Dimensions of Interpersonal Relationships: Corpus and Experiments

Farzana Rashid and **Eduardo Blanco**
Human Intelligence and Language Technologies Lab
University of North Texas
Denton, TX, 76203
farzanarashid@my.unt.edu, eduardo.blanco@unt.edu

Abstract

This paper presents a corpus and experiments to determine dimensions of interpersonal relationships. We define a set of dimensions heavily inspired by work in social science. We create a corpus by retrieving pairs of people, and then annotating dimensions for their relationships. A corpus analysis shows that dimensions can be annotated reliably. Experimental results show that given a pair of people, values to dimensions can be assigned automatically.

1 Introduction

The task of information extraction (IE) consists in creating structured representations from unstructured text. These representations usually consist of relations explicitly stated in text, and involve two or more arguments. For example, IE systems would extract SPOUSE(*John*, *Mary*) or MARRIED(*John*, *Mary*, *1994*) from *John and Mary have been married since 1994*. IE systems have a long history, and became popular after evaluations such as MUC (Grishman and Sundheim, 1996) and ACE (Doddington et al., 2004).

Traditional IE systems are supervised and extract relations defined before training takes place (Peng and McCallum, 2004). More recently, open IE systems have been proposed to extract *all* relations explicitly stated in text in an unsupervised manner and without defining relations a priori (Mausam et al., 2012). Regarding interpersonal relations—relations that take as arguments two people—both IE approaches extract relations such as RELATIVE, FRIEND and COMMUNICATES_WITH. Open IE systems are domain independent and would extract, in principle, relations such as CLASSMATES and ADVISOR from students' diaries or biographies of scientists.

While useful for applications such as question answering (Yao and Van Durme, 2014), these dyadic relations only provide a generic understanding of the relationship between two people. For example, COMMUNICATES_WITH may relate people who have an *intense* or *superficial* relationship (e.g., engaged couples talking about wedding plans vs. home owners discussing remodels with contractors), *pleasure-* or *task-oriented* relationships (e.g., friends planning a backpacking trip vs. software developers discussing the next delivery), and may be *spatially near* or *distant* (e.g., individuals having an in-person meeting vs. those exchanging emails or talking on the phone).

These elemental properties of interpersonal relationships are called *dimensions* in social science, and have been studied for decades (Wish et al., 1976). In those studies, the goal is to understand how relationships are *perceived* by people, not to extract them. Note that unlike interpersonal relationships, their dimensions are usually implicitly stated in text, thus extracting them is challenging. Also, extracting dimensions of interpersonal relationships requires text understanding beyond the event in which two people participate. As shown above, two people who communicate may have different dimension values depending on what they talk about or the communication device.

In this paper, we target dimensions of interpersonal relationships that characterize the nature of relationships beyond a name per relationship. The main contributions are: (a) set of dimensions of interpersonal relationships, including dimensions from previous work outside computational linguistics and novel ones; (b) corpus consisting of pairs of people and values for the dimensions of their relationships (Cohen's kappa: 0.68); (c) detailed corpus analysis; and (d) experimental results showing that the dimensions can be extracted automatically from text.

Dimension	Other descriptors and aliases	Ref.
Cooperative vs. Competitive	friendly vs. hostile, promotive vs. contrient	[1]
Equal vs. Hierarchical	autonomy vs. control, submission vs. dominance	[1]
Intense vs. Superficial	important vs. insignificant, influential vs. trivial	[1]
Pleasure vs. Task Oriented	personal vs. impersonal, emotionally involved vs. detached	[1]
Active vs. Passive	direct vs. indirect, unequivocal vs. equivocal	[2]
Intimate vs. Unintimate	emotionally close vs. distant, indifferent, random	[3]
Temporary vs. Enduring	momentary vs. lasting, provisional vs. permanent	[4]
Concurrent vs. Nonconcurrent	convergent vs. divergent, synchronous vs. asynchronous	New
Spatially Near vs. Distant	nearby vs. faraway, attached vs. detached	New

Table 1: Dimensions of interpersonal relationships targeted in this paper. [1] stands for (Wish et al., 1976), [2] for (Kelley, 2013), [3] for (Adamopoulos, 2012), and [4] for (Deutsch, 2011). *New* indicates a dimension discovered after analyzing several examples and pilot annotations.

2 Related Work

Extracting relations between entities such as people, organizations and locations is the core goal of the task of information extraction. A few competitions have served as evaluation benchmarks (Grishman and Sundheim, 1996; Doddington et al., 2004; Kulick et al., 2014; Surdeanu and Heng, 2014), and include interpersonal relationships such as BUSINESS, SPOUSE and CHILDREN. Aguilar et al. (2014) compare several evaluations, and automated approaches to relation extraction—also referred to as link prediction and knowledge base completion— include (Yu and Lam, 2010; Nguyen et al., 2016; West et al., 2014). Open information extraction (Wu and Weld, 2010; Angeli et al., 2015) has emerged as an unsupervised domain-independent approach to extract relations. Regardless of details, all these previous efforts extract explicit relations, and do not attempt to characterize instances of relations with dimensions.

