Improving Photo Tag Recommendation Accuracy Through the use of Geolocational and Temporal Data

Blair Calderwood

School of Computing Science
Sir Alwyn Williams Building
University of Glasgow
G12 8QQ

A dissertation presented in part fulfilment of the requirements of the Degree of Master of Science at The University of Glasgow

7 September 2016

Abstract

Accurate categorisation of photos hosted online is an extremely worthwhile pursuit due to the exponentially increasing number of images being uploaded to social media websites such as Flickr and Facebook. Textual image annotation, alongside various image processing based methods, can be used unaccompanied or in tandem to categorise photos in order to increase the effectiveness of user image retrieval. Users uploading images, or any other media item, are frequently asked to manually annotate their items using textual tags. This can be a laborious process that leads to an increase of noise in the tag space. In order to combat this tags can be recommended to users based on image content, image context and previously entered tags. This active research area has utilised machine learning techniques such as tag co-occurrence in the past to increase the accuracy of tag recommendations. This paper aims to increase accuracy using a tag co-occurrence based approach that relies on multiple features taken from image content and context. It uses geolocation to recommend tags appropriate to the area in which the photo was taken alongside a technique which analyses temporally similar photos to determine appropriate tag recommendations. The novel technique is evaluated using previously established offline evaluation methods alongside a novel user based online approach. Both Offline and online evaluation show that the novel system outperforms each baseline tested by a large margin.

Education Use Consent

I hereby give my permission for this project to be shown to other University of Glasgow students and to be distributed in an electronic format. **Please note that you are under no obligation to sign this declaration, but doing so would help future students.**

Name: _____ Signature: _____

Acknowledgements

acknowledgements go here

Contents

Chapter 1

Introduction

Tagging photos on social media platforms such as Flickr is designed to add context and content descriptions to images to aid in categorisation and searching. Tagging has been poorly adopted in the past. Many users tag their photos with tags that could be considered "noise" while many others simply do not tag their photos at all. Noisy tags can include device related words or phrases such as "Nikon" or "capturedwithiphone"; self categorisation tags, such as "mybestfriend", which only aid the posting user and apparently nonsensical data that does not aid categorisation or searching. Examples of the latter include geographical co-ordinates and the time the photograph was taken. This data can already be found in the photo's metadata and so adds additional unnecessary noise to the tag space.

Photo Tag Recommendation (PTR) systems aim to reduce noise in the tag space by recommending a series of tags to the user based on one or more tags they have entered. Accurate recommendation not only aids image categorisation and searching but also reduces the mental and temporal strain of the image tagging process. Tag recommendation is typically done through the collection of information on different features of the photo. These features could be content related, such as the dominant colour of the image, or context related, such as the time of day the image was taken.

1.1 Project Novelties

The novel elements introduced in this paper are threefold - the introduction of accurate geolocation based tags via dynamically generated tag co-occurrence matrices; the introduction of temporal based tags using a similar method to the aforementioned and the addition of online user based evaluation to the standard offline evaluation used in many previous works.

1.1.1 Geolocation Based Tag Recommendation

Many images uploaded to photo sharing sites such as Flickr contain landmarks and tourist destinations. These types of images relate closely to the geographical area they were taken in. An example of this is that photographs of the Sistine Chapel will always be taken in the one geographical area. The novel system exploits this trend by recommending tags specific to the geographical area the photo was taken in. Not all tags are related to geolocation - for example, an image of a user's family may be taken anywhere in the world. Due to this limitation this feature is used alongside other more established ones, such as the number of faces in a picture, to accurately recommend tags for any situation. The aim of this novel method is to increase diversification in the tag space. This will mean photos are tagged with less generic tags such as "building" and more with specific tags such as "sistinechapel". This increases the ease of searching for photos of a specific landmark or area.

Geolocation based tag recommendation is performed with the use of a dynamically created tag co-occurrence matrix which utilises tags used previously in pictures taken in the immediate vicinity of the user. The immediate area is used instead of exact location so the user does not need to be standing in the exact same position as a previous user to obtain their tags.

1.1.2 Temporal Based Tag Recommendation

Yearly events such as Christmas and New Year are frequently celebrated and documented through the medium of photographs. Users may take pictures of family gatherings on Christmas Day or of fireworks on New Years Eve. This means that a given picture has a higher likelihood to be related to another picture taken on the same day in a previous year. For example, "family" may be tagged highly in photos taken on Christmas day and so it may be advantageous to recommend the tag for any future photos taken on this day. Furthermore, this feature can be used in conjunction with the aforementioned geolocation based tag recommendation to recommend tags pertaining to location sensitive events such as festivals and local gatherings. Even if an event lasts more than one day, for example, Wimbledon, the system will still provide relevant tags such as "tennis" and "andymurray".

A dynamically created tag co-occurrence matrix similar to the one used in geolocation based tag recommendation was used to determine the best temporally related tags. Each day of the year has its own tag co-occurrence matrix which contains tags used in a selection of photos on that day from the previous three years.

1.1.3 Online User Evaluation

The established evaluation method used in determining photo tag recommendation system accuracy is the offline method. This is where a large dataset consisting of thousands of photos is split

into a training and test set. The training set is used to train the system and the testing set is used to measure the system's recommendation accuracy. This is usually done by feeding one tag from the test set into the recommendation system. The system will then recommend a series of tags that it deems to be relevant to the inputted tag. These tags are then measured against the other tags relating to the test image and accuracy is measured based on how many of the additional tags the system can replicate.

The established method simply attempts to replicate previously entered tags and does not strive to improve them in any way. Without improving tags then the noise in the tag space will remain. A tag recommendation system may not perform well in the established offline evaluation if it aims to improve tags. It could be providing tags that are relevant but not included in the photo's original tags as the user simply did not think to include that tag or did not have enough knowledge about the subject of their photograph to know to include the tag. For example, a user may take a picture of the Glasgow Science Centre without knowing that it is named as such. They may then tag it as "modernbuilding" but would have tagged it as "glasgowsciencecentre" if they had the correct knowledge. The offline evaluation method would penalise any system that provided the more accurate tag. The novel online evaluation method asks real users to add one tag to each of their own photos. The system will then recommend additional tags based on this and the user will implicitly rate the system based on the relevance and accuracy of the tags recommended. This means systems that effectively improve tags and do not simply mimic them will receive higher scores than those that do not aim to improve the tag space in any way. This evaluation method is described in further detail in section 4.2

Chapter 2

Literature Review

Tags can be used to annotate many types of digital items. They describe either the content or context of an item [28] which can help the posting user or other users. A tag is typically added manually by the user as a method of describing the item they have uploaded [31]. An item can be any one of a number of digital media such as images, video [6] or music tracks. A great deal of work has been performed with the aim of accurately recommending tags to users [21, 24, 36] and so this has proven to be a very active area of research.

2.1 The Purpose of Tagging Photos

Photo tag recommendation has become a very active research area in recent years due to the exponential rise in images being uploaded to sites such as Flickr and Instagram. Flickr alone had 1.83 million photos uploaded to it on average for each day in 2014 [3]. These photos need to be categorised and annotated in order to aid recall and support search while supporting new methods of social networking and data mining [16]. This tagging cannot feasibly be performed by an authority in the same way as a librarian organises books [26] and so the public are required to tag their own content. This is known as collaborative tagging [11] and carries the benefit of allowing the task of metadata creation to be distributed amongst the group thus decreasing workload for the individual [16].

Photo tagging can also increase the understanding of the semantic meaning of the image which is an important aspect of building an effective photo retrieval system [19]. As tagging systems tend to be unstructured, lacking any categorisation or ontologisation, they allow for greater flexibility at expressing semantic meaning due to their ability to naturally evolve to reflect emerging and growing data trends [23]. The inherent lack of structure in this system, often described as a folksonomy [2], means semantic meaning is often difficult to extract. PTR systems often aim to

recommend tags in such a way that structure and semantic meaning can be extracted from the resulting annotations.

Manual photo annotation is frequently seen as a tiresome and repetitive task by users which is why the process has not been very widely adopted [5]. Although the users that do tag photos state they do it to help the general public [28]. According to Golder and Huberman [11] tags can be split into several categories. Users often tag items as a method of self categorisation or to identify a quality or category of the image - for example "beautiful". Tags can also serve as a form of self organisation. Phrases such as "toedit" or "todelete" can be considered as part of this category of tagging. Golder and Huberman also identify self reference as a form of tagging. Users will frequently tag items with phrases such as "myfriends" or "myfavouritephoto". Trends seen throughout all of these categories suggest that most users annotate images using either nouns or entity descriptors [19]. Identifying why people tag photos and what they tag them with is a major step towards building an effective recommendation system.

