

# 全词消歧的序列标注方法\*

周云, 王挺, 易绵竹, 王之王

国防科学技术大学 计算机学院, 湖南 长沙 410073

E-mail: zy\_vxd@nudt.edu.cn

**摘要:** 全词消歧(All-Words Word Sense Disambiguation)本质上是一个序列标注问题, 本文提出了两种用于全词消歧的序列标注方法, 它们分别基于隐马尔可夫模型(Hidden Markov Model, HMM)和最大熵马尔可夫模型(Maximum Entropy Markov Model, MEMM)模型。首先, 我们用 HMM 对全词消歧进行建模。然后, 针对 HMM 只能利用观察值的缺点, 我们将上述 HMM 模型推广为 MEMM 模型, 将大量上下文特征集成到模型中。对于全词消歧这类超大状态问题, 在 HMM 和 MEMM 模型中均存在数据稀疏和时间复杂度过高的问题, 我们通过柱状搜索 Viterbi 算法和平滑策略来解决。最后, 我们在 Senseval-2 和 Senseval-3 的数据集上进行了评测, 本文提出的 MEMM 方法的性能超过该评测上所有的序列标注方法。

**关键词:** 全词消歧; 序列标注方法; 隐马尔可夫模型; 最大熵马尔可夫模型; 超大状态问题

## Sequence Labeling Methods for All-words Word Sense Disambiguation

Zhou Yun, Wang Ting, Yi Mianzhu, Wang Zhiyuan

Computer School, National University of Defense Technology, Changsha 410073

E-mail: zy\_vxd@nudt.edu.cn

**Abstract:** All-Words Word Sense Disambiguation (WSD) is intrinsically a sequence labeling problem, and two sequence labeling methods for All-Words WSD are proposed in this paper, which are based on Hidden Markov Model (HMM) and Maximum Entropy Markov Model (MEMM), respectively. First, we model All-Words WSD using HMM. Then HMM is generalized to MEMM by integrating a large number of non-independent features. For All-Words WSD, a typical very large state problem, data sparsity and high time complexity seriously hinder the application of HMM and MEMM models and we solve these problems by beam-search Viterbi algorithm and smoothing strategy. Finally, we test our methods on the dataset of All-Words WSD tasks in Senseval-2 and Senseval-3, and the MEMM method proposed in this paper outperforms all sequence labeling methods ever occurred in above tasks.

**Keywords:** all-words word sense disambiguation; sequence labeling method; hidden Markov model; maximum entropy Markov model; very large state problem

### 1 引言

词义消歧, 即在特定的上下文中确定歧义词的词义。根据词义消歧的范围, 可将其分为词样消歧(Lexical-Sample WSD)和全词消歧(All-Words WSD)。词样消歧对给定文本中的某些指定词进行消歧, 而全词消歧对给定文本中的所有开放词(包括名词、动词、形容词和副词)进行消歧。词样消歧是一个典型的分类问题, 可使用各种成熟的有监督分类算法, 如朴素贝叶斯[1]、最大熵算法[2]和支持向量机[3]等。对于全词消歧, 目前通常的做法是将其当作词样消歧, 对句中出现的每个开放词逐个进行消歧, 各个词之间的消歧是独立的。但是, 全词消歧中前后两个词的消歧实际上是相互关联的, 全词消歧本质上是一个序列标注问题。

序列标注指的是对观察值序列的每个成员指定一个类别标签, 因此序列标注可视一系列的分类任务。由于序列标注利用相邻元素的依赖性对整个序列进行全局优化, 一次性为所有观察值给

