Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 547923, 19 pages
doi:10.1155/2008/547923

Research Article

Hierarchical Fuzzy Feature Similarity Combination for Presentation Slide Retrieval

A. Kushki, M. Ajmal, and K. N. Plataniotis

Multimedia Laboratory, The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada M5S 3G4

Correspondence should be addressed to A. Kushki, azadeh.kushki@alumni.utoronto.ca

Received 18 April 2008; Revised 8 September 2008; Accepted 6 November 2008

Recommended by William Sandham

This paper proposes a novel XML-based system for retrieval of presentation slides to address the growing data mining needs in presentation archives for educational and scholarly settings. In particular, contextual information, such as structural and formatting features, is extracted from the open format XML representation of presentation slides. In response to a textual user query, each extracted feature is used to compute a fuzzy relevance score for each slide in the database. The fuzzy scores from the various features are then combined through a hierarchical scheme to generate a single relevance score per slide. Various fusion operators and their properties are examined with respect to their effect on retrieval performance. Experimental results indicate a significant increase in retrieval performance measured in terms of precision-recall. The improvements are attributed to both the incorporation of the contextual features and the hierarchical feature combination scheme.

1. INTRODUCTION

Retrieval tools have proven to be indispensable for searching and locating relevant information in large repositories. A plethora of solutions has been proposed and successfully applied to document, image, video, and audio collections. Despite this success, bridging the so-called *semantic gap* still remains a key challenge in developing retrieval techniques. This semantic gap refers to the incongruity between the subjective and context-dependent human interpretation of semantic concepts and their low-level machine representations. The ambiguities resulting from the semantic gap can be partially resolved if the application domain is restricted to particular types of repositories (e.g., fingerprint databases, news clips, soccer videos, etc.). In such restricted environments, application-specific knowledge can be utilized to develop custom retrieval solutions. In this paper, we restrict the problem domain to slide presentation repositories and exploit the specific characteristics of slide presentations to propose a retrieval tool geared toward such collections. This tool is developed to provide efficient access to the increasing volumes of slides for the purposes of data mining in scholarly and educational settings, where a large number of slide presentations are archived, processed, and browsed [1, 2]. Compared to traditional text and multimedia retrieval, the slide retrieval problem offers unique opportunities and challenges. First, slides generally contain multimodal content; that is, in addition to text information, images, video, and audio clips may be embedded into a slide. We, thus, need a procedure to extract, process, and combine information from various modalities during retrieval. Second, since slides generally contain summarized points, as opposed to full sentences in traditional document retrieval, the occurrence frequency of a term in a slide is not a direct indication of the slide relevance to the query [3]. Third, slide contents are naturally structured; they consist of various levels of nesting delineated by titles and bullet levels. Thus, the relative positioning of text in this structure can provide hints about the degree of relevance of each term as perceived by the author. Such information can be used in combination with traditional keyword matching to improve retrieval performance [3, 4]. The direct availability of structural information in slides should be contrasted to other multimedia, such as images and video, where the determination of structure (e.g., position of objects and division into shots and scenes) requires significant processing effort.

In this paper, we propose a tool for retrieval of slides from a presentation repository. An outline of the proposed system is depicted in Figure 1. Upon receiving a textual user query term, binary keyword matching is applied to parsed presentation content to generate a subset of candidate slides using the XML representation. The proposed system uses structural and text formatting attributes, such as indentation level, font size, and typeface, to calculate a *relevance* score for occurrences of the query term on each slide. Slides are then ranked and returned to the user in order of descending relevance.

The contributions of this work are threefold. First, the Extensible Markup Language (XML) [5] representation of presentations, based on the standard open format OpenXML [6], is used here for the first time to provide direct access to slide contents. XML tags are used to obtain semantic and contextual information, such as typeface and level of nesting, about the prominence of their enclosed text in addition to slide text. These tags also readily identify nontext components of slides including tables and figures. Lastly, multimedia objects augmented with XML-compatible metadata, such as Exif metadata provided by most digital cameras, can be processed and associated with semantic information. The second contribution of this paper lies in the use of contextual information supplied by XML tags to judge the *relevance* of each slide to the user query. A novel solution is proposed to model the naturally structured contents of slides and their context by constructing a feature hierarchy from the available XML tags. Slide *relevance* with respect to a given user query is then calculated based on leaf nodes (keywords and their context) and the scores are propagated through the hierarchy to obtain the overall slide *relevance* score. The slide scores are computed through a fuzzy framework to model the inherent vagueness and subjectivity of the concept of *relevance*. The third contribution of this paper is the examination of various fuzzy operators for combining feature level scores. The proposed score combination scheme provides a flexible framework to model the subjective nature of the concept of term relevance in varying slide authoring styles.

The rest of this paper is organized as follows. Section 2 outlines the prior art and contributions of this work, Section 3 provides the details of the features used in the proposed system, Sections 4 and 5 present the details of the proposed fuzzy score calculation framework, Section 6 outlines the experiments and results, and Section 7 concludes the paper and provides directions for future work.

2. OVERVIEW OF CONTRIBUTIONS AND RELATED WORK

Figure 2 shows the typical components of a slide retrieval system. The first step is to extract text and multimedia content from slides. This is followed by extraction of features from this content for the purpose of retrieval. Lastly, the extracted features are used to determine relevant slides in response to a user query specified as a textual keyword. The rest of this section outlines the existing efforts with respect to each of these three components.

Direct access to slide contents has traditionally posed a significant challenge because slides generated by popular software applications are generally stored in proprietary formats, such as Microsoft PowerPoint or Adobe Portable Document Format (PDF), and not in plain text. Consequently, an application programming interface (API) is needed for extraction of slide contents [7–9]. For example, the work of [9] translates the Microsoft PowerPoint (PPT) format into an XML file that can then be used for feature extraction. Such APIs, however, may be expensive and must be updated regularly to maintain conformance to these formats. An alternative method of accessing slide content is to rely on additional presentation media, such as audio and video, and to extract slide content using automatic speed recognition (ASR) [7, 10] and optical character recognition (OCR) [4, 11] techniques. While these methods provide a format-independent solution for slide retrieval, their inherent reliance on the existence of additional media limits their utility in existing slide repositories as capturing video and audio recordings requires additional effort and equipment and is not yet common practice in current classrooms, conferences, and business venues. Moreover, transcription errors resulting from the inaccuracy of ASR and OCR are propagated to the retrieval stages, degrading the retrieval effectiveness of the system [11]. Lastly, although OCR can be used to access the text in images, detection of objects on a slide, such as tables, figures, and multimedia clips, and extraction of text features, such as size and indentation level, require further processing.

This paper utilizes the recently standardized open file formats for exchanging and storing documents, such as Microsoft's OpenXML and OASIS' OpenDocument, to overcome the limitations of previous methods in content extraction from slides. In particular, we propose a novel XML-based slide retrieval solution based on the OpenXML format used by Microsoft PowerPoint 2007 to store slide presentations. In contrast to API-based methods discussed previously, the XML method presented herein does not require any proprietary information since OpenXML is an open file format and an Ecma international standard [6]. Since the OpenXML format contains information extraneous to the retrieval process, we have developed a lightweight XML parser to generate a custom XML representation to improve readability and improve efficiency of feature parsing.

As shown in Figure 2, the second step is extraction of features for use during retrieval. Most existing slide retrieval solutions rely on the assumption that the number of occurrences of a keyword in a document is directly proportional to that document relevancy [11, 12]. This leads to the use of *term frequency* as the primary feature used for retrieval. Such an approach is, however, adopted from traditional document retrieval and does not fully utilize the specific characteristics of slides. In particular, slides generally contain a set of brief points and not complete sentences. Therefore, relevant terms may not appear more than once as authors use other techniques to indicate higher degrees

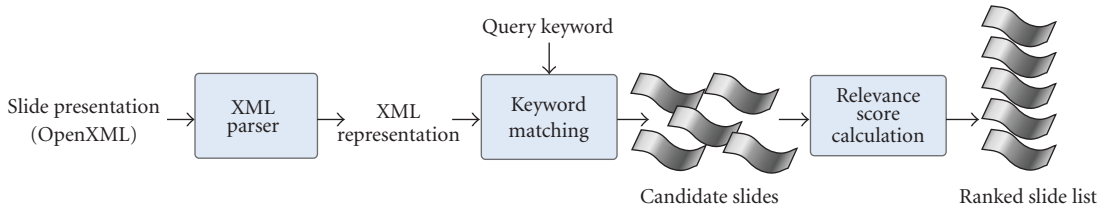


Figure 1: Overview of the proposed system.

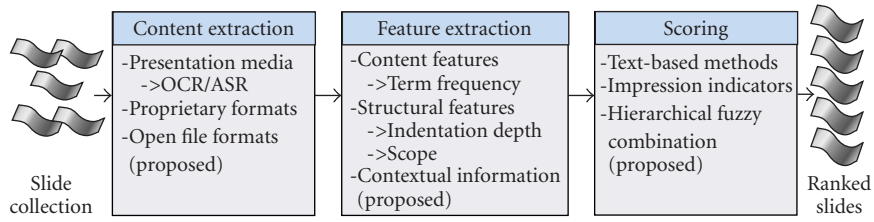


Figure 2: Components of a typical slide retrieval solution.

of relevance, for example, typeface [3]. In this light, recent slide retrieval techniques employ additional hints to calculate a score indicating the degree of relevance of each slide to the user query. For example, UPRISE [9] uses indentation level and slide duration in combination with term frequency.

Extraction of text-related information, such as nesting level, is especially convenient in XML-based formats as such information can readily be obtained from XML tags. The pervasive use of the XML format on the World Wide Web has motivated much research in the area of XML document retrieval, considering both content and structure of documents leading to *structure-aware retrieval* [13, 14]. The nesting level in an XML tree is an example of a structural feature used to express the degree of relevance of a keyword [15, 16]. While the efforts in the area of XML document retrieval do not deal with the unique characteristics of presentation slides, they motivate the incorporation of structural features, such as indentation depth, in slide retrieval. In addition to the use of structural features, we propose the utilization of *contextual features* that may be used by authors to indicate the degree of relevance of keywords. Contextual features, such as font size and typeface characteristics, are easily extractable from the XML representation of slides and can be used to provide hints as to the perceived degree of relevance of a keyword by the presentation author. Moreover, we propose a hierarchical feature representation to mirror the nested structure of slides and their XML representations.

Once the features have been extracted, they are used to generate a score indicating the degree of relevance of each slide to the user query. In text-based approaches, the vector space model [11, 12] is utilized to compute such a relevance score. For the problem of slide retrieval, however, the incorporation of structural and contextual features requires the development of methods for generating a relevance score based on multiple features. In UPRISE [9], a term score is in turn computed as the geometric mean of a position indicator (indentation level), slide duration, and number of query term occurrences. The contribution of adjacent slides is weighed into the slide score through the use of an exponential window and the overall score is the average of scores obtained for each occurrence of the query term in a slide. This work, however, does not provide any justification for the use of the geometric mean for feature combination. We propose a flexible framework based on fuzzy operators to model the subjective human perception of slide relevance based on the combination of term frequency, structural, and contextual features.

3. RETRIEVAL FEATURES

A slide consists of various text lines and possibly other objects, such as tables and figures. Each text line in turn contains multiple terms, a table contains rows, and multimedia objects are comprised of metadata as well as media content. Figure 3 depicts the decomposition of a slide into its constituent components using such a nested structure.

The corresponding XML representation of a slide is also a series of nested tags and each element in this nested structure describes the features of a slide component. An example slide and its XML representation, generated by our custom parser from the OpenXML representation, are shown Figure 4.

