Personalized tag recommendation based on generalized rules

Personalized tag recommendation based on generalized rules

Luca Cagliero[*], Luigi Grimaudo

Dipartimento di Automatica e Informatica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Torino, Italy

Alessandro Fiori

IRC@C: Institute for Cancer Research at Candiolo, Str. Prov. 142 Km. 3.95 10060 - Candiolo (TO) - Italy

---

Abstract

Tag recommendation is focused on recommending useful tags to a user who is annotating a Web resource. A relevant research issue is the recommendation of additional tags to partially annotated resources, which may be based on either personalized or collective knowledge. However, since the annotation process is usually not driven by any controlled vocabulary, the collections of user-specific and collective annotations are often very sparse. Indeed, the discovery of the most significant associations among tags becomes a challenging task.

This paper presents a novel personalized tag recommendation system that discovers and exploits generalized association rules, i.e., tag correlations holding at different abstraction levels, to identify additional pertinent tags to

[*]Corresponding author. Tel.: +39 0110907084 Fax: +39 0110907099.

Email addresses: `luca.cagliero@polito.it` (Luca Cagliero), `luigi.grimaudo@polito.it` (Luigi Grimaudo), `alessandro.fiori@ircc.it` (Alessandro Fiori)

suggest. The use of generalized rules relevantly improves the effectiveness of traditional rule-based systems in coping with sparse tag collections, because (i) correlations hidden at the level of individual tags may be anyhow figured out at higher abstraction levels and (ii) low level tag associations discovered from collective data may be exploited to specialize high level associations discovered in the user-specific context.

The effectiveness of the proposed system has been validated against other personalized approaches on real-life and benchmark collections retrieved from the popular photo-sharing system Flickr.
Keywords: Tag recommendation, Generalized association, rule mining, Flickr

---

1. Introduction

Recommender systems help users find desirable products or services by analyzing user profiles and their similarities, or by finding products that are similar to those the users expressed interest in. The diffusion of the collaborative tagging systems (e.g., Del.icio.us[1], Flickr[2], Zooomr[3]) has recently focused the attention of the research community on the problem of tag recommendation. Tags are keywords that provide meaningful descriptors of a Web resources. Recommending tags to a user who is annotating a resource is a challenging research issue that has been recently investigated in different real-life contexts (e.g., photo annotation [9, 26], blog post tagging [21],

[1] http://delicious.com Last accessed: 25 June 2012
[2] http://www.flickr.com Last accessed: 25 June 2012
[3] http://www.zooomr.com Last accessed: 25 June 2012

bookmark tagging [12]).

Given a set of user-defined tags, a relevant research issue is the recommendation of additional tags to partially annotated Web resources. Accomplishing this task effectively has the twofold aim at automating the annotation process by suggesting to the user an ordered set of pertinent tags and improving the effectiveness and the efficiency of querying retrieval systems (e.g., [2, 6, 7]. Recommendation of additional tags may be either exclusively based on collective knowledge, i.e., independently of the knowledge about the user who annotated the resources [12, 26, 15], or personalized [9, 21]. To figure out valuable correlations between previously annotated and recommendable tags rule-based approaches have shown to achieve fairly good performance against probabilistic and co-occurrence-based machine learning strategies [12]. To enhance the performance of the tag recommendation systems in the context of photo tag recommendation, the combined usage of user-specific and collective knowledge has also been recently addressed [23]. However, the lack of a controlled vocabulary from which tags could be selected during the annotation process makes the sets of previously assigned annotations very sparse [12, 26] and, thus, unsuitable for being successfully coped with most of the information retrieval and data mining techniques.

This paper presents a novel rule-based recommendation system that addresses the task of recommending additional tags to partially annotated Flickr photos by combining the knowledge provided by the personal and collective contexts, i.e, the history of the past personal and collective photo annotations. To address this issue, it discovers and exploits high level tag correlations, in the form of generalized association rules, from the collections of the past user annotations. To the best of our knowledge, this is the first attempt to exploit generalized rules in tag recommendation. Generalized association rules $X \to Y$ represent correlations among tag sets X and Y such that (i) frequently occur in the analyzed dataset, i.e., the observed frequency (the support) of $X \cup Y$ is above a given threshold, (ii) almost hold in the source data, i.e., the strength of the implication between X and Y (the confidence) is higher than a given threshold, and (iii) may also include items belonging to different abstraction levels (i.e., tags may be generalized as the corresponding categories). The use of tag generalization hierarchies allows the discovery of relevant tag associations that may remain hidden at the level of individual tags. Hence, it may effectively counteract the issue of data sparsity, thus, allowing the recommendation of meaningful and pertinent tags, as shown in the experimental evaluation (see Section 4). In the following the use of generalized rules in tag recommendation is explained with the help of a running example.

Motivating example 1. Consider a photo, published on Flickr, of the Guildhall, which is a famous building situated in the center of London (U.K.). Our goal is to recommend to a given user pertinent additional photo tags to annotate, knowing that his first user-specified annotation is *London*. A graphical representation of the considered use-case is shown in Figure 1. To perform tag recommendation, we exclusively consider, as preliminary step, the collection of the past user-specified annotations (i.e., the personal knowledge base) while temporarily disregarding the collective knowledge provided by annotations made by the other system users. A traditional association rule mining process may discover the rule $\{London\} \to \{Guildhall\}$, where

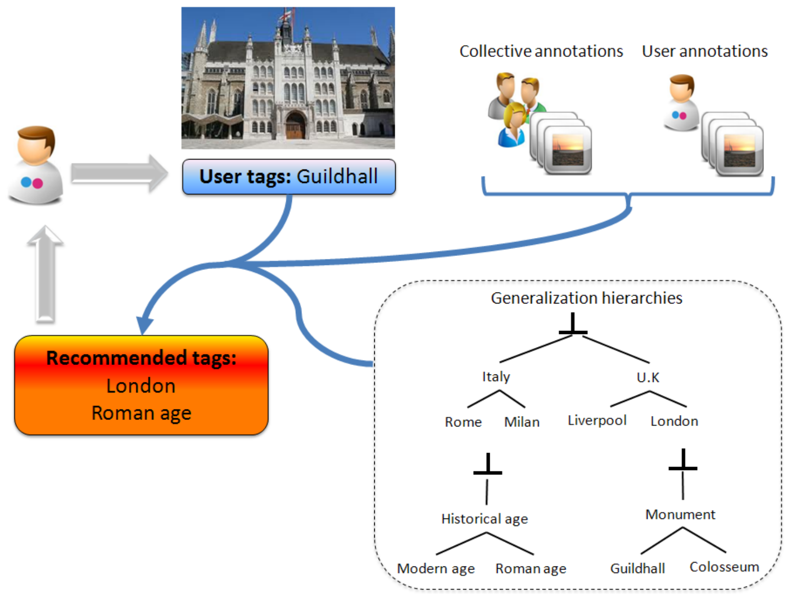


Figure 1: Example of use-case.

London and *Guildhall* are tags. Since the user has already annotated the photo with the tag *London*, *Guildhall* is an example of subsequent tag to recommend. The quality of the proposed recommendation could be evaluated in terms of well-known rule quality indexes (e.g., the rule support and confidence [1]). As discussed in [12], the analysis of the strength of the discovered implications is the core part of rule-based recommendation systems. In particular, frequent and high-confidence rules are deemed the most reliable ones for being used in tag recommendation. Enforcing a minimum frequency of occurrence of the selected rules reduces the sensitivity of the rule-based model to noise and data overfitting, However, data sparsity still makes the discovery of potentially relevant rules a computationally intensive task, because specific rules often occur rarely in the analyzed data [12, 15].

The use of generalization hierarchies built over the history tags, as the ones reported in Figure 1, may allow the generation of high level tag associations that occur more frequently than their low level versions. For instance, by aggregating the tag *London* into the corresponding state *U.K.* the generalized (high level) rule $\{U.K.\} \rightarrow \{Guildhall\}$ may prompt the suggestion of the same annotation while considering a higher level view of the analyzed pattern.

To discriminate among potentially pertinent tags, two distinct rule sets are generated: (i) a *user-specific rule set*, which represents the personalized knowledge base and includes (generalized) rules extracted from the past annotations made by the user to which the recommendation is targeted, and (ii) a *collective rule set*, which represents the collective knowledge and includes (generalized) rules mined from the past annotations made by the other users. Tags mainly referable to user-specific rules are deemed the most suitable ones for additional tag recommendation. However, their significance strictly depends on user activeness and ability in photo tagging [23]. To overcome this issue, in our system we consider tag recommendations based on collective knowledge as well. Collective knowledge also plays a key role in specializing high level associations discovered from the user-specific context, as shown in the following example.

