Human Language Technology Research Institute

# Ensemble-Based Medical Relation Classification

**Jennifer D'Souza** and **Vincent Ng**
Human Language Technology Research Institute
University of Texas at Dallas

# i2b2

- NIH-funded center for biomedical computing
- Organizes an NLP shared task every year
  - focus: information extraction from clinical data

# 2010 i2b2 Shared Task

- Medical relation (MR) classification
  - For each sentence in a given patient discharge summary, determine the relation type between each pair of medical concepts in it

# 2010 i2b2 Shared Task

- Medical relation (MR) classification
  - For each sentence in a given patient discharge summary, determine the relation type between each pair of medical concepts in it

problems, treatments, tests

# 2010 i2b2 Shared Task

- Medical relation (MR) classification
  - For each sentence in a given patient discharge summary, determine the relation type between each pair of medical concepts in it

problems, treatments, tests

Her pain resolved after surgery.

# 2010 i2b2 Shared Task

- Medical relation (MR) classification
  - For each sentence in a given patient discharge summary, determine the relation type between each pair of medical concepts in it

problems, treatments, tests

Her pain resolved after surgery.

# 2010 i2b2 Shared Task

- Medical relation (MR) classification
  - For each sentence in a given patient discharge summary, determine the relation type between each pair of medical concepts in it
  - 

problems, treatments, tests

Her pain resolved after surgery.

?

# 11 Predefined Relation Types

- 6 types involve a Problem and a Treatment
  - Treatment improves Problem
  - Treatment worsens Problem
  - Treatment causes Problem
  - Treatment administered for Problem
  - Treatment not administered because of Problem
  - No relation between Treatment and Problem

# 11 Predefined Relation Types

- 6 types involve a Problem and a Treatment
  - Treatment improves Problem
  - Treatment worsens Problem
  - Treatment causes Problem
  - Treatment administered for Problem
  - Treatment not administered because of Problem
  - No relation between Treatment and Problem

Her pain resolved after surgery.

?

# 11 Predefined Relation Types

- 6 types involve a Problem and a Treatment
  - Treatment improves Problem
  - Treatment worsens Problem
  - Treatment causes Problem
  - Treatment administered for Problem
  - Treatment not administered because of Problem
  - No relation between Treatment and Problem

Her pain resolved after surgery.

Treatment improves Problem

# 11 Predefined Relation Types

- 3 types involve a Problem and a Test
  - Test reveals Problem
  - Test conducted to investigate Problem
  - No relation between Test and Problem

# 11 Predefined Relation Types

- 3 types involve a Problem and a Test
  - Test reveals Problem
  - Test conducted to investigate Problem
  - No relation between Test and Problem

A postoperative MRI revealed no findings.

# 11 Predefined Relation Types

- 3 types involve a Problem and a Test
  - Test reveals Problem
  - Test conducted to investigate Problem
  - No relation between Test and Problem

A postoperative MRI revealed no findings.

Test reveals Problem

# 11 Predefined Relation Types

- 2 types involve two Problems
  - Problem indicates Problem
  - No relation between two Problems

# 11 Predefined Relation Types

- 2 types involve two Problems
  - Problem indicates Problem
  - No relation between two Problems

… with a fixed inferior defect indicative of scar.

# 11 Predefined Relation Types

- 2 types involve two Problems
  - Problem indicates Problem
  - No relation between two Problems

… with a fixed inferior defect indicative of scar.

Problem indicates Problem

# Assumptions

- Intra-sentential relations only
- Concepts and their types (Problems, Tests, Treatments) are provided as part of the input

# i2b2 Relation Classification Task

- 11-class classification task
  - classify each pair of concepts in the same sentence as belonging to one of the 11 classes

# i2b2 Relation Classification Task

- 11-class classification task
  - classify each pair of concepts in the same sentence as belonging to one of the 11 classes
- Natural approach: supervised learning
  - Each training/test instance involves a pair of concepts
  - Adopted by best system in shared task (Rink et al., 2011)
  - Performance limited by the amount of training data

# How to improve the supervised approach?

# How to improve the supervised approach?

- Distant supervision (Mintz et al., 2009)
  - Automatically create annotated relation instances by extracting their labels from relation instances in a knowledge base (e.g., Freebase, YAGO)

