Towards Multimedia Question Answering: From Text to Multimedia QA

Tat-Seng Chua

School of Computing,
National University of Singapore
Table of Content

- From Text-based QA to MM QA
- On Medium Selection for MMQA Answers
- “How-To” MMQA
- Finding Good MM Examples Dynamically
- Other Types of MMQA and Applications
- Summary
Rich Multimedia Contents

• MM contents are growing at exponential rate

• Flickr
  o Receives over 3,000 photos upload every minute, or >4 million per day
  o >4 Billion photos??

• YouTube
  o Statistics of Uploads:
    ▪ > 200,000 per day
    ▪ 10 hours of video every minute
  o Statistics of Downloads:
    ▪ >100 millions per day
    ▪ Over 60% of video watched on Internet
    ▪ Over 20% of Web traffic is for YouTube video
Popular Social Network Sites

- Most so-called text-based sites contain lots of images

- Wikipedia:
  - Has over 1 million topics, most with media contents

- Yahoo!Answers:
  - Over 15% of answers now have links to other sites, mostly for media contents

- Twitter:
  - Many postings have links to other media sources
  - Many startups allow photos/videos to be tweeted

- Facebook:
  - Most popular & biggest photo sharing site in the world
  - 3 billion new photos/month
  - 1 million photos delivered per second
  - However, tagging is sparse and less content-oriented
  - Just imagine when many photos from mobile phones in developing countries get loaded
Information Rich World

- Information growing at exponential rate
  - Large amount of text and multimedia info available
  - Contain answers to almost all questions
  - In many cases, huge data size and content redundancy permit most answers to be found by simple matching??

- Users are bewildered by huge amount of info presented to them
  - Often need to painstakingly browse thru large ranked lists to find answers

- Question-Answering (QA) is evolved to meet user’s precise info need
Text-based QA – Factoid QA

- Has gained popularity following the introduction of QA in TREC in late 1990s

- What can we learn from text community??

- Factoid QA
  - Aim to provide precise fact-based answers
  - “What is the most populous country in Africa?”
  - Has achieved good performance and commercial systems have been deployed

Examples are the Powerset and Quora

Quora: A continually improving collection of questions and answers created, edited, and organized by everyone who uses it.
Text-based QA – Definition QA

- **Definition QA**
  - “What is X?” or “Who is X?”
  - System returns a set of answer sentences that best describe the question topic
  - Equivalent to focused summary

- Did 1933: Who is Vlad the Impaler?
  1 okay 16th century warrior prince
  2 vital Inspiration for Bram Stoker
  3 okay Buried in medieval monastery
  4 vital Impaled opponents on stake
  5 okay Lived in Transylvania (Romania)
  6 okay Fought Turks
  7 okay Called "Prince of Darkness"
  8 okay Possibly related to British royalty
Text-based QA – Others -1

- How-to, Why, Opinion QA
  - “How-to syn Android phone with calendars in Outlook?”
  - “What are people’s opinion on communication on iPhone4?”
  - “Why is iPhone so popular?”
  - Answers require the analysis, synthesis and aggregation of answer candidates from multiple sources
  - They are difficult even in text domain -- in early stage of research

- Current approaches
  - Leverage on crowd sourcing contents available in community QA and forum sites for answers
“How-to” QA on Yahoo!Answers
- YA is a popular online QA service
- Leverage on large QA banks to provide the required answers
- Transform the QA problem to one of finding equivalent questions with readily available answers

“Opinion” QA on Forum Sites
- Retrieve postings on desired topic (e.g., hotel) and aspects (e.g., Price, location ..)
- Mine sentiments of each aspect and organize them into topic hierarchy
- Able to answer questions on opinion on each aspect
Multimedia QA

- Information increasingly available in MM form
  - May use MM contents as supplement to text answers
  - Certain questions are better answered in MM form
  - MM answer is often more direct and intuitive

- Example: “How do I transfer pictures/videos to computer using memory cards?”

- Example: “How do I check engine oil of my car?”

- Aim: to extend text-based QA to MM QA
  - To find multimedia answers from Web-scale media resources such as Flicker and YouTube
  - To tackle a range of QA’s in a common framework
A Text-Oriented Architecture for MMQA
Challenges Facing MMQA

- Appropriate medium to answer the question
  - Many variety of contents -- which is best?
  - Dependent on types of question, answer target, and availability of answers

- Need to analyze visual features
  - Text terms and concepts – for recall
  - Presence of appropriate visual concepts – for precision
  - How to train visual concepts dynamically on the fly?

