao, & Zheng, 2006), social network of users' relationships (Passant & Raimond, 2008), and so on, have been proposed. How to evaluate the effectiveness of the recommendation algorithms is indispensable for performance improvement.

In a typical recommendation application, three key entities including *recommender*, *resource* and *recipient*, are involved. That is, a recommender provides appropriate resources to a recipient (Burke, 2002). An evaluation dataset for recommendation applications consists of a set of triplets, (recommender, resource, recipient), which covers three key entities in a recommendation event. In the past, a famous competition, Netflix Prize, provides a dataset for participants to predict the interesting degree of a user to a movie based on her/his past preferences. However, the dataset only considers resources and recipients so that it lacks of the information of social network.

It is hard to develop an artificial evaluation dataset because it covers different cases, e.g., any two persons may have a relationship in different degrees, a person may be interested in different set of resources, and a resource may be described by different metadata. Social media on the web provides an opportunity to deal with the problem. Since the social relationships are embedded in the friend list of users and users' preferences may be expressed in the interactions implicitly. Moreover, the relationships vary with time. It captures the dynamic phenomena of

real world.

Microblog such as Twitter, Tumblr and Plurk has become one of the most important social media in recent years. In microblog, users make friends, share information and gossip with one another by posting short messages, which are usually limited to 140 characters. Microblog contains various characteristics, including immediacy, quickness and sociability. It can be regarded as a social network because users are linked together through relationship, friendship, common interests, and so on. Various applications on different domains such as entertainment, education, business, etc. are developed via social network. Besides, plenty of ontology such as Freebase, DBpedia, etc. provides metadata of resources.

In this paper, we aim to build an evaluation dataset from microblogs for measuring the performance of recommendation algorithms. The evaluation dataset not only contains three main entities such as users, resources and metadata, but also specifies network information of the entities. Most importantly, the dataset is dynamic. Each link in a social network has different meanings depending on the connected nodes. Links between users denote friendship or fans; links between users and resources denote interests or habits; links between resources and resources denote similar characteristics such as genres, directors and casts. All the links varies with time. It means a person may have new friends and new interests. Besides, her/his interests may be shifted.

To extract messages from microblog posts for resource recommendation is challenging, since these messages are shorter than regular blog articles and other free texts, and the meanings of terms are ambiguous. Metadata recorded in the ontology is usually represented in English. For retrieving metadata from the ontology, a resource must be represented as an English query. Therefore, the names of non-English resources have to be translated to English for consulting the ontology. General

purpose machine system may not be suitable for the specific translation.

This paper is organized as follows. Section 2 reviews recommendation algorithms and their evaluation corpora. Section 3 describes a social media corpus used in this study. Section 4 presents how to extract three types of link pairs such as user-user, user-resource, and resource-metadata from social network and external ontology. How to automatically establish dynamic ground truths with the passage of time is introduced in Section 5. The constructed evaluation dataset called *NTUSocialRec* is illustrated in Section 6. In addition, the potential recommendation applications are introduced. Lastly, Section 7 concludes the remarks.

2. Related Work

In recent years, the analyses of microblogs attract much attention (Jansen, Zhang, Sobel, & Chowdury, 2009). Starbird et al. (2010) investigated the social network in microblogs, and the social life is classified as four classes for evaluation. Qu et al. (2011) analyzed what types of messages are more likely to be discussed after a special event such as an earthquake.

Konstas et al. (2009) developed a music recommendation system which considered the relationships between users, and tags of songs given by users. Chu & Park (2009) proposed a feature based machine learning algorithm for personalized news recommendations and dealt with the cold-start problem. The features cover demographic information, area information and characteristics of user behaviors. Chen et al. (2009) proposed a recommendation algorithm which considered content matching, content plus link, friend of friend and SONAR. Golder et al. (2009) proposed four kinds of explanations to target friends on Twitter, such as reciprocity, shared interests, shared audience, and filtered people. Degirmencioglu et al. (2010) explored the same interest group in microblog. They extracted associated terms in a specific area from labeled metadata, and used them to identify users who have a common interest.

Netflix¹ is one of the largest evaluation dataset for resource recommendation. The dataset is divided into training and test datasets. In training dataset, it consists of a set of quadruplet (user, movie, date of grade, grade). Recommendation systems are asked to predict grades on the test dataset. In addition, several widely used benchmark evaluation datasets such as MovieLens² and EachMovie³ are employed in current researches (Zhou, Yang, & Zha, 2011).