# How to improve the supervised approach?

- Distant supervision (Mintz et al., 2009)
  - Automatically create annotated relation instances by extracting their labels from relation instances in a knowledge base (e.g., Freebase, YAGO)

- But… may not be applicable to MR classification
  - Only one of 11 relation types is represented in the UMLS (most comprehensive medical ontology)

# How to improve the supervised approach?

- ensemble approach
  - uses multiple systems rather than just one system for classification

# How to improve the supervised approach?

- ensemble approach
  - uses multiple systems rather than just one system for classification
- joint inference over pairwise decisions

# How to improve the supervised approach?

- ensemble approach
  - uses multiple systems rather than just one system for classification
- joint inference over pairwise decisions
  - **Motivation**: since each pair of concepts is classified independently of the others, their resulting classifications may not be consistent with each other

# How to improve the supervised approach?

- ensemble approach
  - Identify the members of the ensemble

- joint inference over pairwise decisions
  - Identify global constraints on different concept pairs

# Plan for the Talk

- Corpus
- Baseline system
- Two extensions
  - Ensemble approach
  - Global constraints
- Evaluation

# Plan for the Talk

- Corpus
- Baseline system
- Two extensions
  - Ensemble approach
  - Global constraints
- Evaluation

# Corpus

- 426 discharge summaries
  - 170 (training), 256 (test)

140: classifier training
30: development

# Class Distribution

| | % |
|---|---|
| Treatment improves Problem | 0.6 |
| Treatment worsens Problem | 0.4 |
| Treatment causes Problem | 1.8 |
| Treatment is administered for Problem | 8.9 |
| Treatment is not administered for Problem | 0.6 |
| No relation between Treatment and Problem | 15.2 |
| Test reveals Problem | 10.4 |
| Test conducted to investigate Problem | 1.7 |
| No relation between Test and Problem | 10.1 |
| Problem indicates Problem | 7.5 |
| No relation between two Problems | 42.6 |

# Class Distribution

| | % |
|---|---|
| **Treatment improves Problem** | 0.6 |
| **Treatment worsens Problem** | 0.4 |
| **Treatment causes Problem** | 1.8 |
| **Treatment is administered for Problem** | 8.9 |
| **Treatment is not administered for Problem** | 0.6 |
| **No relation between Treatment and Problem** | 15.2 |
| **Test reveals Problem** | 10.4 |
| **Test conducted to investigate Problem** | 1.7 |
| **No relation between Test and Problem** | 10.1 |
| **Problem indicates Problem** | 7.5 |
| **No relation between two Problems** | 42.6 |

# Class Distribution

| | % |
|---|---|
| Treatment improves Problem | 0.6 |
| Treatment worsens Problem | 0.4 |
| Treatment causes Problem | 1.8 |
| Treatment is administered for Problem | 8.9 |
| Treatment is not administered for Problem | 0.6 |
| No relation between Treatment and Problem | 15.2 |
| Test reveals Problem | 10.4 |
| Test conducted to investigate Problem | 1.7 |
| No relation between Test and Problem | 10.1 |
| Problem indicates Problem | 7.5 |
| No relation between two Problems | 42.6 |

# Class Distribution

| | % |
|---|---|
| Treatment improves Problem | 0.6 |
| Treatment worsens Problem | 0.4 |
| Treatment causes Problem | 1.8 |
| Treatment is administered for Problem | 8.9 |
| Treatment is not administered for Problem | 0.6 |
| No relation between Treatment and Problem | 15.2 |
| Test reveals Problem | 10.4 |
| Test conducted to investigate Problem | 1.7 |
| No relation between Test and Problem | 10.1 |
| Problem indicates Problem | 7.5 |
| No relation between two Problems | 42.6 |

# Class Distribution

| | % |
|---|---|
| Treatment improves Problem | 0.6 |
| Treatment worsens Problem | 0.4 |
| Treatment causes Problem | 1.8 |
| Treatment is administered for Problem | 8.9 |
| Treatment is not administered for Problem | 0.6 |
| No relation between Treatment and Problem | 15.2 |
| Test reveals Problem | 10.4 |
| Test conducted to investigate Problem | 1.7 |
| No relation between Test and Problem | 10.1 |
| Problem indicates Problem | 7.5 |
| No relation between two Problems | 42.6 |

# Plan for the Talk

- Corpus
- Baseline system
- Two extensions
  - Ensemble approach
  - Global constraints
- Evaluation