- Finding precise answers:
  - Depending on QA types- Factoid, How-to, Definition, others..

- Many challenges
Table of Content

- From Text-based QA to MM QA
- On Medium Selection for MMQA Answers
- “How-To” MMQA
- Finding Good MM Examples Dynamically
- Other Types of MMQA and Applications
- Summary
Motivation - 1

- With vast variety of information, one key issue is which information medium is the most appropriate to answer a question.

- For example: “How to tie shoe lace?”

- What about: “What day is the national day of Singapore?”
  - Quantity oriented question → Text is Best?
Motivation - 2

- What about questions like:
  - “What are the colors of the rainbow?”
  - “What are the colors to be mixed to get the color pink?”
  - Is Text sufficient? Or Image, or Text + Image??

- We study the distribution of 5,000 questions:
  - It reveals that around 50% questions are best answered in MM form
Formulation

- Treat media selection as a classification problem

Consider different factors to perform classification:

- Analyze questions (determine askers’ intension)
- Analyze answers (determine best answer medium)
- Analyze media resource (determine availability of answers of certain medium)
Design two-layer hybrid classifier to perform the question string classification

- **Layer 1**: Interrogative-based Classification
- **Layer 2**: Naive Bayes Classification
Layer 1: Interrogative-based Classification

- Use interrogative words (or “wh” words); and
- a set of heuristic phrases mined automatically
- to classify questions into 5 classes

<table>
<thead>
<tr>
<th>category</th>
<th>definition</th>
<th>question phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes/no</td>
<td>the answer is yes or no</td>
<td>be, can, will, have, be there</td>
</tr>
<tr>
<td>choice</td>
<td>the answer is in the question</td>
<td>be, can, will</td>
</tr>
<tr>
<td>quantity</td>
<td>the answer is measure word</td>
<td>how + adj/adv, when</td>
</tr>
<tr>
<td>enumerat</td>
<td>collection of things share some common characters</td>
<td>name, list</td>
</tr>
<tr>
<td>descript</td>
<td>the answer required to illustrate more information</td>
<td>what, where, which, why, how to, who, etc</td>
</tr>
</tbody>
</table>
Question Analysis -3

- Layer 2: Naive Bayes Classification
  - A classifier each for Enumerate type and Description type
  - Features used:
    - N-gram (N=1, 2.)
    - Head words. *(what year did the cold war end ?)*
    - List of class-specific related words (see table2)

  - Result: a 4-D answer type confidence scores
    (text, image, video, text+image)

<table>
<thead>
<tr>
<th>Categories</th>
<th>Class-Specific Related Words List</th>
</tr>
</thead>
<tbody>
<tr>
<td>text</td>
<td>how far, nickname, abbreviation, population, when, date, birthday, how long, how often, number, temperature, scale, duration, rate, synonym, season, period, nationality, etc.</td>
</tr>
<tr>
<td>image</td>
<td>symbol, logo, diagram, architecture, pictures, photos, appearance, image, photograph, table, snapshots, surface, drawing, illustration, show framework, sketch, figure, chart, graph, etc.</td>
</tr>
<tr>
<td>video</td>
<td>how to, how can, film, invented, event, story, steps, ways, stages, first, war, said, volcanoes, earthquake, prime minister, movie, song, kill, music, nuclear, king, president, battle, etc.</td>
</tr>
<tr>
<td>combi</td>
<td>list, who, authors, types, kind, actors, actress, gift, colors, star, product, Nobel Prize, flower, body, plant, singer, birds, dogs, college, wear, birthstone, accessory, currency, poet, etc.</td>
</tr>
</tbody>
</table>
Answer Medium Analysis -1

- Idea: Find answers in Yahoo!Answers, and analyze the answers obtained to determine best answer medium types
  - Employ syntactic tree matching technique to find best question-answer pairs in Y!A (Wang & Chua, SIGIR 2001)
  - Example of Answers for “How to cook Pork Leg?”

```xml
<subject>how to cook fish?</subject>
<content>I want to cook a big fish, but really do not know how to deal with...can you help?</content>
<bestanswer>I can. Fish is delicious. Please refer to http://www.youtube.com/watch?v=5IQ8xhJel8s</bestanswer>
```
Answer Medium Analysis -2

- Another example

A QA pair mined from Y!A corpus with URL embedded in its best answer linking to a video content page.
Answer Medium Analysis -3