2.2 The Purpose of Recommendation Systems

Tag recommendation systems are often used to decrease tag ambiguity [10]. For example, if a user enters the tag "java" then this could refer to the coffee, the island or the programming language and so tags will be recommended based on these three meanings [27]. The user will choose the tags relating to the correct meaning thus decreasing ambiguity and increasing accuracy when categorising photos. Weinberger, Slaney and Zwol [34] attempt to reduce tag ambiguity through examining contextual information such as location. It labels a tag as ambiguous if it falls into more than one distinct category. The example used in their paper surrounds a user who inputs "cambridge" as a tag. The system then looks at geolocation information to determine if it is more likely they meant Cambridge, UK or Cambridge, Massachusetts. Techniques, such as the aforementioned, that reduce tag ambiguity enable much more accurate photo organisation and searching to take place.

Recommendation systems use the 'wisdom of the crowd' by helping people make choices based on the opinion of others [25]. The wisdom of a crowd of non-experts is thought to come to even better decisions than those of experts under the right conditions [30]. Photo tag recommendations utilise crowd wisdom by looking at the opinion of others via the tags they have chosen and the images they have assigned aforementioned tags to. This leads to tag convergence where more people will use the same tags and therefore will find it easier to search for items [4] thus reducing the scope of the Vocabulary Problem [8] where different users describe the same thing in different ways. For instance, if everyone had one word for cat then it would be extremely easy for a user to search for images of one. However, as people use different words for cat, such as "feline", "kitten" or the name of a cat breed, it is more difficult to search for images as searching for the word "cat" would mean images containing the only the aforementioned tags would not be included. The problem of tag redundancy [10], such as this occurring in tags is very problematic

as it has been found that around half of all annotations on an image have one or more synonyms in the same tag set [19].

Redundancy of tags amongst images is also a problem. Many users simply copy the tags they have entered from one image and paste it into the next [19]. This decreases tagging accuracy as these photos will not be identical and so should not have identical tags. It is theorised that by making the tagging process easier that the copying of previous sets of tags would occur less often. Reducing process difficulty will also result in a reduction of the misspelled words problem where tag noise occurs due to wrong or alternate word spellings [22].

2.3 Automatic Image Annotation Versus Photo Tag Recommendation

Much of the work in this paper is based upon McParlane's research into PTR and Automatic Image Annotation (AIA) systems [19]. In this paper the author differentiates between AIA and PTR by describing the relative merits of each type of system. While AIA attempts to annotate images from a cold start without any user interaction PTR recommends tags for the user and waits for confirmation that the tag is relevant. AIA is very useful for annotating images where the user is not immediately available such as when a system is trying to annotate historical images. PTR requires user communication and so is better suited to social networks. AIA was can only annotate image content, i.e. what can be seen in the image, and not the context surrounding the image [19]. Conversely, PTR allows concepts not captured within the image itself to be annotated [15]. This attribute of PTR systems has great influence on the type of evaluation that can be used. As the context surrounding the image needs to be annotated the best person to perform this will always be the person who took the photo as they know the exact circumstances in which it was taken. For example, a user may take a picture at a festival but, short of a few other eagle eyed viewers, most people will not know which festival this is and so will not be able to adequately annotate the image [19].

While AIA systems suffer from the context problem they also hold certain advantages over their PTR peers. Tag ambiguity and redundancy is much less widespread in AIA tagged photos due to each photo being tagged by a computer and so a fixed vocabulary can be strictly adhered to [15].

2.4 Tag Co-Occurrence Versus Other Machine Learning Techniques

There has been a divide in the field between systems which use the tag co-occurrence method and systems which use machine learning approaches such as k-nearest neighbour. Tag co-occurrence based methods often require a large amount of data to accurately recommend tags [28] - this has

been dubbed the small dataset problem in this paper. This means that datasets usually stretch into the hundreds of thousands [19] or even millions

2.5 Features and Feature Extraction

The majority of PTR systems use features to determine which tags should be recommended to the user. These features relate to an aspect of the image, whether that be a content related aspect such as dominant colour, or a context based aspect such as the location in which the image was taken. Users tend to tag both the visual content of an image and the context surrounding the image [28] and so taking both types of features into consideration when recommending tags is of paramount importance. Both content and context based features are most effective when used in a cold start scenario where only one or two tags have been entered [28]. In addition to recommendation based directly upon the image, such as the aforementioned content and context based recommendations, systems often also consider users and other tags when recommending additional annotations [16]. Using inputted tags to recommend additional ones is one of the most established methods of tag recommendation and is used in systems such as Flickr's Get Related Tags API feature [7] alongside many tag co-occurrence methods [33].

2.5.1 Content Based Features

Content based features extract data from the image itself in order to accurately recommend tags. The bulk of this is done by utilising image processing techniques such as Viola and Jones' [32] HAAR based face detection method to establish the amount of faces in an image. Yang et al [35] uses a bag of visual words technique to rank Flickr photos based on what is seen within them. This technique provides promising results but relies on a very large dataset of 500,000 photos in order to provide an adequate source of bags of visual words.

2.5.2 Context Based Features

Geolocation Based Tag Recommendation

Over one quarter of all user tags include some geographical information such as place names [28] and the social media website is being utilised by increasing numbers of tourists who both wish to upload holiday photos and view other's photos in order to plan their trip [14]. Despite this, little has been done to fully exploit the aforementioned trends. McParlane [19] uses location but only recommends tags on a country scale. This is useful for tourists visiting very popular areas on holiday as these are likely to be recommended. However, countries have more than just a handful of attractions and so it was theorised that recommendations based on a smaller

geographical area will provide more accurate results. For example, someone visiting Mount Rushmore may find that the landmark has not been recommended as a tag but landmarks such as The Empire State Building have. If the system provide tags for a smaller geographical area then the likelihood of recommending a relevant landmark is greatly improved.

Systems in the past have utilised more specific geographical location in order to recommend tags based on nearby landmarks. Systems such as ZoneTag [4] create tags based on nearby location data such as city and postcode. The advantage of this system as it does not need a large dataset to recommend tags as they are retrieved directly from mapping systems. However, this means that the system does not take into account the fact that landmarks are more than just their official names. If a user took a picture from the Eiffel Tower the system could recommend "eiffeltower" as a tag. While this recommendation is correct and should be recommended the system may miss out on some other tags that the user wishes to include. Tags such as "greatview" or "cityskyline" would not be included as they are not found on a map anywhere.

Other systems have utilised GPS to determine the popularity of certain locations. Sun et al [29] suggests popular landmarks to the user while also recommending travelling routes between said landmarks. Popularity of landmarks in this system is closely related to the amount of photos containing them.

Temporal Based Tag Recommendation

As with its geolocational counterpart, a great deal of research has gone into detecting events using temporal information. Temporal data has been used in the past in tandem with burst detection techniques to identify events [23]. Not only can this help improve tagging via the recommendation of said event but it can enable categorisation by event. For example, if a user wishes to see pictures of Coachella 2016 then categorisation in this manner can enable this. Often, techniques such as burst detection are not required if the same results can be achieved through means such as co-occurrence. The novel system introduced in this paper attempts to implicitly extract event information without the use of a burst detection algorithm.

Time of year has been used in a less specific manner by McParlane [19]

2.5.3 User Based Features

User based features, while not used in the novel system to aid generalisability, has been used in the past to increase tag recommendation accuracy. User based features were used to great extent by McParlane [19] in conjunction with content and context based features. McParlane shows that user popularity is one of the best features to utilise in a tagging system. User popularity is measured through the number of views a user's photos get. More popular users tend to be professional photographers and so user popularity, alongside context based features such as

camera type, can often determine if a photo was taken professionally or not. The distinction between professional and amateur photographs provides a large boost to tag recommendation accuracy.

2.6 Evaluation Methods

The method of training a system on a dataset and then testing it on another is well established in the machine learning domain [12, 17]. However, using user evaluation to actually improve tags and not just mimic them is novel to the area. In this online evaluation method it is important that the user tags their own photos. Garg and Weber [9] state the need for images to be tagged by the user who captured them: multilingual tags mean that users will typically tag a photo in their own language which can lead to tag redundancy. context tags may be missed if the tagging user is not the image capturing user as they do not know the exact circumstances surrounding the image. For instance, they may not be aware of which city a picture was captured in or which event is the subject of the photo. Additionally, the image capturing user may tag an image to a different level of detail than anyone else. For example, they may tag "europe" if they are American and visiting Berlin while a German may simply tag "berlin" or even the street that it was taken on. Different people with different experiences include different levels of detail in their tags.