Besides extracting relations per se, there have been efforts within computational linguistics involving interpersonal relationships. Voskarides et al. (2015) extract human-readable descriptions of relations in a knowledge graph by ranking sentences that justify the relations. Iyyer et al. (2016) propose an unsupervised algorithm to extract relationship trajectories of fictional characters, i.e., how interpersonal relationships evolve over time in fictional stories. Bracewell et al. (2012) introduce 9 social acts (e.g., agreement, undermining) designed to characterize relationships between individuals exhibiting adversarial and collegial behavior (similar to our cooperative vs. competitive dimension).

Researchers have studied from a computational perspective how people com- municate with each other. For example, Danescu-Niculescu-Mizil et al. (2012) study how power differences affects language style in online communities, and Prabhakaran and Rambow (2014) present a classifier to detect power relationships in email threads. Similarly, Gilbert (2012) explores how people in hierarchical relationships communicate through email, and Bramsen et al. (2011) focus on identifying power relationships in social networks. Politeness in online forums has also been studied (Danescu-Niculescu-Mizil et al., 2013). While power (similar to our equal vs. hierarchical dimension, Section 3) and politeness could be considered dimensions, these works exploit structural and linguistic features derived from communications between two individuals. Unlike all of them, we extract 9 dimensions of interpersonal relationships from sentences describing an event involving two people, and without needing language samples from them.

3 Dimensions of Interpersonal Relationships

Dimensions of interpersonal relationships have been studied for decades outside of computational linguistics, mostly in psychology and social science in general (Wish et al., 1976). The set of dimensions is by no means agreed upon, and neither is the terminology to refer to what apparently is the same dimension. For example, the terms *dominance*, *submission*, *potency*, *autonomy* and *control* are used to describe the distribution of power in a relationship (Deutsch, 2011).

The dimensions we work with in this paper are primarily borrowed from previous works in social science, although we add two new dimensions. Interestingly, the previous works which define these dimensions do so from a theoretical point of view

or after conducting experiments with subjects to reveal how they perceive interpersonal relationships. The latter was done using multidimensional scaling analysis after subjects compared 25 relationships, e.g., between a parent and child, between business partners (Wish et al., 1976).

Table 1 presents the nine dimensions targeted in this paper along with the original references and aliases found in the literature. Social scientists have proposed additional dimensions, e.g., voluntary vs. involuntary, public vs. private and licit vs. illicit (Deutsch, 2011), or self-benefiting vs. service-oriented (Adamopoulos, 2012). We discarded these additional dimensions because we discovered that they are not applicable to most pairs of people we work with (Section 4).

We provide below brief descriptions of the nine dimensions of interpersonal relationships we work with. Note that these dimensions are not completely independent, for example, *enduring* relationships are usually *intense* and *intimate*, and *intense* and *pleasure-oriented* relationship are almost always *intimate*. Section 4.3 presents examples, and Section 4.4 discusses inter-dimensional correlations and inter-annotator agreement. We point the reader to the references in Table 1 for additional details, rationales, and examples.

Cooperative vs. Competitive A relationship is *cooperative* if both people (a) have a common interest or goal, (b) like each other, (c) benefit from the relationship, or (d) think alike or have similar views. Otherwise, the relationship is *competitive*.

Equal vs. Hierarchical A relationship is *equal* if both people (a) have the same social status, (b) are at the same level in the power structure, (c) share similar responsibilities, or (d) have the same role. Otherwise, the relationship is *hierarchical*.

Intense vs. Superficial A relationship is *intense* if both people interact with each other frequently, i.e., they are involved repeatedly. Otherwise, the relationship is *superficial*.

Pleasure vs. Task Oriented A relationship is *pleasure oriented* if both people interact socially and their relationship is not bound by professional rules or regulations. Otherwise, the relationship is *task oriented*.

Active vs. Passive A relationship is *active* if both people are involved in a shared activity or event that grants the relationship. Otherwise,

the relationship is *passive*. For example, individuals commuting to work in the same car have an *active* relationship, but those who happen to take the same subway line to work have a *passive* relationship.

Intimate vs. Unintimate A relationship is *intimate* if both people are emotionally close and warm to each other. Otherwise, the relationship in *unintimate*.

Temporary vs. Enduring A relationship is *temporary* if it lasts less than a day. A relationship is *enduring* if it lasts over a month. Otherwise (if it lasts more than a day and less than a month), this dimension is undefined.

