Построение сочетаемостных ограничений на основе байесовских сетей для разрешения многозначности.

Сочетаемостные ограничения – это закономерности использования глагола относительно семантического класса его параметров (субъект, объект (прямое дополнение) и

косвенное дополнение. Модели автоматического построения сочетаемостных ограничений важны сами по себе и имеют приложения в обработке естественного языка. Сочетаемостные ограничения глагола могут применяться для получения возможных значений неизвестного параметра при известных глаголах. При построении предложения сочетаемостные ограничения позволяют отранжировать варианты и выбрать лучший среди них. Исследование сочетаемостных ограничений могло бы помочь в понимании структуры ментального лексикона. Системы обучения сочетаемостных ограничений без учителя обычно комбинируют статистические подходы и подходы, основанные на знаниях. Компонент базы знаний – это обычно база данных, в которой слова сгруппированы в классы.

Статистический компонент состоит из пар предикат-аргумент, извлеченных из неразмеченного корпуса. В тривиальном алгоритме можно было бы получить список слов (прямых дополнений глагола), и для тех слов, которые есть в WordNet, вывести их семантические классы. Семантическим классом называется синсет тезауруса WordNet, т. е.

класс соответствует одному из значений слова. Таким образом, в тривиальном алгоритме на основе данных WordNet можно выбрать классы (значения слов), с которыми употребляются (встречаются в корпусе) глаголы.

Байесовские сети, или байесовские сети доверия (БСД), состоят из множества переменных (вершин) и множества ориентированных ребер, соединяющих эти переменные.

Такой сети соответствует ориентированный ациклический граф. Каждая переменная может принимать одно из конечного числа взаимоисключающих состояний. Пусть все переменные будут бинарного типа, т. е. принимают одно из двух значений: истина или ложь. Любой переменной А с родителями 𝐵1, ...,𝑛 соответствует таблица условных вероятностей. Иерархия существительных в WordNet представлена в виде ориентированного ациклического графа. Синсет узла принимает значение “истина”, если глагол “выбирает” существительное из набора синонимов. Априорные вероятности задаются на основе двух предположений: во-первых, маловероятно, что глагол будет употребляться только со словами какого-то конкретного синсета, и вовторых, если глагол действительно употребляется только со словами из данного синсета

(например, синсет ЕДА), тогда должно быть правомерным употребление этого глагола с гипонимами этого синсета (например, ФРУКТ).

Те же предположения, что для синсетов, верны и для употреблений слов с глаголами:

1. слово, вероятно, является аргументом глагола в том случае, если глагол употребляется с каким-либо из значений этого слова;
2. отсутствие связки глагол-синсет говорит о малой вероятности того, что слова этого синсета употребляются с глаголом.

Словам “вероятно” и “маловероятно” должны быть приписаны такие числа, сумма которых равна единице.