

Words and Pictures in the News

Jaety Edwards

UC Berkeley

jaety@cs.berkeley.edu

Ryan White

UC Berkeley

ryanw@cs.berkeley.edu

David Forsyth

UC Berkeley

daf@cs.berkeley.edu

Abstract

We discuss the properties of a collection of news photos and captions, collected from the Associated Press and Reuters. Captions have a vocabulary dominated by proper names. We have implemented various text clustering algorithms to organize these items by topic, as well as an iconic matcher that identifies articles that share a picture. We have found that the special structure of captions allows us to extract some names of people actually portrayed in the image quite reliably, using a simple syntactic analysis. We have been able to build a directory of face images of individuals from this collection.

1 Introduction

For the past year we have been building a collection of captioned news photos and illustrated news articles. We believe that, for many applications, words and pictures together provide very rich information on document content. Photographs can link articles in ways that pure textual analysis may overlook or underestimate, and text provides high level descriptions of image contents that current vision techniques cannot obtain.

Our analysis of image captions has revealed various journalistic conventions that we believe make this dataset particularly appealing for a number of applications. Captions act as concise summaries of events, and we believe we can use them to isolate the most pertinent words for different topics. We have implemented various clustering methods to explore this idea.

Captions are also tightly tied to the actual content of the image. The difficulty here on the text side, is in identifying those portions of a caption that refer to tangible objects, physically present in the image. On the image side,

we need to isolate objects of interest and solve the correspondence problem between caption extracts and image regions. We are exploring these issues in the context of trying to build an automated celebrity directory and face classifier.

2 The Collection

We have been collecting news photos with associated captions since Feb. 2002. These photographs come from Reuters and the AP, and we collect between 800 and 1500 per day. To date, we have collected more than 410,000 unique images. The vast majority of these photos are of people, and the majority of that subset focus on one or two individuals. Since November, we have also collected more than 75,000 illustrated news articles from the BBC and CNN. Both CNN and the BBC, along with the majority of all news outlets, make heavy use of news photographs from the AP and Reuters. In fact, the field is dominated by just three agencies, Reuters, the Associated Press and AFP.

Journalistic writing guidelines emphasize clarity. This is certainly necessary in caption writing, where an author has an average of 55 words to convey all salient details about an image. A caption writer's first responsibility is to identify those portrayed. A caption will contain the full names of those represented if they are known. Second, the writer must describe the activity portrayed in the photo. Finally, with whatever space is left, the author may provide some broader context.

Our collection's vocabulary reflects this ordering of priorities. First, there is a heavy bias towards proper names, and thus the number of unique terms in this set is very large (See figure 1). The vocabulary of other word classes, however, reflect the journalistic emphasis on simplicity. People *hug* rather than *embrace*, *talk* rather than *converse*.

Captions are also easy for humans to parse because writers make use of stylized sentence structures. Just

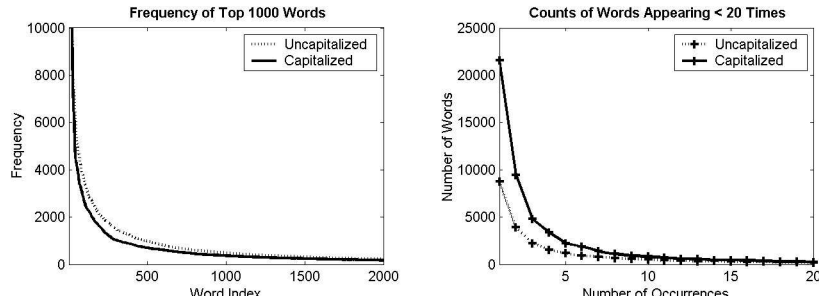


Figure 1: *Caption vocabularies exhibit statistical properties that distinguish them from other collections. A heavy preponderance of proper names creates a very large vocabulary of capitalized words whose frequency distribution has heavy tails. On the other hand, other word classes are somewhat restricted. In a six month period from July to December, 2002, the corpus had 93,457 distinct terms. Of these, 60,182 were capitalized and 32,059 uncapitalized (with 1216 numerical entries). About 1000 terms occur in both forms. Almost a third of all vocabulary terms occur only once. On the **left** we plot term occurrence ordered by frequency for the 1000 most frequent capitalized (solid) and uncapitalized (dotted) terms. On the **right** we analyze the heavy tail, flipping the axes to show the number of words that occur between one and twenty times. Here, capitalized (solid line) words outnumber uncapitalized (dotted line) ones 2-1. We have used capitalization as a proxy for proper names. This obviously misrepresents initial words of sentences, but these only represent on average 2 words out of every 50-60 word caption. The richness of our vocabulary is thus largely due to proper names.*