Concurrent vs. Nonconcurrent A relationship is *concurrent* if both people are involved in an event or action at the same time. Otherwise, the relationship is *nonconcurrent*.

Spatially Near vs. Distant A relationship is *spatially near* (or *near* for short) if both people are at the same location during the event that grants the relationship. Otherwise, the relationship is *spatially distant* (or *distant*).

4 Building a Corpus of Dimensions of Interpersonal Relationships

Existing corpora annotating relations (Section 2) only consider selected interpersonal relationships and do not target dimensions. Our goal is to target dimensions of interpersonal relationships between any two individuals, from weak links (e.g., journalists interviewing celebrities) to strong ties (e.g., close friends). Thus, we create a corpus[1] by first retrieving pairs of people, and then annotating dimensions for their relationships.

We decided to add our annotations to OntoNotes (Hovy et al., 2006). Doing so has several advantages. First, OntoNotes contains texts from several domains and genres (e.g., conversational telephone speech, weblogs, broadcast), thus we not only work with newspaper articles. Second, OntoNotes includes part-of-speech tags, named entities and coreference chains, three annotation layers that allow us to streamline the corpus creation process.

4.1 Retrieving Pairs of People

We retrieve pairs of people within each sentence in OntoNotes following four steps:
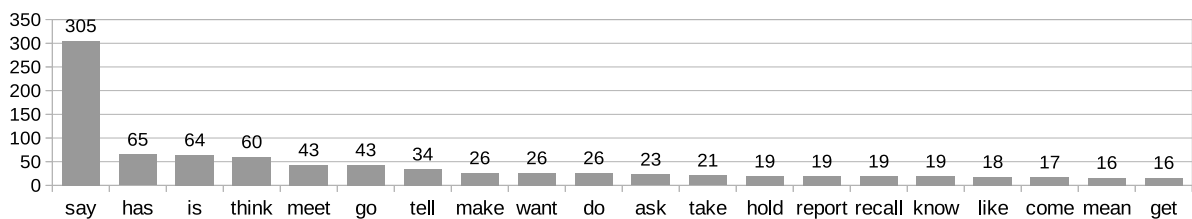
---

[1] Available at http://hilt.cse.unt.edu/

Figure 1: Frequencies of the top 20 most frequent verbs after retrieving pairs of people (Section 4.1). We discard verbs with frequency <4, and randomly select up to 26 pairs per verb for a total of 1,048 pairs.

1. Collect all instances of personal pronouns (part-of-speech tag PRP) *I*, *he* and *she*.
2. Collect all named entities PERSON.
3. Keep one mention per coreference chain, giving priority to named entities over pronouns.
4. Generate combinations of 2 elements from the union of the pronouns and named entities subject to the following constraints: at least
   (a) one is a PERSON named entity, and
   (b) one is the *nsubj* (nominal subject syntactic dependency) of a verb.
   The elements of the pair that satisfy restrictions (4a) and (4b) need not be the same.

Note that removing duplicate mentions (Step 3) does not reduce the number of relationships targeted, it simply avoids duplicate pairs. Also, the only syntactic constraint is that one person in the pair must be the nominal subject of a verb. Thus, we work with relationships between individuals from different clauses, and connected with a variety of syntactic paths (see Examples in Table 2).

The total number of pairs generated skipping Steps 3 and 4 would be 4,886. After removing duplicate mentions (Step 3), the number is reduced to 3,481; restrictions (4a) and (4b) further reduce the number to 3,143 and 2,696 respectively. Executing all steps yields 2,364 pairs.

Figure 1 presents verb frequencies for the top 20 most frequent verbs in the 2,364 pairs. In order to reduce the annotation effort and account for a variety of verbs, we set to annotate 1,000 pairs. After trying several thresholds, we retrieved 1,048 pairs by selecting pairs from verbs that occur at least 4 times, and randomly selecting up to 26 pairs per verb (most verbs occur less than 26 times).

4.2 Annotating Dimensions of Interpersonal Relationships

After generating pairs, annotators determine values for each dimension of interpersonal relationships. The annotation interface shows the sentence from which the pair was generated, and the previous and next sentence to provide some context. The pair of people of interest were highlighted, but no additional information was shown (e.g., the verb of which one person is the subject).

Annotators assign a value to each dimension based on the relationship between the two individuals at the time the verbal event of which one of the individuals is the subject takes place. They were trained to take into account context (previous and next sentences), and to interpret the text as they normally would. Therefore, they assign values using world knowledge that may not be explicitly stated in the text. For example, two people talking on the phone would have a *spatially distant* relationship because (most likely) they are not next to each other while talking. Annotating the changes over time of the dimensions is outside the scope of this paper.

