Supplementary Material

Complete Predicate definitions

In this section, we summarize the predicates defined for the co-reference resolution task in Table 1 and rules excluded from the main manuscript because of the word limitation.

Table 1. Predicate definitions.

Predicate	Description
$backwardNonPersonConcept(i, j, d)$	The distance between the concept i and the
$backwardPersonConcept(i, j, d)$	The distance between the concept i and the person
$conceptCluster(i, c)$	The concept cluster of i is c.
$containPCP(i)$	The concept i contains words like "primary care
$containPatient(i)$	The person concept i contains the word "patient".
$coreference(i, j)$	The two concepts i and j are the same instance.
$exactMatch(i, j)$	The string of the ith concept matches with the jth concept.
$stemMatch(i, j)$	The stem of the ith concept matches with the stem of the jth concept.
$gender(g)$	The discharge summary contains the gender field g, e.g. "sex: M".
$hasBodyPartModifier(l, w)$	The line l mentions parts of the human body w, such as "shoulder" and "liver".
$hasDate(l)$	The line l contains date descriptions like 2013-03-26.
$hasDay(l)$	The line l contains a description referring to a day of the week, e.g. Monday.
$hasDosage(l, d)$	The line l contains a description d related to the amount of medicine that the patient needs to take at one time.
$hasFollowingWord(i, w)$	The ith concept has the following word w.
$hasPrecedingWord(i, w)$	The ith concept has the preceding word w.
$hasQuantity(i, q)$	The concept i has a quantity description q, such as "10 mg".
$hasSpatialModifier(l, w)$	The line l contains a spatial modifier w, such as "lower" and "left".

$jaroWinkerDistance(i, j, s)$	The Jaro-Winker distance between the concept i and j is s.
$line(i, l)$	The concept i is on the line l in the discharge summary.
$mostFrequentPersonalPronoun(i)$	The concept i is the personal pronoun that appears the most in the given discharge summary.
$objective(s_1, s_2)$	The string s_2 (e.g. him) is the objective case of the string s_1 (e.g. he).
$overlap(i, j)$	The two concepts i and j are overlapped.
$personalPronoun(i)$	The concept i is a personal pronoun, such as "he", "she" and "you".
$personalRelativePronoun(i)$	The concept i is a personal relative pronoun, such as who, whom, and whose.
$possessivePronoun(s_1, s_2)$	The string s_1 (e.g. his) is the possessive pronoun of the string s_2 (e.g. he).
$section(i, s)$	The concept i is under the sub-section s.
$string(i, s)$	The string of the ith concept is s.
$wordPosition(i, s, e)$	The first and the last word of the concept i is the sth and the eth words of the text, respectively.
$wordNetSimilarity(i, j, s)$	The WordNet similarity score between the concept i and j is s.

Stage 2: Filtering

The filtering stage ignores the person concept, because in comparison to the other three concept clusters (treatment, problem and test), contents of person concepts remain consistent throughout the text regardless of contextual information. Take Figure 1 in the main manuscript as an example. The concept "Louie A. Keith , M.D." in the text refers to a specific person, and regardless of the context this concept will always indicate the same entity. However, according to the annotation guideline[1], treatments are paired only if they have the same episode and dosage. Therefore, the two treatment concepts "Clozapine" with two different dosages of 300 mg and 25 mg in the Figure 1 are not considered as the same treatment and should be filtered.

$$hasDosage(l_1, d) \wedge line(i, l_1) \wedge \neg hasDosage(l_2, d) \wedge line(j, l_2) \wedge conceptCluster(i, \text{"treatment"}) \wedge conceptCluster(j, \text{"treatment"}) \wedge exactMatch(i, j) \Longrightarrow \neg coreference(i, j)$$

The rule removes the linkage of two treatment concepts if they do not have equivalent dosages.

$$hasSpatialModifier(l_1, w_1) \wedge conceptCluster(i, +ct) \wedge ct \neq \text{"person"} \wedge ct \neq \text{"pronoun"} \wedge line(i, l_1) \wedge line(j, l_2) \wedge hasSpatialModifier(l_2, w_2) \wedge conceptCluster(j, +ct) \wedge exactMatch(i, j) \Longrightarrow \neg coreference(i, j)$$

$$hasBodyPartModifier(l_1, w_1) \wedge conceptCluster(i, +ct) \wedge ct \neq \text{"person"} \wedge ct \neq \text{"pronoun"} \wedge line(i, l_1) \wedge line(j, l_2) \wedge hasBodyPartModifier(l_2, w_2) \wedge conceptCluster(j, +ct) \wedge exactMatch(i, j) \Longrightarrow \neg coreference(i, j)$$

The above rules remove linkages between non-person/-pronoun concepts if they have different modifiers. For example, the problem “pain in the right arm” and “lower leg pain” are not a co-reference.

Ensemble Results

Several works have demonstrated that an “ensemble system,” which combines several systems’ outputs, generally outperforms even the best single system[2]. We ran an experiment to create an ensemble system made up of the best test results generated by the three systems, MR-4, MLN-4 and MR-1+Model$_{1,2,3}$.. The ensemble chains were produced by combining the output of the three systems into a single ranked list. Intuitively, an ensemble system simply asks each system which chain the given concept appears in. However, there is a potential problem. Consider the following two co-reference chains generated by three systems:

System_1: {*a*, *b*, *c*} and {*d*, *e*, *f*};

System_2: {*a*, *b*, *c*, *d*} and {*e*, *f*};

System_3: {*a*, *b*, *d*, *f*} and {*a*, *e*}.

Given the concept *f*, System_1 returns {*d*, *e*, *f*} as *f*'s chain, System_2 is {*e*, *f*} and the chain of System_3 is {*a*, *b*, *d*, *f*}. For an ensemble system, it is difficult to select the best chain that *f* belongs to. Rather than comparing similar components between several possible chains and selecting the best one, we decompose each chain into co-reference pairs. Continuing our example, the co-reference pairs of System_1 are {*a*, *b*}, {*a*, *c*}, {*b*, *c*}, {*d*, *e*}, ***{d, f}*** and ***{e, f}***; System_2 includes {*a*, *b*}, {*a*, *c*}, {*a*, *d*}, {*b*, *c*}, {*b*, *d*}, {*c*, *d*} and ***{e, f}***; System_3 is {*a*, *b*}, {*a*, *d*}, ***{a, f}***, {*b*, *d*}, ***{b, f}***, ***{d, f}*** and {*a*, *e*}. To decide which chain *f* belongs to, we simply ask which one pairs up with *f* as a co-reference. In this case, both {*e*, *f*} and {*d*, *f*} appear twice in three systems, so we can merge them into {*d*, *e*, *f*} as the best chain.

Reference

1. Uzuner O, Forbush T, Shen S, Savova G, Chapman W, Clark C, et al. 2011 i2b2/VA Co-reference Annotation Guidelines for the Clinical Domain, 2011.
2. Dietterich TG. Ensemble Methods in Machine Learning. *Proceedings of the First International Workshop on Multiple Classifier Systems*: Springer-Verlag, 2000.