over 50% of the captions in our collection, for instance, begin with a starting noun phrase that describes the central figure of the photograph followed by a present tense verb that describes the action they are performing (or if in the passive voice, having performed upon them). Textual cues such as “left”, “right” or a jersey number help to clarify any potential correspondence problems between multiple names in the caption and people in the image.

The news photos themselves, like their captions, are also quite stylized with respect to choice of subject, placement within the image, and photographic techniques. A single individual will generally be centered in the photo. Two people will be equally offset to the left and right of center. A basketball player will most likely have the ball in his hands. Each of these photographs will employ a narrow depth of field to blur the background and emphasize their subjects. Like the caption writer, the photo journalist must convey a great deal of information in a small amount of space. These conventions amount to a shared language between photographer and reader that places this single image in a far richer context. We present example captions and their associated images in figure 2.

The textual and photographic conventions we have illustrated are simply trends we have noticed in our dataset, ones that many individual captions and images in our collection break. One of the benefits of scale, however, is that we can throw away a great deal of data and still have meaningful sets with which to work. These simpler sets in turn may act as foundations from which to attack the exceptions.

3 Linking Articles

We have linked and clustered articles in a variety of ways. We have looked at how the specialized vocabulary and statistics of our dataset affect the performance of two different textual clustering methods. We have also examined how images might help to link together articles whose relationships purely text based clustering overlooks or undervalues. We have also used the clusterings built from our captions to provide a topic structure for navigating the news in general, and investigated various interfaces to make this navigation more natural.

3.1 Iconic Matching

Sometimes the same photograph is used to illustrate different articles. This establishes links between articles that might be difficult to discern solely from the text. To establish these links, we built an iconic matcher to find copies of the same image in a database, even if that image has been moderately cropped or given a border.

Images can change in many subtle ways between their distribution and their actual publication. They may be cropped, watermarked, and acquire compression artifacts. The news agency may overlay text or add borders. An iconic matcher needs to be robust to these changes. In addition, given a collection of this size, only minimal image processing is practical.

The iconic matcher we have designed first resizes images to 128x128 pixels in order to accommodate different aspect ratios. It then applies a Haar wavelet transformation separately to each color channel of the image. (Ja-



U.S. President George W. Bush (news - web sites) *ponders* a question at a news conference following his meeting with Czech President Vaclav Havel at Prague Castle, Wednesday, Nov. 20, 2002. Bush urged the NATO (news - web sites) allies to join a U.S.-led 'coalition of the willing' to ensure Iraq disarms. (AP Photo/CTK/ Michal Dolezal) Nov 20 5:52 AM



Sacramento Monarchs' **Ruthie Bolton**, *center*, *grabs* a loose ball from Charlotte Sting's Erin Buescher, bottom, as Monarchs' Andrea Nagy moves in during the first half Friday, July 19, 2002, in Charlotte, N.C. (AP Photo/Rick Havner) Jul 19 8:24 PM

Figure 2: News captions follow many conventions. Individuals are identified with their full names whenever possible. The first sentence describes the actors, and what they are doing. By looking for proper names immediately followed by present tense verbs, we can reliably find names that reference people actually in the image. This is important, as captions are also full of names (i.e. Vaclav Havel in caption 1) that do not actually appear. Often, the captions also help place the person in the frame (i.e. "center" in the second image).

cobs et al., 1995) The first coefficient in each channel, then, represents the average image intensity; the second is the deviation from the mean for the right and left side; the third splits the left hand side similarly, and so on. Rotating through the channels, we use the first 64 largest coefficients as a signature for the image. We say a pair of coefficients match if both are within a given threshold of each other. Images match if their first three coefficients (average color) all match, and if the number of additional matching coefficients is above a certain count. Identical images will of course receive a perfect score using this measure.

To tune the parameters of the matcher, we found sample photos to which CNN or the BBC had added borders and the corresponding borderless originals from AP and Reuters. We then examined how these changes affected the coefficients, and we set the matching thresholds to respond positively even with these changes. This led to fairly broad ranges, ± 5 for the average color coefficients and ± 3 for the rest (out of 0-255 scale). However, we also insisted that at least 42 coefficients actually match.