During the first batch of annotations, we discovered that for a given pair of people, dimensions sometimes cannot be determined because (a) there is not enough evidence in the text provided (i.e., sentence from which the pair was generated, previous and next sentences) or (b) the pair is invalid and assigning dimensions is nonsensical. We use 0 label in the former case, and inv in the latter. For example, in the sentence *[He]$_y$ [criticized]$_{verb}$ [Ken Starr]$_x$*, the value for the dimension *spatially near* (vs. *distant*) was marked 0 as there is not enough information to determine whether *He* and *Ken Starr* are at the same location when *criticized* took place. The most frequent example of inv is when *God* is marked as a PERSON named entity in the gold annotations in OntoNotes.

In the rest of the paper, we refer to dimensions by the first descriptor in Table 1, and use 1 if the first descriptor of a dimension is true, and -1 if the second descriptor is true. For example label -1 applied to dimension *temporary* means that the realtionship is *enduring*.

2310

	Sentence	Cooperative	Equal	Intense	Pleasure Oriented	Active	Intimate	Temporary	Concurrent	Spatially Near
1	[Cheney]$_x$ [got]$_{verb}$ a telephone call from his democratic counterpart [Joseph Lieberman]$_y$ wishing him a speedy recovery.	1	1	-1	-1	1	-1	1	1	-1
2	[...] [I]$_x$ [interviewed]$_{verb}$ one of the nation's top jockies [Shane Sellers]$_y$ about the battle he waged everyday to control his weight.	1	-1	-1	-1	1	-1	1	1	1
3	[I]$_x$ have always remembered the encouragement which Mr. [Yu Youren]$_y$ [gave]$_{verb}$ me as a young reporter. He said that to be a fearless champion of social justice, as is expected of a journalist, the [...]	1	-1	1	-1	1	1	-1	1	1
4	One day, Dingxiang took the opportunity to again urge him to change his ways [...]. After this, [Zhang Sheng]$_x$ [threw]$_{verb}$ out [Dingxiang]$_y$, sold off the family possessions, and spent his days living a life of dissipation.	-1	1	-1	1	1	-1	-1	1	1

Table 2: Annotation examples for pairs of people (x, y). We refer to dimensions by their first descriptor (Section 3); 1 (-1) indicates that the first (second) descriptor is true, and 0 that the value is unknown.
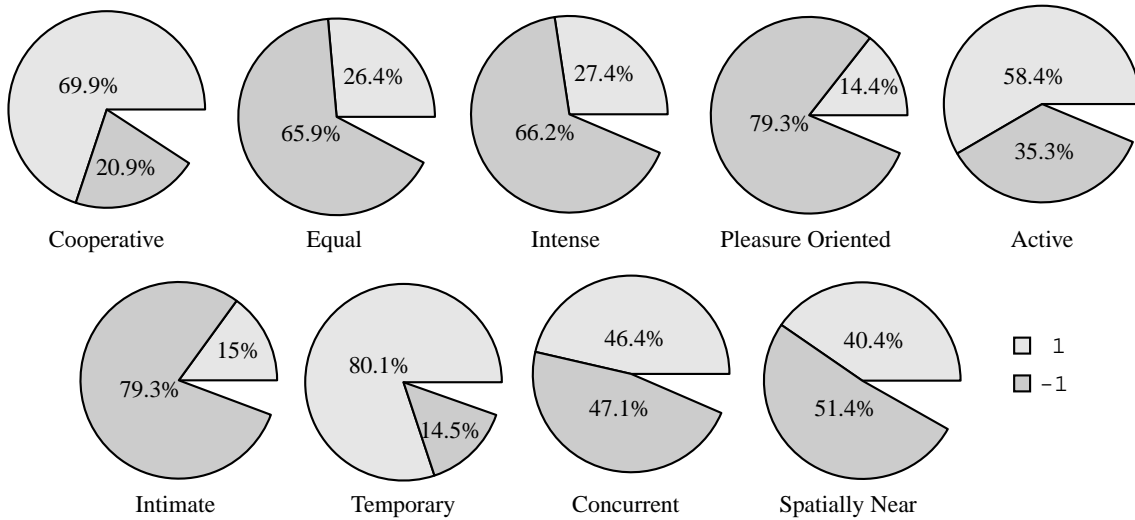

Figure 2: Label distribution per dimension of interpersonal relationships. The missing portion of each pie chart corresponds to labels 0 and inv, which always amount to less than 5% each.

4.3 Annotation Examples

We present annotation examples from our corpus in Table 2, including context if it is relevant. We acknowledge that some annotations are ambiguous, and discuss label distributions and inter-annotator agreement in Section 4.4.