- Analyze 3 types of features in answers to determine the best answer medium
  - Presence of Visual Concepts
    • Defined as concept with concrete visual forms that can be represented by an image, such as a panda
    • Derived from online visual dictionary
  - Presence of Verbs
    • useful for judging whether the question can be answered with video contents
  - Presence and types of Hyperlink
    • Hyperlink types reflect the answer medium of the best answer

- Develop a Bayesian classifier to perform classification
  - The result is again a 4-D answer type confidence scores
Even if a question is judged to be appropriately answered by a medium, such medium answer may not be available on the Web, and we may need to turn to other medium type:
Media Resource Analysis -2

- Idea: Analyze potential resources/sites for availability of answers:
  - For text: we analyze Google text search engine
  - For image: we choose Flickr and Google image search engines
  - For video: we analyze YouTube and Google video search engines
  - For the combination of text and image: we choose Wikipedia

- We estimate the retrieval effectiveness based on the fact that, most frequently, a retrieval is good when the top search results are quite coherent
  - We adopt a proposed method to compute the clarity score for top 10 results of the query based on the relative entropy (Y. Zhou, and W. B. Croft. Sigir 2002)
Dataset Constructions

- **Data set:**
  - 2000 questions from TREC QA;
  - 1000 questions from UIUC dataset;
  - 2000 questions from Y!A.

- **Ground Truth established by 3 users thru voting method:**

- **Overall accuracy**
  - Linear combination of 3 features determined using grid search
  - The weights used are {0.5, 0.2 and 0.3}

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question String Classification</td>
<td>68.10%</td>
</tr>
<tr>
<td>Potential Answer Analysis</td>
<td>50.06%</td>
</tr>
<tr>
<td>Resource Limitation Prediction</td>
<td>53.94%</td>
</tr>
<tr>
<td><strong>Linear Fusion of The Three Components</strong></td>
<td><strong>73.90%</strong></td>
</tr>
</tbody>
</table>
Summary

- Question analysis is essential for QA
  - Help infer user’s intension
  - Help determine best medium and hence Web resource to find good answers
  - Help to determine best analysis and retrieval techniques

- Further analysis needed to:
  - Determine question type such as the “how-to”, “opinion”, “analysis” or “definition” type questions
  - Also need to determine if question expecting simple or sequence answers
Table of Content

- From Text-based QA to MM QA
- On Medium Selection for MMQA Answers
- “How-To” MMQA
- Finding Good MM Examples Dynamically
- Other Types of MMQA and Applications
- Summary
“How-To” QA

- How-to QA is difficult even for text
- Leveraging on large volume of archived text-based question-answer pairs in YA to find the answers
  - Turning how-to QA into problem of similar questions search

- Question: "how do I get my camera to put pics. on the computer?"
- Answer: "...connect your digital camera though USB cable..."

- However, A textual answer may be confusing because user may have no idea how to deal with USB cable!
How-To QA – Basic Concept

- **Target:** Use YouTube or other video sites as Reference Resource to present video answers to questions on: “How to …”

- **Challenges:**
  - Need to locate questions posed in different ways
  - Need to recognize visual relatedness of answer candidates to users

- **Approach:**
  - **STAGE 1:** Find similar questions in YA for query expansion → Recall
  - **STAGE 2:** For top n results, perform visual ranking with opinion voting and content redundancy → Precision
Stage 1: Text Processing

SQS: find questions that express similar information needs
SQS and Key Phrase Extraction

Query Generation:

- Find similar questions in YA – expand coverage
  - Need to handle questions posed in diff. forms with grammatical errors
  - Employ syntactic tree matching technique to find similar question-answer pairs in Y!A (Wang & Chua, SIGIR 2001)

- Example query question: what is the reason that my gums bleeding when I brush my teeth?

<table>
<thead>
<tr>
<th>Rank</th>
<th>Matched Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>qid=20080113135256AAX72kx My gums bleed when i brush my teeth, why?</td>
</tr>
<tr>
<td>2</td>
<td>qid=20080221211719AA5PMD0 why do my gums bleed when i brush my teeth??</td>
</tr>
<tr>
<td>3</td>
<td>qid=2008050503109AAKs4Ba why are my gums bleeding when I brush my teeth?</td>
</tr>
<tr>
<td>4</td>
<td>qid=20080102233315AArC6GR What are the reasons my gums could be bleeding?</td>
</tr>
<tr>
<td>5</td>
<td>qid=20080505223640AAPMuYc what is the reason for bleeding in gums?</td>
</tr>
<tr>
<td>6</td>
<td>qid=20080326043133AAx74HO Do your gums bleed when you wash your teeth?</td>
</tr>
<tr>
<td>7</td>
<td>qid=20080330042256AAYGcpJ bleeding gums whenever brush teeth?</td>
</tr>
<tr>
<td>8</td>
<td>qid=20080518173957AAZiYkN whenever i brush my teeth, my gums start to bleed, is</td>
</tr>
</tbody>
</table>
SQS and Key Phrase Extraction -2

