

GWV – Grundlagen der Wissensverarbeitung

Aufgabenzettel 7 : Version-space learning

Abgabe 14./15.12.2008 Besprechung am 15./16.12.2008.

Übungsaufgabe 7.1 : (Ontologien II (WBS))

In dieser Aufgabe sollen Konzepte aus dem Bereich von Studien- und Prüfungsordnungen gesammelt und die semantischen Bezüge zwischen ihnen spezifiziert werden.

Neben den Konzepten und ihren Relationen (Terminologie) können auch Beispielobjekte spezifiziert werden (Assertionen, Fakten).

Folgende Konzepte sollten in der Modellierung enthalten sein:

Lehrveranstaltung, Note, PrüferIn, Prüfung, Prüfungsprotokoll, Semester, Sommersemester, StudentIn, Studienfach, Studienordnung, Wintersemester.

(Sie dürfen die Liste um eigene Konzepte und Relationen ergänzen.)

Geben Sie Ihre Spezifikationen in natürlicher Sprache an und formalisieren Sie sie durch prädikatenlogischen Formeln.

VON
5

Übungsaufgabe 7.2 : (CEA Segelbeispiel (WM))

Tabelle 1 enthält einen Trainingsatz bestehend aus vier Trainingsbeispielen (Zeileneinträge), die zum Erlernen des Konzepts „Tage, an denen Fred gerne segelt“ benutzt werden sollen. Die Attribute Himmel, Lufttemperatur, Feuchtigkeit, Windstärke, Wassertemperatur und Voraussage werden mit $A_1 - A_6$ bezeichnet. Die Wertebereiche der Attribute seien: $V(\text{Himmel}) = \{\text{sonnig, regnerisch, Schnee}\}$, $V(\text{Lufttemp.}) = \{\text{warm, kalt}\}$, $V(\text{Feuchtigk.}) = \{\text{normal, hoch}\}$, $V(\text{Windstärke}) = \{\text{stark, schwach}\}$, $V(\text{Wassertemp.}) = \{\text{warm, kalt}\}$, $V(\text{Voraussage}) = \{\text{gleichbleibend, Wechsel}\}$

Eine Hypothese wird als 6-Tupel $h = (a_1, \dots, a_6)$ dargestellt. Dabei ist jedes a_i , $1 \leq i \leq 6$, ein Element aus $\{?, \emptyset\} \cup V(A_i)$. Ist $a_i = ?$, bedeutet das, dass unter der Hypothese h keine Beschränkung für die Werte des entsprechenden Attributs A_i existiert; ist $a_i = \emptyset$ bedeutet das, dass unter der Hypothese h kein Wert gestattet ist für A_i ; ist $a_i = v \in V(A_i)$ bedeutet das, dass unter der Hypothese h genau der Wert v für das Attribut gestattet ist. (Beispiel: „sonnig“ für Attribut A_1 (Himmel)).

1. Aus wie vielen syntaktisch verschiedenen Hypothesen besteht der Hypothesenraum H , wenn die obige Sprache zur Repräsentation von Hypothesen vorausgesetzt wird?
Aus wie vielen semantisch verschiedenen Hypothesen besteht H ? (3 Pkt.)
2. Wenden Sie den Candidate-Elimination-Algorithmus (CEA) von Listing 2 auf die Daten von Tabelle 1 (in der Reihenfolge der Zeileneinträge von oben nach unten) an, um die Menge aller konsistenten Konzepthypothesen zu bestimmen. Geben Sie für jeden Schritt dabei die „Specific Boundary“ SB und die „General Boundary“ GB an. (3 Pkt.)

VON
11

3. Es werde wieder der Trainingssatz von Tabelle 1 vorausgesetzt. Der Hypothesenraum H' ist jedoch ein neuer: Er bestehe aus allen Disjunktionen $h_1 \vee h_2$ von Hypothesen h_1, h_2 aus H . Bestimmen Sie wieder mit CEA die Menge aller konsistenten Konzept-hypothesen. (5 Pkt.)

Listing 2: CEA-Algorithmus

```

1 Initialize SB to the empty concept and SG to the most general
2 concept
3 For each training example e do:
4   If e is a positive example then:
5     Remove from GB any hypothesis inconsistent with e
6     For each hypothesis h in SB that is not consistent with e do:
7       Remove h from SB
8       Add to SB all minimal generalizations h' of h such that
9         h' is consistent with e and
10        some member of GB is more general than h'
11       Remove from SB any hypothesis that is more general
12       than another hypothesis in SB
13   else if e is a negative example:
14     Remove from SB any hypothesis consistent with e
15     For each hypothesis h in GB that is consistent with e do:
16       Remove h from GB
17       Add to GB all minimal specializations h' of h such that
18         h' is inconsistent with e and
19         some member of SB is more specific than h' .
20       Remove from GB any hypothesis that is less general
21       than another hypothesis in GB

```

Version: 8. Dezember 2008

Summe der erreichbaren Punkte auf diesem Blatt: 16

Summe der bisher erreichbaren Punkte: 95

Himmel	Lufttemp.	Feuchtig.	Windstärke	Wassertemp.	Voraussage	Segeln?
sonnig	warm	normal	stark	warm	gleichbleibend	ja
sonnig	warm	hoch	stark	warm	gleichbleibend	ja
regnerisch	kalt	hoch	stark	warm	Wechsel	nein
sonning	warm	hoch	stark	kalt	Wechsel	ja

Tabelle 1: Trainingssatz Segeln