Extracting Social Networks and Biographical Facts From Conversational Speech Transcripts

Hongyan Jing
IBM T.J. Watson Research Center
1101 Kitchawan Road
Yorktown Heights, NY 10598
hjing@us.ibm.com

Nanda Kambhatla
IBM India Research Lab
EGL, Domlur Ring Road
Bangalore - 560071, India
kambhatla@in.ibm.com

Salim Roukos
IBM T.J. Watson Research Center
1101 Kitchawan Road
Yorktown Heights, NY 10598
roukos@us.ibm.com

Abstract

We present a general framework for automatically extracting social networks and biographical facts from conversational speech. Our approach relies on fusing the output produced by multiple information extraction modules, including entity recognition and detection, relation detection, and event detection modules. We describe the specific features and algorithmic refinements effective for conversational speech. These cumulatively increase the performance of social network extraction from 0.06 to 0.30 for the development set, and from 0.06 to 0.28 for the test set, as measured by f-measure on the ties within a network. The same framework can be applied to other genres of text — we have built an automatic biography generation system for general domain text using the same approach.

1 Introduction

A social network represents social relationships between individuals or organizations. It consists of *nodes* and *ties*. *Nodes* are individual actors within the networks, generally a person or an organization. *Ties* are the relationships between the nodes. Social network analysis has become a key technique in many disciplines, including modern sociology and information science.

In this paper, we present our system for automatically extracting social networks and biographical facts from conversational speech transcripts by integrating the output of different IE modules. The IE modules are the building blocks; the fusing module depicts the ways of assembling these building blocks. The final output depends on which fundamental IE modules are used and how their results are integrated.

The contributions of this work are two fold. We propose a general framework for extracting social networks and biographies from text that applies to conversational speech as well as other genres, including general newswire stories. Secondly, we present specific methods that proved effective for us for improving the performance of IE systems on conversational speech transcripts. These improvements include feature engineering and algorithmic revisions that led to a nearly five-fold performance increase for both development and test sets.

In the next section, we present our framework for extracting social networks and other biographical facts from text. In Section 3, we discuss the refinements we made to our IE modules in order to reliably extract information from conversational speech transcripts. In Section 4, we describe the experiments, evaluation metrics, and the results of social network and biography extraction. In Section 5, we show the results of applying the framework to other genres of text. Finally, we discuss related work and conclude with lessons learned and future work.

2 The General Framework

For extraction of social networks and biographical facts, our approach relies on three standard IE modules — entity detection and recognition, relation detection, and event detection — and a fusion module that integrates the output from the three IE systems.

2.1 Entity, Relation, and Event Detection

We use the term *entity* to refer to a person, an organization, or other real world entities, as adopted

Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pages 1040–1047,
Prague, Czech Republic, June 2007.

in the Automatic Content Extraction (ACE) Workshops (ACE, 2005). A *mention* is a reference to a real world entity. It can be *named* (e.g. "John Lennon"), *nominal* (e.g. "mother"), or *pronominal* (e.g. "she").

Entity detection is generally accomplished in two steps: first, a mention detection module identifies all the mentions of interest; second, a co-reference module merges mentions that refer to the same entity into a single co-reference chain.

A *relation detection* system identifies (typically) binary relationships between pairs of mentions. For instance, for the sentence "I'm in New York", the following relation exists: *locatedAt (I, New York)*.

An *event detection* system identifies events of interest and the arguments of the event. For example, from the sentence "John married Eva in 1940", the system should identify the marriage event, the people who got married and the time of the event.

The latest ACE evaluations involve all of the above tasks. However, as shown in the next section, our focus is quite different from ACE — we are particularly interested in improving performance for conversational speech and building on top of ACE tasks to produce social networks and biographies.