- Extract noun phrases from these questions
- Generate multiple queries to YouTube

**Example:**

<table>
<thead>
<tr>
<th>Original Verbose Question</th>
<th>Similar Questions found in YH</th>
<th>Extracted Noun Phases</th>
</tr>
</thead>
<tbody>
<tr>
<td>How do you change to different shutter speeds on your digital camera?</td>
<td>How do you change to different shutter speeds of your digital camera?</td>
<td>different shutter speeds of your digital camera</td>
</tr>
<tr>
<td></td>
<td>Change shutter speed on camera?</td>
<td>change shutter speed, camera</td>
</tr>
<tr>
<td></td>
<td>What digital camera has a faster shutter speed?</td>
<td>Digital camera, faster shutter speed</td>
</tr>
</tbody>
</table>
Stage 2: Visual Processing

Similar Question Search

Key Phase Extraction

YouTube

“How do I get my camera to put pics. on the computer?”

Google Image Dataset

Visual Ranking

Opinion Voting

Content Redundancy

Recall-based Video Search

Precision-based Video Ranking
Visual Concept Annotation

- Build vocabulary tree
  - Acquire training image samples from Web
  - Features: Robust features, like 64-d SURF, color & direction histograms
  - Hierarchical K-means clustering, new level will be decided by empirically threshold
  - More discussions later

- Perform concept annotation as a nearest neighbor problem
Opinion Voting and Fusion Ranking

- Indication of video’s popularity and comments:
  - Videos with positive comments tend to be relevant
  - 3 emotional categories: positive, neutral, and negative
  - Overall opinion score

\[
VT(V_j) = \frac{\text{Post}(V_j) + \varepsilon \text{Neu}(V_j) - \text{Neg}(V_j)}{\text{Post}(V_j) + \text{Neu}(V_j) + \text{Neg}(V_j)}
\]

- Perform NDK detection to assess importance of videos based on video redundancy
- Perform fusion ranking of 3 scores using the Bayesian Classifier
Evaluation

- Textual questions:
  - Assembled a collection of questions posted on YA from Mar - Dec 2008, under the Consumer Electronics Category

- Sentimental voting:
  - Comments from YouTube as training set
  - Classification: Bayesian Network classification with K2 search

<table>
<thead>
<tr>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOW TO upload photos to ipod</td>
</tr>
<tr>
<td>How can I copy files from pc to xbox 360 HDD</td>
</tr>
<tr>
<td>How can I install a game from a memory card on ps3</td>
</tr>
<tr>
<td>How can you turn you digital camera in to a web cam</td>
</tr>
<tr>
<td>How do I get my camera to put pics. on the computer</td>
</tr>
<tr>
<td>How do I sync up my Xbox 360 Media remote to my Xbox 360</td>
</tr>
<tr>
<td>How do I take out the hardrive out of my xbox 360</td>
</tr>
<tr>
<td>How do i connect my PS2 to the web</td>
</tr>
<tr>
<td>How do u put pictures/videos on your computer with a memory card</td>
</tr>
<tr>
<td>How do you clean digital camera lens</td>
</tr>
<tr>
<td>How do you put videos from your camera on to Windows Movie Maker</td>
</tr>
<tr>
<td>How to upload Movies from my camera to computer</td>
</tr>
<tr>
<td>When I connect my ipod to my computer, my itunes doesn't show that it's connected. How do I fix that</td>
</tr>
<tr>
<td>how do i get videos from youtube to my ipod</td>
</tr>
<tr>
<td>how do i put my memory card from my camera into my computerlaptop</td>
</tr>
<tr>
<td>how do i save my pics from my digital camera unto my memory card</td>
</tr>
<tr>
<td>how do turn off the flash of a canon powershot a620 camera</td>
</tr>
<tr>
<td>how do u charge an ipod nano</td>
</tr>
<tr>
<td>how do you change to different shutter speeds on your digital camera</td>
</tr>
<tr>
<td>how do you download songs into your ipod</td>
</tr>
<tr>
<td>how do you put dvd movies into ipods</td>
</tr>
<tr>
<td>how to connect ps3 to internet with a router</td>
</tr>
</tbody>
</table>
Evaluation Results

- Compared with 4 methods:
  1. Manually selecting phrases from the original question and then search by YouTube without further analysis,
  2. Only visual ranking,
  3. Only opinion voting, and
  4. Ranking fusion.