Using the given XML representation, slide text is easily accessible and a term frequency-based method can be used for retrieval. As previously discussed, however, such an approach is not sufficient in the case of slides due to the weaker correlation between a term occurrence frequency and its perceived relevance. In this light, the *context* of a keyword can be used to judge its prominence in a slide [9]. We use the term context to refer to text formatting features including font attributes and size as well as structural features such as indentation level. The XML representation of a slide provides a natural means for extracting such context-related features through tags which describe the various elements. In Figure 4, for example, the *level* and *attr* attributes appearing

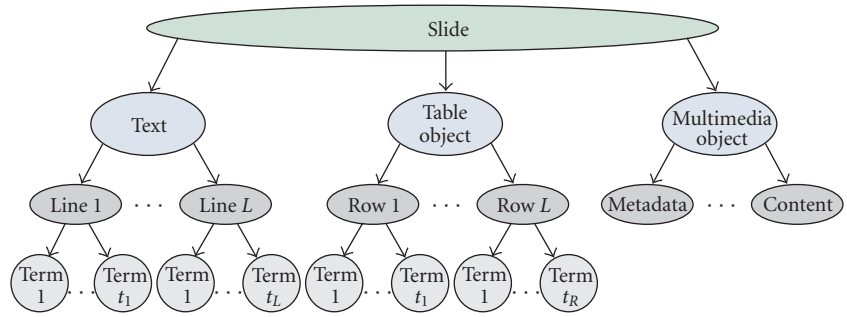


Figure 3: Slide structure.

XML-based retrieval

- Use the **structured** XML representation to access slide contents
- **Contextual** information provided by XML tags
  - Keyword features:
    - Bold, italics, underline
  - Line features:
    - Indentation level

(a) Example slide

```
<slide id="1">
    <title>XML-based retrieval</title>
    <bullet level="1">
        <w attr="n"> Use the  </w>
        <w attr="bi">structured </w>
        <w attr="n"> XML representation to access slide contents </w>
    </bullet>
    <bullet level="1">
        <w attr="bi">Contextual </w>
        <w attr="n"> information provided by XML tags </w>
    </bullet>
    <bullet level="2">
        <w attr="n"> Keyword features: </w>
    </bullet>
    <bullet level="3">
        <w attr="n"> Bold, italics, underline </w>
    </bullet>
    <bullet level="2">
        <w attr="n"> Line features: </w>
    </bullet>
    <bullet level="3">
        <w attr="n"> Indentation level </w>
    </bullet>
</slide>
```

(b) Simplified XML representation

Figure 4: Example slides and their simplified XML representation.

within the *bullet* and *w* tags describe the indentation level and text formatting features. This section describes the details of the structural and contextual features used for score calculation.

3.1. Feature hierarchy

This work proposes the modeling of the nested structure of a slide and its XML representation through a *feature hierarchy*. At the lowest level of this hierarchy reside term specific features such as font typeface characteristics (bold, italics, underline). The next level includes features that describe an entire line of text, that is, a group of terms, as opposed to an individual term. An example of a line feature is indentation level which provides information on the relative placement of a group of terms with respect to the rest of the slide content. The highest level in the hierarchy is used for features that describe a slide as a whole; term frequency, for example, is a slide-level feature as it considers the number of occurrences of a term on a slide and not features of any individual occurrence. We limit the scope of this work to text-based content and structural features, and note that additional feature levels can readily be added to include multimedia metadata and content features.

3.2. Word level features

The features residing on the lowest level of the hierarchy describe the formatting attributes of individual textual terms. The main motivation for the use of these formatting features is that these text effects are often used to add emphasis and distinguish relevant terms from the rest of the text. Typeface features used in this work are boldface, italics, and underline, denoted as $B(t)$, $I(t)$, and $U(t)$ for a term t, respectively. These features are binary in nature, that is, $B(t),I(t),U(t) \in \{0, 1\}$. Mathematically, we define these features as

$$B(t) = \begin{cases} 1, & \text{if } t \text{ appears in bold,} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

The italic and underline features, $I(t)$ and $U(t)$, are defined similarly.

3.3. Line level features

The second level in the feature hierarchy is comprised of those features that describe a group of terms appearing at the same bullet level. We consider indentation level and font size as line features here. Note that font size can also be considered as a word-level feature. The decision to include this feature as a line-level feature was a result of the observation that font size changes are generally applied at the bullet level and not to isolated terms within a sentence.

Since slide contents are generally presented in point form, the indentation or bullet level of a point can be used to indicate the degree of relevance of a group of terms. For this reason, we consider indentation depth, denoted as $\text{ind}(t)$, as a line feature:

$$\text{ind}(t) = d, \quad 0 \leq d \leq D, \tag{2}$$

where the integer d corresponds to the depth of the slide title and D is the maximum indentation level in the slide (in our experiments $D = 5$). Note that while indentation is considered as a line feature, $\text{ind}(t)$ is defined for an individual term t for notational convenience.

The size feature indicates the font size of a term t and is denoted as $sz(t)$. Font size is related to perceived degree of relevance as prominent terms, such as slide titles, are generally marked by an increase in font size. Font size for a term t is defined as

$$sz(t) = s, \quad \text{for } s \in \mathbb{N}, \tag{3}$$

where $\mathbb{N}$ is the set of positive integers. In practice, s is bounded by the minimum and maximum font sizes allowable by the presentation software. Similar to the indentation feature, the size is defined for an individual term t for notational convenience.

Note that for many presentation templates, such as those provided by PowerPoint, the font size decreases with an increase in indentation depth. In this sense, the two line-level features are correlated.

3.4. Slide level features

Slide features are those that describe the slide as a whole and reside on the top-most level of the hierarchy. Term frequency, defined as the number of occurrences of a term within a slide, is used as slide-level feature in this work. We define this feature mathematically as

$$TF(t) = n, \quad 0 \leq n \leq N_{s_i}, \tag{4}$$

where n is the number of times term t appears on the given slide and N_{s_i} is the total number of terms on slide s_i.

4. RELEVANCE CALCULATION

Having described the features used in retrieval, we proceed to present a framework for the calculation of *relevance* scores based on these features. The objective is to calculate a single score for each slide based on the multiple features in the previously discussed hierarchy. To do this, we must consider how the individual features are to be combined to produce such a score [17, 18]. One avenue is to combine the features directly. For example, in the text-based methods the, features of term frequency and inverse document frequency are combined using the product operator to generate a single score. Such a feature-level combination approach, however, is not suitable for use with the proposed feature hierarchy. The difficulty arises from two sources: (1) the proposed features provide values that are on different mathematical scales and quantization levels and (2) features on different levels of hierarchy report on attributes at different resolutions and levels of granularity.

For these reasons, we propose the combination of decisions or opinions formed based on feature values instead of direct combination of features [17, 18]. This approach eliminates the difficulties associated with fusion of features with different dynamic ranges (scales). Secondly, we propose a hierarchical decision combination structure to ensure that decisions are combined at the same granularity level, in this case, word, line, and slide level. The idea of this combination scheme is illustrated graphically in Figure 5. In this section, we detail the calculation of scores on each feature level and dedicate Section 5 to the discussion of decision combination methods.

Since *relevance* is a subjective human concept, we propose to calculate relevance scores through the framework of fuzzy sets [19]. This choice is motivated by the effectiveness of fuzzy sets in modeling vague human concepts and their success in multicriteria decision making applications [18, 20–23]. In [20], for example, the so-called *concept hierarchy* is used to model a complex human concept, such as creditworthiness, through various and possibly correlated low-level concepts. A similar methodology has been applied to the problem of content-based image retrieval in [18] to model the high-level concept of *similarity* between two images in terms of low-level machine features such as color and texture. In a similar manner, we model the high-level concept of term relevance based on the lower level features in the proposed feature hierarchy.

Fuzzy sets provide a way for mathematically representing concepts with imprecisely defined criteria of membership [19]. In contrast to a crisp set with binary membership, the grade of membership to a fuzzy set is gradual and takes on values in the $[0, 1]$ continuum. Formally, a fuzzy set A on a domain χ is defined as the set of ordered pairs $\{(x, \mu_A(x)\}$, where $x \in \chi$ and $\mu_A : \chi \rightarrow [0, 1]$ associates each $x \in \chi$ with a grade of membership to the set A [23].

In order to develop our scoring system, we begin by defining a fuzzy set (or fuzzy goal [23]) *relevant term* denoted as $\mathcal{T}$. A feature score is then the grade of membership of a term t to the fuzzy set $\mathcal{T}$ based on a given feature on a given slide s_i, indicating the degree to which the given feature value satisfies the goal of *relevance*. Denote the kth feature used in retrieval as F_k and the value of this feature for term t as $F_k(t)$. Then, the membership function $\mu_{\mathcal{T},F_k,s_i}(F_k(t))$ maps a feature value $F_k(t)$ into a score or grade of membership to the set $\mathcal{T}$

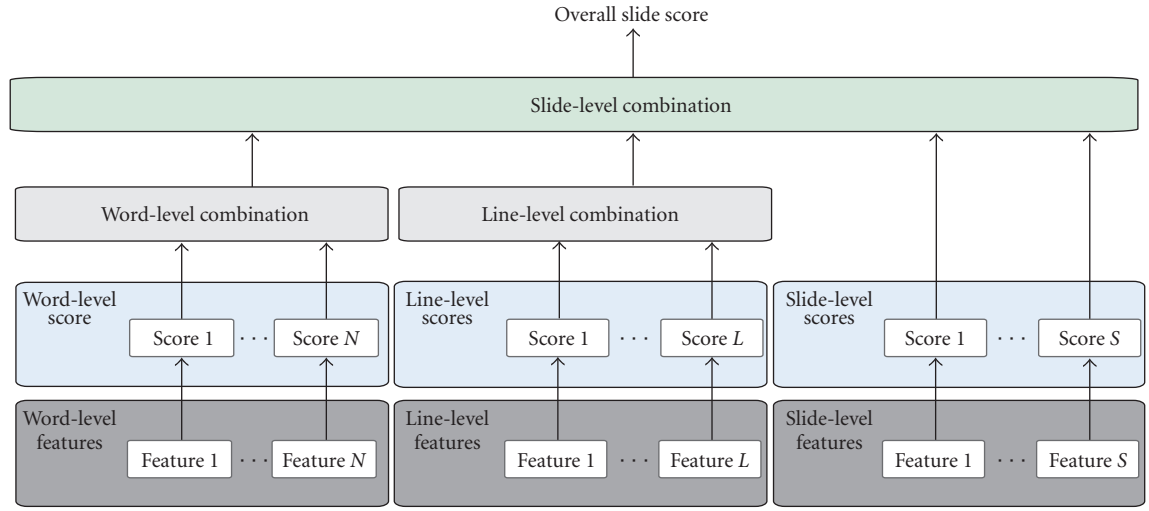
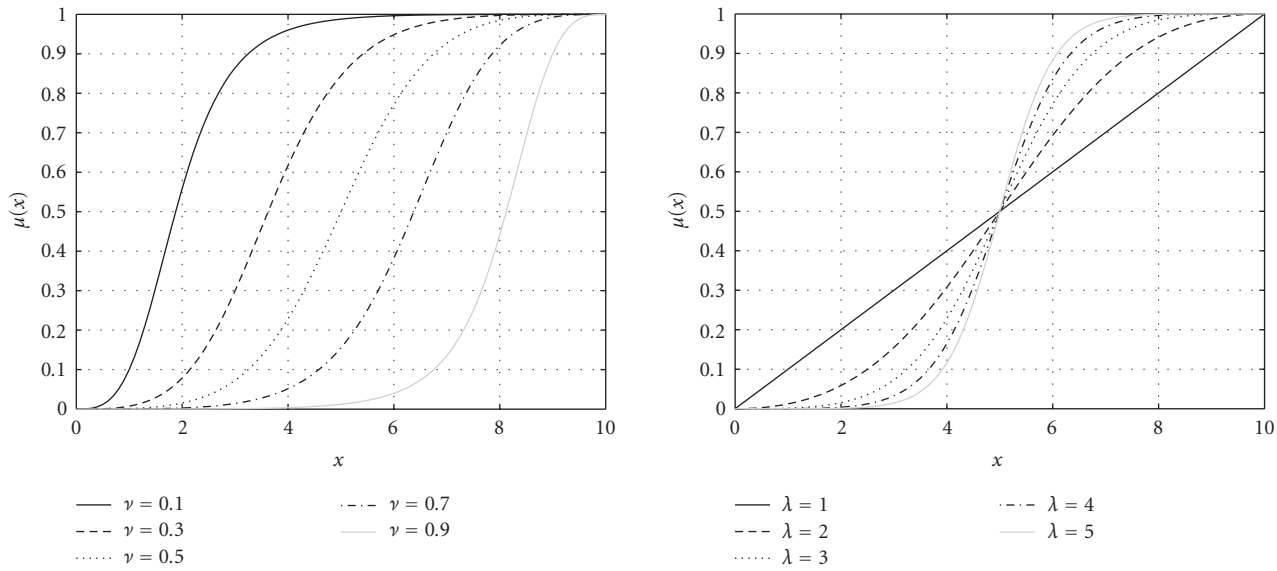


Figure 5: Overview of the relevance calculation model applied to each slide.