Motivating example 2. Consider again the use-case shown in Figure 1. Suppose now that the first user-specified annotations are *London* and *Roman age*. If the rule $\{London, Roman\ age\} \rightarrow \{Monument\}$ is selected from the user-specified rule set, any descendant of *Monument* (e.g., *Colosseum*, *Guildhall*) is an eligible tag to recommend. The presence in the collective rule set of the rule $\{London, Roman\ age\} \rightarrow \{Guildhall\}$ may push the recommendation of the tag *Guildhall* as deemed worthy of notice by the community.

The effectiveness of the proposed system has been validated on real-life and benchmark photo collections retrieved from Flickr. The use of generalized rules allows significantly improving the performance of state-of-the-art approaches.

This paper is organized as follows. Section 2 overviews most relevant related works concerning tag recommendation and generalization rule mining. Section 3 presents the architecture of the proposed recommender system and describes its main blocks. Section 4 assesses the effectiveness of the system in providing personalized tag recommendations based on both user-defined tags and collective knowledge, while Section 5 draws conclusions and presents future developments of this work.

2. Previous work

The success of social networks and online communities has relevantly increased the attention to the problem of recommending Web resource annotations, i.e., the tags. Tag recommendation systems focus on suggesting tags to a user who is annotating a resource by combining the information coming from one or more contexts. In particular, collective tag recommendation analyzes the knowledge provided by the past resource annotations independently of the user who annotated each resource [12, 26, 15], while personalized tag recommendation addresses tag recommendation by considering the user context [9, 21]. This paper addresses tag recommendation by combining both personalized and collective knowledge.

A significant research effort has been devoted to personalized tag recommendation. For instance, in [21] the author presents a collaborative filtering method to address personalized blog post tag recommendation. It analyzes the information about users behaviors, activities, or preferences to predict what users will like based on their similarity to other users. Analogously to most of the collaborative filtering methods (e.g., [25]), it assumes that similar users share similar tastes. Similarity between posts, users, and tags is evaluated by exploiting information retrieval techniques. In [14] the combination of a graph-based and collaborative filtering method is proposed. A User-Resource-Tag (URT) graph is indexed by means of an ad-hoc indexing strategy derived from the popular PageRank algorithm [4]. To reduce the sparsity of the generated graphs, the use of Singular Value Decomposition (SVD) methods has been also investigated [30]. Differently, the application of content-based strategies has been studied in [16, 5, 17]. They focus on recommending tags that are similar to those that a user annotated in the past (or is annotating in the present). For instance, in [5] the authors present an application for large scale automatic generation of personalized annotations. They automatically select from the main Web page keywords personalized tags based on their relevance to the content of both the considered page and the other documents residing on the surfer's Desktop. Similarly, in [16, 10, 18] multimedia content related to the annotated Web resource is analyzed and used to drive the tag recommendation process. For instance, in [10, 18] the information discovered from both Web page content and related annotations is exploited for tag recommendation purposes, while, in [16], the authors analyze interpersonal relations, image text, and visual content together. Differently, in [17] an hybrid collaborative filtering method is proposed and integrated in a scalable architecture. The issue of interactive Flickr tag recommendation is addressed in [9]. Suggested tags are first selected from the set of previously assigned ones based on co-occurrence measures. Next, based on the recommendation, the candidate set is narrowed down to make the suggestion more specific. However, co-occurrence methods are challenged by data sparsity as either the computational complexity may increase exponentially with the number of tags or the score associated with each tag may be not directly comparable. Unlike previous approaches, to counteract the sparsity of the tag collections this paper proposes to exploit generalized rules.

A parallel issue has been devoted to collective tag recommendation [12, 15, 26, 17]. For instance, in [26], additional tags are recommended to partially annotated Flickr photo by using co-occurrence measures to analyze the collective knowledge. The work proposed in [23] extends the previous system by analyzing the knowledge coming from different contextual layers, including the personal and the collective ones. Differently, authors in [12] reformulate the task of content-based tag recommendation as a (supervised) classification problem. Using page text, anchor text, surrounding hosts, and available tag information as training data, they train a classifier for each tag they want to predict. Even though their approach is able to achieve fairly high precision, the overall training time may become significant when the cardinality of the considered tags increases. This work is also the first attempt to address collective tag recommendation by means of association rules. Association rules allow the discovery of strong tag associations that may be profitably exploited in tag recommendation. Similarly, other approaches (e.g., [19]) focus on rule-based collective tag recommendation. However, the commonly high sparsity of the collections of past annotations limits the effectiveness of the proposed approaches as the most specific (and possibly interesting) rules may remain hidden. This paper proposes to overcome the above issue by discovering tag associations at different abstraction levels. To the best of our knowledge, this is the first attempt to exploit generalized rules in tag recommendation. Authors in [15] also address the same issue by adopting an approach based on Latent Dirichlet Allocation (LDA). The proposed strategy is proved to very effective in tackling the cold start problem for tagging new resources for which no tag has been assigned yet. Differently, this paper specifically addresses personalized tag recommendation of partially annotated resources.

In recent years, a notable research effort has been devoted to discovering generalized association rules from (possibly large) data collections. Generalized association rules have been first introduced in [27] in the context of market basket analysis as an extension of the traditional association rule mining task [1]. By evaluating a set of hierarchies of aggregation built over the data items, items belonging to the source data are aggregated based on different granularity concepts. Each generalized rule, which is a high level representation of a "lower level" rule, provides a higher level view of a pattern hidden in the analyzed data. The first generalized association rule mining algorithm [27] follows the traditional two-process for generalized rule mining: (i) frequent generalized itemset mining, driven by a minimum support threshold, and (ii) generalized rule generation, from the previously mined frequent itemsets, driven by a minimum confidence threshold. Candidate frequent itemsets are generated by exhaustively evaluating the generalization hierarchies. To reduce the complexity and improve the efficiency of the mining process, several optimizations strategies and more efficient algorithms have been proposed [20, 22, 28, 27, 11, 29, 3]. This paper discovers and exploits generalized rules in personalized tag recommendation by adopting an Apriori-based strategy [27] that integrates, as itemset mining step, the approach recently proposed in [3].

3. The recommendation system

This paper presents a novel personalized photo tag recommendation system. Given a photo and a set of user-defined tags, the system proposes novel pertinent tags to assign to the photo based on both the user-specific preferences (i.e., the tags already annotated by the same user to any photo) and the remaining part of collective knowledge (i.e., the annotations provided by the other users). Its main architectural blocks are shown in Figure 2. A brief description of each block follows.

Preprocessing. This block aims at making the collections of the previous tag annotations suitable for the generalized rule mining process. The tag set is tailored to a transactional data format, where each transaction corresponds to the annotations performed by a user to a given photo and includes the corresponding set of assigned tags. Over the history tag collection a set of generalization hierarchies is also derived from the established Wordnet lexical database [32].

Generalized association rule mining. This block focuses on discovering high level tag correlations, in the form of generalized association rules, from the transactional representation of the tag set. The available tag general-

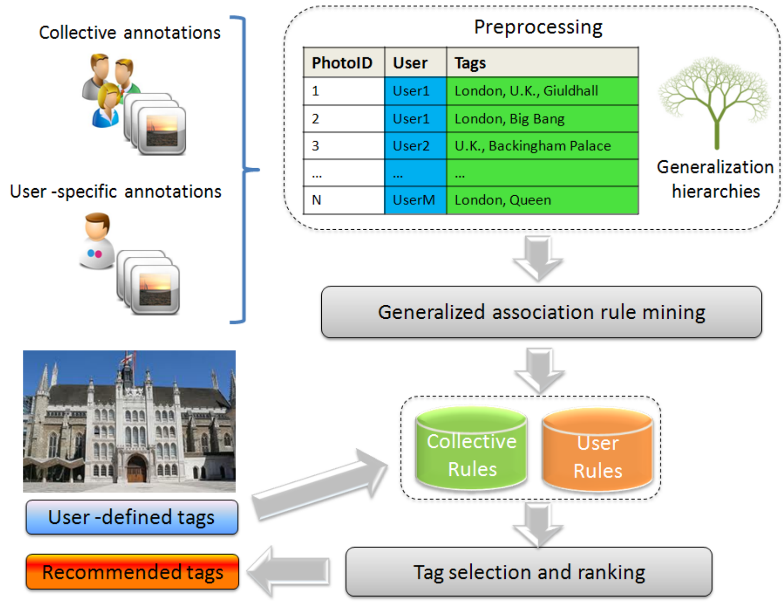


Figure 2: The recommendation system architecture

ization hierarchies are also evaluated to discover tag correlations at different abstraction levels. Two distinct rule sets are generated: (i) a user-specific rule set, which includes generalized rules extracted from the past annotations made by the user to which the recommendation is targeted, (ii) a collective rule set, which includes generalized rules mined from the past annotations made by the other users.