2.2 Fusion Module

The fusion module merges the output from IE modules to extract social networks and biographical facts. For example, if a relation detection system has identified the relation *motherOf (mother, my)* from the input sentence "my mother is a cook", and if an entity recognition module has generated entities referenced by the mentions {*my, Josh, me, I, I, ...*} and {*mother, she, her, her, Rosa......*}, then by replacing *my* and *mother* with the named mentions within the same co-reference chains, the fusion module produces the following nodes and ties in a social network: *motherOf (Rosa, Josh)*.

We generate the nodes of social networks by selecting all the PERSON entities produced by the entity recognition system. Typically, we only include entities that contain at least one *named* mention. To identify ties between nodes, we retrieve all relations that indicate social relationships between a pair of nodes in the network.

We extract biographical profiles by selecting the events (extracted by the event extraction module) and corresponding relations (extracted by the relation extraction module) that involve a given individual as an argument. When multiple documents are used, then we employ a cross-document co-reference system.

3 Improving Performance for Conversational Speech Transcripts

Extracting information from conversational speech transcripts is uniquely challenging. In this section, we describe the data collection used in our experiments, and explain specific techniques we used to improve IE performance on this data.

3.1 Conversational Speech Collection

We use a corpus of videotaped, digitized oral interviews with Holocaust survivors in our experiments. This data was collected by the USC Shoah Foundation Institute (formerly known as the Visual History Foundation), and has been used in many research activities under the Multilingual Access to Large Spoken Archives (MALACH) project (Gustman et al., 2002; Oard et al., 2004). The collection contains oral interviews in 32 languages from 52,000 survivors, liberators, rescuers and witnesses of the Holocaust.

This data is very challenging. Besides the usual characteristics of conversational speech, such as speaker turns and speech repairs, the interview transcripts contain a large percentage of ungrammatical, incoherent, or even incomprehensible clauses (a sample interview segment is shown in Figure 1). In addition, each interview covers many people and places over a long time period, which makes it even more difficult to extract social networks and biographical facts.

> *speaker2* in on that ninth of November nineteen hundred thirty eight I was with my parents at home we heard not through the we heard even through the windows the crashing of glass the crashing of and and they are our can't

Figure 1: Sample interview segment.

3.2 The Importance of Co-reference Resolution

Our initial attempts at social network extraction for the above data set resulted in a very poor score of 0.06 f-measure for finding the relations within a network (as shown in Table 3 as baseline performance).

An error analysis indicated poor co-reference resolution to be the chief culprit for the low performance. For instance, suppose we have two clauses: "his mother's name is Mary" and "his brother Mark went to the army". Further suppose that "his" in the first clause refers to a person named "John" and "his" in the second clause refers to a person named "Tim". If the co-reference system works perfectly, the system should find a social network involving four people: {*John, Tim, Mary, Mark*}, and the ties: *motherOf (Mary, John)*, and *brotherOf (Mark, Tim)*. However, if the co-reference system mistakenly links "John" to "his" in the second clause and links "Tim" to "his" in the first clause, then we will still have a network with four people, but the ties will be: *motherOf (Mary, Tim)*, and *brotherOf (Mark, John)*, which are completely wrong. This example shows that co-reference errors involving mentions that are relation arguments can lead to very bad performance in social network extraction.

Our existing co-reference module is a state-of-the-art system that produces very competitive results compared to other existing systems (Luo et al., 2004). It traverses the document from left to right and uses a mention-synchronous approach to decide whether a mention should be merged with an existing entity or start a new entity.

However, our existing system has shortcomings for this data: the system lacks features for handling conversational speech, and the system often makes mistakes in pronoun resolution. Resolving pronominal references is very important for extracting social networks from conversational speech, as illustrated in the previous example.

3.3 Improving Co-reference for Conversational Speech

We developed a new co-reference resolution system for conversational speech transcripts. Similar to many previous works on co-reference (Ng, 2005), we cast the problem as a classification task and solve it in two steps: (1) train a classifier to determine whether two mentions are co-referent or not, and (2) use a clustering algorithm to partition the mentions into clusters, based on the pairwise predictions.

