5.2 Journal Article Summarization in Non-medical Domains While summaries of medical journal articles are clearly important for physicians, summarization of journal articles from other ﬁelds is also a topic of current research. Given that journal articles are typically organized in a more predictable way than texts from other genres, proposed approaches often exploit that predictable structure to ﬁnd salient information. It is expected, for example, that the paper will haveanintroduction,statementofgoals,comparisonwithrelatedwork, approach description, results and conclusions. In this section, we surveyanapproachusingrhetoricalstatusandothersthatuseinformation about citations.

5.2.1 Using Genre-speciﬁc Rhetorical Status A summarization system for journal articles can take advantage of the rhetorical status of a sentence in producing a summary. Teufel and Moens [200] proposed an annotation scheme for rhetorical status of sentences with seven categories. The categories include : AIM The speciﬁc research goal of the current paper. OWN A neutral description of methodology, results and discussion. CONTRAST Statements of comparison with other work or weaknesses of other work. BASIS Statements of agreement with other work or continuation of other work. Human annotators were able to follow the scheme reliably to annotate each sentence in scientiﬁc articles with one of these rhetorical status categories. The annotated corpus was used to train a Naive Bayesclassiﬁerusingavarietyoffeatures,includingsentencelocationin the article, the section and the paragraph, type of section (conclusion, experiments, etc.), sentence length, number of words in the sentence thatalsoappearinthepapertitle,presenceofwordswithhighTF∗IDF weight, verb tense, voice and presence of modal auxiliaries, presence and nature of citation, rhetorical context, and presence of 644 formulaic expressions. The classiﬁer was able to predict the rhetorical status of novel sentences with accuracy of 73%.

5.2 Journal Article Summarization in Non-medical Domains 181

Listing the sentences that fulﬁll the AIM, CONTRAST, BASIS and BACKGROUND for each paper gives an excellent generic summary of a scientiﬁc article. The explicit indication of the rhetorical status is helpful for users not familiar with the paper that is being summarized. Usually, there are relatively few AIM, CONTRAST and BASIS sentences in the paper and all of them can be included in the summary. BACKGROUND sentences, on the other hand, are more numerous so displaying all of them might become problematic when stricter summary length restrictions are imposed. To solve this problem, one more classiﬁerwastrainedtodistinguishbetweensummary-worthysentences and other sentences. The classiﬁer is used to select only the most appropriateBACKGROUNDsentencesanditsuseincreasestheoverall performance of the summarizer.

5.2.2 Exploiting Citation Links between Papers Genre-speciﬁc summarization of scientiﬁc articles can take advantage of a valuable and powerful source of information not available in most other genres: the references (citations) that link papers. The sentences and paragraphs in which references occur often contain concise summaries of the cited paper or other information relevant to that paper. Consequently, the analysis of related articles becomes an essential knowledge source for summarization. Three types of extractive summaries based on citation link analysis have been proposed: overview of a research area (multi-document scientiﬁc paper summary) [145], impact summary (single document summary of a paper using sentences from the paper itself) [134] and citation summary(amixofmulti-andsingledocumentsummarization, in which a single paper is summarized but the input to the summarizer are the sentences from other papers in which that paper is cited) [169]. Writing an overview of several related scientiﬁc papers is a diﬃcult task, even for people. Nanba and Okumura [145] proposed a system for facilitating writing such overviews, which visualizes connections between papers and provides information on how the papers are related to each other. They develop a rule-based system to identify reference areas in a paper, in which other papers are discussed and cited.

182 Genre and Domain Speciﬁc Approaches

In addition, each of these areas is classiﬁed as belonging to one of three types: (i) describing methods or approaches used in the paper, (ii) discussion of and comparison with related work, and (iii) other. Classiﬁcation is performed using hundreds of manually coded cue words. The usercanrequesttoseeeachtypeofreferenceareaforapaperofinterest. With the ever increasing number of scientiﬁc publications, the need to further develop and improve such aids for paper browsing and access will only increase with time. In fact much progress has already been achieved in the automatic identiﬁcation and classiﬁcations of citations and citation types [191]. Impact summarization is deﬁned by Mei and Zhai [134] as the task of extracting sentences from a paper that represent the most inﬂuential content of that paper. They employ a language model approach to solve the task. For each paper to be summarized, they ﬁnd other papers in a large collection that cite that paper and extract the areas in which the references occur. A language model is built using the collection of all reference areas to a paper, giving the probability of each word to occur in a reference area. This language model gives a way of scoring the importance of sentences in the original article: important sentences are those that convey information similar to that which later papers discussed when referring to the original paper. The measure of similarity between a sentence and the language model was measured by Kullback–Leibler (KL) divergence. In order to account for the importance of each sentence within the summarized article alone, they use word probabilities estimated from the article. The ﬁnal score of a sentence is a linear combination of impact importance coming from KL divergenceandintrinsicimportancecomingfromthewordprobabilities in the input article. The method produces extractive summaries that are vastly superior, as measured by the ROUGE automatic measure, to baselines and to the generic summarization system MEAD [202]. One drawback of impact-based summaries is that, while they contain valuable content, they have low linguistic quality and are hard to read and understand, as can be seen from the example shown in Figure 5.2. Citationsummarization[169]isadiﬀerentapproachtosummarizing a single article, based on the way other papers refer to it. Citation summarization does not use at all the text of the article being summarized.