A visual scan of the sets returned by our iconic matcher shows very few false positives. The matcher frequently accommodates borders or croppings that change up to 10% of the pixels in the original image. Its response to these changes, however, is somewhat unpredictable. Two crops of similar appearance can have very different effects upon the Haar parameters if one pushes many features into different quadrants of the image than the other.

As Table 1 indicates, a significant percentage (10%) of those articles we have collected from the BBC and CNN share an image with another article. This phenomenon seems to stem from three main sources. Across collections, **different authors** may use the same image to illustrate similar stories. Authors may also use the same im-

Table 1: Iconic Matches

Iconic Matches		Collection Stats		
Collection	Total Docs		BBC	CNN
BBC	16990	BBC	3793	223
CNN	8397	CNN		1233

A shared picture is a very good indication that two articles are in some way topically related. Our iconic matcher establishes those links. On the **left**, the figure gives the total number of articles we collected in a three month (Oct-Dec, 2002) period from the BBC and CNN collection, and on the **right**, how often the same image was used for multiple stories. We have split these iconic matches out into inter-agency and intra-agency totals. More than 10 % of the articles in our collection share an image.

age over time to provide **temporal continuity** as the underlying story changes. Finally, the same image may be used to indicate a **broad theme**, while the articles themselves discuss quite different topics. A series of articles on an oil spill of the coast of Spain, for example, moved from simply reporting the incident, to investigating the captain, to discussing the environmental impact, always using an illustration of an oil drenched cormorant.

3.2 Text Clustering

Our second tool for grouping articles comes from automated clustering of the AP and Reuters caption texts. We implemented two different clustering algorithms. Each generates a probability distribution over K topics for each document. Documents are then assigned to the maximum likelihood cluster.

We fit two models to the data. The first was a simple

unigram mixture model. This model posits that caption words are generated by sampling from a multinomial distribution over the corpus vocabulary V . Topics are separate distributions over V . For a given number of topics, we have $|K| \times |V|$ parameters, where K is the set of topics. The probability of a caption in this model is

$$\sum_{k \in K} \left(p(k) \prod_{w \in V} p(w|k) \right) \quad (1)$$

We can fit this model using EM, with the assignments of captions to topics as our hidden variables. In our implementation, we initialize $p(k)$ and $p(w|k)$ with random distributions.

Effectively, we treat each caption as an unordered set of keywords. Although this is a simple language model, we expected it to fit the captions fairly well. Given their extremely short length, we believed captions to have a far higher percentage of topically important (and thus discriminative) words than one would find in a more generic article collection. In longer documents researchers have investigated modelling documents as mixtures of topics (Blei et al., 2001), but we believed captions truly were narrowly focused around a single topic.

Still, our original vocabulary contained over 90,000 terms. To fit this model, we trimmed the tail end of the vocabulary. We applied two heuristics. removing all words that happened less than 200 times. This seems a drastic reduction, but as figure 1 illustrates, the tail of our vocabulary is primarily proper names. Moreover, we collect more than 1000 captions a day. A single word, especially a name, could therefore easily occur 200 times in a single day. Since we are interested in topics of larger temporal extent than a day, this reduction seems at least somewhat justified. We also removed all words of three characters or less. During fitting, we normalized the document word count vectors to a constant length.

Unfortunately, the model was still overwhelmed by common words, and the maximum likelihood configuration invariably driven to make every topic equally likely and every topic distribution almost exactly the same. We therefore were forced to additionally remove very common words, namely the 2000 most frequent words from the web at large.¹ This heuristic almost certainly removes some words strongly associated with specific topics (e.g. “ball” with sports). Still, the remaining middle frequency words do a good job of separating captions into qualitatively good topics, Our contention is that this middle frequency is closely aligned with the true statistics of the entire corpus.