Tag selection and ranking. Given a photo and a set of tags already assigned by the user, this block aims at generating a ranked list of additional tags to suggest. To this aim, from the user-specific and collective rule sets generalized rules pertinent to the already assigned tags are selected. The ranked list of suggested tags is derived from the set of selected rules based on their main quality indexes.

This section is organized as follows. Section 3.1 formally states the recommendation task addressed by this paper, while Sections 3.2, 3.3, and 3.4 thoroughly describe the main blocks of the recommendation system separately.

3.1. Problem statement

Given a set of photos P, a set of tags T, and a set of users U the ternary relation $X = P \times T \times U$ represents the user assignments of tags in T to photos in P. The set $\tau(p_i, u_j) \subseteq T$ includes the tags assigned by user $u_j \in U$ to $p_i \in P$ and could be defined as follows:

$$\tau(p_i, u_j) = \pi_t \sigma_{p_i, u_j} X \tag{1}$$

where π and σ are the commonly used projection and selection primitive operators of the relational algebra [8].

To discriminate between past assignments made by the user u_j and collective ones (i.e., $\neg u_j$), the ternary relation X may be partitioned as follows:

$$X(u_j) = \pi_t \sigma_{u_j} X \tag{2}$$

$$X(\neg u_j) = \pi_t \sigma_{U \setminus u_j} X \tag{3}$$

We denote as user-specific and collective knowledge bases the sets $X(u_j)$ and $X(\neg u_j)$ such that $X(u_j) \cup X(\neg u_j)=X$. Given a set $\tau(p_i, u_j)$ of user-defined tags and the user-specific and collective knowledge bases $X(u_j)$ and $X(\neg u_j)$, the personalized tag recommendation task addressed by this work focuses on suggesting to user u_j new tags in $T \setminus \tau(p_i, u_j)$ for a photo p_i.

3.2. Preprocessing

Flickr is an online photo-sharing system whose resources are commonly annotated by the system users. The analysis of the past photo annotations is crucial for recommending novel tags to users who are annotating a photo. However, data retrieved from the Web is commonly unsuitable for being directly analyzed by means of data mining algorithms. Indeed, a preprocessing step is needed to tailor the retrieved tag sets to a suitable data format.

To enable the association rule mining process, the collection of past Flickr photo annotations is tailored to a transactional data format. A transactional dataset is a set of transactions, where each transaction is a set of items of arbitrary size. To map a tag set to a transactional data format, the annotations made by a user to a given photo are considered as a transaction composed of the set of (not repeated) assigned tags. A more formal definition of the transactional tag set is given in the following.

Definition 1. Transactional tag set. *Let $X = P \times T \times U$ be the ternary relation representing the assignments of tags in T made by users in U to photos in P. Let $\tau(p_i,u_j) \subseteq T$ be the set of all (distinct) tags assigned by user $u_j \in U$ to $p_i \in P$. A transactional tag set $\mathcal{T}$ is a set of transactions, where each transaction corresponds to a set $\tau(p_i, u_j)$ for a certain combination of user $u_j \in U$ and photo $p_i \in P$ occurring in X.*

For instance, if the user u_j assigns to the photo p_i the tags *Guildhall* and *London* the corresponding transaction is $\tau(p_i, u_j)$={*Guildhall, London*}. The transactional tag set $\mathcal{T}$ including the set of all distinct $\tau(p_i, u_j)$ occurring in X is the full list of all past photo annotations.

Given a user u_j to which the personalized tag recommendation is targeted, the transactional tag set $\mathcal{T}$ is partitioned between the annotations made by u_j and not, i.e., distinct transactional representations of $X(u_j)$ and $X(\neg u_j)$, denoted as $\mathcal{T}(\sqcap_{|})$ and $\mathcal{T}(\neg\sqcap_{|})$ throughout the paper, are generated. The separate analysis of $\mathcal{T}(\sqcap_{|})$ and $\mathcal{T}(\neg\sqcap_{|})$ allows the discovery of both user-specific and collective tag associations, in the form of generalized rules.

To enable the process of generalized rule mining from $\mathcal{T}(\sqcap_{|})$ and $\mathcal{T}(\neg\sqcap_{|})$, a set of hierarchies of aggregations (i.e., the generalization hierarchies) is built over the transaction tag set $\mathcal{T}$.

Definition 2. Generalization hierarchy. *Let T be the set of tags occurring in the transactional tag set $\mathcal{T}$. A generalization hierarchy GH built over $\mathcal{T}$ is a predefined hierarchy of aggregations over T. The leaves of GH are all the tags in T. Each non-leaf node in GH is an aggregation of all its children. The root node (denoted as $\perp$) aggregates all the tags occurring in $\mathcal{T}$.*

The Wordnet lexical database [32] is queried to retrieve the most relevant semantic relationships holding between a tag in T and any other term. More specifically, the following semantic relationships are considered: hyponyms (i.e., is-a-subtype-of relationships) and meronyms (is-part-of relationships). Terms to which any selected relationship is directed are considered as generalizations of the original tag. For instance, consider the example tag *London*. If the following semantic relationship is retrieved from the Wordnet database

<London> is-part-of <U.K.>

then the term *London* is selected as the upper level generalization of the tag *U.K.*. Next, the database querying process is deepened to find possible upper level aggregations (e.g., <*U.K.*> *is-part-of* <*Europe*>). The above procedure allows the construction of meaningful generalized hierarchies, according to definition 2, built over a given transactional tag set. Extracts of some example generalization hierarchies are reported in Figure 1. The generalization hierarchies will be used to drive the generalized rule mining process, as described in the following.

3.3. Generalized association rule mining

This block focuses on discovering high level associations, in the form of generalized association rules, from the transactional tag sets $\mathcal{T}(\sqcap_{|})$ and $\mathcal{T}(\neg\sqcap_{|})$. Association rules represent significant correlations among the analyzed data [1]. More specifically, an association rule is an implication $A \Rightarrow B$, where A and B are itemsets, i.e., sets of data items. In the transactional representation of the tag set, items are tags in T associated with any photo included in the collection.

Generalized association rules [27] are rules that may include items at higher levels of abstraction, i.e., the generalized items. By considering the generalization hierarchies built over the transactional tag set (Cf. definition 2), any concept that aggregates one or more tags in T at a higher abstraction level is considered as a generalized item. For instance, consider again the semantic relationship <*London*> *is-part-of* <*U.K.*>. If *London* is a tag (item) that occurs in the transactional tag set, *U.K.* is an example of generalized item. Similarly, generalized itemsets are itemsets (tag sets) including at most one generalized item (e.g., {*Guildhall, U.K.*}). A more formal definition follows.

Definition 3. Generalized itemset. *Let $\mathcal{T}$ be a transactional tag set and T the corresponding item domain, i.e., the set of tags occurring in $\mathcal{T}$. Let $\rho = \{GH_1, \ldots, GH_m\}$ be a set of generalization hierarchies built over $\mathcal{T}$ and E the set of generalized items (high level tag aggregations) derived by all the generalization hierarchies in ρ. A generalized itemset I is a subset of $T \bigcup E$ including at least one generalized item (high level tag aggregation) in E.*

Generalized itemsets are characterized by a notable quality index, i.e., the support, which is defined in terms of the itemset coverage with respect to the analyzed data.

Definition 4. Generalized itemset coverage. *Let $\mathcal{T}$ be a transactional tag set and ρ a set of generalization hierarchies. A (generalized) itemset I covers a given transaction $tr \in \mathcal{T}$ if all its (possibly generalized) items (tags) $x \in I$ are either included in tr or ancestors (generalizations) of items (tags) $i \in tr$ with respect to ρ.*

The support of a (generalized) itemset I is given by the ratio between the number of transactions $tr \in T$ covered by I and the cardinality of $\mathcal{T}$.

A (generalized) itemset I is said to be a descendant of another generalized itemset Y if (i) I and Y have the same length and (ii) for each item $y \in Y$ there exists at least an item $i \in I$ that is a descendant of y.

The concept of generalized association rule extends traditional association rules to the case in which they may include either generalized or not generalized itemsets. A more formal definition follows.

Definition 5. Generalized association rule. *Let A and B be two (generalized) itemsets. A generalized association rule is represented in the form*

$R : A \Rightarrow B$, where A and B are the body and the head of the rule respectively.

A and B are also denoted as antecedent and consequent of the generalized rule $A \Rightarrow B$. Generalized association rule extraction is commonly driven by rule support and confidence quality indexes. While the support index represents the observed frequency of occurrence of the rule in the transactional tag set, the confidence index represents the rule strength.