Sentences (1) and (2) encode a COMMUNICATION relationships between two individuals, and both are *cooperative*, *superficial*, *work oriented*, *active*, *unintimate*, *temporary*, and *concurrent*. The values for two dimensions, however, are different. Two counterparts (Sentence 1) are at the same level in the power structure (*equal*), but interviewer and interviewee are not (Sentence 2). Similarly, talking on the phone entails that the individuals are *spatially distant* (Sentence 1), but inter-

viewing (most likely) means that they were *spatially near* (Sentence 2). One could argue that 0 would be a better label for *spatially near* in Sentence (2), but annotators interpreted that *interviewed* refers to an in-person interview.

Sentence (3) in context describes one person (Yu Youren) encouraging another one (I). Annotators indicate that this relationship, unlike the ones in Sentences (1) and (2), is *intense* (frequent interaction), *intimate* (emotionally close), and *enduring* (lasting over a month). These values are not explicitly stated, but they are understood given the long-lasting impact *Yu Youren* had on *I*.

Finally, Sentence (4) exemplifies a *competitive* relationship. The context describes a struggling relationship between *Dingxiang* and *Zhang Sheng*. When the latter *threw* the former out, the rela-

	Cooperative	Equal	Intense	Pleasure Or.	Active	Intimate	Temporary	Concurrent
Cooperative	–							
Equal	-.06	–						
Intense	.24	.11	–					
Pleasure Or.	.14	.29	.39	–				
Active	.18	.22	.43	.25	–			
Intimate	.18	.17	.59	.62	.31	–		
Temporary	-.17	.10	-.56	-.43	-.27	-.61	–	
Concurrent	.17	.22	.31	.21	.74	.26	-.17	–
Spat. Near	.21	.18	.30	.25	.68	.28	.19	.89

Table 3: Pearson correlations between dimensions of interpersonal relationships in our corpus.

Dimension	Agreement	κ coefficient
Cooperative	86%	0.74
Equal	81%	0.63
Intense	84%	0.73
Pleasure Oriented	86%	0.70
Active	82%	0.59
Intimate	83%	0.68
Temporary	76%	0.61
Concurrent	84%	0.72
Spatially Near	80%	0.67
All	82%	0.68

Table 4: Inter-annotator agreement per dimension of interpersonal relationships. κ values in the 0.60–0.80 range are considered *substantial*, over 0.80 would be *perfect* (Landis and Koch, 1977).

tionship was *superficial* and *unintimate*, but (most likely) existed for longer than a month (*enduring*).

4.4 Corpus Analysis

Label Distribution. Figure 2 shows the percentage of label 1 (first descriptor) and -1 (second descriptor) per dimension in our corpus. The percentage of Label 0 ranges from 0.86% (*temporary*) to 4.3% (*cooperative*) depending on the dimension, and the percentage of inv is 4.6% (not shown in Figure 2). Importantly, annotators assigned a useful value (either 1 or -1) to most pairs of people (>90%) for all dimensions.

The distributions of 1 and -1 clearly show that most dimensions are biased towards one label. For example, few relationships are *pleasure oriented* (14.4%) or *enduring* (14.5%). The exceptions are *concurrent vs. nonconcurrent*, with roughly the same percentages (46.4% and 47.1%), *spatially near vs. distant* (40.4% and 51.4%) and *active vs. passive* (58.4% and 35.3%). These distributions are not a representative sample of all interpersonal relationships, we would expect many *pleasure ori-*

ented and *intimate* relationship if we work with personal diaries instead of OntoNotes.

Inter-Dimension Correlations. While the dimensions we work with have a long tradition in social science (Section 2), to the best of our knowledge, they have not been extensively annotated in text before. Table 3 shows inter-dimensional correlations for all pairs of dimensions in our corpus. Not surprisingly, some dimensions correlate with each other. For example, *enduring* relationships tend to also be *intense* (0.56) and *intimate* (0.61), and *concurrent* relationships tend to be *active* (0.74). The highest correlation is between *concurrent* and *spatially near* (0.89), indicating that if two people participate in a common event at the same time, usually they are at the same location (see counterexample in Table 2, Sentence 1). Note, however, that most correlations are low, and some dimensions (e.g., *cooperative*, *equal*) have low correlations (<0.30) with all dimensions.

Inter-Annotator Agreement. The annotations were done by two graduate students. They started annotating small batches of pairs of people, and discussed disagreements with each other. After several iterations, they annotated independently 10% of all pairs of people generated. Table 4 depicts the inter-annotator agreements obtained. Overall Cohen's kappa coefficient is 0.68, and the coefficients range between 0.59 to 0.74 depending on the dimension. Note that kappa coefficients in the range 0.60–0.80 are considered *substantial*, and over 0.80 would be *perfect* (Landis and Koch, 1977). Given these high agreements, the rest of pairs were annotated once.