\* 本文承国家自然科学基金(批准号:60873097)和新世纪优秀人才支持计划(编号:NCET-06-0926)的资助。

出标签,因而标注性能通常会得到提升。常用的序列标注算法有隐马尔可夫模型(Hidden Markov Model, HMM)[4]、最大熵马尔可夫模型(Maximum Entropy Markov Model, MEMM)[5]和条件随机域[6]等。序列标注在一些自然语言处理任务中取得了很好的效果,如词性标注[7]和语音识别[8]等。虽然 HMM 等序列标注方法在词性标注等领域取得了突出的成绩,但在词义消歧中的结果似乎不太理想[9][10][11][12][13]。这可能是由于:第一, HMM 的观察值只能是词形,难以有效利用各类语言学特征;第二,全词消歧是一个超大状态问题,序列标注方法存在严重的数据稀疏问题,并且具有很高的时间复杂度。

针对上述问题,本文提出两种基于序列标注的全词消歧方法,主要贡献如下:(1)提出了一种基于 HMM 的全词消歧方法;(2)将基于 HMM 的方法推广为基于 MEMM 的全词消歧方法,将观察值由词形扩展为特征向量,引入了大量的语言学特征;(3)通过柱状搜索和平滑策略解决上述模型中存在的数据库稀疏问题,在解决数据库稀疏问题的同时,柱状搜索还显著降低了解码的时间复杂度;(4)在 Senseval 数据集上对序列标注方法进行了系统的评测。

本文结构安排如下。第 1 节,引言。第 2 节,我们给出一种新的基于 HMM 的全词消歧方法。第 3 节,我们将模型推广为 MEMM 模型,将观察值扩展为若干特征构成的向量,这些特征包括邻近词性、局部搭配关系、依存句法关系、WordNet 上位词链、WordNet 语义标签、WordNet 动词框架和词袋等。针对序列标注方法存在的数据库稀疏和很高的时间复杂度等问题,在第 4 节我们设计了平滑策略和柱状搜索 Viterbi 算法加以解决。第 5 节,我们在 Senseval-2 和 Senseval-3 上对 HMM 和 MEMM 方法进行了评测,并与以往的序列标注方法的性能进行了对比。第 6 节,我们总结全文并讨论下一步工作。

## 2 基于 HMM 的全词消歧方法

HMM 是一个经典的数学模型[4],它一般用以解决评估、序列标注(也称为解码)和学习等 3 个问题,这 3 个问题分别可用前向算法(或后向算法)、Viterbi 算法和前向-后向算法解决。在本文中我们主要关心序列标注问题和 Viterbi 算法。

HMM 包含两个随机过程  $Q_t, O_t, 1 \leq t \leq T$ , 和五个模型参数  $S, V, A, B, \pi$  (除特殊说明外,本文均沿用[4]中的符号)。其中, (1)  $S = \{s_1, s_2, \dots, s_N\}$  为状态集合; (2)  $V = \{v_1, v_2, \dots, v_M\}$  为观察值集合; (3)  $A = \{a_{ij}\}, a_{ij} = P(Q_{t+1} = s_j | Q_t = s_i), 1 \leq i, j \leq N$  为状态转移概率矩阵; (4)  $B = \{b_j(k)\}, b_j(k) = P(O_t = v_k | Q_t = s_j), 1 \leq j \leq N, 1 \leq k \leq M$  为状态-观察值发射概率矩阵; (5)  $\pi = \{\pi_i\}, \pi_i = P(Q_1 = s_i), 1 \leq i \leq N$  为状态初始分布概率向量。

令  $\lambda = (\pi, A, B)$ 。在 IIMM 中,序列标注问题为,给定模型参数  $\lambda$  和观察值序列  $o = o_1 o_2 \dots o_T, o_t \in V$ , 找出状态序列  $\hat{q} = \hat{q}_1 \hat{q}_2 \dots \hat{q}_T, \hat{q}_t \in S$ , 使得