The previous evaluation datasets consider information from resources and recipients so that they lack of the information of social network and metadata of resources. Different from their work, we propose strategies to sample users in a social media, apply human-generated knowledge to translate non-English

resources, consult external resources to collect metadata, extract user-to-user friend relationships and user-to-resource interesting relationships in the social media, and construct an evaluation dataset.

3. A Social Media Corpus

The social media corpus in this study was clawed from Plurk, a very popular microblog platform in Taiwan. Figure 1 shows the visitor distributions of Plurk reported from Google Trends in 2010. The visitors from Taiwan are higher than other countries. As shown in Figure 2, the number of visitors is still increased. Therefore, Plurk is a very suitable resource to construct an evaluation dataset for recommendation applications. Table 1 shows numbers of users, post messages and reply messages collected. In total, there are more than 20 millions post messages and 123 millions reply messages among the 111,870 microbloggers from April 1 to October 31, 2009. The messages on Plurk, like those on Twitter and some other microblog platforms, are limited to 140 characters (i.e., maximum 70 Chinese characters). Users can have friends, followers or fans so that the relationships between different users provide a social network of the user.



Figure 1. Visitor Distributions from Different Countries in Plurk

Table 1: Statistics of a Plurk dataset

# of users	# of posts	# of replies
111,870	20,265,405	123,483,033

4. Relationship Extraction

Figure 3 shows the proposed framework to construct an evaluation dataset from Microblogs for recommendation applications in social networks. The framework contains two main phases, such as relationship extraction and ground truth generation. This section introduces how to extract the basic information in the social network and the external ontology. The information includes three types of links at different time, such as U-U (user-to-user) links, U-R (user-to-resource) links and R-M (user-to-metadata) links.