2.6.1 Limitations

There are limitations to both offline and online methods of evaluation. The offline method may flag a tag as irrelevant as it does not appear in the ground truth even if the tag is a synonym of one that does appear in it [10]. Although the tag was technically wrong it has the same meaning of a tag that is correct and so should also be labelled as relevant. Online evaluation also has limitations - users often select tags that are recommended to them even if the tags do not appear immediately relevant [28]. This limitation can potentially have much greater implications on the tag space - the offline limitation only reduced the number of relevant tags in the tag space while this limitation may actually introduce completely irrelevant tags.

2.7 Number of Tags Recommended

Chapter 3

Implementation

3.1 Formal Definition

For a given photo F and tag T the novel system will recommend a series of five related tags. These tags will extract content and context related features from F and use them to build a tag co-occurrence matrix M. Tags are then ranked on how often they co-occur with T within M and the highest ranking tags are returned to the user.

3.2 Flickr

The Flickr web API [7] was used to retrieve photo information such as EXIF data and Flickr Places data which was saved in databases thus enabling faster future dissection of the data. This data was then used to help determine rankings when recommending tags to users. Flickr was chosen as it contains an extensive range of API features which can be used to exploit photo data in a multitude of ways.

3.3 Data Structure

The main data structure utilised throughout the system is a length n array where n is the amount of features used in the system. Each element in this array will hold a length m array of sparse matrices where m is the amount of feature values in the corresponding feature.

The system recommends tags through a tag co-occurrence based method. There are several features in the system (for example, number of faces) each of which have their own set of feature

values. A feature value is any value that could be found in a feature. For example, the system counts the number of faces in an image as either 0, 1, 2 or 3 and above faces. Each of these number of faces is a feature value and will have their own tag co-occurrence matrix. If a photo is found to have two faces in it and contains the tags “party” and “nightclub” then these two tags co-occur in the 2 feature value of the number of faces feature and so 1 co-occurrence mark will be added to the relevant places in that co-occurrence matrix. This data structure is illustrated in figure 3.1.

Features that host string based values (for example, continents are strings and not integers) were switched for numbers in the dataset. This aids spatial efficiency and means each feature will contain a set of distinct integer feature values each of which will refer to a sparse matrix.

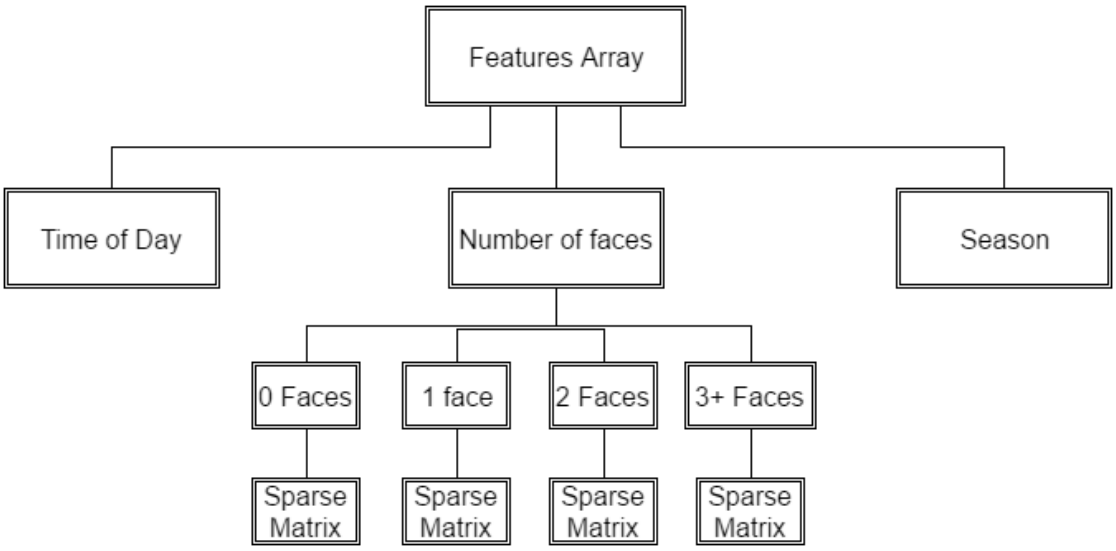

Figure 3.1: Tag Co-Occurrence Data Structure

Each tag co-occurrence matrix will be an *n x n* sparse matrix where *n* is the cumulative amount of tags that occur in each photo in the dataset. Each tag has *n* corresponding values - each relating to the amount of times this tag co-occurs in the same image as the other tag. An example tag co-occurrence matrix is found in table 3.1. The data structure does not technically contain numerous sparse matrices for efficiency reasons - see section 3.7.2 - but acts in a very similar way and so is described as such.

The dynamically generated tag co-occurrence matrices pertaining to the geolocation and temporal based features do not abide by this data structure. As there are a vast number of days in the year and an far vaster number of possible places on earth it would be computationally infeasible to store n x n sparse matrices for all of these. Instead, these are generated upon request of the user and so are separate from the main data structure. More information about this can be found in sections 3.6.3 and 3.6.4.

Table 3.1: A tag co-occurrence matrix where “nightclub” and “party” co-occur in three photos

	Hill	Dog	Nightclub	Party	Jupiter
Hill	0	2	0	0	1
Dog	2	0	0	0	0
Nightclub	0	0	0	3	0
Party	0	0	3	0	0
Jupiter	1	0	0	0	0

3.4 Datasets

The dataset created for the system is based upon the Flickr AIA and PTR datasets [20] which house a collection of photos and photo metadata retrieved from the aforementioned media sharing website. This data needed to be expanded upon as extra features were included in the novel system and so a new dataset was created. This dataset included values relating to each feature used in the system - for example, number of faces. It also holds a places ID which relates to the geographical area in which the photo resides (see section 3.6.3 for details). The dataset includes 50,000 records and so is large enough to base a tag recommendation system upon.

3.5 Features

The following features were used in the novel system:

1. Overall - tags are counted as co-occurring regardless of feature values
2. Geolocation - see section 3.6.3.
3. Continent - can be any one of the Earth’s seven continents and is used to provide more general geolocation based tags
4. Number of Faces - The number of faces found in the image using Viola and Jones’ face detection method [32]. This can be 0, 1, 2 or 3 and above
5. Dominant Colour - This can be white, black, red, green or blue.
6. Image Orientation - Square, portrait or landscape.
7. Time of Day - Can be morning, afternoon, evening or night.
8. Time of Year - Refers to the seasons of the year and can be spring, summer, autumn or winter.
9. Day of Week - Either weekday or weekend.

10. Day of Year - numbered from 1 to 365 (366 in leap years) see 3.6.4
11. Flash - Determines whether the camera was taken with or without flash. It can either be on, off or unknown.

Each of the features mentioned, with the exception of geolocation and day of year, were used by McParlane's tag recommendation system [19].

3.6 The Tag Recommendation Process

The following outlines the process the system follows when creating tag recommendations.

3.6.1 Creating the Dataset

Given a path to a folder of jpeg images named with their corresponding Flickr ID the system will create a dataset. It will compute all feature values so they do not need to be processed each time the system is run. Data concerning items such as image location and time taken will be retrieved from the Flickr AIA dataset if it contains a record relating to the image ID [20]. If it doesn't then this data will be retrieved via the Flickr API [7]. Some images may have been deleted by their posting users since their creation and so it is unlikely the system will retrieve 100% of the photos in a given folder. The dataset creation process is further described in algorithm 1.

The tags for each photo are retrieved from either the Flickr AIA dataset or from the Flickr API and stored in a separate SQL database for later retrieval. The process of creating the dataset is a lengthy one and so has been designed to allow it to be completed in chunks. Preprocessing of the data means the process of creating co-occurrence matrices is much less time consuming.