5.2 Journal Article Summarization in Non-medical Domains 183

1. Figure 5: Interpolation versus backoﬀ for Jelinek-Mercer (top), Dirichlet smoothing (middle), and absolute discounting (bottom). 2. Second, one can decouple the two diﬀerent roles of smoothing by adopting a two stage smoothing strategy in which Dirichlet smoothing is ﬁrst applied to implement the estimation role and Jelinek–Mercer smoothing and Jelinek–Mercer smoothing is then applied to implement the role of query modeling. 3. We ﬁnd that the backoﬀ performance is more sensitive to the smoothing parameter than that of interpolation, especially in Jelinek–Mercer and Dirichlet prior. 4. We then examined three popular interpolation-based smoothing methods (Jelinek–Mercer method, Dirichlet priors, and absolute discounting), as well as their backoﬀ versions, and evaluated them using several large and small TREC retrieval testing collections. 5. By rewriting the query-likelihood retrieval model using a smoothed document language model, we derived a general retrieval formula where the smoothing of the document language model can be interpreted in terms of several heuristics used in traditionalmodels,includingTF-IDFweightinganddocumentlengthnormalization. 6. We ﬁnd that the retrieval performance is generally sensitive to the smoothing parameters, suggesting that an understanding and appropriate setting of smoothing parameters is very important in the language modeling approach.

Fig. 5.2 An example of an impact-based summary.

Instead,QazvinianandRadev[169]proposetosummarizethereference areas in other articles related to the target paper that has to be summarized. The input to the summarizer consists of the sentences or short paragraphs from other papers that discuss the target article. There is a high degree of repetition in the input because many papers refer to the same aspect of the target paper: an approach, result, main contribution, etc. Given this characteristic of the data, a clustering method for summarization seems highly appropriate (see Section 2.1.2). Similarsentencesareclusteredtogetherandarepresentativesentenceischosen to convey the information in that cluster. The best way to ﬁnd the mostrepresentativesentenceineachclusterturnedouttobeapplyingthe graph-basedmethoddiscussedinSection2.1.3forsummarizationtoeach cluster.Theevaluationwasperformedon25articlesfromﬁvesub-areasof computationallinguistics,usingthemanualPyramidevaluationmethod.

184 Genre and Domain Speciﬁc Approaches

1. The Czech parser of Collins et al. (1999) was run on a diﬀerent data set and most other dependency parsers are evaluated using English. 2. More precisely, parsing accuracy is measured by the attachment score, which is a standard measure used in studies of dependency parsing (Eisner, 1996; Collins et al., 1999). 3. In an attempt to extend a constituency-based parsing model to train on dependency trees, Collins transforms the PDT dependency trees into constituency trees (Collins et al., 1999). 4. More speciﬁcally for PDT, Collins et al. (1999) relabel coordinated phrases after converting dependency structures to phrase structures, and Zeman (2004) uses a kind of pattern matching, based on frequencies of the parts-of-speech of conjuncts and conjunctions. 5. In particular, we used the method of Collins et al. (1999) to simplify part-ofspeech tags since the rich tags used by Czech would have led to a large but rarely seen set of POS features.

Fig. 5.3 An example of a citation summary.

Citation summaries are even harder to read than impact summaries because they mix information about the work done in the paper which citesthetargetarticlewithdescriptionsoftheworkdescribedinthetarget article itself, as shown in Figure 5.3. Application of the approaches developedforcategorizationoftypesofreferenceareasshouldbehelpful inovercomingthisprobleminfuturework. All of these approaches reveal that there is much work that remains in the ﬁeld of journal summarization. Very diﬀerent approaches from generic summarization have been proposed that exploit the structure and speciﬁc characteristics of journal articles and this is appealing. The quality of the summaries that are produced, however, still leaves much to be desired. Further work is needed that takes these new approaches to the next level, perhaps by integrating work on ﬂuency, cohesion and sentence ordering.