$$\hat{q} = \arg \max_q P(Q = q | O = o) \quad (1)$$

这个问题可通过称为 Viterbi 算法的动态规划方法加以解决,其时间复杂度为  $O(TN^2)$ 。

基于 HMM 的全词消歧模型需要选择合适的状态集合  $S$  和观察值集合  $V$ 。状态集合  $S$  的选择有多种,我们直接取训练集中出现过的 synset 作为状态集  $S$  的一部分。由于封闭词的个数是有限的,我们令每个封闭词为一个状态。另外,对于某些特殊的词类(子类),我们认为它们的成员对于词义消歧而言是等价的,因而将它们整体作为一个状态,这有利于缓解数据的稀疏性,这包括人名、地名、组织名、其他专有名词、序数词、属格代词、联结词和情态动词。观察值集合  $V$  的选择也有多种,我们以词条(lemma)和词性的字符串连接作为观察值。可用作观察值的有:词、词条、词与词性的字符串连接或词条与词性的字符串连接等。[13]的实验结果表明,词与词性的字符串连接作为输入效果较好,我们沿用这一作法。 $A, B, \pi$  均由训练集作极大似然估计得到。

### 3 基于 MEMM 模型的全词消歧方法

McCallum 等人在[5]中将 HMM 改造为 MEMM, 其不同之处在于: (1)观察值的扩充。HMM 的观察值集合一般只能为一个有限的词表; 而 MEMM 对原始观察值抽取若干个非独立特征, 构成特征向量, 然后根据特征向量计算给定观察序列的状态序列的条件概率。(2)由生成模型到条件模型。在 HMM 中, 当前状态仅依赖于前一状态(马尔可夫性); 在 MEMM 中, 当前状态不仅依赖于前一状态, 还依赖于当前特征向量。

在 MEMM 中, 状态转移概率和状态-观察值发射概率被一个统一的新的概率  $P_{q_{i-1}}(Q_i = q_i | O_i = o_i) = P(Q_i = q_i | O_i = o_i, Q_{i-1} = q_{i-1})$  所代替, 即在前一状态为  $q_{i-1}$  且当前特征向量为  $o_i$  时, 当前状态为  $q_i$  的概率。  $P_{q_{i-1}}(Q_i = q_i | O_i = o_i)$  可用某些带概率输出的机器学习算法进行训练得到, 如最大熵算法、朴素贝叶斯算法和支持向量机等。最早提出该模型的[5]采用了最大熵算法, MEMM 模型因此得名。

MEMM 模型的要素为  $S, V, M, \pi$ 。其中, (1)  $S = \{s_1, s_2, \dots, s_N\}$  为状态集合; (2)  $V = \{v_1, v_2, \dots\}$  为由观察值抽取的特征向量构成的集合; (3)  $M = \{P_{q_{i-1}}(Q_i = q_i | O_i = o_i), q_i \in S, o_i \in V, q_{i-1} \in S\}$  为概率模型集合; (4)  $\pi = \{P_{begin}(Q_1 = q_1 | O_1 = o_1), q_1 \in S, o_1 \in V\}$  为初始概率模型。在 MEMM 中, 序列标注问题为也可通过 Viterbi 算法加以解决, 它与 HMM 的 Viterbi 算法是十分类似的。

在基于 MEMM 的全词消歧模型中, 状态集合  $S$  与 HMM 的状态集合  $S$  相同。特征向量集合  $V$  中的元素为由原始观察值抽取的特征向量构成的集合。我们以[2][14]中的特征为基础, 设计了以下 7 类特征: (1)邻近词的词性; (2)局部词形搭配关系; (3)依存句法关系; (4)特定范围内的 WordNet 上位词(hypemym)链; (5)特定范围内的 WordNet 语义标签; (6)特征范围内的 WordNet 动词框架; (7)句中的各个单词构成的词袋。对于每一个概率模型  $P_{q_{i-1}}(Q_i = q_i | O_i = o_i)$ , 我们通过将训练集中紧接在  $q_{i-1}$  后面的所有状态-特征向量对  $(q_i, o_i)$  收集起来, 然后用最大熵算法进行训练, 就得到了模型  $P_{q_{i-1}}(Q_i = q_i | O_i = o_i)$ 。对于初始概率模型  $P_{begin}(Q_1 = q_1 | O_1 = o_1)$ , 我们则收集句首的状态-特征向量对即可。