(a) $\mu(x)$ for various values of ν ($\lambda = 3$)

(b) $\mu(x)$ for various values of λ ($\nu = 0.5$)

Figure 6: The generalized membership function for different parameter values.

for a slide s_i. This grade of membership can then be viewed as the *feature score*, decision, or opinion formed based on the value of $F_k(t)$. To increase readability, the dependence on the set $\mathcal{T}$ and slide s_i is dropped for the rest of the discussion and $\mu_{\mathcal{T},F_k,s_i}(F_k(t))$ is denoted as $\mu_{F_k}(t)$.

The main challenge in developing the fuzzy scoring scheme is the determination of the membership functions that map a feature value to a score value in $[0,1]$. This corresponds to the *modeling* step in multicriteria decision making [24]. Formally, we seek a membership function $\mu_{\mathcal{T},\mathcal{F}_k,s_i} : \mathcal{F}_k \to [0,1]$. In the simplest case, the membership function normalizes the feature values to lie in the range $[0,1]$:

$$\mu_{F_k}(t) = \frac{F_k(t)}{\max_{\forall t} F_k(t)}. \tag{5}$$

Membership functions can be interpreted in several other ways [25]. Among these is the *likelihood* view, where the membership grade of a term t to a set $\mathcal{T}$ is interpreted as a conditional probability $\mu_{F_k}(t) = P(\mathcal{T}|F_k(t))$. Here, it is assumed that the meaning of $\mathcal{T}$ is objective and fuzziness is a result of error or inconsistency. Experiments, such as polling, can be used to capture the view of fuzziness in such cases [25]. For the intended application, the meaning of $\mathcal{T}$, the set of relevant terms, is subjective and context dependent. This renders the likelihood view inappropriate for the slide retrieval problem.

Fuzzy membership to a set $\mathcal{T}$ can also be viewed as the degree of similarity between t and an ideal or prototype object of $\mathcal{T}$ denoted as t_0 [26]. This membership is a function of the distance between the features of t and those of t_0, denoted as $d(F_k(t), F_k(t_0))$. The following form for the

function has been proposed [27–29]:

$$\mu_{F_k}(t) = \frac{1}{1 + d(F_k(t), F_k(t_0))}. \quad (6)$$

This view requires the existence of an ideal prototype t_0 and the definition of a metric space, where similarity between the features $F_k(t)$ and $F_k(t_0)$ is measured. In [29], a context-dependent standard b is used for a quick evaluation of the above function. Noting the exponential relationship between physical units and perception, the following membership function is then proposed:

$$\mu(F_k(t)) = \frac{1}{1 + \exp\left(-a(F_k(t) - b)\right)}. \quad (7)$$

Equation (7) defines an S-shaped function with the context-dependent standard b and evaluation unit a.

As an alternative to the above approaches, the work of [22] provides a theoretical basis for design of the membership functions. This is done by an examination of previous approaches to membership construction and the consequent postulation of five axioms that lead to the derivation of a general form for membership function. The effectiveness of this form is then verified against the empirical data in [29]. The generalized membership function is as follows [22]:

$$\mu_{F_k}(t) = \frac{(1-\nu)^{\lambda-1}(F_k(t) - a)^{\lambda}}{(1-\nu)^{\lambda-1}(F_k(t) - a)^{\lambda} + \nu^{\lambda-1}(b - F_k(t))^{\lambda}}. \quad (8)$$

Equation (8) defines a parameterized family of S-shaped, monotonically increasing functions with $\mu(a) = 0$ and $\mu(b) = 1$, where a and b represent the range of $F_k(t)$. The parameters λ and ν determine the sharpness and inflection point of the function. Figure 6 shows this function for different values of ν and λ. For the case of $\lambda = 1$, the membership function of (8) reduces to a linear function:

$$\mu_{F_k}(t) = \frac{F_k(t) - a}{b - a}. \quad (9)$$

The monotonically decreasing version of the above membership function can be defined through a linear transformation [22]:

$$\mu_{F_k}(t) = \frac{(1-\nu)^{\lambda-1}(b - F_k(t))^{\lambda}}{(1-\nu)^{\lambda-1}(b - F_k(t))^{\lambda} + \nu^{\lambda-1}(F_k(t) - a)^{\lambda}}. \quad (10)$$

An important consideration in developing membership functions for the application of slide retrieval is the subjectivity and context dependence of the concept of *relevant term*. This is especially evident in slide repositories that include presentations with numerous authoring styles, where each author uses different means to indicate varying degrees of relevance for each term. While some authors use indentation level to indicate the relevance of terms, some vary the typeface or change the font features to achieve the same effect. For the proposed application, therefore, the membership functions are functions not only of the feature $F_k(t)$ but also of the context in which the term appears. We use this observation to generate context-dependent membership functions for slide features. In particular, context dependence is achieved by *contextualizing* the model parameters, a and b, to indicate the context of a term within a slide. Recall that the parameters a and b indicate the range of values for the particular feature. Instead of using global extremities, obtained over the entire database, we consider the range of feature values over a localized context such as a single presentation or a slide. Such localized determination of the feature domain aims to capture the varying author styles among the different presentations. In the rest of this section, this context-dependent formulation is used to develop membership functions for features discussed in Section 3.

4.1. Word-level scores

The typeface features, $B(t)$, $I(t)$, and $U(t)$, are binary in nature. A simple context-independent membership then assigns the highest membership grade to a term when it appears in bold, italics, or is underlined, respectively, and the lowest grade of zero otherwise. The membership function then becomes the identity function

$$\mu_B(t) = B(t), \qquad \mu_I(t) = I(t), \qquad \mu_U(t) = U(t). \quad (11)$$

Note that the above can be obtained from (9) with $a = 0$ and $b = 1$ for the binary features. The main disadvantage of this formulation is the assumption that changes in typeface always indicate changes in degrees of relevance. This, however, is a serious limitation in the slide retrieval application as various authoring styles may use typeface changes for different purposes. Consider, for example, the scenario when the entire presentation is written in italics. In this case, italicizing a term does not add any emphasis and is, therefore, not an indication of the degree of relevance of the term. In order to incorporate the context of a query keyword into the membership function, we propose to use (8) with contextual parameters a_C and b_C, where the parameter C denotes a *context unit*, corresponding to either a slide or an entire presentation. The contextual parameters will be used to indicate the rarity of a given feature, utilizing the intuitive notion that rarely used typeface features carry more information than those that are frequently used. For this purpose, the context parameters are defined as $b_C = \sum_{t_i \in C} B(t_i)$ and $a_C = 0$, where t_i denotes the ith term in context unit C. These parameters are used in (8) to obtain

$$\mu_B(t) = \frac{(1-\nu)^{\lambda-1}B(t)}{(1-\nu)^{\lambda-1}B(t) + \nu^{\lambda-1}\left(\sum_{t_i \in C} B(t_i) - B(t)\right)^{\lambda}}, \quad (12)$$

where we have noted that $B(t)^{\lambda} = B(t)$ since $B(t)$ is a binary feature. This membership function is consistent with the above discussion since $\mu(B(t)) = 0$ when $B(t) = 0$, $\mu(B(t)) = 1$ if $B(t) = 1$ and t is the only bold term on the slide, and $\mu(B(t))$ is a decreasing function with respect to bold terms on the slide. That is, if the query term appears in bold in two different contexts with parameters $\sum_{t_i \in C} B(t_i)$ and $\sum_{t_i \in C'} B(t_i)$, then $\mu_B(t) \geq \mu'_B(t)$ if $\sum_{t_i \in C} B(t_i) \leq \sum_{t_i \in C'} B(t_i)$.

Membership functions of $I(t)$ and $U(t)$ are derived in a similar manner.

4.2. Line-level scores

4.2.1. Indentation

Intuitively, as apparent, relevance of a term decreases as its bullet level on the slide increases. We again consider the context of indentation by taking into account the minimum and maximum indentation depths in the slide and presentation through the context parameters $b_C = \max_{t_i \in C} \text{ind}(t_i)$ and $a_C = \min_{t_i \in C} \text{ind}(t_i)$, where t_i is ith term in context unit C. Section 6 reports on the effectiveness of each of these in terms of retrieval performance.

Noting that the indentation score is inversely proportional to indentation depth, (10) is used to obtain the membership function for this feature:

$$\mu_{\text{ind}}(t) = \frac{(1-\nu)^{\lambda-1}\left(b_C - \text{ind}(t)\right)^{\lambda}}{(1-\nu)^{\lambda-1}\left(b_C - \text{ind}(t)\right)^{\lambda} + \nu^{\lambda-1}\left(\text{ind}(t) - a_C\right)^{\lambda}}. \tag{13}$$

In (13), $\mu_{\text{ind}}(t) = 0$ if $\text{ind}(t) = \max_{t_i \in C} \text{ind}(t_i)$ and $\mu_{\text{ind}}(t) = 1$ if $\text{ind}(t) = \min_{t_i \in C} \text{ind}(t_i)$, as required.

4.2.2. Size

In deriving the membership function for the size feature, we note that an increase in font size can be used to indicate relevance of text segments on a slide. Font size, however, is not absolute and its correlation with perceived relevance is context dependent in the sense that term t is deemed relevant if its font size is larger than that of the surrounding text. The membership function, therefore, must consider $sz(t)$ in relation to the rest of the slide contents. This naturally lends itself to the context parameters $b_C = \max_{t_i \in C} sz(t_i)$ and $a_C = \min_{t_i \in C} sz(t_i)$, corresponding to the minimum and maximum font sizes in context unit C. Using these parameters, the following membership function is obtained:

$$\mu_{sz}(t) = \frac{(1-\nu)^{\lambda-1}\left(sz(t) - a_C\right)^{\lambda}}{(1-\nu)^{\lambda-1}\left(sz(t) - a_C\right)^{\lambda} + \nu^{\lambda-1}\left(b_C - sz(t)\right)^{\lambda}}. \tag{14}$$

As expected, $\mu_{sz}(t) = 0$ if $sz(t) = \min_{t_i \in C} sz(t_i)$ and $\mu_{sz}(t) = 1$ for $sz(t) = \max_{t_i \in C} sz(t_i)$.