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision@1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Only YouTube</td>
<td>0.41</td>
</tr>
<tr>
<td>(b) Only visual ranking</td>
<td>0.45</td>
</tr>
<tr>
<td>(c) Only opinion voting</td>
<td>0.36</td>
</tr>
<tr>
<td>(d) Ranking fusion</td>
<td>0.54</td>
</tr>
</tbody>
</table>
Evaluation – Sample Outputs

- Performance Comparison between YouTube and Visual Reference Output for Question: "How can you turn your digital camera into a web cam?"

- Achieved good average precision of over 0.6
Table of Content

- From Text-based QA to MM QA
- On Medium Selection for MMQA Answers
- “How-To” MMQA
- Finding Good MM Examples Dynamically
- Other Types of MMQA and Applications
- Summary
Remarks on Mining of Good Visual Samples

- One key limitations of MMQA framework is in mining sufficient video samples to train relevant visual concept detectors on the fly.

- Initial investigation focus on consumer electronics product domain
  - Abundance of product videos can be found on the Web.
Product Detection in Web Videos

- **Challenge**
  - How to find sufficient samples to model different aspects of a product in one model

- **Proposed framework**:
  - **Product Information from Amazon Portal**
  - **Collect Visual Examples from Web**
  - **Filter out unrelated Visual Examples**
  - **Mining the salient visual features (create visual signature file)**
  - **Detect the product in video**
Online stores like Amazon provide good visual examples but limited in number.

Good examples from different aspects.
Collect more Product Samples from Image Search Engine

- Use Amazon images as seed to find more visual examples from Web using media search engine

Result is less accurate
Filter out unrelated Visual Examples from Web

Query images (in BoW style feature)

Inverted index structure for all Web images

Product expansion based on text-based image search engine

Re-ranked based on kNN of query images

Filtered product image after expansion

Amazon Product examples
Mining the Salient Visual Features

L1-sparseness

Correlative - sparseness

Original histogram

L₁ regularized Least Square Optimization (with more 0-bins)

Incorporate product info by adding correlation info

\[
\arg\min_{x_i \in X} \|x_i - y_i\|_2^2 + \lambda_1 \|x_i\|_1
\]

(1)

\[
\arg\min_{x_i \in X} \sum_{i=1}^{n} (\|x_i - y_i\|_2^2 + \lambda_1 \|x_i\|_1^i + \lambda_2 \sum_{u=1}^{n} w_{iu} \|x_i - x_u\|_2^2)
\]

(2)
Product Detection in Web Videos

- Use multi-aspects of product sample images to generate matching product instances on video

Product Images for Canon G9 from different aspects

Product in video is in multi-view
Initial Results

- Average Precision for different products over 43000 video frames (Note: no textual info is used)

- Use textual ranking first follow by visual ranking lead to significant improve in performance
Table of Content

- From Text-based QA to MM QA
- On Medium Selection for MMQA Answers
- “How-To” MMQA
- Finding Good MM Examples Dynamically
- Other Types of MMQA and Applications
- Summary
Challenges:
- Explosive growth of web videos
- Need ability to efficiently browse large lists of similar videos
- Assumptions: important info tends to be repeated in different video entries
- Duplicate == key info??
- Form basis to provide overview or summary of main contents of videos at a glance
Definition QA: System Overview
Subjective evaluation showed that this approach is effective.
Application in Life-log Retrieval

- Life-log Search
  - Many projects on capturing life-logs
  - Few works deal with search of life-log data with incomplete metadata

- Recall of life-log data is non trivial
  - Need combination of metadata and visual analysis
  - Recall (or query) of past events is often vague, like: “some where overseas; sunset, sea, drinks?” – but where & when?
  - Or “when was the last time I met person-X and what did we talk about??”