### 4 实现

如前所述, 序列标注一般通过 Viterbi 算法解决。全词消歧作为一个超大状态问题, 上述两个模型均存在严重的数据稀疏问题, 同时还具有过高的时间复杂度。下面提出的柱状搜索 Viterbi 算法和平滑策略, 解决了数据稀疏的问题。另外, 柱状搜索在解决数据稀疏问题的同时, 还显著地降低了解码的时间复杂度。

#### 4.1 柱状搜索

在全词消歧的 HMM 模型中, 发射概率矩阵是十分稀疏的。状态空间  $S$  的非常大, 包括训练集中出现过的 synset 及封闭词等, 其规模约为数万。然而, 一个观察值  $o_i$  (lemma)对应的状态数 (synset 数)却是十分有限的, 至多为数十个。为解决这个问题, 我们采用柱状搜索 Viterbi 算法进行解码。我们将观察值  $o_i$  对应的状态集合记为  $stateSet(o_i)$ , 在 Viterbi 算法迭代的每一步, 只搜索  $stateSet(o_i)$ , 而不是整个状态空间, 即

$$\delta_i(j) = \max_{1 \leq i \leq N} [\delta_{i-1}(i) a_{ij}] b_j(o_i) \quad \Rightarrow \quad \delta_i(j) = \max_{1 \leq i \leq N, s_j \in stateSet(o_{i-1})} [\delta_{i-1}(i) a_{ij}] b_j(o_i)$$

这不仅有效地解决了发射概率矩阵稀疏问题, 还显著地降低了解码的时间复杂度。通常, Viterbi 算法的时间复杂度为  $O(TN^2)$ , 其中,  $T$  为待消歧句子的长度,  $N$  为状态空间  $S$  的大小。在 HMM 方法中, 由于  $N$  的值非常大, 该算法的实际运行时间是难以承受的。使用柱状搜索后, Viterbi 算法

的时间复杂度由  $O(TN^2)$  降为  $O(TS_{\max}^2)$ , 其中  $S_{\max}$  为一个 lemma 最多可能对应的 synset 数。MEMM 的柱状解码与此十分类似, 本文从略。

## 4.2 平滑策略

在全词消歧的 HMM 模型中, 转移概率矩阵也是稀疏的。Viterbi 算法计算概率的乘积, 若转移概率为 0, 则乘积结果为 0, 无法比较结果的大小。为了避免整个乘积的结果为 0, 我们必须采用某种平滑策略。在 HMM 中, 转移概率矩阵  $A$  中元素  $a_{ij}$  的极大似然估计为  $a_{ij} = \frac{C(s_i s_j)}{C(s_i)}$ , 其中  $C(s_i s_j)$  表示在训练集中  $s_i s_j$  出现的次数。我们令  $a_{ij}$  平滑后的值为

$$a_{ij}^* = \begin{cases} \gamma \cdot \frac{C(s_i s_j)}{C(s_i)}, & \text{若 } C(s_i s_j) > 0 \\ (1-\gamma) \cdot \frac{1}{N - \sum_{i=1}^N I(C(s_i s_j) > 0)} \cdot \left( \sum_{s \in \text{Synsets}(v)} 1 \right) \cdot \frac{F(s_j)}{\sum_{s \in \text{Synsets}(v)} F(s)}, & \text{若 } C(s_i s_j) = 0 \text{ 且 } s_j \in \text{Synsets}(v) \\ (1-\gamma) \cdot \frac{1}{N - \sum_{i=1}^N I(C(s_i s_j) > 0)}, & \text{若 } C(s_i s_j) = 0 \text{ 且 } s_j \notin \text{Synsets}(v) \end{cases} \quad (2)$$