Our second algorithm, which we call a “two-level” mixture model, attempts to deal with very common words

¹Berkeley Digital Library “Web Term Document Frequency” <http://elible.berkeley.edu/docfreq/index.html>

in a more principled way. The top level is a single multinomial distribution γ shared by all captions. The second level has K topic distributions, equivalent to the simple unigram model. For each caption word, we sample a Bernoulli random variable λ to decide whether to draw from the top-level or second-level distribution. This model shunts common shared words into the top-level “junk” distribution, leaving the topic distributions to reflect truly distinctive words. The probability of a document in this model is

$$\sum_{k \in K} \left(\prod_{w \in V} (\lambda p(w|k)p(k) + (1 - \lambda)p(w|\gamma)) \right) \quad (2)$$

We used EM once again to estimate these parameters. Regardless of starting position or $|K|$, λ consistently converges to just over .5 (.53 with std .004).

3.3 Quantitative Analysis

We fit each of these models 10 different times each for $K = 10, 20, \dots, 100$ where K is the number of topics. To compare the quality of fit between different values of K , we held out 10% of the captions and compared the negative log likelihood of the held out data for each run. A model which assigns the highest probability to the test set will have the lowest log likelihood. In figure 3, we plot the average negative log likelihood for each of these runs. Across all K , EM converged in just a few iterations.

3.4 Evaluation and Interfaces

Quantitatively our models appear to fit well, but an analysis of their usefulness for interacting with these collections is more difficult. Our collection has no canonical topical structure against which we can compare our results. However, we have built various tools to aid in the qualitative evaluation of our results.

3.4.1 Temporal Structure

The clustering algorithms discussed in the previous section take no account of the time at which an article or caption was published. News topics, however, certainly have temporal structure. Some, like baseball, happen during a specific season. Others, like an election day, are events that are heavily discussed for a brief period of time and then fade. Our first interface illustrates these temporal relationships. We plot each topic over time. Each timestep is approximately a day.² Each element in the plot of a cluster represents the percentage of captions in that cluster collected during each time period. We have constructed a web interface that lays out all K topics as columns in a matrix, moving through time from top to

²A timestep actually represents 1000 chronologically contiguous captions, but we collect an average of 1000 captions a day.

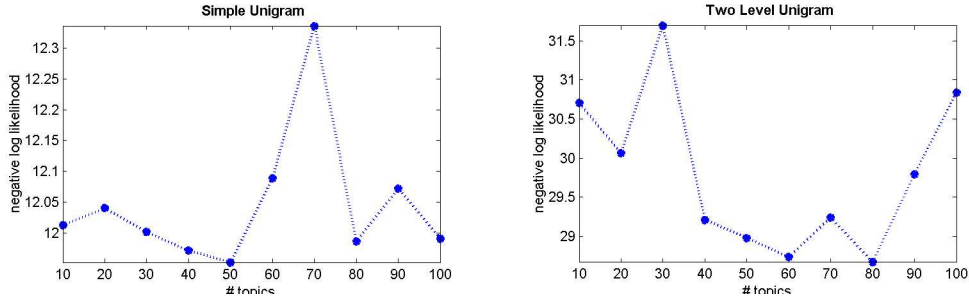


Figure 3: We fit two clustering models. The first was a simple unigram mixture model with K topics. The second “Two Level” model extended the first with a global distribution such that words could choose to come either from the topic distribution associated with that document or from the global one. This figure plots the average and minimum negative log likelihood attained for values of $K = 10, 20, 30, \dots, 100$. The simple unigram model uses a vocabulary filtered of common words, 2000 terms smaller than that used by the “two level” model. This accounts for the higher log likelihood numbers on the right. When fit with the full vocabulary, the simple unigram model invariably creates K identical topics, losing all topic structure to the noise generated from common words.

bottom. One may click on any individual element to bring up a list of captions specific to that period and cluster, as well as the word distribution that characterizes this topic. (Figure 4) The example in this figure utilizes the simple unigram model. One striking aspect of this view is that topics clearly do appear to have time signatures. Some are periodic, others ramp up, others are extremely peaked. We are contemplating methods to integrate these temporal features into our clustering methods.

3.4.2 2D Embedding

Our second interface attempts to lay out topics on the plane in such a way that distances between them are semantically meaningful. First, we use a symmetrized KL divergence to define a distance metric between topic distribution pairs. We define a symmetric KL divergence between two topics distributions t_i and t_j as

$$KL_{symmetric} = \frac{1}{2}(KL(t_i||t_j) + KL(t_j||t_i)) \quad (3)$$

$$KL(t_i||t_j) = \sum_{w \in V} ((p(w|t_i) \log(\frac{p(w|t_i)}{p(w|t_j)}))) \quad (4)$$

where V is the corpus vocabulary.