¹ <http://www.netflixprize.com/>

² <http://www.grouplens.org/node/73>

³ <http://www.grouplens.org/node/76>

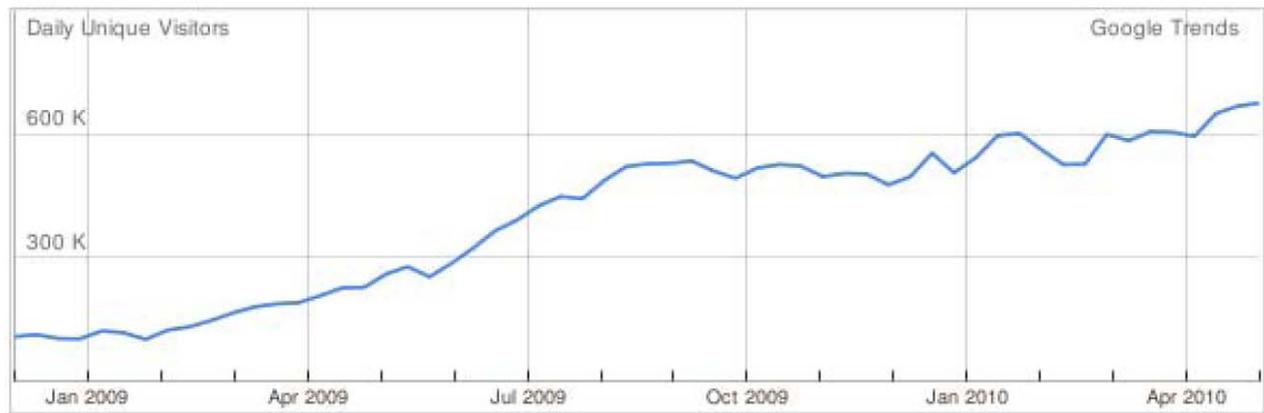


Figure 2. Daily visitors of Plurk

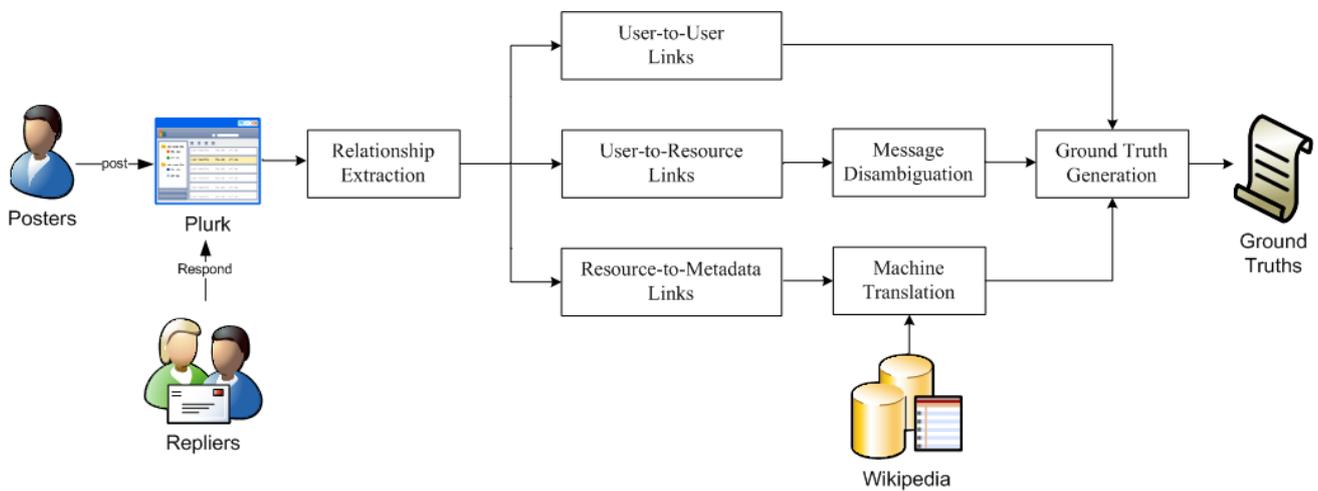


Figure 3. The Framework of Evaluation Dataset Construction

4.1 User-to-User Links at Different Time

In social media, each user owns a set of friends which is incrementally updated through time. Each link signifies a relationship between two individuals. The direct and indirect relationships form a network. Users' activities in the social media can be investigated via the network. The network of each user is recorded in the evaluation dataset.

4.2 User-to-Resource Links at Different Time

We postulate that a user mentions a resource by posting/replying messages if s/he is interested in the resource. Therefore, a relationship between a user and a resource is established if a resource is mentioned in someone's message. In this paper, the resources are restricted to visual entertainment media, movies in particular. A list of movies released in Asia in 2009 is collected. In total, there are 326 movies in the movie list. All the U-R links are extracted from our Plurk dataset.

Some movie names such as "2012" are ambiguous. String "2012" may denote as a movie name or a year. Without disambiguation, it is unclear whether a user is

interested in the movie or just talks about a future year. We formulate the mention disambiguation as a supervised learning problem. A collection of messages mentioning potential movie names are selected. We propose two criteria to select messages with less ambiguity from the collection. The basic idea is that a short-term or a highly frequent term tends to be ambiguous. Thus movie names whose length is longer than the average length of movie names in the movie list, and whose frequency is less than the average mention frequency of movies in the movie list are selected.

After filtering, total 79 movies remain. The messages mentioning the 79 movies are classified as *clear* type. The messages mentioning the other 247 movie names are classified as *ambiguous* type. Messages belonging to *clear* type are defined as movie instances. Messages never mentioning any names in the movie list are defined as *non-movie* instances. We randomly select 10,000 movie instances and non-movie instances to form a training set. A disambiguation model is constructed by integrating *tf-idf* features extracted from the training set with support vector machines (SVM). The model is used

to disambiguate messages belonging to ambiguous type.

For evaluating the quality of the disambiguation model, we randomly sample 4 ambiguous messages for each movie. In total, 988 messages are selected and the class (*movie/non-movie*) of the selected messages is manually annotated. Table 2 shows the precision, recall and F-score for the two classes. The disambiguation model achieves a precision of 0.9412, a recall of 0.7896, and an F-score of 0.8587 in the movie class.

Table 2: Performance of mention disambiguation model

	Precision	Recall	F-score
Movie	0.9412	0.7896	0.8587
Non-movie	0.3908	0.7323	0.5096

4.3 Resource-to-Metadata Links

Internet Movie Database (IMDb) provides an ontology of visual entertainment media such as movies, television shows, and so on. Metadata of movies includes title, director, cast, genre, summary, keywords, etc. However, non-English consulting is not supported in the database. That is, an English query has to be formed before consulting IMDb.

The non-English movie names may not be translated accurately if we use a general purpose machine translation tool like Google Translate. For example, the correct translation of a Chinese movie “梅蘭芳” should be “Forever Enthralled” rather than its transliteration “Mei Lanfang”. Here, we adopt Wikipedia which is a human-generated ontology for translation. Given a non-English movie name, we consult Wikipedia using the movie name as a query. The results which belong to movie category are extracted. As shown in Figure 4, the English name in the info box is extracted as the translation of the Chinese movie name, and used to retrieve the corresponding metadata from IMDb.