Algorithm 1: Creating the Photo Information Dataset

Data: A directory containing an arbitrary number of JPG images
Result: A dataset comprising of data about several thousand images retrieved from Flickr

for *Each photo in the given image directory* **do**
    Retrieve the Flickr ID of the image from the file name - this allows retrieval of information from the Flickr API
    Create a two-dimensional matrix representation of the image (used in image processing)
    if *The data for the retrieved photo is in the Flickr MIR [20] dataset* **then**
        Retrieve this data and save it as *D*
    else
        Retrieve data concerning this image from the Flickr API and save it as *D*
    end
    Retrieve image context details (e.g. dominant colour, number of faces etc.) via image processing techniques
    Retrieve image content details (e.g. time of day and season) from operations performed on *D*
    Retrieve the Flickr Places ID by performing an API lookup using the photo's GPS co-ordinates
    if *The geolocation database already has a record pertaining to this Places ID* **then**
        Save photo details in this record
    else
        Create new record based upon this photo and the top 25 photos retrieved from the Flickr API pertaining to this location
    end
    Save photo details to the correct record in the day of year database
    Save image details alongside day of year and Places ID into the dataset
end

3.6.2 Creating the tag co-occurrence matrices

One overall sparse co-occurrence matrix is created where tags for all photos are included. Alongside this a sparse matrix for each feature value within each feature is created as described in section 3.3. The system then loops through each photo in the dataset and notes all tags that occur in the same photo. If two tags co-occur in one photo then one co-occurrence mark is added to the two relevant values in the overall co-occurrence matrix. The system will then loop through each feature and add one to the values found in the same position in each relevant feature values matrix. Creation of a tag co-occurrence matrix is illustrated in algorithm 2. The co-occurrence matrices are then stored in files which means they do not need to be created each time the system is run.

Algorithm 2: Creating a Tag Co-Occurrence Matrix

Data: A two-dimensional array of photos and their respective tags
Result: An n x n tag co-occurrence matrix

```
for Each photo in photo set do
    for Each tag in the photo do
        for Each other tag in the photo do
            Add one to the co-occurrence matrix for these two tags
        end
    end
end
```

To increase tag diversity and ensure that the system did not simply recommend the most popular overall tags for each input the Term Frequency - Inverse Document Frequency (TF-IDF) was used [13]. This calculation was performed on each value in each matrix and weights each co-occurrence value based upon the overall popularity of its related tags. For example, the tag "sky" will be found in many photos so would have a high co-occurrence rate with most tags. However, a system that recommended "sky" for every input would not increase tag diversity. A more accurate system will look for less popular tags that occur highly. For instance, when given the tag "sheep" instead of recommending the tag "sky" the system may recommend "farm" even though it has a lower number of co-occurrences. The TF-IDF calculation is illustrated as follows:

$$TFIDF = log\frac{T}{O}$$

Where T is the total amount of tags in the system and O is the total number of occurrences of the tag in question.

3.6.3 Creating the Geolocation Tag Co-Occurrence Matrices

The Places ID will be retrieved from the Flickr API for each photo in the dataset. This unique identifier corresponds to the local area of where the image was taken. It does not correspond to the exact co-ordinates of the image as this would mean each subsequent image taken would need to be captured in the exact same geographical position as the first image in order to retrieve relevant tags. If the Places ID is found in the Areas database then there is images already in the dataset corresponding to this real world location. If the places ID is not found in the database then an entry is created with some 'starter' data. The system retrieves a sample of 25 photos from the Flickr API. All of these photos have been taken within the image in question's places ID. These photos are then saved to the database for future use in tag co-occurrence. This means that an image with a valid places ID will always have some data to build a co-occurrence matrix

upon. The tag co-occurrence matrix is then created as seen in algorithm 3. As seen in the aforementioned algorithm the matrix in question is only partially created. This is to increase efficiency as only one row of the matrix, the row of the tag in questions, needs to be created. This leads to a great performance increase over the creation of the full *n* x *n* matrix.

Algorithm 3: Creating a Dynamic Geolocation Based Tag Co-Occurrence Matrix

Data: A photo object containing contextual information such as a Flickr places ID or day of year and *t* - the tag for which recommendations are required
Result: An 1 x *n* geographical or temporal tag co-occurrence matrix

Read photo details from image or from database

if *photo has preprocessed places ID* **then**
  Lookup places ID in locations database and get related photos
else
  Get related photos from Flickr
  Save new place to locations database
end

Remove all photos from the list that do not contain *t* - these are unnecessary in the partially created matrix
Create a sparse matrix of size 1 x *n* where n is the number of tags used in the system

for *Each photo in photo set* **do**
  for *Each tag that isn't t in the photo* **do**
    Add one to the co-occurrence matrix where *t* and this tag meet
  end
end

3.6.4 Creating the Temporal Tag Co-Occurrence Matrices

Temporal tag co-occurrence is performed in a very similar way to its geographical counterpart as seen in section 3.6.3. Instead of checking for the relevant entry and creating a new one if one is not available the temporal matrix has an entry for each one of the 365/366 days of the year present from initialisation. Each of these entries will contain a series of photos retrieved from Flickr for that date from the last three years. These photos will be used to dynamically create a tag co-occurrence matrix upon user request. This matrix will emphasise re-occurring themes and events through higher co-occurrence levels.

3.6.5 Recommending Tags

When given a photo and an initial tag the system recommends five tags based on the co-occurrence matrices relating to the photo content and context. These tags are found by per-

forming a lookup on each co-occurrence matrix, including the two dynamically created matrices, for the inputted tag. The row corresponding to this tag is then extracted and added to the rows extracted from each other matrix. Each value is then added to the corresponding value in each other row. For example given the tag "nightclub" it may occur with "party" 6 times in the basic co-occurrence extracted row, 4 times in the "faces = 3+" row and 0 times in the "time = afternoon" row and so would have an total score of 10. Each row in normalised so all values lie between 0 and 1 and then all rows are added together and sorted into descending order. The highest occurring tags found in this summed row are then returned as recommendations.

3.7 Efficiency

Due to the size of the dataset and the inherent computational intensity of the novel system efficiency had to be considered in great detail. Performance enhancements were implemented in large numbers so that the system not only produced offline evaluation results in adequate time but also ensured that participants in the online evaluation would be able to undertake photo upload and subsequent recommendation analysis in a reasonable amount of time. Multiple small efficiency improvements have been implemented in addition to the major ones described in this section. Smaller improvements such as the removal of some costly division operations and the implementation of a string based binary search lead to a large overall efficiency improvement.

3.7.1 Multiprocessing

The novel system has been built to exploit the benefits of a multi-core CPU system. Two of the most computationally expensive processes in the system utilise multiprocessing techniques. The process of creating the tag co-occurrence matrices for each feature value in each feature of the system (see section 3.6.2) utilises multiprocessing by sharing the workload between all cores in the CPU. It divides the list of features into n lists where n is the number of available cores. These lists will be as equal in size as possible to distribute workload relatively evenly. A similar method was implemented to greatly increase offline evaluation efficiency. The test photo dataset was divided into n lists and each of these lists was processed by a different core. The results obtained from each of these processed was then collated so as to allow further analysis.

3.7.2 Preprocessed Data

As much of the data in the tag co-occurrence matrices will remain static after the training dataset has been processed the data in all co-occurrence matrices except the geolocational and temporal dynamic matrices is saved so it only needs to be created once. This is done as it would be computationally infeasible to create each matrix every time a request for recommendations was

submitted. This is why the matrices have been split into preprocessed static matrices and dynamic matrices created as requests are submitted. As these static matrices require a great deal of storage space they are not stored in RAM as this would be inefficient. They are not stored in files on hard disk either as this would require the whole matrix to be read into RAM simply to retrieve the co-occurrences on one row. Instead, the system stores each matrix in the form of a database. Only records which contain a non-zero value (i.e. the co-occurrence of the two tags in question is more than zero) are stored to save space. One database is created for each matrix in question and so a total of 39 matrix databases are required.

Each matrix database is indexed based upon row number (this will relate to the tag which requires recommendations) and the co-occurrence value. This means the latter can easily be retrieved and sorted in descending order which, in turn, allows the system to retrieve the top n tags. This is more efficient that retrieving records and then sorting them. Many tags will co-occur with thousands of others and so it is inefficient to retrieve all tags. The system combats this problem by retrieving the top 500 tags which co-occur with the inputted tag. Any tags outwith the aforementioned 500 are rarely included in the final outputted recommendation and so they are irrelevant.

3.7.3 Dynamic Co-Occurrence Creation

A reasonable amount of performance efficiency has been sacrificed in order to improve spatial efficiency by dynamically generating the geolocation and temporal co-occurrence matrices upon request. This means that the user must wait for the matrix to be generated before receiving their recommendations. However, it also means that storage of a matrix for every possible place in the world and day in the year is not required thus saving a massive amount of storage space. This is especially important as the dataset size increases which leads to an exponentially growth in matrix size. To increase efficiency in this process the co-occurrence matrices are only partially created. Section 3.6.3 and algorithm 3 explain how this is achieved in detail.