# Baseline System

- Largely modeled after Rink et al. (2011), the best-performing system in the shared task
  - Employs a multi-class SVM classifier trained on a set of flat features

# Baseline: Training Instance Creation

- One training instance for each concept pair
  - **Problem**: class distribution is skewed
    - “no relation” instances outnumber other types of instances
  - **Solution**: downsample the “no relation” instances
    - Downsampling ratio tuned the on development set

# Baseline: 5 Types of Features

- **Context**
  – Based on POS tags, bigrams, word strings, …
- **Vicinity**
  – Encodes relationship with neighboring concepts
- **Wikipedia**
  – Encodes relationship between the Wiki pages retrieved for the two concepts
- **Single concept**
  – Based on one of the two concepts in the instance
    - word, word lemma, concept type, …

# Baseline: 5 Types of Features

- **Similarity**
  - Features that encode the relation types predicted by its nearest neighbors in the training set
  - How to identify the nearest neighbors of a concept pair from the training set?
    - Rink et al. (2011) proposed 5 methods

# Method for Finding Nearest Neighbors

1. Generalize each concept pair and its context in corpus

Postop, her exam only improved slightly in her hyperreflexia

↓

Postop, test1 only improved slightly in problem2

# Method for Finding Nearest Neighbors

1. Generalize each concept pair and its context in corpus

Postop, her exam only improved slightly in her hyperreflexia

↓

Postop, test1 only improved slightly in problem2

2. For each concept pair involved in a training/test instance, find its 20 nearest neighbors in the training set given the generalized representation, using Levenshtein distance as the similarity metric

# Method for Finding Nearest Neighbors

1. Generalize each concept pair and its context in corpus

   Postop, her exam only improved slightly in her hyperreflexia

   Postop, test1 only improved slightly in problem2

2. For each concept pair involved in a training/test instance, find its 20 nearest neighbors in the training set given the generalized representation, using Levenshtein distance as the similarity metric

3. Encode as Similarity features the predictions made by these 20 nearest neighbors

# Baseline: Test Instance Creation

- Test instances are created from all pairs of concepts appearing in the same sentence

# Plan for the Talk

- Corpus
- Baseline system
- Two extensions
  - Ensemble approach
  - Global constraints
- Evaluation

# Ensemble Approach

- **Issue**: identify the members of the ensemble
- The Baseline system will be one of the members
- What should the other members be?

# Member 2: SVM with Structured Feature

- Commonly used in relation extraction (e.g., Zhou et al. (2007), Zhu et al. (2013))

# Member 2: SVM with Structured Feature

- Commonly used in relation extraction (e.g., Zhou et al. (2007), Zhu et al. (2013))

- **Idea**: rather than using a set of flat features, just use one structured feature (a **parse subtree**) in combination with a tree kernel
  - subtree is used to provide better generalization

# Which parse subtree should we use?

- Given two concepts and the associated parse tree, we use the subtree that covers
  - all the nodes lying on the shortest path between the two concepts; as well as
  - all the immediate children of these nodes unless they are leaves

# Which parse subtree should we use?

- Given two concepts and the associated parse tree, we use the subtree that covers
  - all the nodes lying on the shortest path between the two concepts; as well as
  - all the immediate children of these nodes unless they are leaves

Simple Expansion Tree

# Training and Applying Classifiers

- **Training**
  - Existing SVM learners that can handle structured features can only make binary predictions
    - Need to train 11 binary SVM classifiers
      - each classifier predicts exactly one of the 11 classes
- **Testing**
  - Apply each of the 11 binary classifiers to a test instance
  - Class value is determined by the SVM classifier with the maximum classification confidence

# Member 3: 1-Nearest Neighbor

- In the Baseline, the Similarity features encode the predictions made by the 20 nearest neighbors of a concept pair
- As the 3rd member of the ensemble, we employ a 1-nearest neighbor system
  - use the nearest neighbor in the training set to make predictions

# Member 3: 1-Nearest Neighbor

- Question: why employ 1NN as a separate system in the ensemble if it has been used to generate Similarity features in the Baseline?

# Member 3: 1-Nearest Neighbor

- Question: why employ 1NN as a separate system in the ensemble if it has been used to generate Similarity features in the Baseline?
  - There is no control over whether these Similarity features are deemed useful by the learner and subsequently used by the Baseline classifier

# Member 4: Dependencies

- **Hypothesis**: dependency relations and verb information are useful for inferring the relation type

His pain was well controlled with oral medication.