Definition 6. Generalized association rule support. *Let $\mathcal{T}$ be a transactional tag set and ρ a set of generalization hierarchies. The support of a generalized rule $R : A \Rightarrow B$ is defined as the support (i.e., the observed frequency) of $A \cup B$ in $\mathcal{T}$.*

Definition 7. Generalized association rule confidence. *Let $\mathcal{T}$ be a transactional tag set and ρ a set of generalization hierarchies. The confidence of a rule $R : A \Rightarrow B$ is the conditional probability of occurrence in $\mathcal{T}$ of the generalized itemset B given the generalized itemset A.*

For instance, the generalized association rule $\{$*U.K.*$\} \rightarrow \{$*Guildhall*$\}$ (s=10%,c=88%) states that the tag generalization *U.K.* co-occurs with the tag *Guildhall* in 10% of the transactions (annotations) of the collection and the implication holds in 88% of the cases.

To address generalized association rule mining task [27] from the tag history collections $\mathcal{T}(\sqcap_{|})$ and $\mathcal{T}(\neg\sqcap_{|})$, we performed the traditional two-step process: (i) generalized itemset mining, driven by a minimum support threshold *minsup* and (ii) generalized association rule generation, from the set of previously extracted itemsets, driven by a minimum confidence threshold *minconf*. A generalized association rule is said to be *strong* if it satisfies both *minsup* and *minconf*.

Given a set of generalization hierarchies built over the tags in X, a minimum support threshold *minsup*, and a minimum confidence threshold *minconf*, the generalized rule mining process is performed on $\mathcal{T}(\sqcap_{|})$ and $\mathcal{T}(\neg\sqcap_{|})$ separately. More specifically, given a photo p_i, a user u_j, and a set of user-specific tags $\tau(p_i,u_j)$, the main idea behind our approach is to treat strong high level correlations related to the annotations made by the user u_j differently from that made by the other users. To this aim, two distinct rule sets are generated: (i) a *user-specific rule set*, which includes all strong generalized rules extracted from the past annotations made by the user to which the recommendation is targeted, (ii) a *collective rule set*, which includes all strong generalized rules mined from the past annotations made by the other users. To accomplish the generalized itemset mining task efficiently and effectively, we exploit our implementation of a recently proposed mining algorithm, i.e., the GenIO algorithm [3]. A brief description of the adopted algorithm is given in Section 3.3.1.

3.3.1. The GenIO *Algorithm*

GenIO [3] is a generalized itemset mining algorithm that addresses the discovery of a smart subset of all the possible frequent (generalized) itemsets. Given a source dataset, a set of generalization hierarchies ρ, and a minimum support threshold *minsup* it discovers all frequent not generalized itemsets and all frequent generalized itemsets having at least an infrequent descendant, i.e., a descendant that does not satisfy *minsup*. To achieve this goal, the generalization process is support-driven, i.e., it generalizes an itemset only if it is infrequent with respect to the minimum support threshold. A more through description of the main algorithm steps follows.

GenIO is an Apriori-based algorithm [1] that performs a level-wise itemset generation. More specifically, at arbitrary iteration k, the Apriori-based itemset mining steps are the following: (i) candidate generation, in which all possible k-itemsets are generated from the $(k-1)$-itemsets and (ii) candidate pruning, which is based on the property that all the subsets of frequent itemsets must also be frequent in the source data, to early discard candidate itemsets that cannot be frequent. Candidate generation is known to be the most computationally and memory intensive step The actual candidate support value is counted by performing a dataset scan. GenIO follows the same level-wise pattern. However, it manages rare itemsets by lazily evaluating the given generalization hierarchies. The generalization process is performed by applying on each item (tag) contained in an (infrequent) itemset I the corresponding generalization hierarchies. All itemsets obtained by replacing one or more items in I with their generalized versions are generalized itemsets of I. Hence, the generalization process on itemset I potentially generates a set of generalized itemsets. The generalization process of I is triggered if and only if I is infrequent with respect to the minimum support threshold. Since the GenIO algorithm has been first proposed in the context of structured datasets, a few straightforward modifications to the original algorithm have been adopted to make it applicable to transactional data as well.

3.3.2. Rule generation

The generalized rule generation task entails the discovery of all generalized association rules satisfying a minimum confidence threshold *minconf*, starting from the set of frequent (generalized) itemsets discovered by the GenIO algorithm.

The proposed recommendation system accomplishes the rule generation task by performing the second step of the traditional Apriori algorithm [1]. To achieve this goal, we exploited our more efficient implementation of the generalized rule generation procedure first proposed in [27].

3.4. Tag selection and ranking

Given a photo p_i, a set of user-defined tags $\tau(p_i,u_j)$ assigned by user u_j to p_i, and the sets of generalized rules $R_{\mathcal{T}(\sqcap_|)}$ and $R_{\mathcal{T}(\neg\sqcap_|)}$ mined, respectively, from $\mathcal{T}(\sqcap_|)$ and $\mathcal{T}(\neg\sqcap_|)$, this block entails the selection and the ranking of the additional tags to recommend to u_j for p_i. For the sake of clarity, in the following we discuss how to effectively tackle the selection and ranking problems separately.

3.4.1. Selection

The selection step focuses on selecting additional tags to suggest to user u_j for the partially annotated photo p_i from the rules belonging to the user-specific and the collective rule sets $R_{\mathcal{T}(\sqcap_|)}$ or $R_{\mathcal{T}(\neg\sqcap_|)}$. A pseudo-code of the selection procedure is given in Algorithm 1.

To select tags that are strongly associated with the user-specified ones, only a subset of the extracted rules is deemed worth considering for additional tag recommendation. More specifically, the strong generalized rules in $R_{\mathcal{T}(\sqcap_|)}$ and $R_{\mathcal{T}(\neg\sqcap_|)}$ whose rule antecedent covers, at any level of abstraction, the user-specified tag set $\tau(p_i,u_j)$ or any of its subsets are selected and included in the corresponding rule sets $covered\_rules(u_j)$ and $covered\_rules(\neg u_j)$ (see

Algorithm 1 Tag selection

Input: the user-specific rule set $R_{\mathcal{T}(\sqcap_|)}$, the collective rule set $R_{\mathcal{T}(\neg\sqcap_|)}$, and the user-specified tags $\tau(p_i,u_j)$

Output: the tag selection C

1: covered_rules(u_j) = select_pertinent_user-specific_rules($R_{\mathcal{T}(\sqcap_|)}$, $\tau(p_i,u_j)$)
2: covered_rules($\neg u_j$) = select_pertinent_collective_rules($R_{\mathcal{T}(\neg\sqcap_|)}$, $\tau(p_i,u_j)$)
3: **for all** user-specific rules R in covered_rules(u_j) **do**
4:     insert tags in $R.consequent$ into C
5:     **for all** generalized tags g in C **do**
6:         **for all** collective rules R_2 in covered_rules($\neg u_j$) **do**
7:             **if** $R_2.consequent$ includes any tag t^* in $g.leafdescendant$ **then**
8:                 insert t^* in C
9:             **end if**
10:         **end for**
11:     **end for**
12: **end for**
13: remove generalized tags from C
14: **return** C

lines 1-2). According to Definition 4, the coverage of (a portion of) the tag set $\tau(p_i,u_j)$ may be due to the presence in the rule antecedent of either an exact matching (i.e., the same tags) or one of its generalized versions. Any rule that does not fulfill the above-mentioned constraint is not considered in subsequent analysis.

Consider, for instance, a photo p_i annotated by the user u_j with the tag *London*. In Table 1 is reported the selection of generalized rules taken from the set of rules mined from, respectively, the past user annotations $T(u_j)$ and $T(\neg u_j)$ by exploiting the generalization hierarchies reported in Figure 1 and by enforcing, respectively, a minimum support threshold equal to 1% and a minimum confidence threshold equal to 50%. Notice that any selected rule

Table 1: Generalized rules used for recommending to user u_j tags subsequent to *Rome*.

ID	**Generalized rule**	**Support** (%)	**Confidence** (%)
Annotations made by user u_j			
1	$\{London\} \Rightarrow \{Guildhall\}$	2.5%	100%
2	$\{London\} \Rightarrow \{Historical\ age\}$	1.4%	85%
3	$\{U.K.\} \Rightarrow \{Royal\ family\}$	1.8%	91%
Annotations made by the other users			
4	$\{London\} \Rightarrow \{Guildhall, Royal\ family\}$	1.5%	95%
5	$\{U.K.\} \Rightarrow \{Roman\ Age\}$	1.3%	80%
6	$\{London\} \Rightarrow \{Tourism\}$	1.2%	72%

contains the tag *London* or its generalization *U.K.* as rule antecedent. Consider now the case in which the set of user-specified tags $\tau(p_i,u_j)$ is {*London*, *Roman age*}. Rules including either {*London*, *Roman age*}, {*U.K.*, *Roman age*}, {*London*, *Historical age*}, or {*U.K.*, *Historical age*} as rule antecedent are considered as well together with that covering only one of the user-specified tags *London* or *Roman age* or their relative generalizations.