We added many features to our model specifically designed for conversational speech, and significantly improved the agglomerative clustering used for co-reference, including integrating relations as constraints, and designing better cluster linkage methods and clustering stopping criteria.

3.3.1 Adding Features for Conversational Speech

We added many features to our model specifically designed for conversational speech:

Speaker role identification. In manual transcripts, the speaker turns are given and each speaker is labeled differently (e.g. *"speaker1"*, *"speaker2"*), but the identity of the speaker is not given. An interview typically involves 2 or more speakers and it is useful to identify the roles of each speaker (e.g. interviewer, interviewee, etc.). For instance, "you" spoken by the interviewer is likely to be linked with "I" spoken by the interviewee, but "you" spoken by the third person in the interview is more likely to be referring to the interviewer than to the interviewee.

We developed a program to identify the speaker roles. The program classifies the speakers into three categories: interviewer, interviewee, and others. The algorithm relies on three indicators — number of turns by each speaker, difference in number of words spoken by each speaker, and the ratio of first-person pronouns such as "I", "me", and "we" vs. second-person pronouns such as "you" and "your". This speaker role identification program works very well when we checked the results on the development and test set — the interviewers and survivors in all the documents in the development set were correctly identified.

Speaker turns. Using the results from the speaker role identification program, we enrich certain features with speaker turn information. For example, without this information, the system cannot distinguish "I" spoken by an interviewer from "I" spoken by an interviewee.

Spelling features for speech transcripts. We add additional spelling features so that mentions such as "Cyla C Y L A Lewin" and "Cyla Lewin" are considered as exact matches. Names with spelled-out letters occur frequently in our data collection.

Name Patterns. We add some features that capture frequent syntactic structures that speakers use to express names, such as "her name is Irene", "my cousin Mark", and "interviewer Ellen".

Pronoun features. To improve the performance on pronouns, we add features such as the speaker turns of the pronouns, whether the two pronouns agree in person and number, whether there exist other mentions between them, etc.

Other miscellaneous features. We also include other features such as gender, token distance, sentence distance, and mention distance.

We trained a maximum-entropy classifier using these features. For each pair of mentions, the classifier outputs the probability that the two mentions are co-referent.

We also modified existing features to make them more applicable to conversational speech. For instance, we added pronoun-distance features taking into account the presence of other pronominal references in between (if so, the types of the pronouns), other mentions in between, etc.

3.3.2 Improving Agglomerative Clustering

We use an agglomerative clustering approach for partitioning mentions into entities. This is a bottom-up approach which joins the closest pair of clusters (i.e., entities) first. Initially, each mention is placed into its own cluster. If we have N mentions to cluster, we start with N clusters.

The intuition behind choosing the agglomerative method is to merge the most confident pairs first, and use the properties of existing clusters to constrain future clustering. This seems to be especially important for our data collection, since conversational speech tends to have a lot of repetitions or local structures that indicate co-reference. In such cases, it is beneficial to merge these closely related mentions first.

Cluster linkage method. In agglomerative clustering, each cycle merges two clusters into a single cluster, thus reducing the number of clusters by one. We need to decide upon a method of measuring the distance between two clusters.

At each cycle, the two mentions with the highest co-referent probability are linked first. This results in the merging of the two clusters that contain these two mentions.

We improve upon this method by imposing *minimal distance criteria* between clusters. Two clusters C_1 and C_2 can be combined only if the distance between all the mentions from C_1 and all the mentions from C_2 is above the minimal distance threshold. For instance, suppose $C_1 = \{he, father\}$, and $C_2 = \{he, brother\}$, and "he" from C_1 and "he" from C_2 has the highest linkage probability. The standard single linkage method will combine these two clusters, despite the fact that "father" and "brother" are very unlikely to be linked. Imposing minimal distance criteria can solve this problem and prevent the linkage of clusters which contain very dissimilar mentions. In practice, we used multiple minimal distance thresholds, such as minimal distance between two named mentions and minimal distance between two nominal mentions.