3.8 Limitations

Several limitations have been identified in the system. Some of these limitations have been addressed in section 6. Tag noise reduction was not implemented in the novel system as this could mean that participants in the online evaluation receive no recommendations for an uploaded image-tag pair. If a user enters a tag deemed by the system to be noise then it will have already been removed from each co-occurrence matrix and therefore will not have any co-occurrences. As noisy tags are usually outlying data, i.e. only a small number of users will tag a photo with the same noisy tag, the nature of tag co-occurrence, which relies on a tag being tagged a large a number of times, means that it will inherently weed out noisy tags. The addition of the novel geolocational and temporal features further assist in the removal of noisy tags. There may still

be some noisy tags recommended to users but the amount of these is hypothesised to be much lower due to the aforementioned nature of tag co-occurrence.

3.8.1 Dataset Size

Compared to some datasets, such as Flickr AIA's 309,000 record set [20], the one created for evaluation is relatively small. This is due to the need for access to the photo data itself and not just the data provided by the Flickr AIA dataset upon which the novel dataset was based. The Flickr AIA dataset provides photo data for 70,000 of these records. Photo metadata pertaining to geolocation and the time the photo was taken is also needed which means that some of these 70,000 photos are unusable due to missing metadata.

The amount of data for each location and day of year in the dynamically created matrices is also reasonably small. The top 25 geographically nearby photos are retrieved when a new location is entered while the top 50 photos were retrieved for each day in the year. These amounts of photos were retrieved as the data needs to be downloaded from the Flickr servers which is computationally expensive. Furthermore, storing this data in their related databases is spatially intensive and so only retrieving a small number of photos means efficiency increases greatly. This, however, hinders the performance of the geographical and temporal features as the created matrix is very sparse and so any entered tag has a lower chance of having any co-occurrences in a matrix built with so few photos. The sparsity of the dynamically created matrices has an inverse relationship with the size of the training dataset as each photo in the training dataset is added to its affiliated entry in the geolocation and day of year databases meaning this photo's tags will also be taken into consideration in future tag recommendations.

3.8.2 Inability to Learn

To keep training and testing datasets completely separate the system can only 'learn' from the training set. This means that no changes are made to the co-occurrence matrices after training is complete. If the system were to be deployed into a real world scenario then it should learn from item-tag sets that users input i.e. given a photo and a tag the system should recommend tags to a user and then add any tags the user chooses into the relevant co-occurrence matrices. This means that system recommendation accuracy increases alongside the amount of photos uploaded.

Chapter 4

Evaluation

The evaluation of the novel system was performed in two stages. Firstly, offline evaluation was performed where a set of data mined from Flickr was used to train the system and populate the matrices and a testing set was used to evaluate the system. Secondly, online evaluation was performed where participants were asked to upload their own photos and select all recommended tags that were relevant to them.

4.1 Offline Evaluation

In offline evaluation the tags that were inputted by the user when posting the image on Flickr were extracted. One tag from this set was chosen at random and removed from the original tags. This random tag was then inputted into each of the baselines and the novel system. A random tag was chosen instead of simply choosing the first tag in each set as the sets are in alphabetical order and so would not provide an accurate view of overall system accuracy. Each system would then return five recommended tags which would be compared to the original tags (ground truth) to measure recommendation accuracy. Accuracy was measured using the metrics described in section 4.3.

Offline evaluation took place by extracting 45,500 photos from Flickr. 40,000 of these photos were used in the training set while the 5,500 were used in the testing set to measure recommendation accuracy. A total of 39 co-occurrence matrices were created based on the 40,000 training photos and their respective tags. These matrices were then used to provide the recommendations for each of the 5,500 test photos.

4.2 Online Evaluation

Online evaluation was performed via a web app. Participants visited the web app where they would upload several photos each and enter one tag for each photo. Upon receival of the initial tag the web app chooses a system (either one of the baselines or the novel system) at random and provides recommendations received from this system after inputting the relevant photo and initial tag. This is a blind experiment as the user does not know which system they are using nor do they know that there is more than one system involved in the experiment at all. No participant will be told that they are receiving recommendations from one of a number of systems. This significantly reduces bias.

The users will implicitly rate the system which has provided recommendations by selecting all recommended tags that are relevant to them. No explicit rating system is provided and so the online evaluation uses the same metrics as its offline counterpart - see section 4.3. When each photo the user has uploaded has been tagged the photos and tags are deleted from the system (only the amount of chosen tags remain) and the user is thanked for their time. Further details of this evaluation method is seen in algorithm 4. The web pages used in online evaluation can be seen in appendix A

Algorithm 4: The Online Evaluation Procedure

Display evaluation instructions to user
Ask user to upload a number of their own photos
Ask user to enter a tag for each photo

for *each uploaded photo* **do**
    Extract content and context details for each feature (found in section 3.5)
    if *all features are present* **then**
        Choose a system (baseline or novel) at random
        Display recommendation retrieved from this system
        Ask user to select all tag recommendations that are relevant to them
        Save these results
    end
end

Debrief user

4.3 Evaluation Metrics

Accuracy was measured in both evaluation methods using metrics that are well established in the field [9, 19]. The three methods used are as follows:

1. Precision at One (P@1) - The percentage of tag recommendations where the first tag recommended is found in the ground truth
2. Precision at Five (P@5) - The percentage of the five recommended tags that are found in the ground truth
3. Success at Five (S@5) - The percentage of times at least one relevant tag is found in the ground truth
4. Mean Reciprocal Rank (MRR) - 1 / the rank of the first relevant tag returned (if any)

4.4 Baselines

Each evaluation method was employed to compare the performance of the novel system against three baselines. As well as baselines comparison of the novel system as a whole the score for location and time alone was also measured using the aforementioned metrics.

4.4.1 Flickr Get Recommended

The Flickr API [7] has its own tag recommendation system. Users can enter a tag and they will receive a number of recommendations based on this input. No image content or context is used in this baselines - only the inputted tag is utilised by the system.

4.4.2 TF-IDF Tag Co-Occurrence

This well established system creates one tag co-occurrence matrix based on overall co-occurrences. It does not rely on any feature based matrices and therefore only uses the tag as an input instead of an image-tag pair.

4.4.3 Combined Tag-Co-Occurrence

The unweighted combined tag co-occurrence method created by McParlane et al [18] was used as a baselines as this method has been shown to yield positive results in the past. The system uses a tag co-occurrence based method with the following four features:

1. Continent - any one of the Earth's seven continents
2. Time of Day - Can be morning, afternoon, evening or night.

3. Time of Year - Refers to the seasons of the year and can be spring, summer, autumn or winter.

4. Day of Week - Can be either weekday or weekend

A more complex version of this system was introduced by McParlane [19] but this system was not used as it involved the use of user based features such as the previous tags the user had chosen. This reliance on user-item-tag could not be re-implemented as the online evaluation method dictated that users had no previous photo uploads. It also means that the baselines are a wider representation on image based social media as not all image based social media sites will allow for past user data to be utilised and so the used baseline is more easily generalisable.

Chapter 5

Results and Discussion

5.1 Offline Evaluation Results

As seen in table 5.1 the novel system outperformed each baseline by a considerable amount. In addition to this the temporal dynamically generated co-occurrence method also outperformed each baseline. Its location counterpart performed considerably worse. This result is likely down to the amount of photos used in the temporal and geolocation features. While the temporal feature downloaded information from 50 images from the Flickr API and constructed a co-occurrence matrix from this data the geolocation feature only downloaded the data of 25 images. This design choice was taken due to efficiency constraints as there are much more locations in the world than days in the year and so this database would take up far more storage space if 50 images were used. Furthermore, as there were less possible values in the temporal feature more images in the training set were distributed amongst fewer values, i.e. amongst the 365/366 days of the year and not amongst thousand of possible locations, each day of the year contained a large amount of photo data. This is compounded by the fact that many locations did not have 25 photos taken within them and so had less photo data to use in matrix creation. This last problem could be eliminated through the use of variable area sizes as discussed in section 6.3.

Table 5.1: Offline Evaluation Results

	Flickr Get Recommended	TD-IDF	Combined Co-occurrence	Novel System	Location Only	Time Only
P@1	31.09%	28.64%	12.98%	38.35%	3.24%	26.44%
P@5	16.27%	15.30%	9.40%	27.17%	2.00%	20.16%
S@5	48.82%	45.55%	31.84%	66.58%	8.93%	55.07%
MRR	46.79%	44.47%	31.73%	55.27%	20.60%	44.94%

It is worth noting that the Flickr Get Recommended baselines has surpassed scores it has achieved in the past [19] which suggests that they have improved upon their recommendations since it was

last analysed. The Flickr system also has an advantage over the other systems when using a small dataset as the others are based on tag co-occurrence so the smaller the dataset the less accurate it is. One of the reasons the dynamically created matrices were created were to combat the effects of a small dataset. As the geolocation and temporal features load data from Flickr to provide some 'starter' data to build the matrix upon the novel system is much less susceptible to the small dataset problem.