Not generalized tags belonging to the consequent of the selected user-specific or collective rules in $R_{\mathcal{T}(\sqcap_|)}$ or $R_{\mathcal{T}(\neg\sqcap_|)}$ are eligible tags to recommend. Since we consider the tag associations mainly referable to the user-specific context the most reliable ones for personalized tag recommendation, we first select the collection C of generalized and not generalized tags contained in the consequent of any rule in $R_{\mathcal{T}(\sqcap_|)}$ (line 4). Then, we refine the selection by replacing generalized tags with the most pertinent not generalized descendants derivable from the collective knowledge base (line 8).

Recalling the previous example, the set C of candidate tags is first initialized as follows: $\{$*Guildhall*, *Historical age*, *Royal family*$\}$. Readers could notice that *Guildhall* and *Royal family* are tags, while *Historical age* is an upper level generalization. Since the generalization *Historical age* could not directly recommended, it is replaced with one (or more) of its low level tags. The selection of the eligible descendants of any generalization in C is driven by the collective knowledge. For instance, since, among the two low level descendants of *Historical age* (i.e., the tags *Roman Age* and *Modern age*), only *Roman Age* occurs at least once in the consequent of any of the selected collective rules in $R_{\mathcal{T}(\neg\sqcap_{|})}$ (see Table 1), the tag *Historical age* is exclusively replaced by its leaf descendant *Roman Age*, as it is strongly recommended by the community.

The selection procedure performs two nested loops. The outer loop (lines 5-11) iterates over the generalizations occurring in the candidate set C, while the inner one (lines 6-10) iterates over the collective rule sets and selects the leaf descendants of any generalization in C. While leaf descendants are included in C as pertinent additional tags to recommend (line 8), any generalization in C is discarded (line 13). Finally, the updated set C of selected candidate tags is returned (line 14).

3.4.2. Ranking

The last but not the least task in tag recommendation is the ranking of the candidate recommendable tags in C. Tag ranking should reflect (i) the tag significance with respect to the user-defined tags in $\tau(p_i,u_j)$, (ii) the tag relevance according to the past user-specific preferences, and (iii) the tag relevance based on the past collective knowledge related to other system users.

To evaluate the significance with respect to $\tau(p_i,u_j)$ we propose a tag ranking strategy that considers the interestingness of the rules in $R_{\mathcal{T}(\sqcap_|)}$ and $R_{\mathcal{T}(\neg\sqcap_|)}$ from which they have been selected. Generalized rule interestingness is evaluated in terms of its confidence index value [1], i.e., the rule strength in the analyzed dataset (Cf. Definition 7) in both the personal and collective knowledge base.

Formally speaking, let $c \in C$ be an arbitrary candidate tag and $R^c_{\mathcal{T}(\sqcap_|)} \subseteq R_{\mathcal{T}(\sqcap_|)}$, $R^c_{\mathcal{T}(\neg\sqcap_|)} \subseteq R_{\mathcal{T}(\neg\sqcap_|)}$ be, respectively, the subsets of rules in $R_{\mathcal{T}(\sqcap_|)}$ and $R_{\mathcal{T}(\neg\sqcap_|)}$ whose antecedent covers c (at any level of abstraction). The ranking score of c in $\mathcal{T}(\sqcap_|)$ and $\mathcal{T}(\neg\sqcap_|)$ is defined as the average confidence of the rules in $R^c_{\mathcal{T}(\sqcap_|)}$ and $R^c_{\mathcal{T}(\neg\sqcap_|)}$, respectively.

$$rankscore(c, \mathcal{T}(\sqcap_|)) = \frac{\sum_{\nabla_{\sqcap_|} \in \mathcal{R}^{\rfloor}_{\mathcal{T}(\sqcap_|)}} \rfloor \wr \backslash \{ (\nabla_{\sqcap_|})}{|\mathcal{R}^{\rfloor}_{\mathcal{T}(\sqcap_|)}|}$$

$$rankscore(c, \mathcal{T}(\neg\sqcap_|)) = \frac{\sum_{\nabla_{\neg\sqcap_|} \in \mathcal{R}^{\rfloor}_{\mathcal{T}(\neg\sqcap_|)}} \rfloor \wr \backslash \{ (\nabla_{\sqcap_|})}{|\mathcal{R}^{\rfloor}_{\mathcal{T}(\neg\sqcap_|)}|}$$

Roughly speaking, the ranking scores $rankscore(c, \mathcal{T}(\sqcap_|))$ and $rankscore(c, \mathcal{T}(\neg\sqcap_|))$ reflect the average significance of the tag c in the personal and collective contexts. To combine the individual tag ranks achieved in different contexts in a unified ranking list we adopted an aggregation method based on the Borda Count group consensus function [31]. The chosen approach first assigns descending integer scores to the elements of each individual rank and then combines the voting scores to generate a unique ranking. To effectively deal with ranking lists of different lengths, in our Borda Count implementation we assign to the first element of each rank the same value equal to the length of the longest of all the input ranks.

The recommendation system returns the ranked list of candidate tags in C produced by the Borda Count method.

4. Experimental results

We performed a large set of experiments addressing the following issues: (i) a performance comparison between our system and a set of recently proposed methods, (ii) a discussion about the impact of the generalization process on the recommendation performance, (iii) an analysis of a real-life use-case for our system and the discovered generalized tag associations, and (iv) the analysis of the impact of the main system parameters on the recommendation performance.

This section is organized as follows. Section 4.1 describes the characteristics of the photo collections exploited in the experimental evaluation. Section 4.2 presents the experimental design and introduces the evaluation metrics adopted for performance evaluation. Section 4.3 compares the results achieved by our system with both different versions of the recently proposed approach [23] and a baseline version of our system that does not exploit generalized rules. Section 4.4 validates the applicability of our approach on an example of real-life use-case. Finally, Section 4.5 analyzes the impact of the main system parameters and the data distribution on the recommendation performance.

4.1. Photo collections

To test recommendation system performance, we used a benchmark and a real-life dataset.

The used benchmark dataset is the MIR Flickr 2008 image collection, which was offered by the LIACS Medialab at Leiden University and introduced by the ACM MIR Committee in 2008 [13]. It collects 25,000 images and the related annotating users and tags.

The real-life collection is generated by retrieving, by means of the Flickr APIs, 5,000 real photos. The selected photos were chosen based on a series of high level geographical topics, i.e., *New York*, *San Francisco*, *London*, and *Vancouver*. The retrieved dataset is made available for research purposes.

Since for both the benchmark and the real-life datasets the majority (i.e, around 80%) of the contained photos have at least 5 tags, to perform a fair performance evaluation (see Section 4.2) we focus our analysis on this photo subset.

By following the strategy described in Section 3.2 a set of generalization hierarchies is derived from the Wordnet lexical database over the collected photo tags. A portion of one of the generated generalization hierarchies is reported in Figure 3.

4.2. Experimental design

Our system retrieves a ranked list of pertinent additional tags based on the extracted frequent generalized rules to tackle the tag recommendation ranking problem. Given a photo p_i and a set of user-defined tags $\tau(p_i,u_j)$, the system has to recommend tags that describe the photo based on both

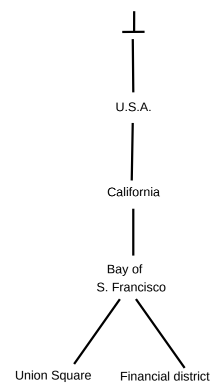


Figure 3: Portion of an example generalization hierarchy built over the photo collection tags

user-specific and collective past annotations. To perform personalized recommendation, from both the tested photo collections the user-specific annotations made by 10 users who annotated at least 15 photos are considered separately. Once a user-specific annotation subset is selected, the rest of the collection is considered as the collective set. For each analyzed user collection, the evaluation process performs a hold-out train-test validation, i.e., the user-specific collection is partitioned in a training set, including the 75% of the whole annotations, whereas the remaining part is chosen as test set. To evaluate the additional tag recommendation performance of our system, for each test photo two random tags are selected as initial (user-specified) tag set and the recommended tag list is compared with the held-out test tags. A recommended tag is judged as correct if it is present in the held-out set. Since held-out tags need not to be the only tags that could be assigned to the photo, the evaluation method actually gives a lower bound of the system performance.

To evaluate the performance of both our recommendation system and its competitors, we exploited three standard information retrieval metrics, previously adopted in [26, 23] in the context of additional Flickr tag recommendation. The selected measures are deemed suitable for evaluating the system performance at different aspects. Let Q be the set of relevant tags, i.e. the tags really assigned by the user to the test photo, and C the tag set recommended by the system under evaluation. The adopted evaluation measures are defined as follows.