- Similar to “known item search” in TRECVID with vague descriptions
Table of Content

- From Text-based QA to MM QA
- On Medium Selection for MMQA Answers
- “How-To” MMQA
- Finding Good MM Examples Dynamically
- Other Types of MMQA and Applications
- Summary
Discussions

- **MMQA is an emerging field:**
  - Many problems -- Rich research opportunities and applications

- **Key Research Issues:**
  - Query Analysis – what is the answer class & answer type
  - Multimedia Query Expansion – mine visual concepts and examples for visual re-ranking
  - Domain Specific Knowledge - and Web-scale concept annotation
  - Modeling of verbs (motions) and sequences
  - Efficiency and Scalability - how to do it fast
  - Life-log search and other applications

- **Related activities in TRECVID:**
  - Known-Item Search
  - Instance Search
Thank You
Question & Answer
Towards New Search Paradigm on Web Scale Info Collections -2

- Overall Goal, WebQA = wQA + cQA + vQA + iQA
  - Wikipedia (w), Yahoo!Answers (c), YouTube (v), Flickr (image); and equivalent in Chinese

- Target: Multilingual & Multimedia web-QA
Factoid Multimedia QA

- **TRECVID video retrieval task is an example of factoid QA**
  - Specifically looking for relevant image or video of object
  - For example: *“Find shots of Condoleezza Rice”*

- **Most approaches utilize multi-modal features**
  - Low-level features: color, texture, visual keywords
  - ASR & OCR text
  - High-level visual concepts, e.g. face, car, road..

- **Perform precise shot retrieval in 2 stages:**
  - Shot retrieval based on text metadata & high-level visual features
  - Shot re-ranking by employing multi-modal analysis
Experiments -1

- Accuracy of Question Analysis:

<table>
<thead>
<tr>
<th>Feature Space</th>
<th>Accuracies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unigram</td>
<td>64.76%</td>
</tr>
<tr>
<td>Unigram+Verb</td>
<td>62.98%</td>
</tr>
<tr>
<td>Unigram+Head</td>
<td>67.52%</td>
</tr>
<tr>
<td>Unigram+Related</td>
<td>66.40%</td>
</tr>
<tr>
<td>Unigram+Head+Related</td>
<td>68.10%</td>
</tr>
<tr>
<td>Bigram</td>
<td>67.44%</td>
</tr>
<tr>
<td>Bigram+Verb</td>
<td>65.92%</td>
</tr>
<tr>
<td>Bigram+Head</td>
<td>69.52%</td>
</tr>
<tr>
<td>Bigram+Related</td>
<td>68.50%</td>
</tr>
<tr>
<td><strong>Bigram+Head+Related</strong></td>
<td><strong>69.78%</strong></td>
</tr>
<tr>
<td>Bigram+Verb+Head+Related</td>
<td>69.28%</td>
</tr>
</tbody>
</table>

Observations:
- Using all features lead to better results
- Use of Verbs not effective here
- Use of Head words and class-specific related words are effective
Experiments -2

- Accuracy of Answer Medium Analysis:

<table>
<thead>
<tr>
<th>Feature Space</th>
<th>Accuracies</th>
</tr>
</thead>
<tbody>
<tr>
<td>UnigStopStem</td>
<td>46.12%</td>
</tr>
<tr>
<td>UnigStopStem+Verb</td>
<td>47.27%</td>
</tr>
<tr>
<td>UnigStopStem+Link</td>
<td>47.00%</td>
</tr>
<tr>
<td>UnigStopStem+Visual</td>
<td>46.39%</td>
</tr>
<tr>
<td>UnigStopStem+Verb+Link+Visual</td>
<td><strong>50.06%</strong></td>
</tr>
<tr>
<td>BigStopStem</td>
<td>44.21%</td>
</tr>
<tr>
<td>BigStopStem+Verb</td>
<td>46.07%</td>
</tr>
<tr>
<td>BigStopStem+Link</td>
<td>45.69%</td>
</tr>
<tr>
<td>BigStopStem+Visual</td>
<td>44.94%</td>
</tr>
<tr>
<td>BigStopStem+Verb+Link+Visual</td>
<td>48.77%</td>
</tr>
</tbody>
</table>

Observations:
- Using all features lead to better results
- Use of Verbs is very effective
- Combination of Visual Concepts, Verbs and Link are essential
Experiments -3

- Overall accuracy
  - Linear combination of three features determined using grid search
  - The weights used are {0.5, 0.2 and 0.3}

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question String Classification</td>
<td>68.10%</td>
</tr>
<tr>
<td>Potential Answer Analysis</td>
<td>50.06%</td>
</tr>
<tr>
<td>Resource Limitation Prediction</td>
<td>53.94%</td>
</tr>
<tr>
<td>Linear Fusion of The Three Components</td>
<td>73.90%</td>
</tr>
</tbody>
</table>