4.3. Slide-level scores

In traditional text retrieval techniques, the term frequency-inverse document frequency (TD-IDF) weight is used to evaluate the relevance of a document in a collection to a query term. This weight indicates that the relevance of a document is directly proportional to the number of times the query term appears within that document, and inversely proportional to the number of occurrences of the term in the collection. Term frequency is generally normalized by the length of the document to avoid any bias. In the interest of space and for reasons discussed in Section 5, we limit the scope of this work to single-term queries. Consequently, inverse document frequency remains constant for a given query and is ignored.

In an approach analogous to the normalized TD scheme, we define the context of term frequency to be the total number of words in a context unit C. Consequently, $b_C = N_C$ and $a_C = 0$, where N_C denotes the number of terms in C, and the membership function can be written as

$$\mu_{tf}(t) = \frac{(1-\nu)^{\lambda-1} tf(t)^{\lambda}}{(1-\nu)^{\lambda-1} tf(t)^{\lambda} + \nu^{\lambda-1}\left(N_C - tf(t)\right)^{\lambda}}. \tag{15}$$

It can be seen from (15) that $\mu_{tf}(t) = 0$ when $tf(t) = 0$ and $\mu_{tf}(t) = 1$ when $tf(t) = b_C$. At the same time, for two documents with the same query term frequency but different lengths b_C and b'_C, $\mu_{tf}(t) \geq \mu'_{tf}(t)$ if $b_C \leq b'_C$. Lastly, note that this formulation of the membership function is equivalent to the application of (8) to term frequency normalized by the length of the context unit C.

5. RELEVANCE AGGREGATION

The aim of the aggregation process is to combine information from the various features to increase completeness and make a more accurate decision regarding the relevance of each term [30]. This step is referred to as *aggregation* in multicriteria decision making [24]. As previously mentioned, the proposed scheme combines feature scores, obtained in the Section 4, instead of feature values directly. In doing so, two issues must be addressed, namely, the aggregation structure or the order in which the feature scores are combined, and the choice of aggregation operators used to form a single score from multiple feature scores.

To address the first issue, we propose a hierarchical aggregation scheme, shown in Figure 5, to exploit the characteristics specific to each feature granularity. An example of such a characteristic is complementarity of the typeface attributes in the sense that a high score in one of the bold, italic, and underline features is sufficient to indicate a high word-level score. In contrast, the line-level features, size and indentation, are correlated as previously noted. Such feature characteristics are important in the choice of the aggregation operators used to combine the scores. While the scope of the aggregation scheme presented in this section is limited to text-related features on a slide, scores obtained from multimedia objects and their metadata on a given slide can be combined with text-related scores at the slide level.

As previously mentioned, we have limited the scope of this paper to single-word queries. We note here that the well-known standard technique of combining multiple-word queries using the logical connectives AND, OR, and NOT can be used to extend the proposed methodology to multiple-term queries. Since such an extension does not provide any novel contributions, the rest of the manuscript focuses on single-term queries to highlight the novel aspects of this work with respect to the XML-based features and the fuzzy aggregation framework.

Before presenting the details of the proposed aggregation scheme, we briefly discuss relevant examples and properties of aggregation operators. These properties are then used to guide our choices for feature score combination.

5.1. Aggregation operators: overview

An aggregation operator is a mapping $\mathcal{A} : [0,1]^n \rightarrow [0,1]$, where n is the number of elements being combined. The choice of aggregation operators is dependent on the application and the nature of the values to be combined.

The well-known operation of AND and OR in bivariate logic is extended to fuzzy theory to result in two classes of operators known as triangular norms (t-norms) and triangular conorms (t-conorms), respectively [31, 32]. The min operator is an example of a t-norm and the max operator belongs to the class of t-conorms. Further examples of aggregation operators include the various mean operators, ordered weighted averages [33], and Gamma operators [29, 34]. While weighting schemes can be used to indicate the relative relevance of each features, the determination of weights is not trivial and is beyond the scope of this work.

Aggregation operators can be classified with respect to their *attitudes* in aggregating various criteria as conjunctions, means, and disjunctions [30, 31], as discussed below.

5.1.1. Conjunctive operators

An operator $\mathcal{A}(\mu_i, \mu_j)$ is conjunctive if $\mathcal{A}(\mu_i, \mu_j) \leq \min(\mu_i, \mu_j)$. The aggregation result is dominated by the worst feature score, and in this sense, a conjunction provides a pessimistic or severe behavior, requiring the simultaneous satisfaction of all criteria [30]. The family of t-norms is an example of conjunctive operators. Conjunctive operators do not allow for any compensation among the criteria.

5.1.2. Mean operators

An operator $\mathcal{A}(\mu_i, \mu_j)$ is a compromise if $\min(\mu_i, \mu_j) \leq \mathcal{A}(\mu_i, \mu_j) \leq \max(\mu_i, \mu_j)$. Mean-type operators exhibit a compromising behavior, where the aggregation result is a tradeoff between various criteria (feature scores, in this case). In other words, mean operators are compensative in that they allow for the compensation of one low feature score with a high score in another feature. An example of mean operators is the family of quasilinear means, $\mathcal{A}(x, y) = ((x^\alpha + y^\alpha)/2)^{1/\alpha}$ [31]. For $\alpha \rightarrow -\infty$, $\alpha = -1$, $\alpha = 0$, $\alpha = 1$, and $\alpha \rightarrow \infty$, the min operator, harmonic mean, geometric mean, arithmetic mean, and the max operator are obtained. Another example of mean operators is the symmetric sums [31]. Examples of mean operators and symmetric sums are shown in Table 1.

5.1.3. Disjunctive operators

An operator $\mathcal{A}(\mu_i, \mu_j)$ is disjunctive if $\mathcal{A}(\mu_i, \mu_j) \geq \max(\mu_i, \mu_j)$. Consequently, the aggregation of two feature scores results in a score that is at least as high as the highest of the two scores. Disjunctive operators, therefore, exhibit an optimistic or indulgent behavior, requiring satisfaction of at least one goal [30]. T-conorms are examples of disjunctive operators. These operators allow for full compensation among criteria.

An aggregation operator may have a constant characterization as a disjunction, mean, or conjunction for all values of its arguments or express hybrid attitudes depending on the values of its arguments and operator parameters [30, 31]. For example, t-norms always behave as conjunctions whereas symmetric sums act as conjunctions, means, or disjunctions based on the values being combined. The work of [30] provides an ordering of the above aggregation operators. Such an ordering is shown in Figure 7 and provides a guideline for choice of aggregation operators in what follows.

In selecting appropriate aggregation operators for each feature level, we consider mathematical properties of aggregation operators in addition to the aggregation attitude discussed above. Some of the properties of aggregation operators pertinent to the problem of slide retrieval are

Table 1: Example of quasilinear means and symmetric sums. HM: harmonic mean, GM: geometric mean, AM: arithmetic mean.

Quasilinear means	Symmetric sums
$\mathrm{HM}(x, y) = \dfrac{2xy}{x+y}$	$\sigma_0(x, y) = \dfrac{xy}{1-x-y+2xy}$
$\mathrm{GM}(x, y) = \sqrt{xy}$	$\sigma_{\min}(x, y) = \dfrac{\min(x, y)}{1-\lvert x-y\rvert}$
$\mathrm{AM}(x, y) = \dfrac{x+y}{2}$	$\sigma_{\max}(x, y) = \dfrac{\max(x, y)}{1+\lvert x-y\rvert}$

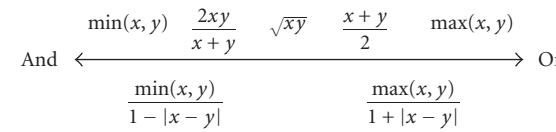


Figure 7: Ordering of aggregation operators adopted from [30].

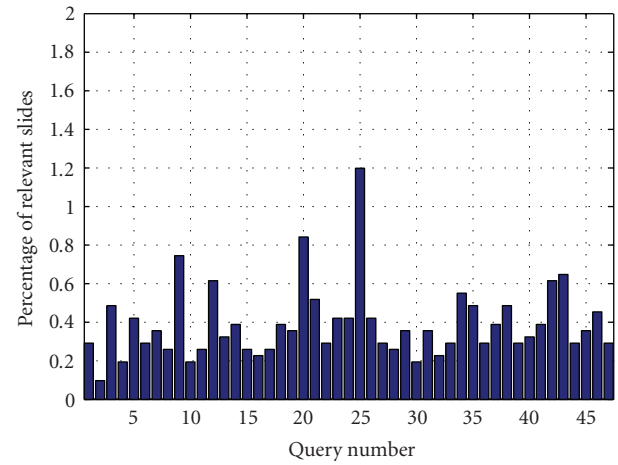
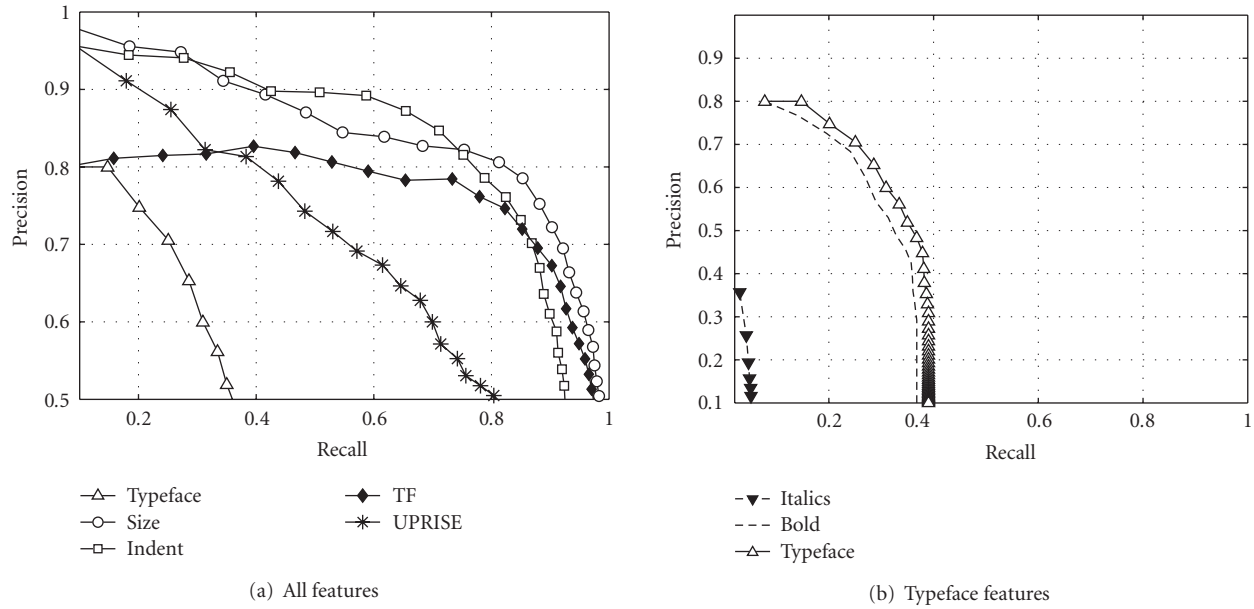


Figure 8: Percentage of the slides relevant to each query with respect to the database size.