# Member 4: Dependencies

- **Hypothesis**: dependency relations and verb information are useful for inferring the relation type

His pain was well controlled with oral medication.

nsubjpass prep pobj

- The dependency relations tell us that
  - The PROBLEM (His pain) is being controlled
  - The TREATMENT (oral medication) is doing the controlling

# Member 4: Dependencies

- **Hypothesis**: dependency relations and verb information are useful for inferring the relation type

His pain was well <u>controlled</u> with oral medication.

nsubjpass prep pobj

- The dependency relations tell us that
  - The PROBLEM (His pain) is being controlled
  - The TREATMENT (oral medication) is doing the controlling
- The verb control allows us to infer that
  - Relation type should be TREATMENT improves PROBLEM

# Member 4: Dependencies

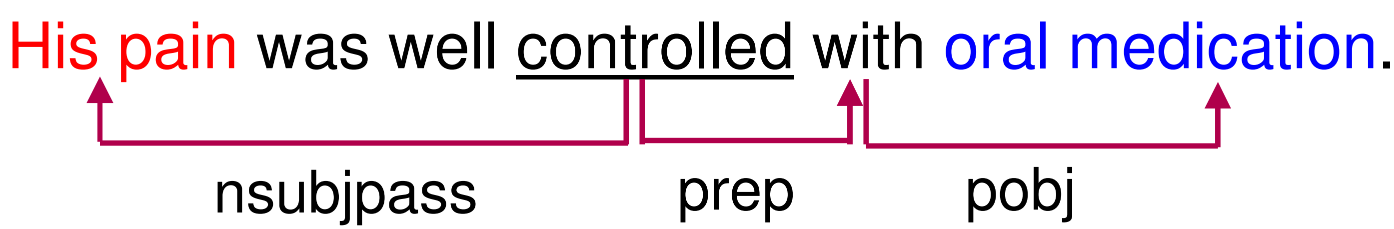
- We make use of dependency relations and verbs by using them to encode a concept pair

# Member 4: Dependencies

- We make use of dependency relations and verbs by using them to encode a concept pair
  - Represent each concept pair by two dependency paths
    - Path 1: between the first concept and its closest verb
    - Path 2: between the second concept and its closest verb

# Member 4: Dependencies

- We make use of dependency relations and verbs by using them to encode a concept pair
  - Represent each concept pair by two dependency paths
    - Path 1: between the first concept and its closest verb
    - Path 2: between the second concept and its closest verb




- Path 1: controlled → nsubjpass → NN
- Path 2: controlled → prep → with → NN

# Member 4: Dependencies

- Classify a test instance given this path representation using 1-nearest neighbor
  - Given two instances a and b (each represented by two dependency paths),

    Similarity(a,b)

    = CosineSimilarity(<path_1a,path_1b>) *
    CosineSimilarity(<path_2a,path_2b>)

# Member 5: Hand-Crafted Rules

- The 5th member relies on a set of hand-crafted rules to make predictions
  - Two humans were asked to
    - identify based on the training data the contexts that are strongly indicative of a relation type
    - design classification rules, each of which is composed of two concepts and their surrounding context
      - due to PROBLEM by TREATMENT
        → TREATMENT causes PROBLEM

# Member 5: Hand-Crafted Rules

- 136 rules were identified
  - ordered in decreasing order of accuracy on the training set

- A test instance is classified using the first applicable rule

# How to classify a test instance using this 5-member ensemble?

- Majority voting
  - Presumes each member is equally important
  - In practice, some members are more important and should be given a higher weight

# How to classify a test instance using this ensemble?

- So.. we do **weighted voting** instead
  - Combine the probabilistic votes of the members in a weighted fashion:

$$P_{combined}(c) = w_1 \cdot P_{tree}(c) + w_2 \cdot P_{flat}(c) + w_3 \cdot P_{dependency}(c) \\ + w_4 \cdot P_{word}(c) + w_5 \cdot P_{rules}(c)$$

# How to classify a test instance using this ensemble?

- So.. we do **weighted voting** instead
  - Combine the probabilistic votes of the members in a weighted fashion:

$$P_{combined}(c) = w_1 \cdot P_{tree}(c) + w_2 \cdot P_{flat}(c) + w_3 \cdot P_{dependency}(c) + w_4 \cdot P_{word}(c) + w_5 \cdot P_{rules}(c)$$

  - The 5 weights are tuned on the development set
  - Details on how to convert the members' predictions into probabilities are omitted

# Plan for the Talk

- Corpus
- Baseline system
- Two extensions
  - Ensemble approach
  - Global constraints
- Evaluation

# Employing Global Constraints

- **Motivation**: pairs are classified independently, resulting classifications may not be globally consistent
- How to identify constraints on the relation types of different concept pairs?