5 Experiments and Results

We conduct experiments using standard supervised machine learning. Each pair of people become an instance, and we split instances into training (80%) and test (20%). As a learning algorithm, we use SVM with RBF kernel as implemented in scikit-learn (Pedregosa et al., 2011).

We report results in the test set after tuning the SVM parameters (C and γ) using 10-fold cross-validation with the training set. More specifically, we train one classifier per dimension, and experiment with all instances but the ones annotated inv. Thus, each classifier predicts 3 labels: 1 (the first descriptor applies), -1 (the second descriptor applies), and 0 (neither descriptor applies).

	Feature	Description
Verb	word, tag	Word form and part-of-speech tag of the verb
	dep_out	Outgoing syntactic dependency from verb
	deps_in	Flags indicating incoming syntactic dependencies to the verb
	lex_name	Name of the WordNet lexical file of the verb
	token_before	Word form and part-of-speech tag of the token before the verb
	token_after	Word form and part-of-speech tag of the token after the verb
Person	words, tags	Concatenation of word forms and part-of-speech tags
	type	Flag indicating whether the person is a pronoun or named entity
	dep_out	Outgoing syntactic dependency
	distance_verb	Number of tokens between the person and the verb
	first_token	Word form and part-of-speech tag of the first token in the person
	last_token	Word form and part-of-speech tag of the last token in the person
	token_before	Word form and part-of-speech tag of the token before the person
	token_after	Word form and part-of-speech tag of the token after the person
$Person_x$_$Person_y$	direction	Flag indicating whether x occurs before or after y
	type	Flag indicating whether x and y are PERSON NEs, or either one is a pronoun

Table 5: Feature set used to determine dimensions of interpersonal relationships between pairs of people (x, y). *Verb* features are extracted from the verb of which either x or y is the subject, *Person* features are extracted from x and y independently, and *Persons* features are extracted from x and y.

5.1 Feature Set

The features we work with are summarized in Table 5. Most features are standard and have been used before to extract relations from text (Section 2). Following the notation in Table 2, we refer to the pair of people as x and y.

Verb features capture information about the verb to which x or y attach. We include words and part-of-speech tags (verb, and tokens before and after), the name of the WordNet lexical file to which the verb belongs, and dependencies.

Person features are extracted from x and y independently, and consists mostly of words and part-of-speech tags. We also include a flag indicating whether the person is a pronoun or named entity (type feature), and the number of tokens between the person and the verb (distance_verb).

$Person_x$_$Person_y$ features capture information of both x and y. They capture (a) whether x occurs before or after y in the sentence, and (b) whether they are both named entities or one is a pronoun and the other one a named entity (type feature).

5.2 Results

We present overall results (averages of the classifiers for each dimension) using the majority baseline and with several feature combinations in Table 6. Then, we present detailed results per dimension with the best feature combination in Table 7. We only present results obtained in the test set.

Overall Results and Feature Ablation (Table 6). The majority baseline obtains 0.53 average F-measure. Recall that we build a classifier per dimension, thus the combination of the nine

Features	Label	P	R	F
Majority Baseline	1	0.43	0.76	0.55
	0	0.00	0.00	0.00
	-1	0.57	0.59	0.64
	Avg.	0.45	0.65	0.53
Verb	1	0.62	0.72	0.64
	0	0.00	0.00	0.00
	-1	0.71	0.78	0.65
	Avg.	0.62	0.71	0.65
Verb, $Person_x$	1	0.68	0.75	0.69
	0	0.00	0.00	0.00
	-1	0.78	0.79	0.78
	Avg.	0.70	0.74	0.71
Verb, $Person_y$	1	0.64	0.72	0.67
	0	0.00	0.00	0.00
	-1	0.74	0.76	0.74
	Avg.	0.66	0.70	0.67
Verb, $Person_x$, $Person_y$, $Person_x$_$Person_y$	1	0.69	0.74	0.70
	0	0.00	0.00	0.00
	-1	0.77	0.80	0.76
	Avg.	**0.71**	**0.76**	**0.72**

Table 6: Results obtained for all dimensions with several combinations of features.

majority-baseline classifiers predict two labels: 1 (0.55 F-measure) and -1 (0.64 F-measure).

Models trained with any combination of features outperform the majority baseline, but they never learn to predict label 0. Since 0 occurs between 0.86% and 4.3% depending on the dimension (Section 4.4), this limitation does not affect overall performance substantially.