We chose not to use complete or average linkage methods. In our data collection, the narrations contain a lot of pronouns and the focus tends to be very local. Whereas the similarity model may be reasonably good at predicting the distance between two pronouns that are close to each other, it is not good at predicting the distance between pronouns that are furthur apart. Therefore, it seems more reasonable to use single linkage method with modifications than complete or average linkage methods.

Using relations to constrain clustering. Another novelty of our co-reference system is the use of relations for constraining co-reference. The idea is that two clusters should not be merged if such merging will introduce contradictory relations. For instance, if we know that person entity A is the mother of person entity B, and person entity C is the sister of B, then A and C should not be linked since the resulting entity will be both the mother and the sister of B.

We construct co-existent relation sets from the training data. For any two pairs of entities, we collect all the types of relations that exist between them. These types of relations are labeled as co-existent. For instance, "motherOf" and "parentOf" can co-exist, but "motherOf" and "sisterOf" cannot. By using these relation constraints, the system refrains from generating contradictory relations in social networks.

Speed improvement. Suppose the number of mentions is N, the time complexity of simple linkage method is $O(N^2)$. With the minimal distance criteria, the complexity is $O(N^3)$. However, N can be dramatically reduced for conversational transcripts by first linking all the first-person pronouns by each speaker.

4 Experiments

In this section, we describe the experimental setup and present sample outputs and evaluation results.

	Train	Dev	Test
Words	198k	73k	255k
Mentions	43k	16k	56k
Relations	7K	3k	8k

Table 2: Experimental Data Sets.

4.1 Data Annotation

The data used in our experiments consist of partial or complete English interviews of Holocaust survivors. The input to our system is transcripts of interviews.

We manually annotated manual transcripts with entities, relations, and event categories, specifically designed for this task and the results of careful data analysis. The annotation was performed by a single annotator over a few months. The annotation categories for entities, events, and relations are shown in Table 1. Please note that the event and relation definitions are slightly different than the definitions in ACE.

4.2 Training and Test Sets

We divided the data into training, development, and test data sets. Table 2 shows the size of each data set. The training set includes transcripts of partial interviews. The development set consists of 5 complete interviews, and the test set consists of 15 complete interviews. The reason that the training set contains only partial interviews is due to the high cost of transcription and annotation. Since those partial interviews had already been transcribed for speech recognition purpose, we decided to reuse them in our annotation. In addition, we transcribed and annotated 20 complete interviews (each interview is about 2 hours) for building the development and test sets, in order to give a more accurate assessment of extraction performance.

4.3 Implementation

We developed the initial entity detection, relation detection, and event detection systems using the same techniques as our submission systems to ACE (Florian et al., 2004). Our submission systems use statistical approaches, and have ranked in the top tier in ACE evaluations. We easily built the models for our application by retraining existing systems with our training set.

The entity detection task is accomplished in two steps: mention detection and co-reference resolution. The mention detection is formulated as a labeling problem, and a maximum-entropy classifier is trained to identify all the mentions.

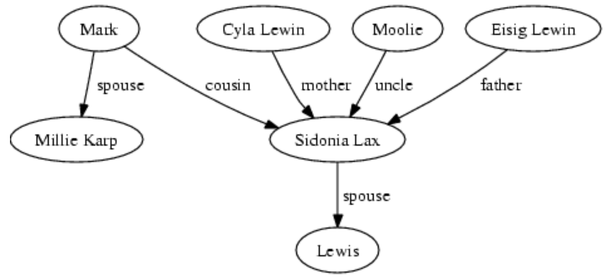

Figure 2: Social network extracted by the system.

Similarly, relation detection is also cast as a classification problem — for each pair of mentions, the system decides which type of relation exists between them. It uses a maximum-entropy classifier and various lexical, contextual, and syntactic features for such predications.

Event detection is accomplished in two steps: first, identifying the event anchor words using an approach similar to mention detection; then, identifying event arguments using an approach similar to relation detection.