# Employing Global Constraints

- **Motivation**: pairs are classified independently, resulting classifications may not be globally consistent

- How to identify constraints on the relation types of different concept pairs?
  - By inspection on the training data

# Inter-Pair Constraints: Example

**Treatment i improves Problem j**

**→Treatment i cannot worsen Problem k**

# Inter-Pair Constraints: Example

**Treatment i improves Problem j**

**→Treatment i cannot worsen Problem k**

This is a constraint on the predicted values of two test instances, one involving i,j and the other i,k

# Inter-Pair Constraints: Example

**Treatment i improves Problem j**

**→Treatment i cannot worsen Problem k**

Doesn't seem intuitive, but can be attributed to the way discharge summaries are written

→ Constraint can be violated across sentences, but always holds within a sentence

# Inter-Pair Constraints: Example

**Treatment i improves Problem j**

**→Treatment i cannot worsen Problem k**

Doesn't seem intuitive, but can be attributed to the way discharge summaries are written

→ Constraint can be violated across sentences, but always holds within a sentence

After identifying a set of inter-pair constraints, how can we enforce them?

# Enforcing Constraints

- Use Integer Linear Programming
  - One ILP program per sentence
  - One binary indicator variable for each possible combination of concept pairs and relation types
    - Let the ILP solver re-assign relation types to concept pairs in the sentence so that no constraints are violated
    - The objective function is constructed based on the probabilities returned by members of the ensemble

# Plan for the Talk

- Corpus
- Baseline system
- Two extensions
  - Ensemble approach
  - Global constraints
- Evaluation

# Evaluation

- **Goal**: evaluate the effectiveness of our two extensions in improving the baseline

# Experimental Setup

- Following the i2b2 evaluation scheme, assume
  - concepts and their types are given
  - system evaluated on all but the “no relation” types
    - no reward for correctly classifying “no relation” instances
    - will be penalized for misclassifying “no relation” instances

# Results: Ensemble Members

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | 64.3 | 55.6 | 59.6 |
| 3 | 1 nearest neighbor | 63.9 | 59.2 | 61.4 |
| 4 | Dependencies | 4.3 | 82.9 | 8.2 |
| 5 | Rules | 11.9 | 84.4 | 9.1 |

- Flat outperforms Tree significantly

# Results: Ensemble Members

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | 64.3 | 55.6 | 59.6 |
| 3 | 1 nearest neighbor | 63.9 | 59.2 | 61.4 |
| 4 | Dependencies | 4.3 | 82.9 | 8.2 |
| 5 | Rules | 11.9 | 84.4 | 9.1 |

- Flat outperforms Tree significantly

# Results: Ensemble Members

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | 64.3 | 55.6 | 59.6 |
| 3 | 1 nearest neighbor | 63.9 | 59.2 | 61.4 |
| 4 | Dependencies | 4.3 | 82.9 | 8.2 |
| 5 | Rules | 11.9 | 84.4 | 9.1 |

- Flat outperforms Tree significantly

# Results: Ensemble Members

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | 64.3 | 55.6 | 59.6 |
| 3 | 1 nearest neighbor | 63.9 | 59.2 | 61.4 |
| 4 | Dependencies | 4.3 | 82.9 | 8.2 |
| 5 | Rules | 11.9 | 84.4 | 9.1 |

- 1-nearest neighbor outperforms Tree significantly but underperforms Flat significantly

# Results: Ensemble Members

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | 64.3 | 55.6 | 59.6 |
| 3 | 1 nearest neighbor | 63.9 | 59.2 | 61.4 |
| 4 | Dependencies | 4.3 | 82.9 | 8.2 |
| 5 | Rules | 11.9 | 84.4 | 9.1 |

- 1-nearest neighbor outperforms Tree significantly but underperforms Flat significantly