其中,  $I(\bullet)$  为指示函数,  $\sum_{s \in \text{Synsets}(v)} 1$  表示观察值  $v$  (lemma) 对应的状态 (synset) 数,  $F(s_j)$  表示  $s_j$  对应的 synset 在 WordNet 中出现的频数。上述公式的直观含义是, 我们假设  $s_i$  的转移概率中有  $\gamma$  出现在训练集中, 有  $1-\gamma$  未出现在训练集中; 对于未出现在训练集中的  $s_j$ , 若  $s_j \in \text{Synsets}(v)$ , 我们令其概率与其对应的 synset 在 WordNet 中出现的频数  $F(s_j)$  成正比; 对于所有未出现在训练集中的  $s_j$ , 若  $s_j \notin \text{Synsets}(v)$ , 我们令其概率均相等。在我们的实验中,  $\gamma = 0.999$ 。

对于 MEMM 模型, 我们令  $P_{s_i}(s_j | v)$  平滑后的值为

$$P_{s_i}^*(s_j | v) = \begin{cases} \gamma \cdot P_{s_i}(s_j | v), & \text{若 } C(s_i s_j) > 0 \\ (1-\gamma) \cdot \frac{1}{N - \sum_{i=1}^N I(C(s_i s_j) > 0)} \cdot \left( \sum_{s \in \text{Synsets}(v)} 1 \right) \cdot \frac{F(s_j)}{\sum_{s \in \text{Synsets}(v)} F(s)}, & \text{若 } C(s_i s_j) = 0 \\ & \text{且 } s_j \in \text{Synsets}(v) \\ (1-\gamma) \cdot \frac{1}{N - \sum_{i=1}^N I(C(s_i s_j) > 0)}, & \text{若 } C(s_i s_j) = 0 \\ & \text{且 } s_j \notin \text{Synsets}(v) \end{cases} \quad (3)$$

(3)的直观含义与(2)类似。

## 5 实验

在 Senseval/Semeval 出现之前,曾有少数学者尝试用 HMM 等序列标注方法进行全词消歧,如 Second 等人[9]采用 HMM 进行全词标注,其标记为 WordNet 的 45 个语义标签; Loupy 等人[10]采用一种集成了语义标签和词义的混合 HMM 进行全词标注。但由于上述研究并未使用通用测试集,其结果并不具有可比性。

Senseval/Semeval(2007 年之前称为 Senseval)是目前国际上最权威的词义消歧评测,我们的测试集来自 Senseval-2(2001)[15]、Senseval-3(2004)[16]的 English All Words 任务,表 1 给出了这些任务的基本信息。

表 1 英语全词消歧任务基本信息

	Senseval-2	Senseval-3
WordNet 版本	1.7	1.7.1
测试集规模(标注词数)	2473	2037

English All Words 为开放测试,只提供测试集,对训练集没有限制。我们仅使用了 SemCor[17]的 Brown1 和 Brown2 作为训练集,包含 359,732 个词(不含标点),其中 192,639 个词有词义标注。对于训练集中不存在的观察值,我们使用 WordNet 中的最常用词义进行标注。MEMM 方法采用 GIS 算法训练最大熵模型,迭代次数为 100。表 2 给出了参加英语全词消歧任务系统的性能(F1 值)。

表 2 参加英语全词消歧任务的系统性能(F1 值)

	Senseval-2	Senseval-3
第 1 名	0.690[18]	0.652[20]
第 2 名	0.636[19]	0.646[21]
基准系统	N/A	0.609
本文 HMM 方法	0.622	0.625
本文 MEMM 方法	0.648	0.654
[11]中 HMM 方法	0.618	N/A
[12]中 HMM 方法	0.602	N/A
[13]中 HMM 方法	N/A	0.609

English All Words 对训练集没有限制,因此对不同系统的比较是十分困难的,下面我们试图对实验结果作出一些说明。

首先,本文提出的 MEMM 方法的性能远远高于相同数据集上的其他序列标注方法,包括本文提出的 HMM 方法。这证明了将简单观察值扩展为复杂的特征向量确实是有效的,这得益于大量语言学特征的引入。