5. Ground Truth Generation

We propose an algorithm to generate ground truths from the dataset for evaluating recommendation applications. Given the three types of extracted relationships, i.e., user-to-user friend relationships and user-to-resource interesting relationships in social media, and resource-to-metadata descriptions in ontology, we can generate temporal ground truths from dynamic social networks for evaluating three main entities in a recommendation event, i.e., *who* recommends *what* to *whom*.

Given a user u and her/his two sets of friend and interesting relationships at time t_i , a ground truth is defined as: the potential recipient c should mention r at some later time (i.e., after t_i). Algorithm 1 generates a sequence of ground truths, where $U(u,t)$ is a list of users communicating with u at time t ; $CU(u,t)$ is a set of candidate recipients to be recommended by user u at time t (i.e., $CU(u,t)=CU(u,t-1) \cup U(u,t)$); $R(u,t)$ means a set of resources mentioned by u in posts or replies at time t ; $CR(u,t)$ is a set of resources which can be recommended by user u at time t (i.e., $CR(u,t) = CR(u,t-1) \cup R(u,t)$).

For each iteration denoting a time t_i , the algorithm attempts to generate a set of recipient-resource pairs (c, r) for recommender u . A person c who discussed her/his interesting resources with u before or at t_i is a candidate recipient for u . A ground truth, i.e., a resource r is recommended to a recipient c by a recommender u at time t_i , must meet the criterion: c mentioned r at some later time t_j , i.e., $r \in CR(c,t_j)-CR(c,t_i)$. In other words, c expresses an interest to r after t_i .

Algorithm 1. Generating a set of ground truths

Input: A user u who mentions resources in a sequence of times t_1, t_2, \dots, t_n

Output: A set of ground truths $GT(u, t_1), GT(u, t_2), \dots, GT(u, t_n)$

```

1:  $CU(u,t_0) \leftarrow \emptyset, CR(u,t_0) \leftarrow \emptyset, i \leftarrow 1$ 
2: while  $i \leq n$ 
3:    $CU(u,t_i) = CU(u,t_{i-1}) \cup U(u,t_i)$ 
4:    $CR(u,t_i) = CR(u,t_{i-1}) \cup R(u,t_i)$ 
5:    $GT(u, t_i) = \{(c,r) \mid c \in CU(u,t_i), r \in CR(u,t_i),$ 
       $\exists t_j, t_i < t_j, r \in CR(c,t_j)-CR(c,t_i)\}$ 
6:    $i \leftarrow i + 1$ 
7: end while
8: return  $GT(u, t_1), GT(u, t_2), \dots, GT(u, t_n)$ 

```

Consider an example shown in Figure 5. The user with id 16116 posts a message at eight o'clock on May 12, 2009 and the user with id 21117 replies this message about the movie “Inkheart” (墨水心). We add the user with id 21117 and the movie “Inkheart” to $CU(16116, 2009-05-15 08:00)$ and $CR(16116, 2009-05-15 08:00)$, respectively, that is, $CU(16116, 2009-05-15 08:00)=\{21117\}$ and $CR(16116, 2009-05-15 08:00)=\{\text{Inkheart}\}$. The movie “Inkheart” is added to $CR(21117, 2009-05-15 08:00)$ as well due to that fact that the user with id 22217 replied the movie. Within the next three months, the user with id 22217 mentioned movies “Invitation Only” (絕命派對) and “Star Trek” (星際爭霸戰) at 2009-06-02 18:00 and 2009-08-07 10:00, respectively, so that $CR(21117, 2009-08-07 10:00)=\{\text{“Inkheart”}, \text{“Invitation Only”}, \text{“Star Trek”}\}$. A relative complement of $CR(21117, 2009-05-12 08:00)$ with respect to a set $CR(21117, 2009-08-07 10:00)$ is $\{\text{“Invitation Only”}, \text{“Star Trek”}\}$. Therefore, the ground truth $GT(16116, 2009-05-15 08:00)$ is $\{(21117, \text{“Invitation Only”}), (21117, \text{“Star Trek”})\}$.

6. NTUSocialRec: An Evaluation Dataset

An evaluation dataset called *NTUSocialRec* includes three types of fundamental relationships and ground truths shown as follows. Two recommendation scenarios can be done: (1) given a recommender u , a time t , and a resource r , a system predicts to whom r can be recommended; and (2) given a recommender u_1 , a time t , and a recipient u_2 , a system predicts which resources can be recommended to u_2 .