5.2 Online Evaluation Results

Online evaluation was carried out with 16 participants uploading a total of 184 photos. Due to the nature of the online evaluation the number of photos evaluated was much smaller than its offline counterpart. This introduces a larger probability of inaccuracy. To counter this an empirical study with a large number of participants could be undertaken in the future. Ten of the participants were male and 6 were female. The majority of the photos were taken in Scotland but a reasonable percentage were taken abroad in countries found in Europe, Asia and Oceania.

The results of the online based evaluation can be seen in table 5.2. These results are far less decisive than the ones seen in offline evaluation. While the Flickr Get Recommended system performed best in regards to precision at one and at five the novel system eclipsed all other results in terms of success at 5 by achieving 100% success rate. Flickr Get Recommended also had the highest mean reciprocal rank score which suggests that this system is best overall in online evaluation. However, it should be noted that each system only had about 20 ratings each and so the score displayed in table 5.2 could change dramatically if a much larger sample size was used.

Table 5.2: Online Evaluation Results

	TD-IDF	Flickr Recommended	Combined Co-occurrence	Novel System
P@1	60.00%	73.33%	40.00%	40.00%
P@5	48.00%	58.67%	40.00%	46.67%
S@5	78.57%	85.71%	57.14%	100.00%
MRR	69.44%	80.00%	52.78%	61.00%

5.3 Tagging Trends

The process of creating the co-occurrence matrices, creating the dynamic co-occurrence databases and performing offline evaluation has discovered many tagging trends. Appendix B shows the top co-occurring tags in each static feature. While many of the features used have reasonably generic top co-occurring tags, such as 'clouds' and 'sky', some features have co-occurring tags

specific to their feature and value. For example, when in Europe users tend to tag 'england' and 'uk' in the same photo more often than any other tag pair. Similarly, 'new' and 'zealand' were tagged together more often than any other pair in Oceania. This shows that using continent data as a feature can provide specific and potentially more relevant recommendations than a feature with more generic tags. However, as these features are never used on their own one feature should not be discarded for having generic tags as it may be very useful when utilised in conjunction with other features.

Table 5.3 shows several real world locations alongside the two tags which co-occur most in that locations matrix. It shows that, although all locations contained world famous landmarks, the two top co-occurring tags mainly referred to the location on a larger scale. The three examples have co-occurring tags that refer to the city and the state or country but do not refer to the landmarks directly. This suggests that the areas in which the world is divided into to create these matrices is too large which has led to photos that were not taken in close proximity of the landmark to be included. As the most apparent commonality between images taken of the landmark and images taken elsewhere in the city is the city name itself it is logical to surmise that the reason for highly co-occurring tags lacking in geolocational specificity is due to the areas used in matrix creation being too large. It is hypothesised that the system described in section 6.3 would counteract this problem by introducing variable area sizes.

Table 5.3: Top Co-Occurring Tags for Select Locations

Location	**Top Co-occurring Tags**	
Central Park, New York City	newyork	nyc
Pyramids of Giza, Cairo	cairo	egypt
Eiffel Tower, Paris	france	paris
Coachella Music Festival, Indio, California	california	indio

Appendix C shows the number of photos which occur in each feature value. These percentage values will not always add up to 100% as some photos contain unknown feature values. Europe and North America were the most popular continents to take photos in. This is expected as most of Flickr's user base lives in these continents. The ratio of weekday pictures to weekend pictures was also expected as roughly 40% of days in the week lie within the weekend. Black was the dominant colour in 60% of the photos which correlates with the fact that 'night' occurred highly in many feature values. This also correlates with 63% of all photos being taken without flash enabled. Most images taken of 'night' would be outdoors and so a flash would not be of use. Similarly, many photos were tagged with 'sky' and 'clouds' which would suggest they were taken outdoors during the day and would therefore also not require a flash.

The majority of pictures were taken in landscape orientation which furthers the argument that many Flickr images are taken outdoors with nature being the main subject. This highly correlates with the vast majority of images not featuring any faces and most pictures being taken in the

afternoon. Finally, there is a relatively even split between seasons which shows that Flickr usage does not change throughout the year.

5.4 An Example Image

This is where I will put an example image and I will list recommendations from each system and discuss the results

Chapter 6

Future Work

The system introduced in this paper could be extended in many ways. Some of these ways are highlighted in this chapter while others, such as increasing the amount of dynamic matrix related photos as seen in section 3.8.1, have been described previously in the paper.

6.1 Recommendation From Zero Inputted User Tags

One of the main issues with PTR is that it requires the input of an initial tag in order to recommend others. If this initial tag is considered noise (for example, if it is the photographer's name or the camera model) then the resulting recommended tags are less likely to be accurate. A system that recommends tags from zero initial inputted annotations is hypothesised to be more accurate due to the removal of initial tag noise.

A system has been devised that could help recommend initial tags by using content and context features without the need for inputted tags. For each unique tag T in the tag space the system computes the relevance to the current image I based upon the average value of features found in previous photos containing T. The system will collect the features of I and then compare them to the average features in T by measuring the Euclidean Distance. Each feature in T will be weighted based upon the distribution of the values used to calculate that feature. For instance, the tag "london" would have a very low standard deviation in the geolocation feature as most of the photos will occur in the same geographical area. It will, however, have a high standard deviation in the season feature as photos will be taken at any time of the year in London. A worked example of this can be seen in appendix D. In this example a photo taken near London, during the afternoon in summer with 0 faces in it and a dominant colour of grey would result in the tag "london" being recommended. Although it does not closely match on every feature the ones it does match on have a higher weighting due to their standard deviation and therefore produce a higher result. Details of this process can be seen in algorithm 5.

Algorithm 5: Recommending a Tag With Only Photo Input

for *each tag in the tag space* **do**
    for *Each feature used in the system* **do**
        Find the average value for this feature based on all photos that contain the tag e.g. find all photos tagged with "london" and take the average value of each feature
        Find the standard deviation of the values in each feature
        Weight each feature based on this standard deviation - a higher standard deviation means a lower weighting e.g. the photo was not taken in one but in many locations
    end
end

When a user uploads a new photo extract the features
Measure the distance from these feature values to each one in the tag space
return the tag with the lowest distance

This system has the disadvantage of only having one value per feature per tag which means that London, Ontario would produce "london" as a recommended tag. However, as this system is simply meant to compute the first tag "london" could still be recommended by using the novel system discussed in this paper for subsequent recommendations.

6.2 Live API Data

In addition to taking many holiday photos of landmarks and tourist destinations users may also wish to take pictures while attending events. Tags relating to these events can sometimes be accurately recommended by established PTR systems if they take into account temporal information alongside previously entered tags. For example, someone who takes a picture on December 31st and enters the tag "party" may be recommended the tag "newyear" or "hogmanay". However, many events may not occur on a regular basis. Events such as concerts may occur in a similar location (e.g. an established concert venue) but the band or singer performing will typically change each night. This means that tags would only be accurately recommended for events such as these if a large amount of people had already uploaded photos from the night in question.

A novel system could consult live data obtained from online APIs to determine any events that occurred geographically and temporally close to the image in question and recommends tags based upon this. For example, if the user visits London's O2 Arena the system will decipher that they are attending a concert through geographical information and previously entered tags - known as trigger words. These trigger words will be ones that suggest they are at an event. Examples include "concert", "livemusic" and "nightout".

The system would then consult the online API and determine which act was playing at that venue on that occasion. It will then recommend tags based on this such as "beyonce" or "biffyclyro".

This novel method is similar to the geolocation recommendation method introduced in this paper as it is designed to reduce the reliance on generic tags such as “concert” and increase the amount of event specific tags. This will improve categorisation and searching as users will typically wish to search for photos of a particular band or singer and not simply photos of a concert.

This feature was not introduced in this paper as it requires a great deal of historical concert information which is not currently available to the general public. A database could be created by crawling event websites such as Bandsintown [1] or by creating an app for users to take photos with. This app could then save the relevant concert data upon picture capture.