Mean Reciprocal Rank (MRR). This measure captures the ability of the system to return a relevant tag (i.e., a held-out tag) at the top of the ranking. The measure is averaged over all the photos in the testing collection and is computed by:

$$MRR = max_{q \in Q} \frac{1}{c_q} \tag{4}$$

where c_q is the rank achieved by the relevant tag q.

Success at rank k (S@k). This measure evaluates the probability of finding a relevant tag among the top-k recommended tags. It is averaged over all the test photos and is defined as follows:

$$S@k = \begin{cases} 1 & \text{if } Q \cap C_k \neq \emptyset, \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $q \in Q$ is a relevant tag and C_k is the set of the top-k recommended tags.

Precision at rank k (P@k). This metric evaluates the percentage of relevant tags over the set of retrieved ones. The measure, averaged over all test photos, is defined as follows:

$$P@k = \frac{|Q \cap C_k|}{|Q|} \tag{6}$$

Notice that the combined use of precision and success highlights the system ability to get a set of tags that is globally appreciable from the user's point of view, while MRR measures the quality of the top tag selection. To perform a fair evaluation, on each test photo measure estimates are averaged over several runs, where, within each run, a different (randomly generated) held-out tag set ranking is considered.

4.3. Performance comparison

The aim of this section is twofold. First, it experimentally demonstrates the effectiveness of our system against a state-of-the-art approach. Secondly, it evaluates the impact of the generalization process on the recommendation performance. To achieve these goals, we compared the performance of our system, in terms of the evaluation metrics described in Section 4.2, on both benchmark and real-life datasets with: (i) five different variants of the recently proposed personalized Flickr tag recommendation system [23], which specifically addresses the problem of additional photo tag recommendation given a set of user-specified tags, and (ii) a baseline version of our approach, which does not exploit generalized knowledge.

The system presented in [23] is a personalized recommender system that proposes additional photo tags, pertinent to a number of different user contexts, among which the personal and the collective ones. The system generates a list of recommendable tags based on a probabilistic co-occurrence measure for each context and then aggregates the results achieved within each context in a final recommended list by exploiting the Borda Count group consensus function [31]. To the best of our knowledge, it is the most recent work proposed on the topic of personalized additional Flickr tag recommendation. To perform a fair comparison, we evaluated the performance of the approach presented in [23] (denoted as *Probabilistic prediction* in the following) when coping with the combination of collective and personalized contexts. Moreover, within each context (personalized or collective), we tested different co-occurrence measures as well. More specifically, we also integrated and tested four co-occurrence measures, i.e., Sum, $Vote$, Sum^+ (Sum + Promotion), and $Vote^+$ (Vote + Promotion), previously proposed by the same authors in [26] in the context of collective additional tag recommendation. The additional measures are taken as representatives of different co-occurrence measures that could be adopted to aggregate and select tags pertinent to each context.

To demonstrate the usefulness of generalized rules in tag recommendation, we also compared the performance of our system with that of a baseline version, which exploits traditional (not generalized) association rules [1] solely. More specifically, the baseline method performs the same steps of the proposed approach, while disregarding the use of tag generalizations in discovering significant tag associations (see Section 3.4.1).

To test the performance of our approach we consider as standard configurations for the tested datasets the following settings: $minsup$=50% and $minconf$=40% for the real-life dataset and $minsup$=20% and $minconf$=35% for the benchmark dataset. A more detailed analysis of the impact of the above-mentioned parameters on the proposed recommendation performance is reported in Section 4.5. Even for the baseline version of our system we tested several support and confidence threshold values. For the sake of brevity, in the following we select as representative and report just the configuration that achieved the best results in terms of MMR measure (i.e., minimum support and confidence thresholds equal to 50%).

The overall results achieved by the performance evaluation session on the real-life and the benchmark datasets are summarized in Tables 2 and 3, respectively. They report the success and the precision at ranks from 1 to 5 (i.e., S@k, P@k k $\in$ [1,5]) as well as the Mean Reciprocal Rank (MRR) achieved by both our system and all the tested competitors. Similarly to what previously done in [26, 23], for the sake of brevity we choose not to report ranks with k higher than 5. To validate the statistical significance of the achieved performance improvements the Student t-test has been adopted [24] by using as p-value 0.05. Significant worsening in the comparisons between our system and the other tested competitors are starred in Tables 2 and 3. For each tested measure, the result(s) of the best system(s) is written in boldface.

Our recommendation system significantly outperforms both its baseline version and all the other tested competitors in terms of MRR, S@1, S@2, and P@k (for any tested value of k) on the real-life dataset and in terms of MRR, S@k, and P@k for $k > 1$ on the benchmark dataset. Furthermore, it performs as good as *Probabilistic prediction* [23], $Vote^+$, Sum, and Sum^+ in terms of S@k for $k \geq 3$ on the real-life dataset and as good as *Probabilistic prediction* in terms of P@1/S@1 on the benchmark dataset. Performance improvements in terms of P@k remain statistically significant for for any

$k \leq 9$ on the real-life dataset, while in terms of P@k and S@k they are significant for any tested value of k in the range [2,10] on the benchmark dataset.

To have a deep insight into the achieved results, in Figures 4 and 5 we also plot the variations of the precision and the success at rank k by varying k in the range [1,5] for the real-life and the benchmark datasets, respectively. Results achieved on the real-life crawled data show that our approach performs best for any tested value of k in terms of precision at rank k (see Figure 4(b)). Furthermore, it also performs best for k equal to 1 and 2 in terms of success, while its performance is comparable to the one of the other approaches for $k \geq 3$. A slightly different performance trend comes out on the benchmark dataset. Our system is slightly less accurate than its best competitor in first tag prediction, while it performs significantly better than all the others (including *Probabilistic prediction*) in recommending all the subsequent tags.

In summary, results show that our approach, on average, selects the most suitable recommendable tags at the top of the ranking and precisely identify the potential user interests.

4.4. Real-life use-case

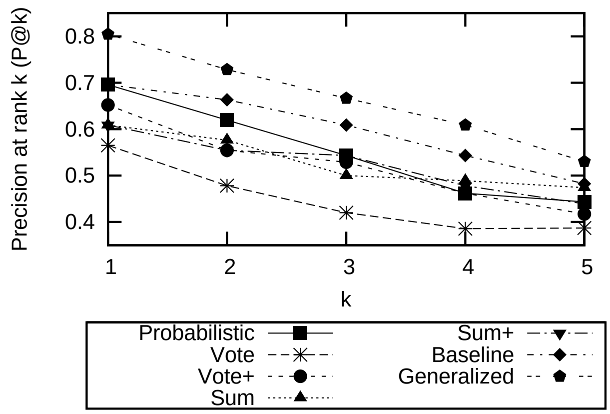
In this section we analyze the results achieved by our system in a real-life use-case. Consider a user that is annotating a Flickr photo of the St.

Table 2: Real-life dataset. Performance comparison in terms of S@k, P@k, and MRR metrics. Statistically relevant worsening in the comparisons between our system and the other approaches are starred.

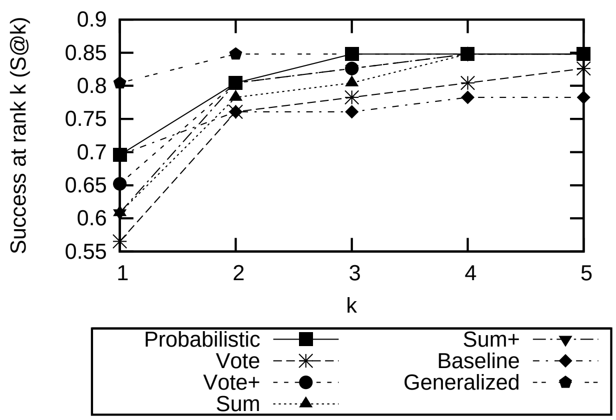
	Probabilistic Prediction	**Vote**	**Vote^{+}**	**Sum**	**Sum^{+}**	**Baseline**	**General**
				Precision at rank k			
P@1	0.6956*	0.5652*	0.6521*	0.6086*	0.6086*	0.6956*	
P@2	0.6195*	0.4782*	0.5543*	0.5760*	0.5543*	0.6630*	
P@3	0.5434*	0.4202*	0.5289*	0.5000*	0.5434*	0.6086*	
P@4	0.4619*	0.3858*	0.4619*	0.4891*	0.4782*	0.5434*	
P@5	0.4434*	0.3869*	0.4173*	0.4739*	0.4391*	0.4826*	
				Success at rank k			
S@1	0.6956*	0.5652*	0.6521*	0.6086*	0.6086*	0.6956*	
S@2	0.8043*	0.7608*	0.8043*	0.7826*	0.8043*	0.7608*	
S@3	**0.8478**	0.7826*	0.8260	0.8043*	0.8260	0.7608*	
S@4	**0.8478**	0.8043	**0.8478**	**0.8478**	**0.8478**	0.7826*	
S@5	**0.8478**	0.8260	**0.8478**	**0.8478**	**0.8478**	0.7826*	
				MRR			
	0.7681*	0.6837*	0.7429*	0.7159*	0.7219*	0.7337*	

Table 3: Benchmark dataset. Performance comparison in terms of S@k, P@k, and MRR metrics. Statistically relevant worsening in the comparisons between our system and the other approaches are starred.