# Results: Ensemble Members

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | 64.3 | 55.6 | 59.6 |
| 3 | 1 nearest neighbor | 63.9 | 59.2 | 61.4 |
| 4 | Dependencies | 4.3 | 82.9 | 8.2 |
| 5 | Rules | 11.9 | 84.4 | 9.1 |

- Dependencies and Rules: high precision, low recall

# Results: Ensemble Method & ILP

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | 64.3 | 55.6 | 59.6 |
| 3 | 1 nearest neighbor | 63.9 | 59.2 | 61.4 |
| 4 | Dependencies | 4.3 | 82.9 | 8.2 |
| 5 | Rules | 11.9 | 84.4 | 9.1 |
| | Ensemble(1+2) | 69.2 | 61.3 | 65.0 |
| | Ensemble(1+2+3) | 70.4 | 63.1 | 66.6 |
| | Ensemble(1+2+3+4) | 70.0 | 64.7 | 67.2 |
| | Ensemble(1+2+3+4+5) | 71.1 | 64.8 | 67.8 |
| | Ensemble(1+2+3+4+5) + ILP | 72.9 | 66.7 | 69.4 |

# Results: Ensemble Method & ILP

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | 64.3 | 55.6 | 59.6 |
| 3 | 1 nearest neighbor | 63.9 | 59.2 | 61.4 |
| 4 | Dependencies | 4.3 | 82.9 | 8.2 |
| 5 | Rules | 11.9 | 84.4 | 9.1 |

| | | | |
|---|---|---|---|
| Ensemble(1+2) | 69.2 | 61.3 | 65.0 |
| Ensemble(1+2+3) | 70.4 | 63.1 | 66.6 |
| Ensemble(1+2+3+4) | 70.0 | 64.7 | 67.2 |
| Ensemble(1+2+3+4+5) | 71.1 | 64.8 | 67.8 |
| Ensemble(1+2+3+4+5) + ILP | 72.9 | 66.7 | 69.4 |

# Results: Ensemble Method & ILP

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | 64.3 | 55.6 | 59.6 |
| 3 | 1 nearest neighbor | 63.9 | 59.2 | 61.4 |
| 4 | Dependencies | 4.3 | 82.9 | 8.2 |
| 5 | Rules | 11.9 | 84.4 | 9.1 |

| | | | |
|---|---|---|---|
| Ensemble(1+2) | 69.2 | 61.3 | 65.0 |
| Ensemble(1+2+3) | 70.4 | 63.1 | 66.6 |
| Ensemble(1+2+3+4) | 70.0 | 64.7 | 67.2 |
| Ensemble(1+2+3+4+5) | 71.1 | 64.8 | 67.8 |
| Ensemble(1+2+3+4+5) + ILP | 72.9 | 66.7 | 69.4 |

# Results: Ensemble Method & ILP

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | 64.3 | 55.6 | 59.6 |
| 3 | 1 nearest neighbor | 63.9 | 59.2 | 61.4 |
| 4 | Dependencies | 4.3 | 82.9 | 8.2 |
| 5 | Rules | 11.9 | 84.4 | 9.1 |
| | Ensemble(1+2) | 69.2 | 61.3 | 65.0 |
| | Ensemble(1+2+3) | 70.4 | 63.1 | 66.6 |
| | Ensemble(1+2+3+4) | 70.0 | 64.7 | 67.2 |
| | Ensemble(1+2+3+4+5) | 71.1 | 64.8 | 67.8 |
| | Ensemble(1+2+3+4+5) + ILP | 72.9 | 66.7 | 69.4 |

# Results: Ensemble Method & ILP

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | 64.3 | 55.6 | 59.6 |
| 3 | 1 nearest neighbor | 63.9 | 59.2 | 61.4 |
| 4 | Dependencies | 4.3 | 82.9 | 8.2 |
| 5 | Rules | 11.9 | 84.4 | 9.1 |
| | Ensemble(1+2) | 69.2 | 61.3 | 65.0 |
| | Ensemble(1+2+3) | 70.4 | 63.1 | 66.6 |
| | Ensemble(1+2+3+4) | 70.0 | 64.7 | 67.2 |
| | Ensemble(1+2+3+4+5) | 71.1 | 64.8 | 67.8 |
| | Ensemble(1+2+3+4+5) + ILP | 72.9 | 66.7 | 69.4 |