(a) All features

(b) Typeface features

Figure 9: Precision-recall curves for the proposed features: size, indentation, typeface, and term frequency.

briefly reviewed below and subsequently used for operator selection. For brevity, the properties are presented for aggregation of two values only, but these can be extended to the general case with n arguments [34].

(i) Continuity: this property requires the operator to be continuous with respect to each of its arguments to ensure that the aggregation does not respond chaotically to small changes in its arguments.

(ii) Monotonicity: mathematically, we require that $\mathcal{A}(a,b) \geq \mathcal{A}(c,d)$ if $a \geq c$ and $b \geq d$. This property is needed to ensure that a slide receives a higher score than any other slide with lower scores in the individual features.

(iii) Commutativity: this property states that $\mathcal{A}(a,b) = \mathcal{A}(b,a)$, ensuring that the ordering of feature scores does not change the result of aggregation.

(iv) Associativity: this property requires that $\mathcal{A}(x, \mathcal{A}(y, z)) = \mathcal{A}(\mathcal{A}(x,y),z)$, ensuring the order in which multiple features are aggregated does not affect the aggregation results.

(v) Neutral element: an operator has a neutral element e if $\exists e \in [0,1]$ such that $\forall a \in [0,1], \mathcal{A}(a,e) = a$. The neutral element does not affect the aggregation result.

(vi) Idempotency: this property states that the aggregation of identical elements results in the same element. That is, $\mathcal{A}(x,x) = x$.

We now proceed to select aggregation operators at each feature level by stating the required properties for combing each set of feature scores.

5.2. Aggregation of word-level scores

The objective of this section is to combine the scores obtained from bold, italic, and underline features to obtain a word-level score $\mu_{\text{word}}(t_{i,j})$, where $t_{i,j}$ corresponds to the ith term on slide s_j. As previously noted, the typeface features are complementary and a high score in either of the bold, italic, or underline features should result in a high word-level score. This observation indicates a need for a disjunctive operator. The operator must also be commutative and associative as the order of combination of the three features should not influence the word-level score. In addition, the operator must be idempotent, as having two typeface features does not increase the relevance of a term. Lastly, the chosen operator must have zero as a neutral element as a score of zero in the typeface features is not an indication of irrelevance but rather of absence of information regarding the relevance of the term [30]. This neutral element requirement indicates the need for a T-conorm. The max operator is the only idempotent choice among the T-conorms [26]. Since the max operator is also associative, it is chosen for combination of the word-level features:

$$\mu_{\text{word}}(t_{i,j}) = \max\left(\mu_B(t_{i,j}), \mu_I(t_{i,j}), \mu_U(t_{i,j})\right). \tag{16}$$

5.3. Aggregation of line-level scores

We now turn the attention to combining the line-level score, size, and indentation to obtain a line-level score $\mu_{\text{line}}(t_{i,j})$ for a slide. As a result of the correlation between the two line-level features, dissonant feature scores are indicative of possible feature unreliability. A possible scenario for obtaining conflicting size and indentation is when a nonbulleted text box is used on a slide. In the absence of a bullet, the indentation level is set to the default value of zero in

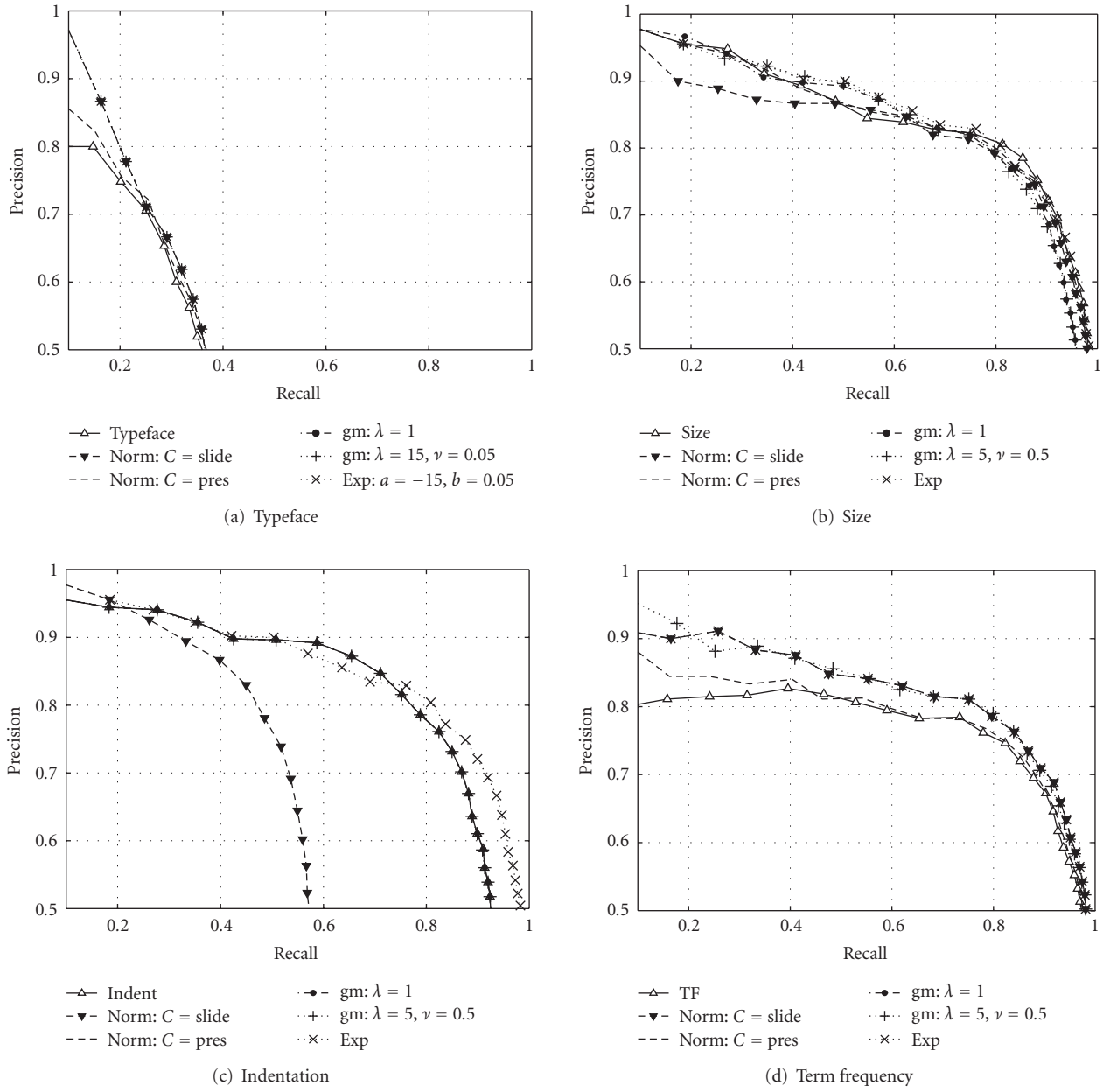


Figure 10: Precision-recall curves for various membership functions and context units.

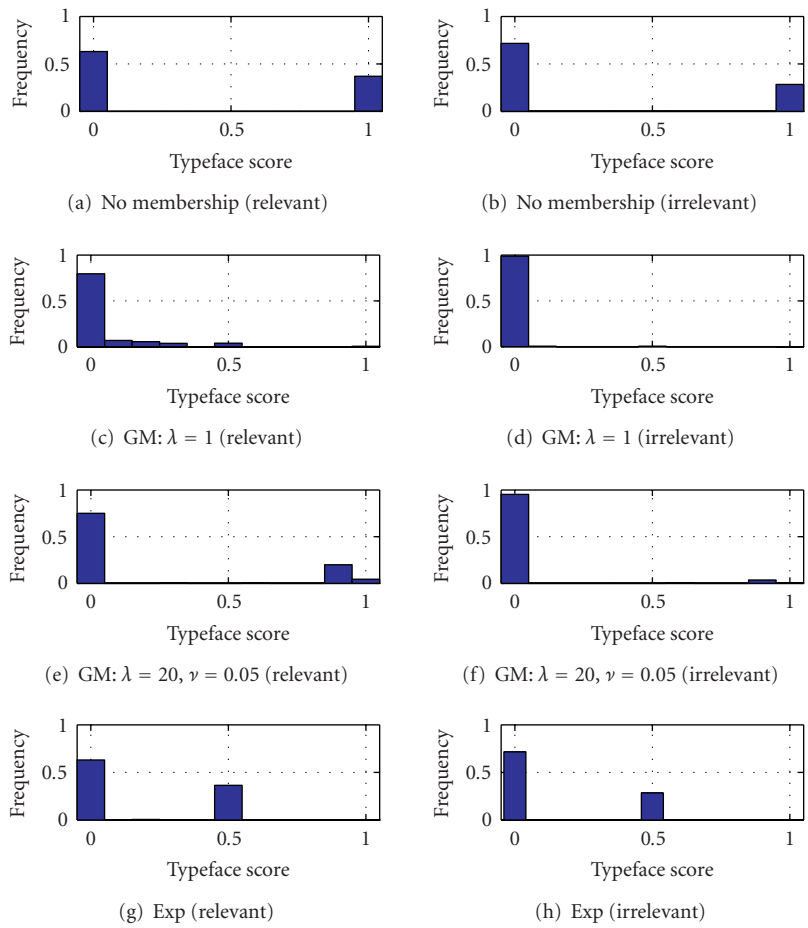
XML representation. In this case, a high-indentation score should be offset by low-size score. While the operator is required to be commutative, associativity is not an issue here since only two values are combined. Lastly, a neutral element is not needed in this case as both feature scores influence the aggregation result. These requirements indicate the need for a mean type or variable behavior operator allowing for some compensation between the criteria, but do not limit the choice among the operators listed in Table 1. We denote the aggregation operator used for combination of line-level features as $\mathcal{A}_{\text{line}}$ and examine the effectiveness of the various means and symmetric sums listed in Table 1 in the experiments of Section 6. The line-level score is then computed as

$$\mu_{\text{line}}(t_{i,j}) = \mathcal{A}_{\text{line}}(\mu_{sz}(t_{i,j}), \mu_{sz}(t_{i,j})), \tag{17}$$

where $\mathcal{A}_{\text{line}}$ denotes an aggregation operator from Table 1.

5.4. Aggregation of slide-level scores

The top-most level of aggregation in the proposed hierarchy is the combination of slide-level scores, where the information obtained from all feature levels is combined to result in an overall score for the given slide. Slide-level combination, however, requires feature scores to be on a slide-level granularity. We must, therefore, transform the word-level and line-level scores into a global slide-level score.



(a) No membership (relevant) (b) No membership (irrelevant)

(c) GM: $\lambda = 1$ (relevant) (d) GM: $\lambda = 1$ (irrelevant)

(e) GM: $\lambda = 20$, $\nu = 0.05$ (relevant) (f) GM: $\lambda = 20$, $\nu = 0.05$ (irrelevant)

(g) Exp (relevant) (h) Exp (irrelevant)

Figure 11: Typeface score distribution using no membership, generalized, and exponential memberships for user-labeled relevant and irrelevant slides.