The co-reference resolution system for conversational speech and the fusion module were developed anew.

4.4 The Output

The system aims to extract the following types of information:

- The social network of the survivor.
- Important biographical facts about each person in the social network.
- Track the movements of the survivor and other individuals in the social network.

Figure 2 shows a sample social network extracted by the system (only partial of the network is shown). Figure 3 shows sample biographical facts and movement summaries extracted by the system. In general, we focus more on higher precision than recall.

4.5 Evaluation

In this paper, we focus only on the evaluation of social network extraction. We first describe the metrics for social network evaluation and then present the results of the system.

Entity (12)	Event (8)	Relation (34)		
		Social Rels (12)	Event Args (8)	Bio Facts (14)
AGE	CUSTODY	aidgiverOf	affectedBy	bornAt
COUNTRY	DEATH	auntOf	agentOf	bornOn
DATE	HIDING	cousinOf	participantIn	citizenOf
DATEREF	LIBERATION	fatherOf	timeOf	diedAt
DURATION	MARRIAGE	friendOf	travelArranger	diedOn
GHETTOORCAMP	MIGRATION	grandparentOf	travelFrom	employeeOf
OCCUPATION	SURVIVAL	motherOf	travelPerson	hasProperty
ORGANIZATION	VIOLENCE	otherRelativeOf	travelTo	locatedAt
OTHERLOC		parentOf		managerOf
PEOPLE		siblingOf		memberOf
PERSON		spouseOf		near
SALUTATION		uncleOf		partOf
				partOfMany
				resideIn

Table 1: Annotation Categories for Entities, Events, and Relations.

<u>Sidonia Lax:</u>
date of birth: June the eighth nineteen twenty seven

Movements:
Moved To: Auschwitz
Moved To: United States
... ...

Figure 3: Biographical facts and movement summaries extracted by the system.

To compare two social networks, we first need to match the *nodes* and *ties* between the networks. Two nodes (i.e., entities) are matched if they have the same canonical name. Two ties (i.e., edges or relations) are matched if these three criteria are met: they contain the same type of relations, the arguments of the relation are the same, and the order of the arguments are the same if the relation is unsymmetrical.

We define the the following measurements for social network evaluation: the *precision for nodes (or ties)* is the ratio of common nodes (or ties) in the two networks to the total number of nodes (or ties) in the system output, the *recall for nodes (or ties)* is the ratio of common nodes (or ties) in the two networks to the total number of nodes/ties in the reference output, and the *f-measure for nodes (or ties)* is the harmonic mean of precision and recall for nodes (or ties). The *f-measure for ties* indicates the overall performance of social network extraction.

F-mea	Dev		Test	
	Baseline	New	Baseline	New
Nodes	0.59	0.64	0.62	0.66
Ties	0.06	**0.30**	0.06	**0.28**

Table 3: Performance of social network extraction.

Table 3 shows the results of social network extraction. The new co-reference approach improves the performance for f-measure on ties by five-fold on development set and by nearly five-fold for test set.

We also tested the system using automatic transcripts by our speech recognition system. Not surprisingly, the result is much worse: the nodes f-measure is 0.11 for the test set, and the system did not find any relations. A few factors are accountable for this low performance: (1) Speech recognition is very challenging for this data set, since the testimonies contained elderly, emotional, accented speech. Given that the speech recognition system fails to recognize most of the person names, extraction of social networks is difficult. (2) The extraction systems perform worse on automatic transcripts, due to the quality of the automatic transcript, and the discrepancy between the training and test data. (3) Our measurements are very strict, and no partial credit is given to partially correct entities or relations.

We decided not to present the evaluation results of the individual components since the performance of individual components are not at all indicative of the overall performance. For instance, a single pronoun co-reference error might slighlty change the co-reference score, but can introduce a serious error in the social network, as shown in the example in Section 3.2.