We then use Principal Coordinate Analysis to project into the plane in a way that in various senses “best” preserves distances. We finally calculate the likelihood for all captions in a given cluster and illustrate the topic with the image associated with the maximum likelihood caption.

In our interface, (Figure 5) one may click on any token to bring up a list of all images and articles associated with this topic. In this example, we have actually used the topic structure generated with captions to organize the BBC and CNN combined dataset. The clusters in

this figure were built using the simple unigram clustering model.

One aspect of Principal Coordinate Analysis that is undesirable in our case is that it tends to emphasize accurately representing large distances at the expense of small ones. In topic space, distances between topics only seem to have semantic meaning up to some threshold. Beyond this threshold they are simply unrelated. We are actively working on implementing a modified version of Principal Coordinate Analysis that will lend more weight to smaller distances for this interface.

4 A Celebrity Face Detector

Our second area of investigation with this dataset is automated methods for establishing correspondences between textual words or phrases and image regions. To this end, we are investigating the creation of an automated celebrity classifier from the AP and Reuters photographs.

Brown et al. (P. Brown and Mercer, 1993) have effectively used co-occurrence statistics to build bilingual translation models. Duygulu, Barnard and Forsyth (Duygulu et al., 2002) have effectively used these machine translation algorithms to establish correspondences between keywords and types of image regions in the Corel image collection.

Unlike Corel, our photos are primarily of people, and so it seemed natural to focus first on proper names as opposed to more general noun classes. Proper nouns are relatively easy to extract from our captions. Caption writers are quite consistent in how they record individual’s names and titles. Simply selecting strings of capitalized words with just a few heuristics to accommodate the beginnings of sentences, middle initials, etc. performs well. Common names like the President’s will be written in multiple

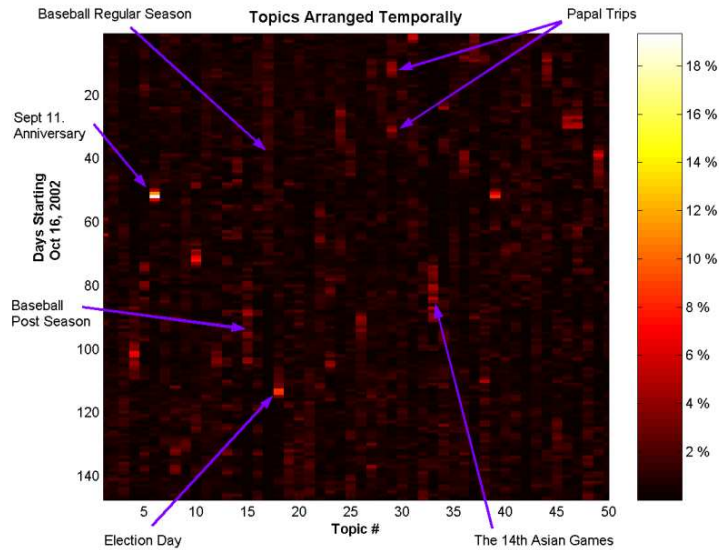


Figure 4: In this figure we see a clustering of 50 topics laid out temporally. The captions are clustered without respect to time using EM and the simple unigram language model. In this example we have used 50 clusters. Each column represents a single topic, and each row is approximately a one day time slice starting July 19, 2002 at top and ending on Dec 8, 2002 at the bottom. The brightness of the entry reflects the percentage of all captions in this topic that occurred during this time slice. To illustrate, we have labeled certain portions of the figure with the realworld topics or events with which certain topics/time periods seem most associated. Topics appear to have time signatures. Some, like football or championship baseball, are periodic. Others, like election day, slowly build to a peak and then rapidly fade. Other, unexpected events, such as the arrest of the D.C. area snipers and Moscow hostage situation ramp up suddenly and then slowly fade over time. We are investigating adding this temporal information into our clustering models.

ways, but even in these cases a single form is overwhelmingly predominant. As for the images, face detectors are a relatively mature piece of vision technology and we can reliably extract a large set of face cutouts.

We could not directly apply the co-occurrence methods used in previous work. First, our captions are full of proper nouns (institutions, locations, and other people) that have no visual counterpart in the image. Duygulu et al. faced a similar problem for certain keywords in the Corel dataset, but to a far smaller degree. They could treat it as a noise problem. Here, it overwhelms the actual signal.