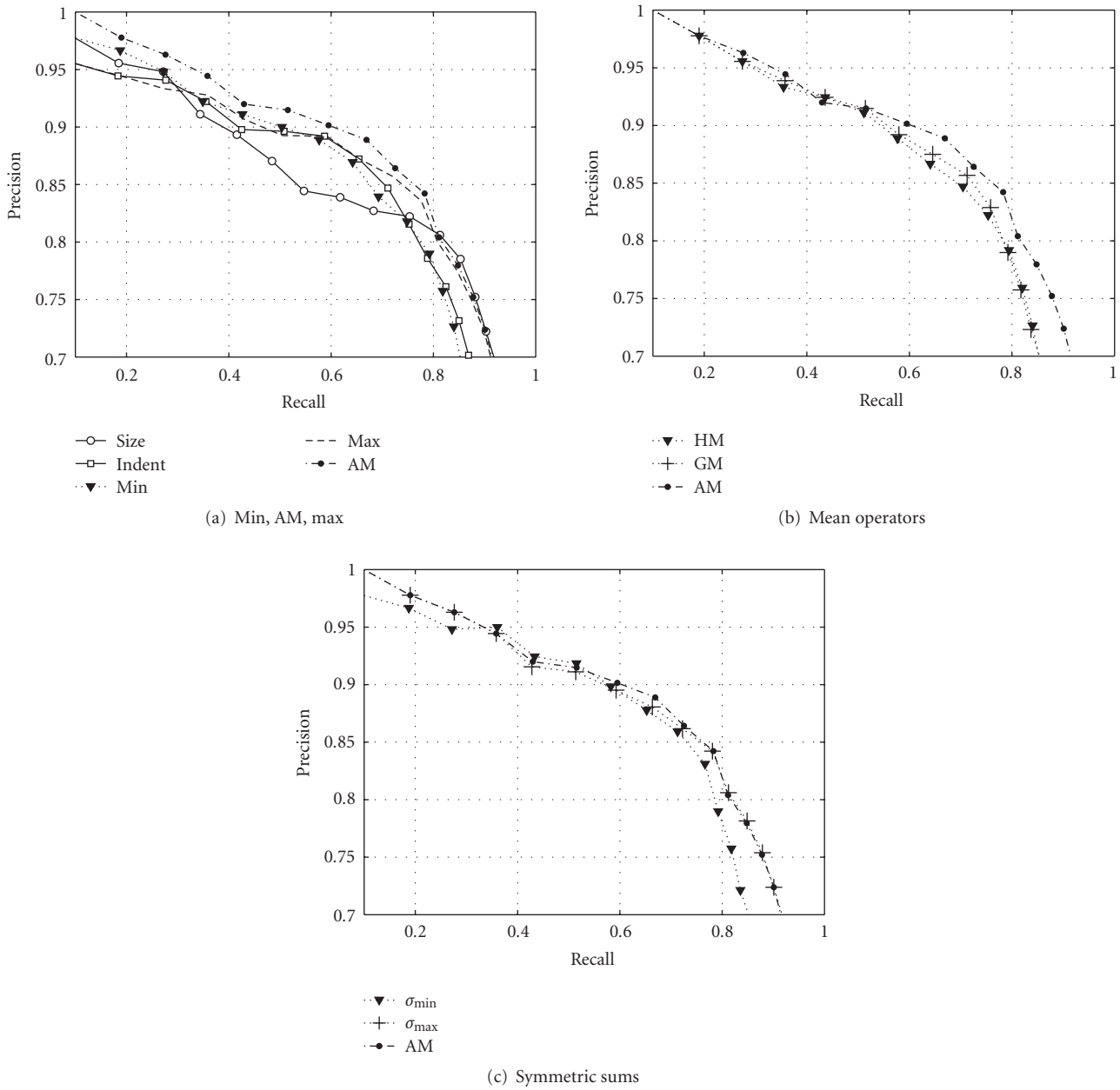
Word-level and line-level scores are local scores in the sense that they report on the feature of particular components of a slide, namely, words and lines. If a query keyword occurs more than once on a slide, each occurrence of the keyword will be associated with a word-level and line-level scores. In order to compute a slide-level score for each feature, the scores over multiple occurrences of the query term must be aggregated. This aggregation naturally lends itself to a disjunctive attitude as the occurrence of a single high-scoring term is sufficient to indicate the relevance of a particular slide. We further require the aggregation operator to be idempotent since the term frequency feature already reinforces the scores of slides with multiple occurrences of a query term. The global word-level and line-level scores for a slide s_j are denoted as $\mu^g_{\text{word}}(s_j)$ and $\mu^g_{\text{line}}(s_j)$ and are computed using the max operator

$$\begin{aligned} \mu^g_{\text{word}}(s_j) &= \max_i \mu_{\text{word}}(t_{i,j}), \\ \mu^g_{\text{line}}(s_j) &= \max_i \mu_{\text{line}}(t_{i,j}). \end{aligned} \tag{18}$$

The last aggregation level combines the slide level scores $\mu^g_{\text{word}}(s_j)$, $\mu^g_{\text{line}}(s_j)$, and $\mu_{tf}(t_{i,j})$ to obtain the score for slide s_j. An important issue to consider is the order of operations. Recall that a low-word-level score is merely an indication of lack information and not of lack of relevance and that the aggregation operator applied to $\mu^g_{\text{word}}(s_j)$ should have zero as its neutral element, leading to the choice of the max operator. Note also that word- and line-level features both report on appearance-based text attributes whereas term frequency reports a purely content-related feature. For this reason, we have chosen to first combine the appearance-based features and then aggregate the result with the term frequency score. The final aggregation operator is expected to have a disjunctive attitude to deal with the missing information in the attributes score. In light of these observations, the final slide score is computed as

$$\mu_{\mathcal{T}}(s_j) = \mathcal{A}_{\text{slide}}\left(\max\left(\mu^g_{\text{word}}(s_j), \mu^g_{\text{line}}(s_j)\right), \mu_{tf}(t_{i,j})\right), \tag{19}$$

where $\mathcal{A}_{\text{slide}}$ denotes the operator used to combine feature scores at the slide level. This above operation is clearly not associative. The experiments of Section 6, however, indicate that the order of operations does not significantly alter the aggregation results. With respect to Figure 1, the slide-level score, $\mu_{\mathcal{T}}(s_j)$, is used to rank the slides in the candidate set.



(a) Min, AM, max

(b) Mean operators

(c) Symmetric sums

FIGURE 12: Precision-recall curves for aggregation of size and indent features using various aggregation operators.

6. RESULTS

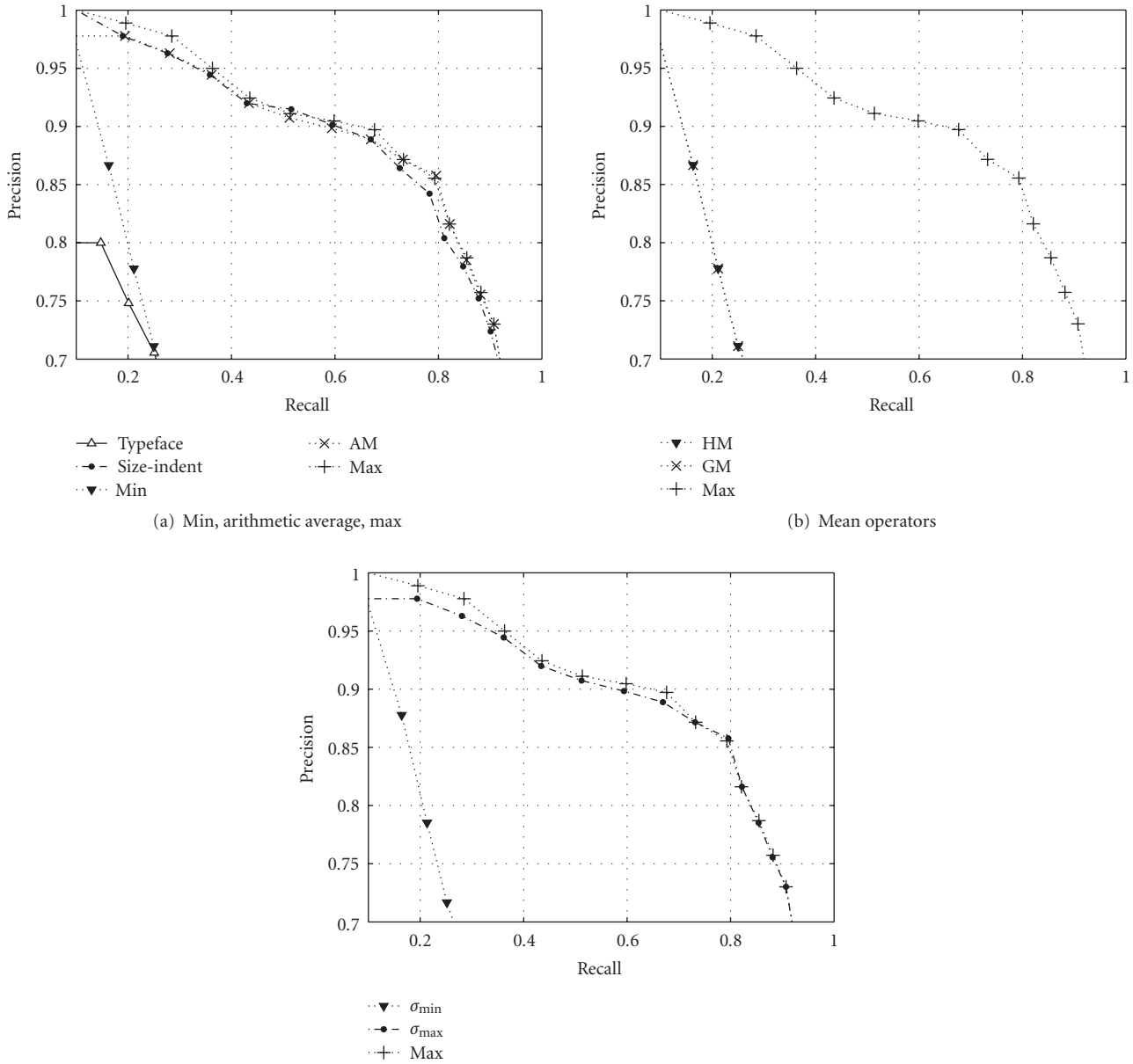
This section evaluates the retrieval effectiveness of the proposed features, membership functions, and aggregation scheme.

6.1. Experiment setup

6.1.1. Dataset

The evaluation dataset includes 142 presentations with a total of 3087 slides. These presentations include lecture material from undergraduate and graduate engineering courses, engineering conference presentations, and other engineering-related material.

A total of 47 single-term query keywords have been manually extracted from the presentation set, corresponding to key concepts in signal processing and pattern recognition courses taught at undergraduate and graduate levels. Examples of query keywords include *convolution*, *Kalman*, *transcoding*, *wavelet*, and *encryption*. For each of the query keywords, the ground truth set is created and corroborated by three users in a manner similar to that of [11]. Percentage of the slides relevant to each queries with respect to the total database size is shown in Figure 8.



(a) Min, arithmetic average, max

(b) Mean operators

(c) Symmetric sums

Figure 13: Precision-recall curves for aggregation of size, indent, and attribute features using various aggregation operators.