Dimension	1 (1st descriptor)			0 (unknown)			-1 (2nd descriptor)			All		
	P	R	F	P	R	F	P	R	F	P	R	F
Cooperative	0.73	0.96	0.83	0.00	0.00	0.00	0.60	0.19	0.29	0.66	0.72	0.65
Equal	0.56	0.10	0.17	0.00	0.00	0.00	0.74	0.97	0.84	0.68	0.74	0.66
Intense	0.39	0.30	0.34	0.00	0.00	0.00	0.78	0.85	0.82	0.67	0.71	0.69
Pleasure	0.40	0.28	0.33	0.00	0.00	0.00	0.87	0.93	0.90	0.79	0.82	0.80
Active	0.69	0.85	0.76	0.00	0.00	0.00	0.68	0.51	0.58	0.67	0.69	0.67
Intimate	0.44	0.17	0.24	0.00	0.00	0.00	0.88	0.98	0.93	0.81	0.86	0.83
Temporary	0.85	0.96	0.91	0.00	0.00	0.00	0.33	0.10	0.16	0.77	0.83	0.79
Concurrent	0.72	0.80	0.76	0.00	0.00	0.00	0.77	0.75	0.76	0.71	0.74	0.73
Spat. Near	0.66	0.68	0.67	0.00	0.00	0.00	0.73	0.79	0.76	0.66	0.70	0.68
Average	0.69	0.74	0.70	0.00	0.00	0.00	0.77	0.80	0.76	0.71	0.76	0.72

Table 7: Results obtained for each dimension with the best combination of features for all dimensions ($Verb + Person_x + Person_y + Person_x\_Person_y$, boldfaced in Table 6)

Verb features alone yield a 0.65 average F-measure (1: 0.64, -1: 0.65). Adding features derived from x (*Verb + Person$_x$*) improves performance (0.71 average F-measure), and adding features derived from y (*Verb + Person$_y$*) slightly improves performance (0.67 average F-measure). In both cases, -1 is predicted more accurately than 1 (0.78 vs. 0.69 and 0.74 vs. 0.67).

Finally, adding all features (*Verb + Person$_x$ + Person$_y$ + Person$_x$_Person$_y$*) yields the best results (average F-measure: 0.72), although by a minimal margin with respect to *Verb + Person$_x$*.

Detailed Results per Dimension. Table 7 presents results per dimension with the best overall combination of features (*Verb + Person$_x$ + Person$_y$ + Person$_x$_Person$_y$*).

All dimensions obtain overall F-measures between 0.65 and 0.83 (last column). Results per label are heavily biased towards the most frequent label per dimension (Figure 2), although it is the case that the models we experiment with predict both 1 and -1 for all dimensions. As stated above, none of them predict 0, but this limitation does not substantially penalize overall performance because of the low frequency of this label.

The model obtains the same F-measures for 1 and -1 with *concurrent* dimension (0.76), and the labels of this dimension are virtually distributed uniformly (46.4% vs. 47.1%, Figure 2). Similarly, F-measures for 1 and -1 with *spatially near* and *active* dimensions are similar (0.67 vs. 0.76 and 0.76 vs. 0.58), and the labels are distributed relatively evenly in our corpus (40.4% vs 51.4% and. 58.4% vs. 35.3%).

Finally, F-measures per label with other dimen-

sion are biased towards the most frequent label. For example, only 15% of all pairs of people have an *enduring* relationship (Figure 2), and the F-measure for 1 with *temporary* dimension is much higher (0.91) that for -1 (0.16).

6 Conclusions

We have presented a set of nine dimensions of interpersonal relationships, including dimensions with a long tradition in social science and new ones. These dimensions allow us to differentiate core characteristics of the relationship between two individuals. For example, people that communicate may be *spatially near* or *spatially distant* (asking questions in class vs. chatting online), and have a pleasure-oriented or work-oriented relationship (somebody wishing good luck to a friend vs. interviewer and interviewee).

Our annotations show that assigning values to dimensions can be done reliably (Cohen's kappa: 0.68), and that useful values (1 and -1 labels) are assigned to dimensions in most pairs of people (>90%). Experimental results following a standard supervised machine learning approach show that assigning values to dimensions can be automated (0.72 overall F-measure), and that results per label and dimensions are biased towards the most frequent label.

We believe that extracting dimensions of interpersonal relationships complements previous efforts that extract relationships. Our future plans include studying values of dimensions for selected relationships (e.g., COWORKER), and investigating changes on the dimensions of the relationship over time. The latter would allow us to, for exam-

ple, analyze how the relationship between two individuals changes over time, and determine which events make a relationship go from *superficial* to *intense* and vice versa.

References

John Adamopoulos. 2012. The emergence of social meaning: A theory of action construal. In *Handbook of Social Resource Theory: Theoretical Extensions, Empirical Insights, and Social Applications*, pages 255–272. Springer New York, New York, NY.

Jacqueline Aguilar, Charley Beller, Paul McNamee, Benjamin Van Durme, Stephanie Strassel, Zhiyi Song, and Joe Ellis. 2014. A comparison of the events and relations across ace, ere, tac-kbp, and framenet annotation standards. In *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 45–53, Baltimore, Maryland, USA. Association for Computational Linguistics.

Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D. Manning. 2015. Leveraging linguistic structure for open domain information extraction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 344–354, Beijing, China. Association for Computational Linguistics.

David B. Bracewell, Marc T. Tomlinson, Mary Brunson, Jesse Plymale, Jiajun Bracewell, and Daniel Boerger. 2012. Annotation of adversarial and collegial social actions in discourse. In *Proceedings of the Sixth Linguistic Annotation Workshop*, LAW VI '12, pages 184–192, Stroudsburg, PA, USA. Association for Computational Linguistics.

Philip Bramsen, Martha Escobar-Molano, Ami Patel, and Rafael Alonso. 2011. Extracting social power relationships from natural language. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 773–782, Portland, Oregon, USA. Association for Computational Linguistics.

Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of power: Language effects and power differences in social interaction. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 699–708, New York, NY, USA. ACM.

Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 250–259, Sofia, Bulgaria. Association for Computational Linguistics.

Morton Deutsch. 2011. Interdependence and psychological orientation. In *Conflict, Interdependence, and Justice*, pages 247–271. Springer.

George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. 2004. The automatic content extraction (ace) program tasks, data, and evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC-2004)*, Lisbon, Portugal. European Language Resources Association (ELRA). ACL Anthology Identifier: L04-1011.

Eric Gilbert. 2012. Phrases that signal workplace hierarchy. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pages 1037–1046, New York, NY, USA. ACM.

Ralph Grishman and Beth Sundheim. 1996. Message understanding conference-6: A brief history. In *Proceedings of the 16th Conference on Computational Linguistics - Volume 1*, COLING '96, pages 466–471, Stroudsburg, PA, USA. Association for Computational Linguistics.

Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. OntoNotes: the 90% Solution. In *NAACL '06: Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers on XX*, pages 57–60, Morristown, NJ, USA. Association for Computational Linguistics.

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, and Hal Daumé III. 2016. Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1534–1544, San Diego, California. Association for Computational Linguistics.

Harold H Kelley. 2013. *Personal relationships: Their structures and processes*. Psychology Press.

Seth Kulick, Ann Bies, and Justin Mott. 2014. Interannotator agreement for ere annotation. In *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 21–25, Baltimore, Maryland, USA. Association for Computational Linguistics.

J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1).

Mausam, Michael Schmitz, Robert Bart, Stephen Soderland, and Oren Etzioni. 2012. Open language learning for information extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL '12, pages 523–534, Stroudsburg, PA, USA. Association for Computational Linguistics.

Dat Quoc Nguyen, Kairit Sirts, Lizhen Qu, and Mark Johnson. 2016. Stranse: a novel embedding model of entities and relationships in knowledge bases. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 460–466, San Diego, California. Association for Computational Linguistics.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Fuchun Peng and Andrew McCallum. 2004. Accurate information extraction from research papers using conditional random fields. In *HLT-NAACL 2004: Main Proceedings*, pages 329–336, Boston, Massachusetts, USA. Association for Computational Linguistics.

Vinodkumar Prabhakaran and Owen Rambow. 2014. Predicting power relations between participants in written dialog from a single thread. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 339–344, Baltimore, Maryland. Association for Computational Linguistics.

Mihai Surdeanu and Ji Heng. 2014. Overview of the english slot filling track at the tac2014 knowledge base population evaluation. In *Proceedings of the TAC-KBP 2014 Workshop*.

Nikos Voskarides, Edgar Meij, Manos Tsagkias, Maarten de Rijke, and Wouter Weerkamp. 2015. Learning to explain entity relationships in knowledge graphs. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 564–574, Beijing, China. Association for Computational Linguistics.

Robert West, Evgeniy Gabrilovich, Kevin Murphy, Shaohua Sun, Rahul Gupta, and Dekang Lin. 2014. Knowledge base completion via search-based question answering. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 515–526, New York, NY, USA. ACM.

Myron Wish, Morton Deutsch, and Susan J Kaplan. 1976. Perceived dimensions of interpersonal relations. *Journal of Personality and Social Psychology*, 33(4):409.

Fei Wu and Daniel S. Weld. 2010. Open information extraction using wikipedia. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 118–127, Stroudsburg, PA, USA. Association for Computational Linguistics.

Xuchen Yao and Benjamin Van Durme. 2014. Information extraction over structured data: Question answering with freebase. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 956–966, Baltimore, Maryland. Association for Computational Linguistics.

Xiaofeng Yu and Wai Lam. 2010. Jointly identifying entities and extracting relations in encyclopedia text via a graphical model approach. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, COLING '10, pages 1399–1407, Stroudsburg, PA, USA. Association for Computational Linguistics.