	Probabilistic Prediction	**Vote**	**Vote$^+$**	**Sum**	**Sum$^+$**	**Baseline**	**General**
				Precision at rank k			
P@1	**0.7660**	0.4468*	0.4894*	0.4681*	0.4894*	0.5319*	
P@2	0.6809	0.3936*	0.4255*	0.3936*	0.4255*	0.4787*	
P@3	0.6170*	0.3333*	0.3759*	0.3475*	0.3789*	0.4468*	
P@4	0.5638*	0.2979*	0.3298*	0.3032*	0.3298*	0.3989*	
P@5	0.4978*	0.2681*	0.2851*	0.2638*	0.2851*	0.3574*	
				Success at rank k			
S@1	**0.7660**	0.4468*	0.4894*	0.4681*	0.4894*	0.5319*	
S@2	0.7872*	0.4894*	0.5106*	0.4894*	0.5106*	0.5957*	
S@3	0.8298*	0.5106*	0.5319*	0.5106*	0.5319*	0.5957*	
S@4	0.8298*	0.5106*	0.5319*	0.5106*	0.5319*	0.5957*	
S@5	0.8298*	0.5106*	0.5319*	0.5106*	0.5319*	0.5957*	
				MRR			
	0.7908*	0.4752*	0.5071*	0.4858*	0.5071*	0.5638*	

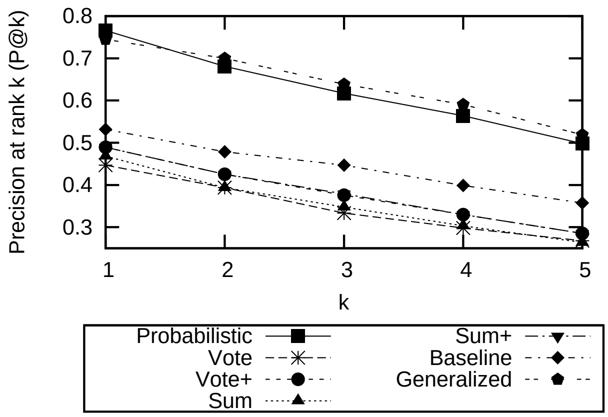


(a) Precision at rank k (P@k).

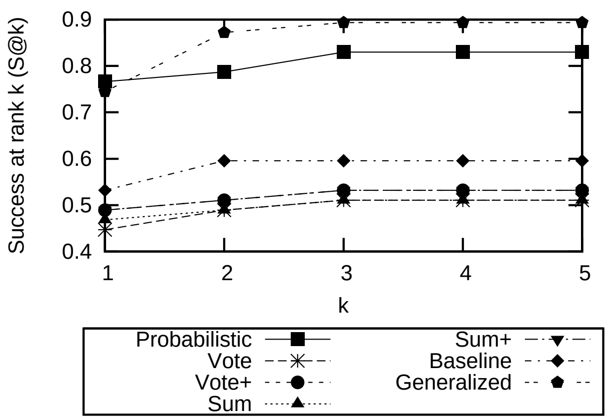


(b) Success at rank k (P@k).

Figure 4: Real-life dataset. Performance comparison by varying the reference rank k.



(a) Precision at rank k (P@k).



(b) Success at rank k (P@k).

Figure 5: Benchmark dataset. Performance comparison by varying the reference rank k.

Mary Church, located at the Financial District of San Francisco (California, U.S.A.) nearby the Financial Center. The photo is taken from the real-life photo collection described in Section 4.1. Over the photo annotations a set of generalization hierarchies, whose extract is shown in Figure 3, is built by our recommendation system (see Section 3.2).

The user is interested in tagging the photo with good descriptors so that the Flickr querying system may effectively retrieve its content based on the user-provided information. Suppose that the user has already annotated the photo with the following tags $\tau(p_i, u_j)$={*St. Mary Square, Financial District*}. The system analyzes the user-specific and collective knowledge bases to suggest additional tags to recommend. By setting the standard configuration (minimum support threshold *minsup*=50%, minimum confidence threshold $minconf$=40%) the following strong rule is discovered by our system from the collective transactional tag set:

A) {*St. Mary Square, Financial District*} $\Rightarrow$ {*Financial center*} (support = 40%, confidence = 100%).

Hence, *Financial center* is a candidate additional tag to recommend suggested by the community. However, due to the sparsity of the user-specific knowledge base none of the not generalized rules includes {*St. Mary Square, Financial District*} as rule antecedent since the combination of the two tags rarely occurs in the analyzed collection. Nevertheless, the following strong generalized rules are extracted:

B) {*San Francisco Bay, Financial District*} $\Rightarrow$ {*St. Mary*} (support = 42%, confidence = 99%)

C) {*San Francisco Bay, Financial District*} $\Rightarrow$ {*Business*} (support = 55%, confidence = 100%)

Both rules represent correlations between the previously assigned and the potentially relevant future tag annotations at a higher abstraction level. Rule (A) suggests the recommendation of the pertinent tag *St. Mary* as additional tag, while rule (B) highlights a high level tag category that is worth considering in the recommendation process. In particular, the latter rule states that, among the past user annotations, a correlation between the category *Business* and the previously annotated tags holds. Indeed, the user would willingly annotate the photo with a tag belonging to that category. The knowledge about the community behavior addresses the system to recommend the tag *Financial center* as it is a lower level descendant of the category *Business*.

4.5. Parameter analysis

We also analyzed the impact of the main system parameters on the tag recommendation performance. To this aim, in Figures 7(a) and 7(b) we plot the average MRR, S@1/P@1, and P@5 measures, as representatives among all the tested measures (see Section 4.2), achieved by our system on the real-life collection by varying the minimum support and confidence threshold enforced during the generalized rule mining process, respectively. Curves, not reported here for the sake of brevity, relative to different evaluation measures and dataset show similar trends.

When relatively high support thresholds (e.g., 70%) are enforced, the percentage of not generalized rules is quite limited (e.g., 13% of the user-specific rule set mined from the training photo collection described in Section 4.1) and many informative rules (generalized and not) are discarded. Nevertheless, the use of generalizations may prevent the discarding of the most informative recurrences thanks to the extraction of high level associations from the user-specific knowledge base. In the opposite case, i.e., when relatively low support thresholds (e.g., 20%) are enforced, many low level tag associations become frequent (e.g., 1.0% of the user-specific rule set from the same training data) and, thus, are extracted by our system. However, the sparsity of the analyzed tag collections still left some of the most peculiar associations among tags hidden. Aggregating tags into high level categories allows achieving the best balancing between specialization and generalization of the discovered associations and, thus, improves recommender system performance.

The confidence threshold may slightly affect the recommendation system performance. By enforcing very low confidence threshold values (e.g., 30%), a large amount of (possibly misleading) low-confidence rules is selected. Indeed, the quality of the rule-based model, at the top of which the recommendation system is built, worsens. Differently, when increasing the confidence threshold a more selective pruning of the low quality rules may allow enhancing the recommender system performance. As an extreme case, when enforcing very high confidence thresholds (e.g., 90%), rule pruning selectivity becomes too high to generate a considerable amount of interesting patterns.

Best values of support and confidence threshold actually depend on the analyzed data distribution. For instance, when coping with the benchmark dataset the best minimum support threshold values are around 20%, because the analyzed dataset is relatively sparse.

Success at rank k (e.g., see S@5 in Figures 7(a) and 7(b)) is shown to be, on average, less affected by support and confidence thresholds than precision at rank k, because the probability of finding a relevant tag in the top-k recommended tags is more weakly influenced by the rule-based model quality than the percentage of retrieved relevant tags.

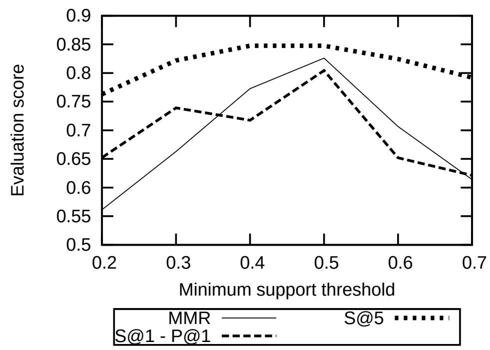
5. Conclusions and future work

This paper presents a novel personalized tag recommendation system that performs additional tag recommendations to partially annotated Flickr photos by exploiting generalized association rules extracted from the collections of the past personal and collective annotations. The use of high level associations is focused on counteracting the impact of data sparsity as it may highlight correlations among tags that could remain hidden at the level of individual tags.