The image side compounds our problem. Duygulu et al. were able to cluster image regions into equivalence classes based on a set of extracted features. We have no similarity metric for faces. Typically, one induces a metric by fitting a parametric model to the data and exploring differences in parameter space. Parametric models perform poorly on faces, however, due to the variability of facial expressions.

If one were to manually label our collection with names and locations of the individuals portrayed in each photo, we believe we could generate hundreds or thou-

sands of unique face images for many individuals. Given this supervised dataset, it might be possible to generate a non-parametric model of faces, fitting a set of models for each expression. Manually annotating half a million photographs is impractical. Instead we have leveraged the special linguistic structure of our captions to create a “photographic entity” detector. In other words, a proper name finder optimized to return only those names that actually occur in the photograph.

4.1 Who Is In The Picture?

We are confronted with the problem of trying to identify those proper names that actually identify people in the photo. A small amount of linguistic analysis has been extremely helpful. We tried various proper name detectors, but it proved difficult to return only people, let alone people who actually appear in the images. Once again, however, we can exploit journalistic conventions. With few exceptions, the first sentence of a caption describes the activity and participants in the picture itself. In 51% of our captions, the beginning of the first sentence follows the pattern [Noun Phrase][Present Tense Action Verb], or less commonly [Noun Phrase][Present Tense Passive

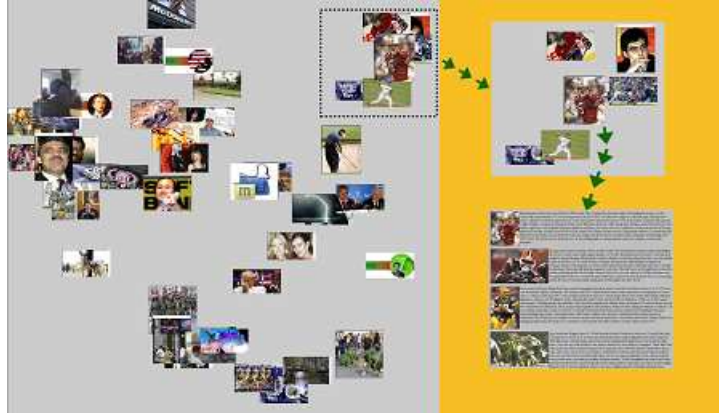


Figure 5: In this interface we have defined a two dimensional distance metric between topics and laid them out on the plane, illustrated with representative photos from the collection. The right hand side illustrates a closeup of the upper right corner of this space, an area dominated by sports related topics. In our interface one may click on any token to bring up a list of additional images and articles associated with this topic. In this example, the topics have been generated using captions from the AP and Reuters photos, but we are actually using this topic structure to navigate articles from CNN and the BBC. We contend that clustering with these captions generates topic distributions that focus on words highly relevant to the topic. Topics are defined as probability distributions across the corpus vocabulary. Our 2D embedding is derived by calculating the symmetrized KL divergence between each pair of topics and using principle coordinate analysis to project into two dimensions in a manner that “best” preserves the distances in the original high dimensional space. (Ripley, 1996)

Verb]. (see figure 2)

The detector we have built identifies potential proper names as strings of two or more capitalized words followed by a potential present tense verb. Words are classified as possible verbs by first applying a list of morphological rules to possible present tense singular forms, and then comparing these to a database of known verbs. Both the morphological rules and the list of verbs are from WordNet. (Wor, 2003)

When there is more than one person in the photo, the author often inserts a term directly following the proper name to help disambiguate the correspondence. This will either be a position such as left or right, or an identifying characteristic. This second form is most frequently used with sports figures and gives a jersey number. Our proper name finder returns the name, the disambiguator if it exists, and the verb.

Our classifier either accepts a caption and returns a proposed name or rejects the caption. We tested it on the same sample we used for clustering (146,870 captions). The name finder accepts 47% of these captions. We manually examined 400 rejected captions. Of these, 50% were true misses, where the caption contained a name that matched a face in the image. Another 35% were images of people but the caption either contained no proper names or only proper names of people that did not appear in the image. The final 15% were images that contained no people.