# Results: Ensemble Method & ILP

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | 64.3 | 55.6 | 59.6 |
| 3 | 1 nearest neighbor | 63.9 | 59.2 | 61.4 |
| 4 | Dependencies | 4.3 | 82.9 | 8.2 |
| 5 | Rules | 11.9 | 84.4 | 9.1 |

| | | | |
|---|---|---|---|
| Ensemble(1+2) | 69.2 | 61.3 | 65.0 |
| Ensemble(1+2+3) | 70.4 | 63.1 | 66.6 |
| Ensemble(1+2+3+4) | 70.0 | 64.7 | 67.2 |
| Ensemble(1+2+3+4+5) | 71.1 | 64.8 | 67.8 |
| Ensemble(1+2+3+4+5) + ILP | 72.9 | 66.7 | 69.4 |

# Results: Ensemble Method & ILP

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | 64.3 | 55.6 | 59.6 |
| 3 | 1 nearest neighbor | 63.9 | 59.2 | 61.4 |
| 4 | Dependencies | 4.3 | 82.9 | 8.2 |
| 5 | Rules | 11.9 | 84.4 | 9.1 |
| | Ensemble(1+2) | 69.2 | 61.3 | 65.0 |
| | Ensemble(1+2+3) | 70.4 | 63.1 | 66.6 |
| | Ensemble(1+2+3+4) | 70.0 | 64.7 | 67.2 |
| | Ensemble(1+2+3+4+5) | 71.1 | 64.8 | 67.8 |
| | Ensemble(1+2+3+4+5) + ILP | 72.9 | 66.7 | 69.4 |

# Results: Ensemble Method & ILP

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | 64.3 | 55.6 | 59.6 |
| 3 | 1 nearest neighbor | 63.9 | 59.2 | 61.4 |
| 4 | Dependencies | 4.3 | 82.9 | 8.2 |
| 5 | Rules | 11.9 | 84.4 | 9.1 |

| | | | |
|---|---|---|---|
| Ensemble(1+2) | 69.2 | 61.3 | 65.0 |
| Ensemble(1+2+3) | 70.4 | 63.1 | 66.6 |
| Ensemble(1+2+3+4) | 70.0 | 64.7 | 67.2 |
| Ensemble(1+2+3+4+5) | 71.1 | 64.8 | 67.8 |
| Ensemble(1+2+3+4+5) + ILP | 72.9 | 66.7 | 69.4 |

# Results: Ensemble Method & ILP

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | | | |
| 2 | Tree | | | |
| 3 | 1 nearest neighbor | | | |
| 4 | Dependencies | | | |
| 5 | Rules | | | |

Except for the addition of the dependency-based system and the hand-crafted rules, the addition of all other components yielded significant improvements

| | | | |
|---|---|---|---|
| Ensemble(1+2) | 69.2 | 61.3 | 65.0 |
| Ensemble(1+2+3) | 70.4 | 63.1 | 66.6 |
| Ensemble(1+2+3+4) | 70.0 | 64.7 | 67.2 |
| Ensemble(1+2+3+4+5) | 71.1 | 64.8 | 67.8 |
| Ensemble(1+2+3+4+5) + ILP | 72.9 | 66.7 | 69.4 |

# Results: Ensemble Method & ILP

| | System | R | P | F |
|---|---|---|---|---|
| 1 | Flat (Baseline) | 66.7 | 58.1 | 62.1 |
| 2 | Tree | | | |
| 3 | 1 nearest neighbor | | | |
| 4 | Dependencies | | | |
| 5 | Rules | | | |
| | Ensemble(1+2) | | | |
| | Ensemble(1+2+3) | 70.4 | 63.1 | 66.6 |
| | Ensemble(1+2+3+4) | 70.0 | 64.7 | 67.2 |
| | Ensemble(1+2+3+4+5) | 71.1 | 64.8 | 67.8 |
| | Ensemble(1+2+3+4+5) + ILP | 72.9 | 66.7 | 69.4 |

Best-performing system is the one that comprises all the components

Yielded a relative error reduction of 19.9% over the Baseline

# Summary

- Improved a state-of-the-art supervised medical relation classification system by proposing two extensions
  - Ensemble
  - Enforcing global constraints in an ILP framework

- Resulting system yielded a relative error reduction of 19.9% over the Baseline