6.2.1 Visible Sky Tags

If an app were to be created it could also measure items not included in EXIF data. Items such as accelerometer measurements could open up new methods of recommending tags to users through the use of ‘trigger conditions’. These conditions would examine the context of an image and, if the conditions were right, trigger a live data lookup. One situation in which this could be useful is when the user is stargazing. If the system deems the conditions to be right - user situated outside, at night and the accelerometer data shows their phone is pointing upwards - it could consult an API to determine which celestial bodies were visible from the user’s location at that point in time. It could then recommend tags based on this - for example, “venus” or “meteorshower”.

6.3 Adaptable Geolocation Area Sizes

Currently, the system’s geolocation based tag recommendation subsystem is reliant on Flickr’s Places IDs. These IDs normally correspond to a neighbourhood and so their geographical size tends not to change dramatically at any point. This means that users in less populous areas may receive less accurate recommendations as they will have less photos occurring in their area. Similarly, users in very populous areas may find that they are being recommended tags that are not relevant to them as the Flickr Place they captured the image in has many other landmarks or places of interest. If the size of the geographical area for which a co-occurrence matrix is created was dynamically adaptable then this problem would be minimised. For places where many photos are taken, such as the centre of London, the size of each geographical area could be reduced. Conversely, if the user is in the countryside then there may not be many nearby photos and so the size of the geographical area should be increased. The system should not attempt to ensure every area has the same number of photos as this could result in city centre areas being minuscule while their countryside counterparts are too big. There should be minimum and maximum area sizes to ensure this does not happen.

6.4 User-Item Tag Recommendations

User-item-tag recommendations could not be implemented in the novel system due to the online evaluation method and due to the wish for the system to be generalisable for use amongst many social networks. However, if it were to be implemented in the future it could be used to great effect alongside the location features. If the system looked at previous picture activity or a user profile and determined in which area they live then each time they venture outside of this area they could be deemed to be on holiday. Users who are at home will typically tag things in a completely different manner to those who are on holiday. For instance, a user at home is much likely to tag local landmarks while a user on holiday may tag landmarks, tourist destinations and even the destination country. It is hypothesis that by providing an 'on holiday' Boolean user feature that much more accurate recommendations could be made by dividing images up into ones taken on holiday and ones taken at home.

6.5 Dynamic Feature Weighting

It is hypothesised that features could be weighted depending on the tag inputted into the recommendation system. These weights would be dynamically altered depending on the perceived meaning of the tag. For instance, the geolocation feature would be given a higher weighting, and therefore more influence over the final recommendations, if the inputted tag was deemed to be closely related to the location in which the image was taken. If the inputted tag were a place name or a description of a geographical feature such as "building" then it is more likely that a geographical based tag would be deemed as an accurate recommendation. Conversely, location based tags should be recommended less often to tags such as those surrounding concepts, such as "family", or those which have no geographical relevance - for instance the sky can be seen from anywhere and so is not a geographically relevant tag.

Dynamic feature weighting could be implemented by looking for semantic meaning using services such as WordNet or by gathering a list of landmarks and other geographical features from Google Maps. This could be used in conjunction with, or instead of, a machine learning technique. This technique would investigate which tags were selected alongside a given tag in the past and which features (e.g. geolocation, time) gave the matching tags the highest score. These features would then receive a higher weighting in the future when given the same tag as input as they have been proven to provide accurate results.

6.6 Suffix Extraction

One problem with the system of tagging in place in social networks such as Flickr is that no spaces are allowed in tags and so the concatenation of two or more words is frequently employed

by users. This can lead to a smaller amount of recommendations as the tag is too specific. For instance, 'happyoldman' may be tagged by a user instead of using a separate tag for each constituent word. This means that the system will only search for occurrence of 'happyoldman' with other tags and so will most probably return a fewer amount of recommendations. Zero recommendations may be returned in some instances if the tag has never been used before. One way to reduce the effect of this problem is to split the tag into all of its possible constituent words, searching for recommendations for each word, and then combining all recommendations and then sorting by highest co-occurrence. This could be done by extracting all suffixes from the tag and then testing each of these suffixes to see if its a word or proper noun. Resources such as WordNet can help with this process. If its a word or proper noun and does not overlap with any other word in the tag then recommendations should be retrieved for it. If there are words overlapping, for example ,in the tag 'goldengate', 'golden' and 'den' overlap, the longest word should always be taken. Similarly 'gold' and 'den' do not overlap but can be concatenated to form a larger word. This larger word should be the one searched upon.

While this should hypothetically return a greater number of accurate results for tags like 'happyoldman' it could mean that tags such as "goldengate" yield less accurate results. As the system would search for "golden" and "gate" separately it may return recommendations related to the colour or the object and not the landmark. However, the effect of this problem could be reduced by giving preference to tags that co-occur with all constituent words e.g. "bridge" will be recommended by both "golden" and "gate" due to many users uploading images with the trio of tags "golden", "gate" and "bridge".

While the novel system discussed in this paper is already relatively computationally expensive it can be deduced that a suffix extraction system that searches upon each constituent word would be n times more expensive where n is the number of words in the tag. If this system were to be build a great deal of effort should go in to making the process as efficient as possible.

Chapter 7

Conclusion

Bibliography

[1] Bandsintown api documentation. http://www.bandsintown.com/api/overview. Accessed: 31 Aug 2016.

[2] The definition of folksonomy. http://www.vanderwal.net/random/entrysel.php?blog=1750. Accessed: 28 Aug 2016.

[3] How many photos are uploaded to flickr every day, month, year. https://www.flickr.com/photos/franckmichel/6855169886/. Accessed: 14 Aug 2016.

[4] Shane Ahern, Marc Davis, Dean Eckles, Simon King, Mor Naaman, Rahul Nair, Miriana Spasojevic, and Jeannie Yang. Zonetag: Designing context-aware mobile media capture to increase participation. In *Proceedings of the Pervasive Image Capture and Sharing, 8th Int. Conf. on Ubiquitous Computing, California*, 2006.

[5] Morgan Ames and Mor Naaman. Why we tag: motivations for annotation in mobile and online media. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 971–980. ACM, 2007.

[6] Zhineng Chen, Juan Cao, YiCheng Song, Junbo Guo, Yongdong Zhang, and Jintao Li. Context-oriented web video tag recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 1079–1080. ACM, 2010.

[7] Flickr. Flickr api documentation. https://www.flickr.com/services/api/. Accessed: 15 Aug 2016.

[8] George W. Furnas, Thomas K. Landauer, Louis M. Gomez, and Susan T. Dumais. The vocabulary problem in human-system communication. *Communications of the ACM*, 30(11):964–971, 1987.

[9] Nikhil Garg and Ingmar Weber. Personalized tag suggestion for flickr. In *Proceedings of the 17th international conference on World Wide Web*, pages 1063–1064. ACM, 2008.

[10] Jonathan Gemmell, Maryam Ramezani, Thomas Schimoler, Laura Christiansen, and Bamshad Mobasher. The impact of ambiguity and redundancy on tag recommendation in folksonomies. In *Proceedings of the third ACM conference on Recommender systems*, pages 45–52. ACM, 2009.

[11] Scott A Golder and Bernardo A Huberman. Usage patterns of collaborative tagging systems. *Journal of information science*, 32(2):198–208, 2006.

[12] Thorsten Joachims. Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning*, pages 137–142. Springer, 1998.

[13] Lyndon Kennedy, Mor Naaman, Shane Ahern, Rahul Nair, and Tye Rattenbury. How flickr helps us make sense of the world: context and content in community-contributed media collections. In *Proceedings of the 15th ACM international conference on Multimedia*, pages 631–640. ACM, 2007.

[14] Ickjai Lee, Guochen Cai, and Kyungmi Lee. Exploration of geo-tagged photos through data mining approaches. *Expert Systems with Applications*, 41(2):397–405, 2014.

[15] Stefanie Lindstaedt, Roland Mörzinger, Robert Sorschag, Viktoria Pammer, and Georg Thallinger. Automatic image annotation using visual content and folksonomies. *Multimedia Tools and Applications*, 42(1):97–113, 2009.

[16] Cameron Marlow, Mor Naaman, Danah Boyd, and Marc Davis. Ht06, tagging paper, taxonomy, flickr, academic article, to read. In *Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 31–40. ACM, 2006.

[17] Julian McAuley and Jure Leskovec. Image labeling on a network: using social-network metadata for image classification. In *European Conference on Computer Vision*, pages 828–841. Springer, 2012.

[18] Philip J McParlane, Yashar Moshfeghi, and Joemon M Jose. On contextual photo tag recommendation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 965–968. ACM, 2013.