其次,在仅采用 SemCor 作为训练集的情况下,本文提出的 MEMM 方法也超过了 Senseval-2 的第 2 名和 Senseval-3 的第 1 名。其中, Senseval-2 的第 2 名[19]采用了基于记忆的方法,并使用多个分类器进行投票; Senseval-3 的第 1 名[20]也采用了基于记忆的方法,它的训练集包括 SemCor, Senseval-2 的数据以及 WordNet 中的例子。这说明序列标注方法完全可以用于全词消歧,并且性能与最好的有监督方法相当。我们看到,本文 MEMM 方法与 Senseval-2 的第 1 名[18]还有一定的差距,部分原因是[18]采用了很多 SemCor 以外的数据,如 WordNet 中的定义和互联网收集的数据等。

在 Senseval-2 中, Crestan 等人[11]采用两阶段 HMM 进行全词消歧。第一阶段先通过 HMM 确定词形的语义标签,第二阶段再用 HMM 确定词形+语义标签的词义,另外还对高频词采用词样消

歧的方法进行消歧。就性能而言,这种两阶段 HMM 方法与与单一 HMM 方法相当[10],在本实验中也得到了证实。

在 Senseval-2 和 Senseval-3 中, Molina 等人[12][13]采用 HMM 进行全词消歧,其状态为 SemCor 中的 `lex_sense`[17]。这相当于对 `synset` 进行了压缩,其好处在于使状态数减少,从而缓解矩阵稀疏的问题。但是,我们认为这种压缩并不具备语言学基础,有可能造成算法的不稳定性,而且难以将 HMM 扩展为 MEMM。实验表明,本文的 HMM 方法以 `synset` 为状态是合适的,其性能要略好于以 `lex_sense` 为状态的[12][13],且具备良好的可扩展性。

## 6 结束语

全词消歧本质上是一个序列标注问题。然而,现有序列标注方法在全词消歧上的表现却不尽如人意,主要原因在于特征的有效利用和数据稀疏问题。本文从全词消歧的特点出发,针对上述不足,提出了两种新的全词消歧序列标注方法,并采用柱状搜索和平滑策略解决了数据稀疏和高时间复杂度等问题,其中基于 MEMM 的全词消歧方法的性能较大幅度地超过了文献中已有的序列标注算法。由于本文仅使用了基本的训练语料,本文提出的方法的实际性能还有进一步提高的空间。实际上,本文提出的方法可适用于一般的超大状态序列优化问题。

无论是一般的分类问题还是序列标注,特征选择都是至关重要的,下一步我们将研究各种特征对整体性能的影响,以便更好地改进算法。另外,条件随机域(CRF)是近来出现、在多个领域均有突出表示的序列标注方法。CRF 是 HMM 和 MEMM 的推广,一般认为 CRF 的效果要比 HMM 和 MEMM 好。但是,CRF 训练的时间复杂度比 HMM 和 MEMM 要高得多,一般只用于状态数较少(一般数百个状态以内)的场合。对于全词消歧这类含有数万个状态的问题,即使在采用柱状搜索算法后,CRF 的训练仍然不可能在普通的工作站上完成,我们下一步将尝试采用大规模集群计算的方式来解决这个问题。