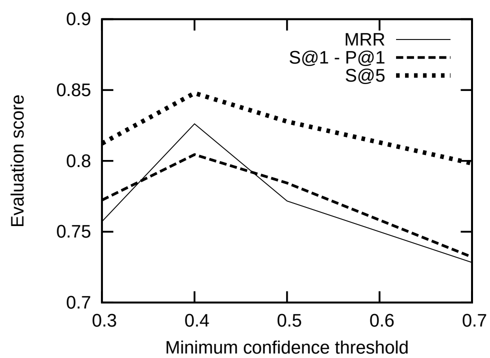
A set of experiments has been conducted on real-life Flickr photo collections. The effectiveness of the proposed approach has been validated against a recently proposed tag recommendation system. Experiments show that the use of the generalizations in rule-based tag recommendation yields significant performance improvements.

Our system has so far not been concerned with the analysis of the textual content related to the annotated Web resources (e.g., photo descriptions, related blogs or articles). We plan to extend it by also considering the user-generated textual content coming from social networks and online com-

Figure 6: Parameter analysis. MRR, S@1/P@1, and P@5 measures.



(a) Impact of the support threshold. $minconf$=40%.



(b) Impact of the confidence threshold. $minsup$=50%

munities. Furthermore, to take the evolution of photo annotations over time we will investigate the integration of incremental rule mining approaches as well.

References

[1] Agrawal, R., Imieliński, T., and Swami, A. 1993. Mining association rules between sets of items in large databases. In *ACM SIGMOD Record*. Vol. 22. ACM, New York, 207–216.

[2] Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., and Su, Z. 2007. Optimizing web search using social annotations. In *Proceedings of the 16th international conference on World Wide Web*. WWW '07. ACM, New York, NY, USA, 501–510.

[3] Baralis, E., Cagliero, L., Cerquitelli, T., D'Elia, V., and Garza, P. 2010. Support driven opportunistic aggregation for generalized itemset extraction. In *IEEE Conference of Intelligent Systems*. 102–107.

[4] Brin, S. and Page, L. 1998. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the seventh international conference on World Wide Web 7*. 107–117.

[5] Chirita, P. A., Costache, S., Nejdl, W., and Handschuh, S. 2007. P-tag: large scale automatic generation of personalized annotation tags for the web. In *Proceedings of the 16th international conference on World Wide Web*. WWW '07. ACM, New York, NY, USA, 845–854.

[6] Datta, R., Ge, W., Li, J., and Wang, J. Z. 2007. Toward bridging the annotation-retrieval gap in image search. *IEEE MultiMedia 14*, 24–35.

[7] Dmitriev, P. A., Eiron, N., Fontoura, M., and Shekita, E. 2006. Using annotations in enterprise search. In *Proceedings of the 15th international conference on World Wide Web*. WWW '06. ACM, New York, NY, USA, 811–817.

[8] Elmasri, R. and Navathe, S. B. 2006. *Fundamentals of Database Systems (5th Edition)*. Addison Wesley.

[9] Garg, N. and Weber, I. 2008. Personalized, interactive tag recommendation for flickr. In *Proceedings of the 2008 ACM conference on Recommender systems*. RecSys '08. ACM, New York, NY, USA, 67–74.

[10] Guan, Z., Bu, J., Mei, Q., Chen, C., and Wang, C. 2009. Personalized tag recommendation using graph-based ranking on multi-type interrelated objects. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. SIGIR '09. ACM, New York, NY, USA, 540–547.

[11] Han, J. and Fu, Y. 2002. Mining multiple-level association rules in large databases. *IEEE Transactions on Knowledge and Data Engineering 11,* 5, 798–805.

[12] Heymann, P., Ramage, D., and Garcia-Molina, H. 2008. Social tag prediction. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. SIGIR '08. ACM, New York, NY, USA, 531–538.

[13] Huiskes, M. J. and Lew, M. S. 2008. The mir flickr retrieval evaluation. In *MIR '08: Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval.* ACM, New York, NY, USA.

[14] Jäschke, R., Marinho, L., Hotho, A., Lars, S.-T., and Gerd, S. 2007. Tag recommendations in folksonomies. In *Proceedings of the 11th European conference on Principles and Practice of Knowledge Discovery in Databases*. PKDD 2007. Springer-Verlag, Berlin, Heidelberg, 506–514.

[15] Krestel, R., Fankhauser, P., and Nejdl, W. 2009. Latent dirichlet allocation for tag recommendation. In *Proceedings of the third ACM conference on Recommender Systems*. RecSys '09. ACM, New York, NY, USA, 61–68.

[16] Lindstaedt, S., Pammer, V., Mörzinger, R., Kern, R., Mülner, H., and Wagner, C. 2008. Recommending tags for pictures based on text, visual content and user context. In *Proceedings of the 2008 Third International Conference on Internet and Web Applications and Services*. IEEE Computer Society, Washington, DC, USA, 506–511.

[17] Lipczak, M. and Milios, E. 2011. Efficient tag recommendation for real-life data. *ACM Trans. Intell. Syst. Technol. 3,* 1, 2:1–2:21.

[18] Lu, Y. T., Yu, S. I., Chang, T. C., and jen Hsu, J. Y. 2009. A Content-Based Method to Enhance Tag Recommendation. In *IJCAI*, C. Boutilier and C. Boutilier, Eds. 2064–2069.

[19] Menezes, G. V., Almeida, J. M., Belém, F., Gonçalves, M. A., Lacerda, A., De Moura, E. S., Pappa, G. L., Veloso, A., and

Ziviani, N. 2010. Demand-driven tag recommendation. In *Proceedings of the 2010 European conference on Machine learning and knowledge discovery in databases: Part II*. ECML PKDD'10. Springer-Verlag, Berlin, Heidelberg, 402–417.

[20] Mennis, J. and Liu, J. 2005. Mining association rules in spatio-temporal data: an analysis of urban socioeconomic and land cover change. *Transactions in GIS 9,* 1, 5–17.

[21] Mishne, G. 2006. Autotag: a collaborative approach to automated tag assignment for weblog posts. In *Proceedings of the 15th international conference on World Wide Web*. WWW '06. ACM, New York, NY, USA, 953–954.

[22] Pramudiono, I. and Kitsuregawa, M. 2004. Fp-tax: Tree structure based generalized association rule mining. In *Proceedings of the 9th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. ACM, New York, 63.

[23] Rae, A., Sigurbjörnsson, B., and van Zwol, R. 2010. Improving tag recommendation using social networks. In *Adaptivity, Personalization and Fusion of Heterogeneous Information*. RIAO '10. Le centre de Haute Etudes Internationales d'Informatique Documentaire, Paris, France, France, 92–99.

[24] Sanderson, M. and Zobel, J. 2005. Information retrieval system evaluation: effort, sensitivity, and reliability. In *Proceedings of the 28th*

annual international ACM SIGIR conference on Research and development in information retrieval. SIGIR '05. ACM, New York, NY, USA, 162–169.

[25] Sarwar, B., Karypis, G., Konstan, J., and Reidl, J. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*. WWW '01. ACM, New York, NY, USA, 285–295.

[26] Sigurbjörnsson, B. and van Zwol, R. 2008. Flickr tag recommendation based on collective knowledge. In *Proceedings of the 17th international conference on World Wide Web*. WWW '08. ACM, New York, NY, USA, 327–336.

[27] Srikant, R. and Agrawal, R. 1995. Mining generalized association rules. In *International Conference on Very Large Data Bases*. Morgan Kaufmann, San Fransisco, 407–419.

[28] Srikant, R., Vu, Q., and Agrawal, R. 1997. Mining association rules with item constraints. In *Conference on Knowledge Discovery and Data Mining*. Vol. 97. AAAI Press, USA, 67–73.

[29] Sriphaew, K. and Theeramunkong, T. 2002. A new method for finding generalized frequent itemsets in generalized association rule mining. In *Seventh International Symposium on Computers and Communications*. IEEE, Washington, D.C., 1040–1045.

[30] Symeonidis, P., Nanopoulos, A., and Manolopoulos, Y. 2008. Tag recommendations based on tensor dimensionality reduction. In *Pro-*

ceedings of the 2008 ACM conference on Recommender systems. RecSys '08. ACM, New York, NY, USA, 43–50.

[31] van Erp, M. and Schomaker, L. 2000. Variants of the borda count method for combining ranked classifier hypotheses. In *7th International Workshop on frontiers in handwriting recognition.* 443–452.

[32] Wordnet. 2012. Wordnet Lexical Database. Retrievable from http://wordnet.princeton.edu. Last accessed: 25-06-2012.