5 Biography Generation from General Domain Text

We have applied the same framework to biography generation from general news articles. This general system also contains three fundamental IE systems and a fusion module, similar to the work presented in the paper. The difference is that the IE systems are trained on general news text using different categories of entities, relations, and events.

A sample biography output extracted from TDT5 English documents is shown in Figure 4. The numbers in brackets indicate the corpus count of the facts.

<u>**Saddam Hussein:**</u>
Basic Information:
citizenship: Iraq [203]
occupation: president [4412], leader [1792], dictator [664],...
relative: odai [89], qusay [65], uday [65],...

Life Events:
places been to: bagdad [403], iraq [270], palaces [149]...
Organizations associated with: manager of baath party [1000], ...
Custody Events: Saddam was arrested [52],
Communication Events: Saddam said [3587]
... ...

Figure 4: Sample biography output.

6 Related Work

While there has been previous work on extracting social networks from emails and the web (Culotta et al., 2004), we believe this is the first paper to present a full-fledged system for extracting social networks from conversational speech transcripts.

Similarly, most of the work on co-reference resolution has not focused on conversational speech. (Ji et al., 2005) uses semantic relations to refine co-reference decisions, but in a approach different from ours.

7 Conclusions and Future Work

We have described a novel approach for extracting social networks, biographical facts, and movement summaries from transcripts of oral interviews with Holocaust survivors. We have improved the performance of social network extraction five-fold, compared to a baseline system that already uses state-of-the-art technology. In particular, we improved the performance of co-reference resolution for conversational speech, by feature engineering and improving the clustering algorithm.

Although our application data consists of conversational speech transcripts in this paper, the same extraction approach can be applied to general-domain text as well. Extracting general, rich social networks is very important in many applications, since it provides the knowledge of who is connected to whom and how they are connected.

There are many interesting issues involved in biography generation from a large data collection, such as how to resolve contradictions. The counts from the corpus certainly help to filter out false information which would otherwise be difficult to filter. But better technology at detecting and resolving contradictions will definitely be beneficial.

Acknowledgment

We would like to thank Martin Franz and Bhuvana Ramabhadran for their help during this project. This project is funded by NSF under the Information Technology Research (ITR) program, NSF IIS Award No. 0122466. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

2005. Automatic content extraction. http://www.nist.gov/speech/tests/ace/.

Aron Culotta, Ron Bekkerman, and Andrew McCallum. 2004. Extracting social networks and contact information from email and the web. In *CEAS*, Mountain View, CA.

Radu Florian, Hany Hassan, Abraham Ittycheriah, Hongyan Jing, Nanda Kambhatla, Xiaoqiang Luo, Nicolas Nicolov, and Salim Roukos. 2004. A statistical model for multilingual entity detection and tracking. In *Proceedings of. HLT-NAACL 2004*.

Samuel Gustman, Dagobert Soergeland Douglas Oard, William Byrne, Michael Picheny, Bhuvana Ramabhadran, and Douglas Greenberg. 2002. Supporting access to large digital oral history archives. In *Proceedings of the Joint Conference on Digital Libraries*, pages 18–27.

Heng Ji, David Westbrook, and Ralph Grishman. 2005. Using semantic relations to refine coreference decisions. In *Proceedings of HLT/EMNLP'05*, Vancouver, B.C., Canada.

Xiaoqiang Luo, Abe Ittycheriah, Hongyan Jing, Nanda Kambhatla, and Salim Roukos. 2004. A mention-synchronous coreference resolution algorithm based on the bell tree. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL2004)*, pages 135–142, Barcelona, Spain.

Vincent Ng. 2005. Machine learning for coreference resolution: From local classification to global ranking. In *Proceedings of ACL'04*.

D. Oard, D. Soergel, D. Doermann, X. Huang, G.C. Murray, J. Wang, B. Ramabhadran, M. Franz, S. Gustman, J. Mayfield, L. Kharevych, and S. Strassel. 2004. Building an information retrieval test collection for spontaneous conversational speech. In *Proceedings of SIGIR'04*, Sheffield, U.K.