## 参考文献

- [1] Mooney, R. J. Comparative experiments on disambiguating word senses: An illustration of the role of bias in machine learning. In: Proceedings of the 1996 Conference on Empirical Methods in Natural Language Processing (EMNLP). 1996. 82-91.
- [2] Tratz, S., Sanfillippo, A., Gregory, M., Chappell, A., Posse, C., Whitney, P. PNNL: A supervised maximum entropy approach to word sense disambiguation. In: Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007). Stroudsburg, PA, USA, 2007. 264-267.
- [3] Escudero, G, Márquez, L., Rigau, G On the portability and tuning of supervised word sense disambiguation. In: Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP). 2000. 172-180.
- [4] Lawrence R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. In: Proceedings of the IEEE. 1989. 257-286. [doi: 10.1109/5.18626]
- [5] Andrew McCallum, Dayne Freitag, Fernando Pereira. Maximum Entropy Markov Models for Information Extraction and Segmentation. In: Proceedings of the 17th International Conference on Machine Learning. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000. 591-598.
- [6] John Lafferty, Andrew McCallum, Fernando Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: Proceedings of the 18th International Conference on Machine Learning. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001. 282-289.
- [7] El-Bèze M. & Mérialdo B.. HMM Based Taggers. In: H. Van Halteren eds. Syntactic Wordclass Tagging. Kluwer Academic Publishers, 1999.
- [8] F. Jelinek. Statistical Methods for Speech Recognition. Cambridge: MIT Press, 1998.

- [9] Segond, F., Schiller, A., Grefenstette, G., Chanod, J.P. An Experiment in Semantic Tagging using Hidden Markov Model Tagging. In: Proceedings of the Joint ACL/EACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources. Stroudsburg, PA, USA, 1997. 78-81. [doi: 10.3115/1220355.1220359]
- [10] Claude de Loupy, MarcEl-Beze, Pierre-François Marteau. Word Sense Disambiguation using HMM Tagger. In: Proceedings of the 1st International Conference on Language Resources and Evaluation. Granada, Spain, 1998. 1255-1258.
- [11] E. Crestan, M. El-Beze, C. De Loupy. Improving WSD with Multi-Level View of Context Monitored by Similarity Measure. In: Proceedings of the 2nd International Workshop on Evaluating Word Sense Disambiguation Systems. Toulouse, France, 2001. 67-70.
- [12] Antonio Molina, Ferran Pla, Encarna Segarra. A Hidden Markov Model Approach to Word Sense Disambiguation. In: Proceedings of the 8th Ibero-American Conference on AI: Advances in Artificial Intelligence. Longdon, UK: Springer-Verlag. 2002. 655-663.
- [13] Antonio Molina, Ferran Pla, Encarna Segarra. WSD system based on Specialized Hidden Markov Model. Proceedings of the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, 2004.
- [14] Yoong Keok Lee, Hwee Tou Ng. An Empirical Evaluation of Knowledge Sources and Learning Algorithms for Word Sense Disambiguation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA, USA, 2002. 41-48. [doi: 10.3115/1118693.1118699]
- [15] Edmonds P., Cotton S.. Senseval-2: Overview. In: Proceedings of the 2nd International Workshop on Evaluating Word Sense Disambiguation Systems. 2001. 1-6.
- [16] Benjamin Snyder, Martha Palmer. The English All-Words Task. In: Proceeding of Senseval-3: The 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text. Barcelona, Spain, 2004. 41-43.
- [17] Miller, G A., Chodorow, M., Landes, S., Leacock, C., Thomas, R. G. Using a Semantic Concordance for Sense Identification. In: Proceedings of the ARPA Workshop on Human Language Technology. Stroudsburg, PA, USA, 1994. 240-243. [doi: 10.3115/1075812.1075866]
- [18] Mihalcea, R. Word sense disambiguation with pattern learning and automatic feature selection. *Natural Language Engineering*, 2002, 8(4): 348-358.
- [19] Hoste, V., Hendrickx, I., Daelemans, W., van den Bosch, A. Parameter optimization for machine learning of word sense disambiguation. *Natural Language Engineering*, 2002, 8(4): 311-325.
- [20] Decadt, B., Hoste, V., Daelemans, W., and van den Bosch, A. GAMBL, genetic algorithm optimization of memory-based WSD. In: Proceedings of the 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text. 2004. 108-112.
- [21] Mihalcea, R., Faruque, E. Senselearner: Minimally supervised word sense disambiguation for all words in option text. In: Proceedings of the 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text. 2004. 155-158.