6.1.2. Figure of merit

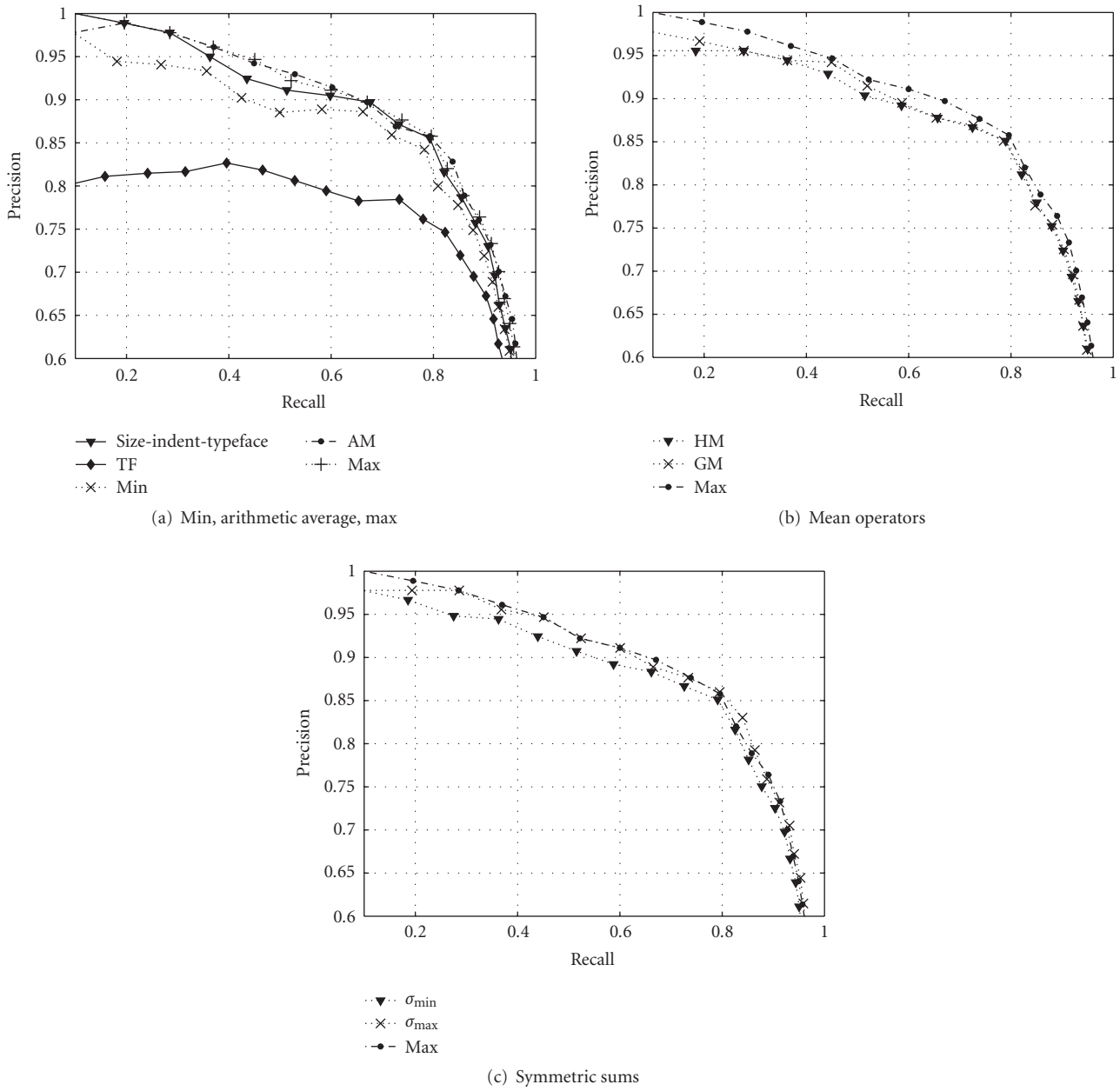
Retrieval performance is measured through precision-recall curves [11, 18]. Precision is defined as the ratio of relevant retrieved slides to the total retrieved slides and is an indication of the efficiency of the retrieval. Recall is the proportion of desired results retrieved within the retrieved set. Mathematically, precision and recall after k slides have been retrieved is defined as

$$\text{Recall}(k) = \frac{RR_k}{N}, \qquad \text{Precision}(k) = \frac{RR_k}{k}, \tag{20}$$

where RR_k is the number of retrieved slides that are part of the ground truth set, and N is the total number of slides in the ground truth set. The results of this section report precision and recall values averaged over the 45 query terms.

6.1.3. Comparison to other methods

The retrieval performance of the proposed method is compared to the UPRISE method [9]. This method incorporates indentation level as well as term frequency to compute slide scores through a geometric mean. This method also proposes the optional inclusion of slide duration as a feature. However, the feature requires access to timing information which is not available in the intended application. For this reason, a value of $\theta = 0$ is used to eliminate the effect of slide duration as suggested in [9].



(a) Min, arithmetic average, max

(b) Mean operators

(c) Symmetric sums

FIGURE 14: Precision-recall curves for aggregation of size, indent, attribute, and term frequency features using various aggregation operators.

6.2. Choice of features

In this section, we report on the effectiveness of the proposed features. To this end, precision-recall plots for the typeface features, size, indentation, and term frequency are shown in Figure 9.

The size, indentation, and term frequency perform reasonably well when compared to UPRISE (note here that UPRISE includes both structural and content features). The aggregated typeface features, however, perform poorly. Figure 9(b) shows the performance of bold and italic features separately. The bold feature outperforms italic while the aggregation of the two features using the max operator improves the retrieval performance. Lastly, note that the underline feature has not been included in these results to the poor performance on the test set. This ineffectiveness of typeface is partially because of the lack of knowledge of relevance of a term in the absence of typeface features, as previously noted.

6.3. Choice of membership functions

The precision-recall plots for three membership functions, namely, simple normalization (5), exponential (7), and generalized membership functions (8), applied to the four-proposed features are depicted in Figure 10 for context units of slide and presentation. For the generalized membership

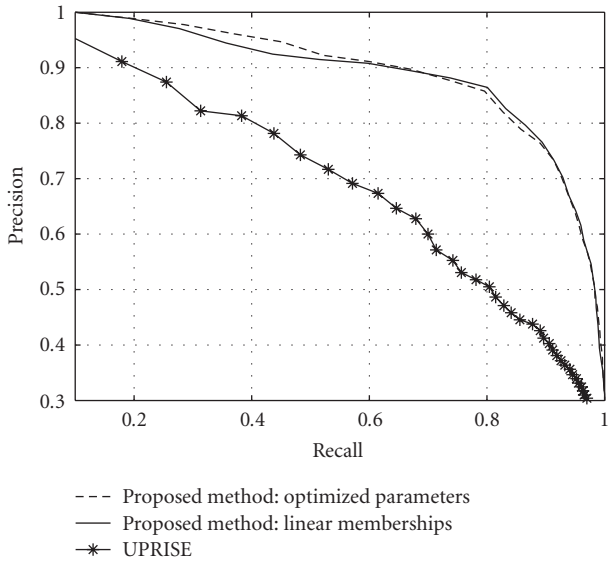


Figure 15: Precision-recall curves for the proposed method and UPRISE.

function, the results are shown for the linear case $\gamma = 1$ as well as for the manually optimized parameter choices. These figures evaluate three issues with respect to precision-recall performance: (1) the effect of addition of context-dependent information, (2) the effect of the form of the membership function (e.g., S-shape versus simple normalization), and (3) the effect of membership function parameters.

The plots of Figure 10 indicate that the addition of context information can improve the precision-recall performance in all four features. The best-performing context unit C, however, varies among the features. For the typeface and term frequency features, a slide is the best performing context unit whereas for indentation and size, a presentation unit provides the best results. Intuitively, this can be attributed to the fact that indentation and size styles generally remain the same over a single presentation. In contrast, typeface and term frequency vary within the same presentation depending on the concepts presented on a given slide.

Figure 10 indicates that while the precision-recall performance is affected by the choice of the context unit C, it is relatively insensitive to the choice of membership functions and their parameters. This is because the precision-recall measure considers the *slide rankings* produced by the scores and not numerical score values. However, since the feature scores generated by the membership function are further aggregated, it is important to consider not just the precision-recall performance, but also the distribution of scores within the interval $[0, 1]$. To illustrate this point, Figure 11 shows the distribution of feature scores when different membership functions are applied. In each case, the score distribution for relevant and irrelevant slides as deemed by the user are shown.

These results show that the generalized membership function produces the best separation between relevant and irrelevant classes. This is important as each feature score is further combined with other scores through the aggregation hierarchy. For example, the choice of the linear version of the generalized membership function results in the maximum score value of 0.5 for relevant slides. In contrast, the maximum line-level score is unity. Thus, the typeface scores inherently receive a lower weight once combined with line-level scores. In light of this observation, membership function parameters should be selected by considering the score distribution as well as retrieval performance.

6.4. Choice of aggregation operators

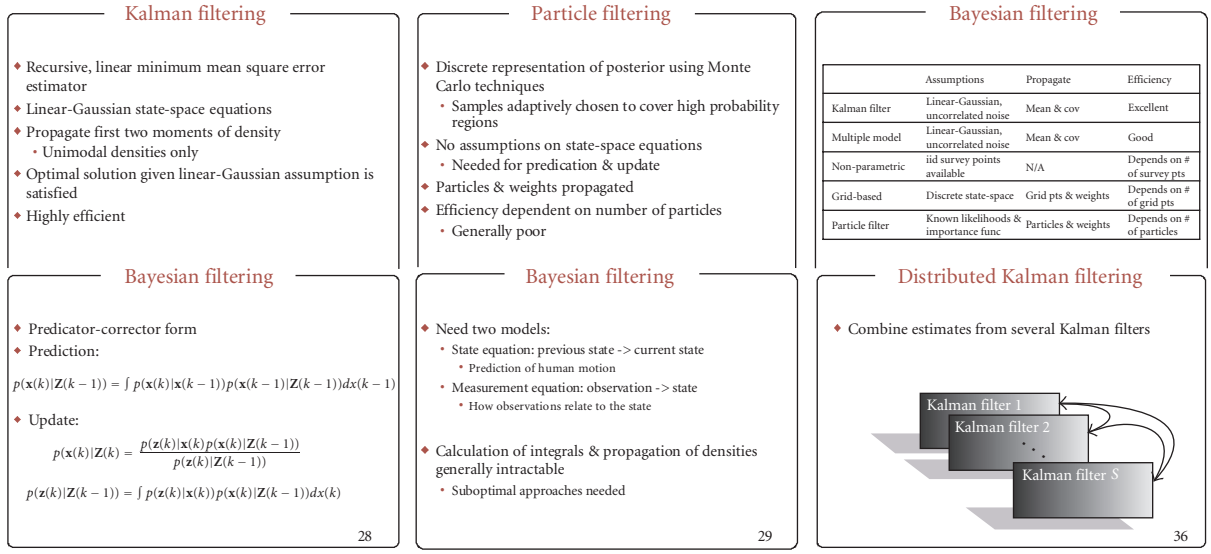
This section examines the effectiveness of the aggregation operators, quantified through precision-recall, in fusing the features scores at various levels of the aggregation hierarchy.

Figure 12 depicts the PR curves obtained from aggregation of size and indentation features, using the generalized membership function with $\lambda = 1$, for various classes of aggregation operators. In particular, the precision-recall plot of Figure 12(a) indicates that a compromise operator outperforms both disjunctive max and conjunctive min operators, as expected. Figures 12(b) and 12(c) show the precision-recall performance of the various mean operators and symmetric sums listed in Table 1. With reference to the ordering of operators shown in Figure 7, compromise operators closer to the middle and right extreme provide the best performance. The variable behavior of symmetric sums does not seem to provide an advantage over the constant behavior quasilinear means in aggregating these correlated features.