(1) User-to-User relationships

These relationships are represented in the form of a quadruple: $\langle \text{Post-timestamp}, \text{Poster-id}, \text{Reply-timestamp}, \text{Replier-id} \rangle$. For privacy concern,

the real users are anonymous with unique ids. Consider an example <2009-05-10 17:06, 16116, 2009-05-10 18:44, 3291439>. It means the user with id 16116 posts a message on May 10, 2009 and the user with id 3291439 replies this message about one hour later.

(2) User-to-Resource relationships

Such a relationship is represented as a quadruple: < Post/Reply-timestamp, Poster/Reply-id, Movie-id, Movie-name>. Consider an example <2009-05-10 17:06, 16116, 796366, “Star Trek” (星際爭霸戰)>. It indicates that the user with id 16116 mentioned the movie “Star Trek” with movie id 796366 on May 10, 2009.

(3) Resource-to-Metadata relationships

The metadata of a resource is in terms of the following format: <Movie-id, Title, Year, Genres, Director, Writer, Cast, Runtime, Country, Language, Rating, Vote, Plot>. An attribute may have a single value or multiple values. For example, the “Cast” attribute of the movie “Star Trek” has multiple values: Chris Pine (Kirk), Zachary Quinto (Spock), Leonard Nimoy (Spock Prime), Eric Bana (Nero), and Bruce Greenwood (Pike).

(4) The ground truths

The ground truth which varies with time is in terms of a quadruple: <Recommendation-time, Recommender-id, Resource, Recipient-id>. Note that a user who knows a resource before recommendation time should not be a recipient of this resource according to step (5) of Algorithm 1.

Table 3 shows the descriptions of this dataset. Metadata of 120 movies are retrieved from IMDb. After message disambiguation, 11,254 unique users mentioning one of the 120 movies by posting messages and 32,997 unique users reply the messages containing these movies.

According to the three main entities in the social network, total 160,502 user-to-user (U-U) relationships, 20,618 user-to-resource (U-R) relationships, and 120 resource-metadata (R-M) relationships are extracted. After Algorithm 1, 512 quadruples (recommendation-time, recommender, resource, recipient) are generated as ground truths.

Table 3: Descriptive of the evaluation dataset

Entities	Numbers
Unique Posters	11,254
Resources	120
Unique Repliers	32,997
Link Types	Numbers
User-to-User (U-U)	160,502
User-to-Resource (U-R)	20,618
Resource-to-Metadata (R-M)	120

7. Conclusion and Future Work

This paper proposes a method to generate an evaluation dataset from microblogs for recommendation applications automatically. The evaluation dataset is composed of quadruples, *who* recommend *what* to *whom* at *when*. Ground truths varying with time are extracted from U-U links, U-R links and R-M links. In this way, any recommendation algorithms can use this evaluation dataset to estimate and improve their performance.

8. Acknowledgements

This research was partially supported by Institute for Information Industry, Taiwan under the grant (100)-0816, and the Excellent Research Projects of National Taiwan University.

The image shows a screenshot of the Wikipedia page for the movie "Forever Enthralled" (梅蘭芳). The page is in Chinese and includes a sidebar with navigation links, a main content area with a description of the film, and a right-hand section with a cast list and a movie poster image.

WIKIPEDIA
The Free Encyclopedia

Forever Enthralled
From Wikipedia, the free encyclopedia

Forever Enthralled (simplified Chinese: 梅兰芳; traditional Chinese: 梅蘭芳; pinyin: *Méi Lánfāng*) is a Chinese biographical film directed by **Chen Kaige**; the film marks Chen's eleventh feature film as a director. *Forever Enthralled* follows the life of **Mei Lanfang**, one of China's premiere opera performers. It stars **Leon Lai** as Mei, **Zhang Ziyi**, **Sun Honglei** and **Masanobu Ando**.