[19] Philip James McParlane. *The Role of Context in Image Annotation and Recommendation*. PhD thesis, University of Glasgow, 2015.

[20] McParlane J. Philip, Moshfeghi Yashar, and Jose M. Joemon. Collections for automatic image annotation and photo tag recommendation. In *ACM MultiMedia Modeling*, 2014.

[21] Adam Rae, Börkur Sigurbjörnsson, and Roelof van Zwol. Improving tag recommendation using social networks. In *Adaptivity, Personalization and Fusion of Heterogeneous Information*, pages 92–99. LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, 2010.

[22] Maryam Ramezani, Jonathan Gemmell, Thomas Schimoler, and Bamshad Mobasher. Improving link analysis for tag recommendation in folksonomies. *Recommender Systems and the Social Web*, page 33, 2010.

[23] Tye Rattenbury, Nathaniel Good, and Mor Naaman. Towards automatic extraction of event and place semantics from flickr tags. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 103–110. ACM, 2007.

[24] Steffen Rendle, Leandro Balby Marinho, Alexandros Nanopoulos, and Lars Schmidt-Thieme. Learning optimal ranking with tensor factorization for tag recommendation. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 727–736. ACM, 2009.

[25] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 175–186. ACM, 1994.

[26] Jennifer Rowley and John Farrow. Organizing knowledge. *An Introduction to Managing Access to Information*, 3:2000–404, 1995.

[27] Rodrygo LT Santos, Craig Macdonald, and Iadh Ounis. Exploiting query reformulations for web search result diversification. In *Proceedings of the 19th international conference on World wide web*, pages 881–890. ACM, 2010.

[28] Börkur Sigurbjörnsson and Roelof Van Zwol. Flickr tag recommendation based on collective knowledge. In *Proceedings of the 17th international conference on World Wide Web*, pages 327–336. ACM, 2008.

[29] Yeran Sun, Hongchao Fan, Mohamed Bakillah, and Alexander Zipf. Road-based travel recommendation using geo-tagged images. *Computers, Environment and Urban Systems*, 53:110–122, 2015.

[30] James Surowiecki. *The wisdom of crowds*. Anchor, 2005.

[31] Karen HL Tso-Sutter, Leandro Balby Marinho, and Lars Schmidt-Thieme. Tag-aware recommender systems by fusion of collaborative filtering algorithms. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1995–1999. ACM, 2008.

[32] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[33] Christian Wartena, Rogier Brussee, and Martin Wibbels. Using tag co-occurrence for recommendation. In *2009 Ninth International Conference on Intelligent Systems Design and Applications*, pages 273–278. IEEE, 2009.

[34] Kilian Quirin Weinberger, Malcolm Slaney, and Roelof Van Zwol. Resolving tag ambiguity. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 111–120. ACM, 2008.

[35] Yi Hsuan Yang, Po Tun Wu, Ching Wei Lee, Kuan Hung Lin, Winston H Hsu, and Homer H Chen. Contextseer: context search and recommendation at query time for shared consumer photos. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 199–208. ACM, 2008.

[36] Zi-Ke Zhang, Tao Zhou, and Yi-Cheng Zhang. Personalized recommendation via integrated diffusion on user–item–tag tripartite graphs. *Physica A: Statistical Mechanics and its Applications*, 389(1):179–186, 2010.

Appendix A

Online Evaluation Webpage

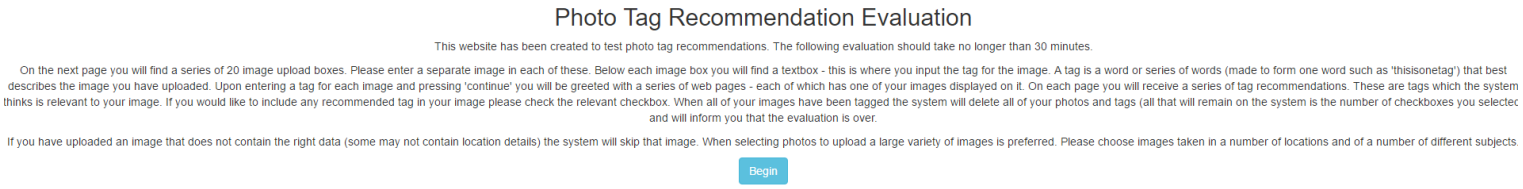

Figure A.1: The Instructions Screen

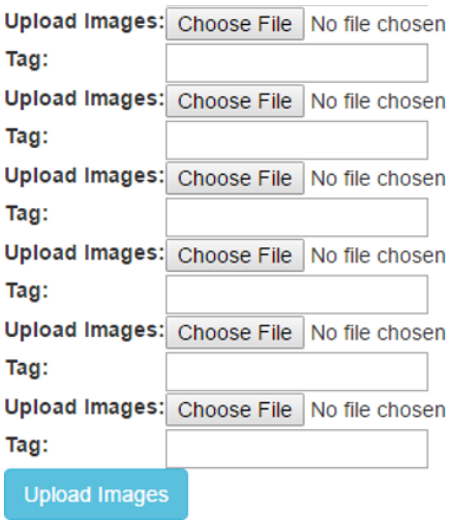

Figure A.2: Upload Images and Enter Tags

Figure A.3: The Recommendations Screen

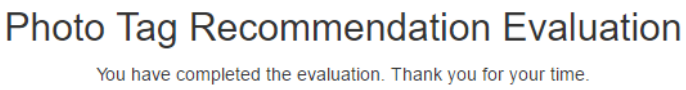

Figure A.4: Confirmation That Evaluation is Complete

Appendix B

Top Tags For Feature Values

Feature	Feature Value	Top Co-occurring tags	
Continent	Europe	england	uk
	Africa	portrait	africa
	Asia	people	asia
	North America	light	night
	South America	brasil	brazil
	Oceania	new	zealand
	Antarctica	No photos taken here	
Day of Week	Weekday	sky	clouds
	Weekend	eos	canon
Dominant Colour	White	black	white
	Black	light	night
	Red	closeup	macro
	Green	nature	green
	Blue	blue	sky
Flash	On	light	night
	Off	nikkor	nikon
Image Orientation	Landscape	clouds	sky
	Portrait	girl	woman
	Square	white	black
Number of Faces	0	sky	clouds
	1	face	portrait
	2	portrait	face
	3+	bar	establishment
Overall	Overall	sky	clouds
Season	Spring	street	urban
	Summer	clouds	sky
	Autumn	fall	autumn
	Winter	night	light
Time of Day	Morning	sky	blue
	Afternoon	fall	black
	Evening	sky	clouds
	Night	black	white

Appendix C

Percentage of Feature Values in Dataset

Feature	Feature Value	Percentage of Photos
Continent	Europe	41.48%
	Africa	1.84%
	Asia	2.35%
	North America	42.14%
	South America	11.38%
	Oceania	0.78%
	Antarctica	0%
Day of Week	Weekday	60.16%
	Weekend	39.84%
Dominant Colour	White	20.10%
	Black	59.82%
	Red	12.74%
	Green	2.32%
	Blue	5.01%
Flash	On	34.12%
	Off	63.78%
Image Orientation	Landscape	68.08%
	Portrait	27.27%
	Square	4.65%
Number of Faces	0	96.32%
	1	3.34%
	2	0.27%
	3+	0.07%
Season	Spring	27.55%
	Summer	26.36%
	Autumn	22.61%
	Winter	23.42%
Time of Day	Morning	16.75%
	Afternoon	35.02%
	Evening	20.34%
	Night	8.93%

Appendix D

Initial Tag Recommendation Table

	Tag		
	London	Sunny	T in the Park
Location	AV - 51.517320, -0.129064 SD - low Almost all tags taken in London	AV -14.587841, -0.718596 SD - high Could be taken anywhere on planet	AV - 56.319497, -3.749423 SD - low Festival that has fixed location
Time	AV - Afternoon SD - medium Most tourist photos taken in afternoon	AV - Afternoon SD - low Sun is at its peak in afternoon	AV - Evening SD - high Photos evenly spread throughout day
Number Faces	AV - 2 faces SD - high Could be of landmark or friends	AV - 1 face SD - medium Usually a nature shot	AV - 3 faces SD - high Could be image of stage or friends
Season	AV - summer SD - high Pictures taken in London year round	AV - summer SD - low Summer is sunniest season	AV - Summer SD - low Festival annually held in July
Dominant Colour	AV - grey SD - medium Most city images are grey	AV - blue SD - medium Could be blue skies but greenery also likely	AV - Green SD - medium Festival set in fields but photos could be of black stage