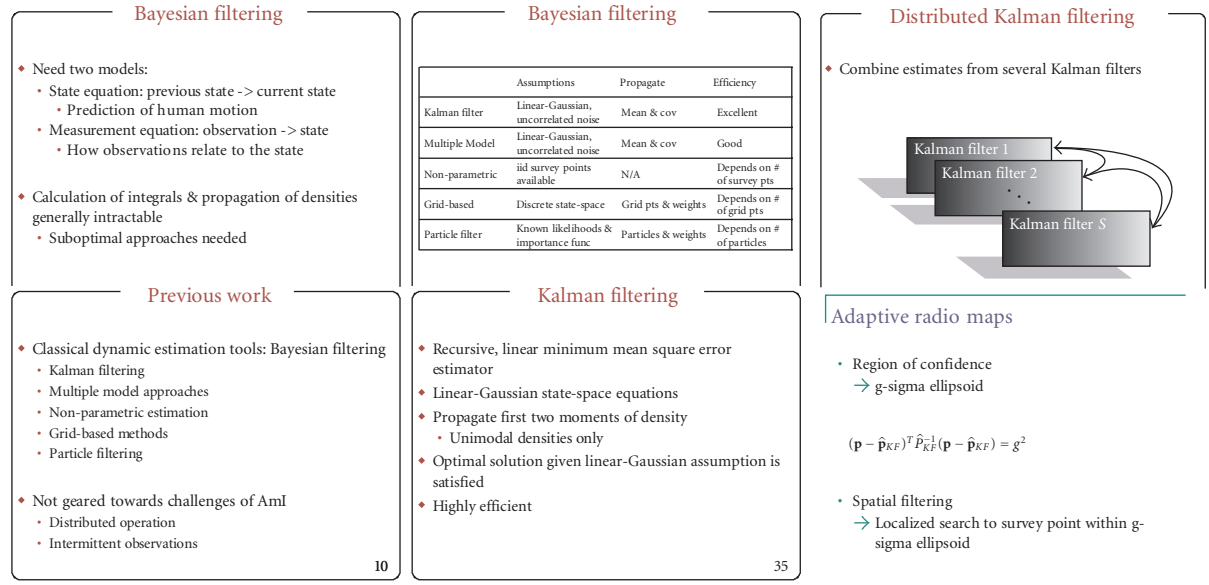
Figure 13 shows the PR curves when the combination of size and indentation scores (using the AM operator) is aggregated with the typeface score. The plots were generated with parameter values of $\lambda = 1$ for size and indentation features and $\lambda = 15$ and $\nu = 0.05$ for the typeface features as discussed previously. As expected, the max operator provides the best performance due to the existence of the neutral element of zero. In fact, as seen in this figure, the min operator and those toward the conjunctive end of the spectrum perform particularly poorly due to their severe behavior emphasizing the cases, where the typeface score is missing.

Figure 14 shows the PR curves when the result of aggregation of size, indentation, and typeface features using the max operator is aggregated with the term frequency score. The parameters used for the membership functions are selected by considering the score distributions for each feature and are $\lambda = 1$ for size and indentation features, $\lambda = 15$ and $\nu = 0.05$ for the typeface features, and $\lambda = 2$ and $\nu = 0.1$ for the term frequency. It is seen from the plots of Figure 14 that the max operator provides the best retrieval results. Such a behavior is expected again because of the existence of the neutral element.

Lastly, Figure 15 compares the precision-recall performance of the proposed method, using all four features, with that of UPRISE [9]. The performance of the proposed method is shown for both manually optimized membership function parameters as well as the case where linear memberships ($\lambda = 1$) are used for all features. It can be seen that



(a) Proposed method



(b) UPRISE

Figure 16: Example of retrieval results for the query term *filtering*: (a) the proposed method, (b) UPRISE.

the simple choice of $\lambda = 1$ does not significantly degrade the retrieval performance. This insensitivity eliminates the concerns of parameter selection for the membership functions. The excellent performance of the proposed scheme can be attributed to the additional features, namely, size and typeface, and to the careful selection of aggregation operators and membership functions.

To illustrate the effectiveness of the system visually, the top 6 retrieval results for the query term *filtering* are shown in Figure 16 for both the proposed method and the UPRISE. The effect of typeface and size features is evident in differences between the two methods in the fourth and sixth retrieval positions.

7. CONCLUSION

The existence of large slide presentation repositories in education and scholarly settings has necessitated the development of effective search and retrieval tools. This paper has examined the unique characteristics of slide presentations, as compared to traditional text and multimedia documents, and has proposed a retrieval tool geared specifically toward such repositories. In particular, the recently standardized XML open file format is used to extract content and contextual features from slides. The traditional term frequency feature used in document retrieval is combined with contextual features, including the appearance-based

attributes such as typeface, font size, and indentation levels, to judge the relevance of each term as intended by the presentation authors. The paper has proposed a feature hierarchy to mirror the naturally nested nature of slides and a hierarchical fuzzy scheme for the combination of scores obtained from each feature. The hierarchical nature of the proposed aggregation scheme allows for identification and future incorporation of features extracted from slide multimedia objects and their related metadata information.

An important avenue for future research is the incorporation of user feedback for the determination of membership function parameters as well as aggregation operators. An interactive design can be used to infer the required aggregation attitude as well as feature weights used during aggregation.

ACKNOWLEDGMENT

This work has been partially supported by the National Research Council of Canada under the Network for Effective Collaboration Technologies through Advanced Research (NECTAR) project.

REFERENCES

[1] D. Hilbert, D. Billsus, and L. Denoue, "Seamless capture and discovery for corporate memory," in *Proceedings of the 15th International World Wide Web Conference (WWW '06)*, pp. 1311–1318, Edinburgh, UK, May 2006.

[2] G. D. Abowd, "Classroom 2000: an experiment with the instrumentation of a living educational environment," *IBM Systems Journal*, vol. 38, no. 4, pp. 508–530, 1999.

[3] W. Hürst, "Indexing, searching, and skimming of multimedia documents containing recorded lectures and live presentations," in *Proceedings of the 11th ACM International Conference on Multimedia (MM '03)*, pp. 450–451, Berkeley, Calif, USA, November 2003.

[4] D. M. Hilbert, M. Cooper, L. Denoue, J. Adcock, and D. Billsus, "Seamless presentation capture, indexing, and management," in *Multimedia Systems and Applications VIII*, vol. 6015 of *Proceedings of SPIE*, Boston, Mass, USA, October 2005.

[5] The World Wide Web Consortium (W3C), "Extensible Markup Language (XML) 1.0 (Fourth Edition)," September 2006, http://www.w3.org/TR/REC-xml.

[6] "Ecma 376," Tech. Rep., Ecma International, Geneve, Switzerland, 2006.

[7] W. Hürst and N. Deutschmann, "Searching in recorded lectures," in *Proceedings of the World Conference on E-Learning in Corporate Government, Healthcare and Higher Education (E-Learn '06)*, pp. 2859–2866, Chesapeake, Va, USA, 2006.

[8] W. Niblack, "SlideFinder: a tool for browsing presentation graphics usingcontent-based retrieval," in *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL '99)*, pp. 114–118, Fort Collins, Colo, USA, June 1999.

[9] H. Yokota, T. Kobayashi, T. Muraki, and S. Naoi, "UPRISE: unified presentation slide retrieval by impression search engine," *IEICE Transactions on Information and Systems*, vol. E87-D, no. 2, pp. 397–406, 2004.

[10] A. Haubold and J. R. Kender, "Augmented segmentation and visualization for presentation videos," in *Proceedings of the 13th Annual ACM International Conference on Multimedia (MM '05)*, pp. 51–60, Singapore, November 2005.

[11] A. Vinciarelli and J.-M. Odobez, "Application of information retrieval technologies to presentation slides," *IEEE Transactions on Multimedia*, vol. 8, no. 5, pp. 981–995, 2006.

[12] D. A. Grossman and O. Frieder, *Information Retrieval: Algorithms and Heuristics*, Springer, Dordrecht, The Netherlands, 2004.

[13] M. Hassler and A. Bouchachia, "Searching XML documents—preliminary work," in *Proceedings of the 4th International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX '05)*, vol. 3977 of *Lecture Notes in Computer Science*, pp. 119–133, Dagstuhl Castle, Germany, November 2006.

[14] N. Fuhr and M. Lalmas, "Introduction to the special issue on INEX," *Information Retrieval*, vol. 8, no. 4, pp. 515–519, 2005.

[15] M. I. M. Azevedo, K. V. R. Paixão, and D. V. C. Pereira, "Processing heterogeneous collections in XML information retrieval," in *Proceedings of the 4th International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX '05)*, vol. 3977 of *Lecture Notes in Computer Science*, pp. 388–397, Dagstuhl Castle, Germany, November 2006.

[16] M. R. Amini, A. Tombros, N. Usunier, and M. Lalmas, "Learning-based summarisation of XML documents," *Information Retrieval*, vol. 10, no. 3, pp. 233–255, 2007.

[17] P. S. Aleksic and A. K. Katsaggelos, "Audio-visual biometrics," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 2025–2044, 2006.

[18] A. Kushki, P. Androutsos, K. N. Plataniotis, and A. N. Venetsanopoulos, "Retrieval of images from artistic repositories using a decision fusion framework," *IEEE Transactions on Image Processing*, vol. 13, no. 3, pp. 277–292, 2004.

[19] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[20] H.-J. Zimmermann and P. Zysno, "Latent connectives in human decision making," *Fuzzy Sets and Systems*, vol. 4, no. 1, pp. 37–51, 1980.

[21] R. Thomopoulos, P. Buche, and O. Haemmerlé, "Fuzzy sets defined on a hierarchical domain," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 10, pp. 1397–1410, 2006.

[22] J. Dombi, "Membership function as an evaluation," *Fuzzy Sets and Systems*, vol. 35, no. 1, pp. 1–21, 1990.

[23] R. E. Bellman and L. A. Zadeh, "Decision-making in a fuzzy environment," *Management Science*, vol. 17, no. 4, pp. B-141–B-164, 1970.

[24] J. Marichal, *Aggregation operators for multicriteria decision aid*, Ph.D. dissertation, University of Liège, Liège, Belgium, 1998.

[25] T. Bilgic and I. B. Turksen, *Measurement of Membership Functions: Theoretical and Empirical Work*, Fundamentals of Fuzzy Sets, Kluwer Academic Publishers, Boston, Mass, USA, 2000.

[26] D. Dubois and H. Prade, Eds., *Fundamentals of Fuzzy Sets*, Kluwer Academic Publishers, Boston, Mass, USA, 2000.

[27] K. N. Plataniotis, D. Androutsos, and A. N. Venetsanopoulos, "Adaptive fuzzy systems for multichannel signal processing," *Proceedings of the IEEE*, vol. 87, no. 9, pp. 1601–1622, 1999.

[28] K. Plataniotis and A. N. Venetsanopoulos, *Color Image Processing and Applications*, Springer, Dordrecht, The Netherlands, 2000.

[29] H.-J. Zimmermann and P. Zysno, "Quantifying vagueness in decision models," *European Journal of Operational Research*, vol. 22, no. 2, pp. 148–158, 1985.

[30] I. Bloch, "Information combination operators for data fusion: a comparative review with classification," *IEEE Transactions on*

Systems, Man and Cybernetics, Part A, vol. 26, no. 1, pp. 52–67, 1996.

[31] D. Dubois and H. Prade, "A review of fuzzy set aggregation connectives," *Information Sciences*, vol. 36, no. 1-2, pp. 85–121, 1985.

[32] M. Mizumoto, "Pictorial representations of fuzzy connectives—part I: cases of t-norms, t-conorms and averaging operators," *Fuzzy Sets and Systems*, vol. 31, no. 2, pp. 217–242, 1989.

[33] R. R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 18, no. 1, pp. 183–190, 1988.

[34] T. Calvo, A. Kolesárová, M. Komorníková, and R. Mesiar, "Aggregation operators: properties, classes and construction methods," in *Aggregation Operators: New Trends and Applications*, pp. 3–104, Physica, Heidelberg, Germany, 2002.