Known during production as *Mei Lanfang*, the film's English title was changed in November 2008, shortly before its release, to *Forever Enthralled*.^[1]

The film was shown in competition at the 2009 **Berlin International Film Festival** for the **Golden Bear** award.^[2]

Contents [hide]

- 1 Production
- 2 Cast
 - 2.1 Previous casting
- 3 Reviews
- 4 References
- 5 External links

Production [edit]

Forever Enthralled was produced on a budget of \$15 million (US) by the **China Film Group** and Taiwan's **China Magnetics Corporation Company** (CMC Company), which had previously helped fund **John Woo's** *Red Cliff* and **Alexi Tan's** *Blood Brothers*.^[3] Unlike Chen's previous film, *The Promise*, Chen is not re-teaming with his usual producers, **Etschie Stroh** and **Moonstone Entertainment**.^[3]

Just prior to the film's release in December 2008, *Forever Enthralled* producer, the **China Film Group**, ran into legal hurdles after it was sued by another production company, **Milimeter**. **Milimeter** claimed that it held the **copyright** to the film based on a contract signed between the two companies in 2004.^[4]

Forever Enthralled

梅蘭芳

梅蘭芳

Méi Lánfāng

Directed by
Chen Kaige

Produced by
Han Sanping
Chen Hong
Du Jiyi

Figure 4. Wikipedia Info Box of the Movie “梅蘭芳”

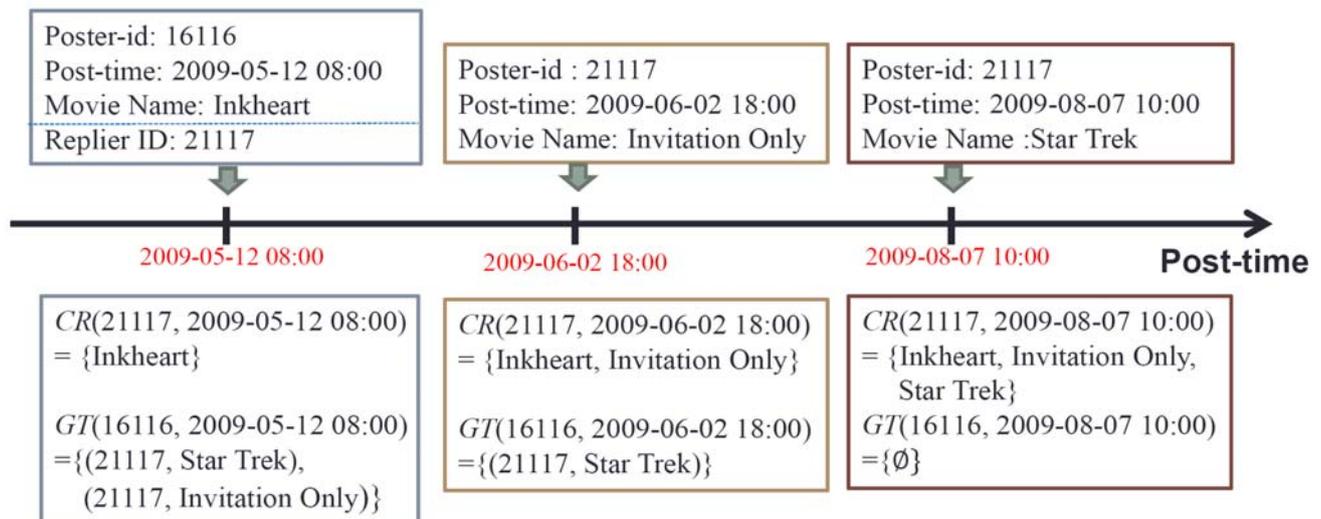


Figure 5. An Example of Ground Truth

9. References

- Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modelling and User-Adapted Interaction*, 12(4), pp. 331-370.
- Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Micro-blogging as Online Word of Mouth Branding. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems*. pp. 3859-3864.
- Konstas, L., Stathopoulos, v., & Jose, J. M. (2009). On Social Networks and Collaborative Recommendation. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 195-202.
- Passant, A. (2010). dbrec: Music Recommendations Using DBpedia. In *Proceedings of the 9th International Semantic Web Conference*, pp. 209-224
- Passant, A., & Raimond, Y. (2008). Combining Social Music and Semantic Web for Music-related Recommender Systems. In *Proceedings of the ISWC 2008 Workshop on Social Data on the Web*. pp. 1-12.
- Qu, Y., Huang, C., Zhang, P., & Zhang, J. (2011). Microblogging After a Major Disaster in China: a Case Study of the 2010 Yushu Earthquake. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, pp. 25-34.
- Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S. (2010). Chatter on the Red: What Hazards Threat Reveals about the Social Life of Microblogged Information. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, pp. 241-250.
- Yu, H., Zheng, D., Zhao, B. Y., & Zheng, W. (2006). Understanding User Behaviour in Large-scale Video-On-Demand Systems. *SIGOPS Operating Systems Reviews*. 40(4), 333-344.
- Zhou, K., Yang, S., & Zha, H. (2011). Functional Matrix Factorizations for Cold-start Recommendation. In

Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 315-324.