We also examined 1000 accepted captions. In 85% of this sample, the classifier accurately extracted the name of at least one person in the image. Of these errors, the vast majority still followed our pattern of [Noun Phrase][Present Tense Verb], and the subject of the noun phrase actually did appear in the picture. Our rules simply failed to accurately parse the noun phrase. Over half of these mistakes, for instance, were due to the phrase “[Proper Name] of the United States [Verb],” and our classifier returned “United States” instead of the correct individual. More robust proper name finders should largely eliminate these sources of error. The more important point is that captions are so carefully structured that very simple rules accurately model most of the collection. We could conceivably even use our face classifier to learn some of these structural rules. If we could use the images to posit names, and even positions, we might be able to use this as a handle for learning certain types of caption structure. If a photo were to have two strong face responses, for instance, we might learn to look for those “left”, “right” indicators in the text. We would also like to investigate the possible clustering of verbs from image data.

4.2 Iterative Improvements

With aggressive pruning of questionable outputs from the name and face finders, we believe we can generate an effectively supervised dataset of faces for thousands of

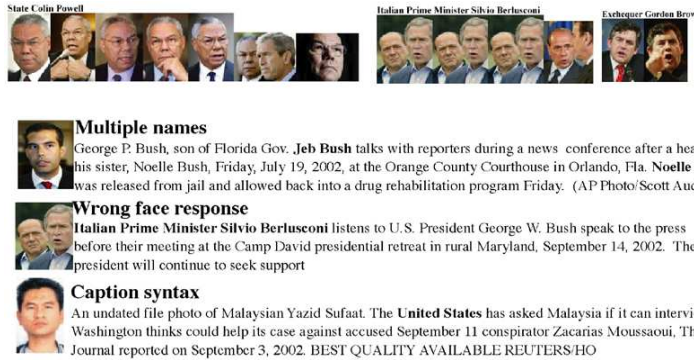


Figure 6: *Not every person named in a caption appears in a picture. However, quite simple syntactic analysis of captions yields some names of persons whose faces are very likely to appear in the picture. By looking for items where the picture has a single, large face detector response — so there is only one face present — and analysis of the caption produces a single name, we produce a directory of news personalities that is quite accurate. The top row shows some entries from this directory: there are relatively few instances of each face, because our tests are quite restrictive, but the faces corresponding to each name are correct and are seen from a variety of angles, meaning we may be able to build a non-parametric face model for some individuals by a completely automatic analysis. The next three rows show some possible failure modes of our approach: First, our analysis of the caption could yield more than one name. Second, there may be more than one large face in the image, with only the wrong face producing a face detector response. Third, the syntax may occasionally follow a syntactic pattern our algorithm does not handle. We are able to extract proper names from 68,496 of 146,870 captions, with an estimated 85% of these actually naming a person in the image. Restricting ourselves solely to large face responses, we are able to produce a gazetteer of 452 distinct names (621 images total), containing only 60 incorrectly filed images.*

individuals, many of whom will have hundreds or even thousands of distinct face images. From these we hope to build a non-parametric model of faces. The next interesting task will be to investigate whether we can then return to the textual side, using our face models to learn more about linguistic structures of the captions. Bootstrapping by alternating between the two sides of a mixed dataset seems a very powerful model.

5 Conclusion

News photo captions are an interesting dataset both for their unique textual properties, and for the opportunities they provide to exploit relationships between the text and image contents. We have used these captions to illuminate underlying topical structure in the collection. This research indicates that captions act much like very short summaries, emphasizing words that are strongly associated with underlying themes in the news. We are investigating how well this topical structure translates to more general collections of news articles. We have also shown that images can provide links between articles that are missed by textual analysis alone. Separately, we are investigating the possibility of building non-parametric models of celebrity faces. This line of research indicates that by combining a face detector with an analysis of the linguistic conventions of the text, captions can be used as

an almost supervised dataset of people in the news.

References

- David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2001. Latent Dirichlet Allocation. In *Neural Information Processing Systems 14*.
- P. Duygulu, K. Barnard, J.F.G. De Freitas, and D.A. Forsyth. 2002. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary.
- Charles E. Jacobs, Adam Finkelstein, and David H. Salesin. 1995. Fast multiresolution image querying. *Computer Graphics*, 29(Annual Conference Series):277–286.
- Christopher D. Manning and Hinrich Schutze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press.
- V.J. Della Pietra P. Brown, S.A. Della Pietra and R.L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 32(2):263–311.
- B. D. Ripley. 1996. *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge.
- 2003. Wordnet, a lexical database for the English language.